EMA/95098/2010 Rev.11

The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology

(Revision 11)

KEYWORDS	*methodological standards, pharmacoepidemiology, pharmacovigilance, ENCePP, research, guidance, real-world data (RWD), real-world evidence (RWE)*

This document should be cited as follows: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 11). EMA/95098/2010.

Available at http://www.encepp.eu/standards_and_guidance

Table of contents

Foreword to the ENCePP Guide on Methodological Standards in Pharmacoepidemiology, Revision 11

The increasing ability and expertise to reuse electronic real-world data (RWD) from routine healthcare systems continues to open up opportunities for investigators to conduct non-interventional studies on the utilisation, safety and effectiveness of medicinal products, and stimulate the development or improvement of design strategies to mitigate bias and generate valid real-world evidence (RWE), especially when the study aims to conclude on causal effects.

In this 11th Revision of the ENCePP Guide, the table of contents has been restructured to better reflect the evidence generation flow and provide greater emphasis on important methodology such as the causal inference target trial emulation approach, which is recommended by ENCePP for non-interventional causal inference studies, to improve internal validity and increase transparency on design characteristics. Chapter 4.2.6 on this topic has been further updated and expanded, and this methodology is integrated, with new references, in Chapters 4 on Study design, 16.1 on Comparative Effectiveness Research, 16.2 on Vaccine safety and effectiveness, and 16.6 on RWE and pharmacoepidemiology.

The HARPER protocol template was published in early 2023 to guide the structure and content of RWE study protocols, and serve as a tool to promote transparency, reproducibility and harmonisation of non-interventional study protocols, and facilitate their assessment by regulators and other stakeholders. HARPER is compatible with the legal format and content of GVP Module VIII and is therefore recommended for the development of post-authorisation study protocols.

On 5 May 2023, the World Health Organisation (WHO) announced that COVID-19 is now an established and ongoing health issue, which no longer constitutes a public health emergency of international concern (PHEIC). Nevertheless, the COVID-19 pandemic still has an impact on the work performed by pharmacoepidemiologists, due to the continued need to evaluate the use, safety and effectiveness of COVID-19 vaccines and therapeutics, and appraise the published evidence. As in Revisions 9 and 10, this 11th Revision builds on lessons from COVID-19 that can be applied in routine pharmacoepidemiological/RWE practice, as well as considered for pandemic preparedness. Several challenges remain, such as the availability of adequate information on vaccine exposure or variant-specific epidemiological data, the widespread use of self-testing that is not recorded in healthcare data sources, the low frequency of data source updates, and the lag times to access recent data, all of which still affects the speed of safety and effectiveness evaluation or its feasibility. ENCePP will continue to evaluate the very rich methodological work done by researchers during the pandemic and support drawing and addressing lessons learned. This includes considering the impact on studies in other areas that include the pandemic in their study period, since profound and varying changes in healthcare utilisation are also reflected in the data sources available for secondary use.

A large amount of literature has started accumulating on post-acute COVID-19 syndrome, or long COVID, in parallel to heightened awareness of this new public health challenge. Its prevalence is estimated to be around 45% among people who have been infected with SARS-CoV-2; based on an estimate of around 760 million infections worldwide at the time of publication of this 11th Revision, this translates into more than 300 million individuals who may have had, or are currently experiencing long COVID. Valid studies are needed to evaluate the incidence and risk factors of long COVID, characterise the multiple components of this syndrome, and identify the possible effects of treatments of the initial infection as well as of post-acute symptoms. This 11th Revision therefore continues to provide examples of COVID-19 related studies illustrating good practice and methodological developments in pharmacoepidemiology.

Finally, some chapters have been revised to reflect a fast changing environment in the field of pharmacoepidemiology and RWE: Chapter 9 on Research networks for multi-database studies addresses the expansion and use of DARWIN EU®; Chapter 12 on Statistical analyses includes a recommendation to consider the estimand framework to inform study design and data analysis choices; and Chapters 16.5 on Artificial intelligence in pharmacoepidemiology and 16.6 on RWE and pharmacoepidemiology have been extensively updated. Another emphasis is placed, in Chapters 3 on Protocol development and 4.1 on Overview of study designs, on the importance of design diagrams to foster transparency, enhance understanding of the study design, and support the evaluation of study protocols and the interpretation of study results.

We hope that this 11th Revision of the ENCePP Guide will continue to support good practices in pharmacoepidemiology/RWE, in Europe and elsewhere.

Susana Perez-Gutthann
Catherine Cohet
Co-Chairs of the ENCePP Steering Group

1. Introduction

Epidemiology is the study of the occurrence of health phenomena in the population, their frequency and their relationship with determinants (O. Miettinen. *Theoretical Epidemiology: Principles of Occurrence Research in Medicine*, Wiley, 1985, vii-viii). Pharmacoepidemiology is the application of epidemiological methods to study the use and effects of medicines. Many textbooks describe methodological standards in pharmacoepidemiology, but they cannot incorporate all new developments. ENCePP therefore considered there was a need for a regularly updated resource providing recommendations on the practical implementation of pharmacoepidemiological principles and innovative methods based on published guidance and illustrative examples.

The ENCePP Guide aims to offer a dynamic and publicly available web resource for methodological English language guidance in pharmacoepidemiology. For each topic covered, recommendations are provided with links to selected published articles, guidelines and public documents. Where relevant, gaps in existing guidance are addressed with what ENCePP considers good practice. The Guide is updated annually in order to maintain its dynamic nature and to ensure up-to-date information is provided. It may also be amended as necessary in response to comments received. For this purpose, any comment and additional relevant guidance documents may be forwarded to ENCePP_Secretariat@ema.europa.eu.

The Guide only briefly addresses general methods of pharmacoepidemiology, as study designs stemming from traditional epidemiological research are already fully described in existing textbooks. It rather discusses important aspects of more recent designs and specific analytical approaches to traditional designs. It also discusses methods that may be used in studies with different objectives, be they safety, effectiveness or drug utilisation. For some specific topics, recommendations and references are provided where they differ from general principles.

Chapter 16 provides guidance on specific topics: comparative effectiveness research; vaccine safety and effectiveness; pharmacogenomic studies; methods for pharmacovigilance impact research; artificial intelligence in pharmacoepidemiology; and real-world evidence and pharmacoepidemiology. Annexes provide more comprehensive guidance: Annex 1 on systematic review and meta-analyses; and Annex 2 on methods for the evaluation of medicines in pregnancy and breastfeeding.

Overall, general guidance on the conduct of pharmacoepidemiology studies can be found in the ISPE Good Pharmacoepidemiology Practices (GPP) and the IEA Good Epidemiology Practice (GEP). The GPP guidance is especially useful for its recommendations on aspects rarely covered by guidelines, such as data quality issues and archiving. The Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP): a guideline developed by the German Society for Epidemiology (Eur J Epidemiol. 2019;34(3):301-17) provides detailed recommendations for the planning, preparation, execution, analysis, and evaluation of epidemiological research.

The Guideline of good pharmacovigilance practices (GVP) Module VIII - Post-authorisation safety studies (2017) provides general guidance on the development, conduct and reporting of post-authorisation safety studies (PASS) conducted by marketing authorisation holders, voluntarily or pursuant to the EU legislation (Directive 2001/83/EC). The Scientific guidance on post-authorisation efficacy studies (2017) provides general scientific guidance in the context of EU regulatory decision-making with regards to the need for such studies and methodological considerations.

Where to begin? Thirty must-read papers for newcomers to pharmacoepidemiology (Pharmacoepidemiol Drug Saf. 2022;31(2):257-9) provides a list of 30 articles that the authors consider touching upon core principles of pharmacoepidemiology and RWE.

Textbooks on standard methods in pharmacoepidemiology that are considered useful are listed below. This list is not aimed at being exhaustive, and researchers may find other textbooks more appropriate to their specific needs. Others are cited in specific chapters.

- Modern Epidemiology, 4th ed. (T. Lash, T.J. VanderWeele, S. Haneuse, K. Rothman. Wolters Kluwer, 2020) is a comprehensive textbook on methods in epidemiology.
- Epidemiology: Study Design and Data Analysis, Third Edition (M. Woodward, Chapman & Hall, 2014) focuses on the quantitative aspects of epidemiological research.
- A Dictionary of Epidemiology, Sixth Edition (M Porta, Editor. Oxford University Press, 2014), sponsored by the International Epidemiological Association (IEA), provides a definition and concise explanation of epidemiologic terms and is a key to understanding epidemiological concepts.
- Dictionary of Pharmacoepidemiology (B. Bégaud, Wiley, 2000) is the most complete dictionary devoted specifically to terms currently used in pharmacoepidemiology.
- Clinical epidemiology: practice and methods, Second Edition (PS Parfrey, BJ Barret, Human Press, 2015) focuses on the diagnosis, prognosis and management of human disease using appropriate research design, measurement and evaluation.
- Causal Inference: What If (Hernán MA, Robins JM, Chapman & Hall/CRC, 2020) aims to help scientists generate and analyse data to make causal inferences that are explicit about both the causal question and the assumptions underlying the data analysis.
- Pharmacoepidemiology, Sixth Edition (B. Strom, S.E. Kimmel, S. Hennessy, Wiley, 2019) provides a comprehensive guidance on pharmacoepidemiology addressing data sources, applications and methodologies.
- Pharmacoepidemiology and Therapeutic Risk Management, First Edition (A.G. Hartzema, H.H. Tilson and K.A. Chan, Editors, Harvey Whitney Books Company, 2008) illustrates practical issues with a large number of real-life examples in addition to a general review of drug-specific methodologies.
- Practical Statistics for Medical Research, Second Edition (D. Altman. Chapman & Hall, 2020) presents a problem-based statistical text for medical researchers.
- Drug Utilization Research. Methods and Applications (M Elseviers, B Wettermark, AB Almarsdóttir, et al. Editors. Wiley Blackwell, 2016) provides a comprehensive manual of methodology and applications of drug utilisation research.
- Mann's Pharmacovigilance, Third Edition (EB Andrews, N Moore, Editors, Wiley-Blackwell, 2014) is a reference for the detection, assessment, understanding and prevention of the adverse effects of medicines, including vaccines and biologics.
- Post-Authorization Safety Studies of Medicinal Products. The PASS Book, 1st Edition (Ayad Ali, Abraham Hartzema, Ed., Academic Press, 2018) covers the use of observational studies in post-marketing drug safety assessment, presents various types of post-authorisation safety studies and discusses challenges and solutions in the design and conduct of these studies.

2. Formulating the research question and objectives and assessing study feasibility

Generating adequate evidence involves formulating the right research question(s), identifying and collecting fit-for-purpose data, applying suitable study designs, and conducting the appropriate analyses. Asking the right question is crucial as it is the stepping-stone to the development and conduct of a meaningful study.

The articles Research: Articulating Questions, Generating Hypotheses, and Choosing Study Designs, Setting a research question, aim and objective (CJHP 2014;67(1):31-34) and Formulating Answerable Questions: Question Negotiation in Evidence-based Practice (JCHLA/JABSC 2013;34:55-60) suggest stepwise approaches to the generation of the research question.

In an initial step, a flow of research topics is generated from clinical practice, patient experience, unmet medical need, pharmaceutical companies' development plans, public health issues, and, overall, during regulatory and health technology assessment (HTA) processes. It is recommended to include all relevant stakeholders in the ideation process. A parallel critical and thorough review of the literature forms the basis for the theoretical framework of the research question and should be included in the background section of the study protocol. Such a review aims at evaluating the current evidence and identifying gaps in knowledge that the study is intended to address. This process should allow the researcher(s) to select the most relevant question(s) for research, and transform the question into study objectives. The study can be hypothesis-generating, or include a testable hypothesis if it involves hypothesis-testing. In Posing the research question: not so simple (Can J Anaesth. 2009;56(1):71-9), the FINER criteria (Feasible, Interesting, Novel, Ethical, and Relevant) are proposed to verify the desirable properties of an appropriate, meaningful and purposeful research idea.

Research questions relevant to regulatory authorities and HTA bodies regarding the utilisation, safety, efficacy (or effectiveness) and impact of medicines are detailed in the European Public Assessment Report (EPAR) available for each centrally authorised product on the EMA website, with general pharmacovigilance-related aspects being described in Modules of the Good Pharmacovigilance Practices (GVP). The European Network of Health Technology Assessment (EUnetHTA) describes The criteria to select and prioritise health technologies for additional evidence generation (2012) and discusses clinical evidentiary requirements to support the study design strategy in Strengthening the Interface of Evidence-Based Decision Making Across European Regulators and Health Technology Assessment Bodies (Value Health 2022:S1098-3015(22)00104-8).

In a second step, a research question is formulated through "*a logical statement that progresses from what is known or believed to be true to that which is unknown and requires validation*", as described in Developing great research questions (Am J Health Syst Pharm. 2008;65(17):1667-70). A poorly defined research question can jeopardise the development of a clear and meaningful protocol and can impact the interpretation of the results of the study and its further publication. How to formulate research recommendations (BMJ. 2006;333(7572):804-6) proposes the EPICOT format with 5 core elements for research recommendations on the effects of treatments: Evidence (source of the current evidence), Population (population characterised by any diagnosis, disease stage, comorbidity, risk factor, sex, age, ethnic group, specific inclusion or exclusion criteria, clinical setting), Intervention (type, frequency, dose, duration, prognostic factor), Comparison (placebo, routine care, alternative treatment/management), Outcome (which clinical or patient related outcomes will the researcher need to measure, improve, influence or accomplish; which methods of measurement should be used), and Time stamp (date of literature search or recommendation). This format was adopted by EUnetHTA in its Position paper on how to formulate research recommendations (2015).

In a further step, the objectives of the study are defined. They are more operational than the research question and can be divided into primary, secondary and exploratory. The primary objective corresponds to the most important objective of the study, which drives the study design, the estimation of the sample size and the methods to address the research question. It can sometimes be a composite objective. Secondary objectives can provide additional details to support the research question, to add new knowledge or comparison, and can potentially use other study designs or methods to complement the evidence. The objectives should be closely related to the research question, cover all its aspects, and be ordered in a logical sequence. The SMART criteria (Specific, Measurable, Appropriate (aligned with the research question), Realistic and Time specific) can be used to formulate the study objectives (There's a S.M.A.R.T. way to write management's goals and objectives, 1981).

Assessing the feasibility of the study constitutes a key preparatory step and is recommended to ensure that sufficient information is available to apply the proposed study design, for example, knowledge about the information available in the data sources. The aim is not to answer the research question, but to determine whether the proposed methodology could answer the research question with the expected statistical power and within the proposed timelines. A feasibility assessment can provide information on the number of subjects with a specific exposure or outcome, the availability of covariates and the follow-up period needed. It can also provide insights into the potential difficulties which may be encountered in the conduct of the study, or which may introduce bias. Importance of feasibility assessments before implementing non-interventional pharmacoepidemiologic studies of vaccines: lessons learned and recommendations for future studies (Pharmacoepidemiol Drug Saf. 2016;25(12):1397-406) illustrates a pragmatic approach for conducting feasibility assessments for post-authorisation studies, which can be applied for any medicinal product. The ISPE Good pharmacoepidemiology practice (GPP) explains how a data collection method or data source can answer a research question with justifications based on feasibility when relevant. Linking electronic health data in pharmacoepidemiology: Appropriateness and feasibility (Pharmacoepidemiol Drug Saf. 2020;29(1):18-29) provides guidance to assess the feasibility of data linkage based on key areas including the design of the research question for study objectives addressed using secondary data collection. Other insights on formulating the research question and evaluating study feasibility are provided in Evaluating the Feasibility of Electronic Health Records and Claims Data Sources for Specific Research Purposes (Ther Innov Regul Sci. 2020;54(6):1296-1302).

Building on existing guidance and frameworks, the SPACE framework (A Structured Preapproval and Postapproval Comparative Study Design Framework to Generate Valid and Transparent Real-World Evidence for Regulatory Decisions, Clin Pharmacol Ther. 2019;106(1):103-115) describes a step-by-step process for identifying valid design elements and minimal feasibility criteria. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies (BMJ 2021;372:m4856) provides detailed templates to capture the final design and implementation details (e.g., specific algorithms for each study variable). In addition to these processes and templates, The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework (Clin Pharmacol Ther. 2022;111(1):122-134) provides a systematic approach to determine if a data source is fit for regulatory decision-making, helping ensure justification and transparency throughout study development, from articulation of a specific and meaningful research question to identification of fit-for-purpose data and study design, illustrated by use cases. A Structured Process to Identify Fit-for-Purpose Study Design and Data to Generate Valid and Transparent Real-World Evidence for Regulatory Uses (Clin Pharmacol Ther. 2023;113(6):1235-1239) provides further guidance to support the generation of valid, fit-for-purpose evidence.

It is to be noted that formulating a research question is an iterative process. For example, the outcome of a feasibility assessment may reveal that the research question is not feasible (e.g., due to

insufficient number of patients who meet the eligibility criteria, unavailability of adequate data sources or data linkages) and consequently lead to the adaptation of the objectives or design of the study.

3. Development of the study protocol

The study protocol is the core document of a study, to be developed as a key step in any study once the research question has been clearly defined. It is strongly recommended to assess feasibility of answering the research question ahead of developing the protocol (see Chapter 2). The final version must precisely describe all study objectives and design characteristics to ensure reproducibility of the study. The protocol should be amended as needed, and amendments should be justified.

The GVP Module VIII - Post-authorisation safety studies (PASS) has been available since 2012. For PASS described in this module, the Commission Implementing Regulation (EU) No 520/2012 provides legal definitions of the start of data collection (the date from which information on the first study subject is first recorded in the study dataset, or, in the case of secondary use of data, the date from which data extraction starts) and end of data collection (the date from which the analytical dataset is completely available). These dates provide a timeline supporting the planning of the overall study and the submission of the final study report to competent authorities. Module VIII of the GVP also details the required format of protocols, abstracts and final study reports for imposed PASS. Based on these requirements, the EMA published detailed templates for the protocol and final study report which it recommends to be used for all PASS, including meta-analyses and systematic reviews. Although these templates have been developed to address research questions related to the safety of medicinal products, they can be applied to any type of pharmacoepidemiological study.

Derived from an international consensus, the HARPER protocol template (HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force, Pharmacoepidemiol Drug Saf. 2023;32(1):44-55) became available in 2023 to guide structure and content of real-world evidence (RWE) study protocols, with a focus on providing information on operational study parameters used to create analytical datasets from the data collected to address the study objectives. It can serve as a tool to promote transparency, reproducibility and harmonisation of non-interventional study protocols and can facilitate design and assessment of high-quality protocols. HARPER is compatible with the legal format and content of GVP Module VIII and can be used in PASS protocols (or the protocol of any pharmacoepidemiological study) without change of structure.

The ISPE Guidelines for Good Pharmacoepidemiology Practices (GPP) provides guidance on contents of a pharmacoepidemiology study protocol and the different contents to be covered. It states that the protocol should include a description of the data quality and integrity, including abstraction of original documents, extent of source data verification, and validation of endpoints. The FDA's Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Health Care Data Sets includes a description of the design elements that should be addressed, including the choice of data sources and study populations, the study design and statistical analyses. The ENCePP Checklist for Study Protocols seeks to stimulate researchers to consider important epidemiological concepts when designing a pharmacoepidemiological study and writing a study protocol. The Agency for Healthcare Research and Quality (AHRQ) published Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide (2013) including best practice, principles and checklists on a wide range of topics that are also applicable to observational studies outside the scope of comparative effectiveness research.

A key component for design transparency, also included in the HARPER template, is visualisation through study design diagrams. Graphical Depiction of Longitudinal Study Designs in Health Care Databases (Ann Intern Med. 2019;19;170(6):398-406) provides a simple framework to help

understanding how the design will be implemented, especially in relation to the definition of time periods for data collection. Such graphical frameworks for presenting study designs in the protocol are recommended to foster transparency, enhance understanding of the design, and support the evaluation of the protocols and the interpretation of study results, as also illustrated in A Framework for Visualizing Study Designs and Data Observability in Electronic Health Record Data (Clin Epidemiol. 2022;14:601-8) and Visualizations throughout pharmacoepidemiology study planning, implementation, and reporting (Pharmacoepidemiol Drug Saf. 2022;31(11):1140-52).

For consent process and ethical guidelines related to human subject research, see Chapter 15.2. HARPER also provides considerations on protection of human subjects based on GDPR and use of anonymised or pseudo-anonymised data sources.

GVP Module VIII - Post-authorisation safety studies provides a structure for study protocols, which should cover at least the following aspects:

- The research question that the study is designed to answer, which might be purely descriptive, exploratory or explanatory (hypothesis-driven) (see Chapter 2). The protocol should include a background description that explains the rationale (scientific, regulatory, etc.) and current knowledge on the research question. It will also explain the context of the research question, including what data are currently available and how these data can or cannot contribute to answering the question. The context will also be defined in terms of what information sources can be used to generate appropriate data and how the proposed study methodology will be shaped around these data.
- The main study objective and possible secondary objectives, which are operational definitions of the research question. In defining secondary objectives, consideration could be given to time and cost, which may impose constraints and choices, for example in terms of feasibility, sample size, duration of follow-up, sensitivity analyses, or data collection (see Chapter 2).
- The source and study population to be used to answer the research question. The protocol should describe whether this population is already identified, and whether data are already available (secondary data collection) or whether data needs to be generated de novo (primary data collection). The boundaries of the desired population will be defined, including inclusion/exclusion criteria, timelines (such as index dates for inclusion in the study) and any exposure or events defining the population.
- Exposures of interest that need to be pre-specified and defined, including duration and intensity of exposure, source of data and methods of ascertainment (see Chapter 4.3).
- Outcomes of interest that need to be pre-specified and defined, including data sources, operational definitions and methods of ascertainment such as data elements in field studies or appropriate codes in database studies (see Chapter 4.3).
- Adverse events/reactions that will or will not be collected and reported and the procedures put in place for this purpose. In the EU, the collection and reporting of adverse events or reactions by companies sponsoring a post-authorisation study should follow the recommendations specified in Module VI of the Guideline on good pharmacovigilance practice (GVP) - Management and reporting of adverse reactions to medicinal products. If the study qualifies as an interventional trial, the reporting criteria laid down in Clinical Trial Regulation (EU) 536/2014 and the draft Volume 10 - Guidance documents applying to clinical trials should be followed.
- The covariates and potential confounders that need to be pre-specified and defined, including how they will be measured (see Chapter 4.3).

- The statistical analysis plan, including statistical methods and software used, adjustment strategies, and how the results are going to be presented (see Chapter 12).
- The identification and way of minimisation of potential biases (see Chapter 6).
- Major assumptions, critical uncertainties and challenges in the design, conduct and interpretation of the results of the study given the research question and the data used.
- Ethical considerations, as described in Chapter 15.
- The study protocol should also explain how the results will be interpreted, avoiding misuse of p-values and statistical significance (see Chapter 4.1).

The HARPER template also recommends including in the protocol:

- A rationale, context and table for choices relating to selection of time zero, inclusion/exclusion criteria.
- Structured tables for exposure, outcome, follow-up and covariates, as well as validation, with a description of algorithms used for data collection.
- An evaluation of the fitness-for purpose of the data source(s) used.
- A structured table detailing all sensitivity analyses.

Various data collection forms including the Case Report Form (CRF) for primary data collection, and list of disease codes or descriptions of the data elements for secondary data collection, may be appended to the protocol, providing an exact representation of how the data will be collected. The study protocol could include a section specifying ways in which the CRF will be piloted, tested and finalised. Amendments of final CRFs should be justified. For field studies, physician or patient forms could be included depending on the data collection methodology. Other forms may be included as needed, such as patient information, consent form or patient-oriented summaries.

Registration of the study protocol before the start of data collection provides information to other researchers about the study, improves transparency and, especially for studies based on secondary use of data, provides assurance that the stated hypotheses have not been influenced by the results. The EU PAS Register is a public register open to any researcher for the registration of non-interventional studies. In addition, study protocols can be registered and posted on other platforms: ClinicalTrials.gov now includes specific guidelines for the posting of non-interventional research, while since 2020, the Open Science Forum has a specific registration portal for observational studies.

4. Study design

4.1. Overview

An epidemiological study measures a parameter of occurrence (generally incidence, prevalence or risk or rate ratio) of a health phenomenon (e.g., a disease) in a specified population and with a specified time reference (time point or time period). Epidemiological studies may be descriptive or analytic. Descriptive studies do not aim to evaluate a causal relationship between a population characteristic and the occurrence parameter and generally do not include formal comparisons between population groups. Analytic studies (also called causal inference studies), in contrast, use study populations assembled by the investigators to assess relationships that may be interpreted in causal terms. In pharmacoepidemiology, analytic studies generally aim to quantify the association between exposure to a medicine and a health phenomenon, and test the hypothesis of a causal relationship. They are

comparative by nature, e.g., comparing the occurrence of an outcome between subjects being users of the medicine or non-users, or users of a different medicinal product.

Studies can be interventional or non-interventional (observational). In interventional studies, the subjects are assigned by the investigator to be either exposed or unexposed. Most often, in these studies, exposure is assigned randomly and are known as randomised clinical trials (RCTs), and are typically conducted to test the efficacy of treatments such as new medications. In RCTs, randomisation is used with the intention that the only difference between the exposed and unexposed groups will be the treatment itself. Thus, any differences in the outcome can be attributed to the effect of such treatment. In contrast to experimental studies where exposure is assigned by the investigator, in observational studies the investigator plays no role with regards to which subjects are exposed and which are unexposed. The exposures are either chosen by, or are characteristics of, the subjects themselves. Observational Studies: Cohort and Case-Control Studies (Plast Reconstr Surg. 2010;126(6):2234-42) provides a simple and clear explanation of the different types of observational studies and of their advantages and disadvantages (see also Chapter 4.2. Study designs).

In order to obtain valid estimates of the effect of a determinant on a parameter of disease occurrence, analytic studies must address three factors: random error (chance), systematic error (bias) and confounding. It is important to understand that error is defined as the difference in the measured value to the true value of a particular observation.

- *Random error (chance)*: the observed effect estimate is a numerical value which may be explained by random error because of the underlying variation in the population. The confidence interval (CI) allows the investigator to estimate the range of values within which the actual effect is likely to fall.
- *Systematic error (bias)*: the observed effect estimate may be due to systematic error in the selection of the study population or in the measurement of the exposure or disease. Two main types of biases need to be considered, selection bias and information bias. Selection bias results from procedures used to select subjects and from factors that influence study participation. For example, a case-control study may include non-case subjects with a higher prevalence of one category of the exposure of interest than in the source population for the cases. External factors such as media attention to safety issues may also influence healthcare seeking behaviours and measurement of the incidence of a given outcome. Information biases can occur whenever there are errors in the measurement of subject characteristics, for example a lack of pathology results leading to outcome misclassification of certain types of tumours, or lack of validation of exposure, leading to misclassify the exposed and non-exposed status of some study participants. For example, mothers of children with congenital malformations will recall more instances of medicine use during pregnancy than mothers of healthy children. This is known in epidemiology as “recall bias”, a type of information bias. The consequences of these errors generally depend on whether the distribution of errors for the exposure or disease depends on the value of other variables (differential misclassification) or not (nondifferential misclassification).
- *Confounding*: Confounding results from the presence of an additional factor, known as a confounder or confounding factor, which is associated with both the exposure of interest and the outcome. As a result, the exposed and unexposed groups will likely differ not only with regards to the exposure of interest, but also with regards to a number of other characteristics, some of which are themselves related to the likelihood of developing the outcome. Confounding distorts the observed effect estimate for the outcome and the exposure under study. As there is not always a firm distinction between bias and confounding, confounding is also often classified as a type of bias.

There are many different situations where bias may occur, and some authors attribute a name to each of them. The number of such situations is in theory illimited. ENCePP recommends that, rather than being able to name each of them, it is preferable to understand the underlying mechanisms of information bias, selection bias and confounding, be alert to their presence and likelihood of occurrence in a study, and recognise methods for their prevention, detection, and control at the analytical stage if possible - such as restriction, stratification, matching, regression and sensitivity analyses. Chapter 6 on methods to address bias nevertheless treats time-related bias (a type of information bias with misclassification of person-time) separately, as it may have important consequences on the result of a study and may be dealt with by design and time-dependent analyses.

The role of chance (random error) in the interpretation of evidence in epidemiology has often relied on whether the p-value is below a certainty threshold and/or the confidence interval excludes some reference value. The ASA statement on *P* values: context, process, and purpose (Am Statistician 2016;70(2),129-33) of the American Statistical Association emphasised that a p-value, or statistical significance, does not provide a good measure of evidence regarding a model or hypothesis, nor does it measure the size of an effect or the importance of a result. It is therefore recommended to avoid relying only on statistical significance, such as p-values, to interpret study results (see, for example, Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations, Eur J Epidemiol. 2016;31(4):337-50; Scientists rise up against statistical significance, Nature 2019;567(7748):305-7; It's time to talk about ditching statistical significance, Nature 2019;567(7748):283; Chapter 15. Precision and Study size in *Modern epidemiology*, Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, 4th edition, Philadelphia, PA, Wolters Kluwer, 2021). This series of articles led to substantial changes in the guidelines for reporting study results in manuscripts submitted to medical journals, as discussed in Preparing a manuscript for submission to a medical journal (International Committee for Medical Journal Editors, 2021). Causal analyses of existing databases: no power calculations required (J Clin Epidemiol. 2022;144:203-5) encourages researchers to use large healthcare databases to *estimate* measures of association as opposed to systematically attempting at *testing* hypotheses (with sufficient power). The ENCePP also recommends that, instead of a dichotomous interpretation based on whether a p-value is below a certain threshold, or a confidence interval excludes some reference value, researchers should rely on a more comprehensive quantitative interpretation that considers the magnitude, precision, and possible bias in the estimates, in addition to a qualitative assessment of the relevance of the selected study design. This is considered a more appropriate approach than one that ascribes to chance any result that does not meet conventional criteria for statistical significance.

Given that the large number of observational studies performed urgently with existing data and in sometimes difficult conditions in early times of the COVID-19 pandemic has raised concerns about the validity of many studies published without peer-review, we recommend to balance urgency and use of appropriate methodology. Considerations for pharmacoepidemiological analyses in the SARS-CoV-2 pandemic (Pharmacoepidemiol Drug Saf. 2020;29(8):825-83) provides recommendations across eight domains: (1) timeliness of evidence generation; (2) the need to align observational and interventional research on efficacy (3) the specific challenges related to "real-time epidemiology" during an ongoing pandemic; (4) which design to use to answer a specific question; (5) considerations on the definition of exposures and outcomes and what covariates to collect ; (6) the need for transparent reporting; (7) temporal and geographical aspects to be considered when ascertaining outcomes in COVID-19 patients, and (8) the need for rapid assessment. The article Biases in evaluating the safety and effectiveness of drugs for covid-19: designing real-world evidence studies.(Am J Epidemiol. 2021;190(8):1452-6) reviews and illustrates how immortal time bias and selection bias were present in several studies evaluating the effects of drugs on SARS-CoV-2 infection, and how they can be addressed. Although these two examples specifically refer to COVID-19 studies, such considerations are applicable to research questions with other types of exposures and outcomes.

COVID-19 pandemic-related disruptions in healthcare are likely to have impacted the design of current as well as future non-interventional, real-world studies. Changes in access to healthcare and healthcare seeking behavior during the pandemic will create and exacerbate the challenges inherent to observational studies when using real-world data from this period. The article Noninterventional studies in the COVID-19 era: methodological considerations for study design and analysis (J Clin Epidemiol. 2023;153:91-101) presents a general framework for supporting study design of non-interventional studies using real-world data from the COVID-19 era.

Finally, graphical frameworks for presenting study designs are increasingly recommended, to foster transparency, enhance understanding of the design, and support the evaluation of study protocols and the interpretation of study results, as illustrated in A Framework for Visualizing Study Designs and Data Observability in Electronic Health Record Data (Clin Epidemiol. 2022;14:601-8) and Visualizations throughout pharmacoepidemiology study planning, implementation, and reporting (Pharmacoepidemiol Drug Saf. 2022;31(11):1140-52).

4.2. Types of study design

This chapter briefly describes the main types of study designs. Specific aspects or applications of these designs are presented in Chapter 4.4. These designs are fully described in several textbooks cited in the Introduction, for example, *Modern Epidemiology* 4th ed. (T. Lash, T.J. VanderWeele, S. Haneuse, K. Rothman. Wolters Kluwer, 2020).

The choice of the study design should be primarily driven by the need to obtain valid evidence regarding the objective(s) of the study by mitigating the risk of selection bias, information bias and confounding (see Chapter 6). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available (Am J Epidemiol. 2016;183(8):758-64) has proposed target trial emulation as a strategy that uses existing tools and methods to formalise the design and analysis of observational studies. It stimulates investigators to identify potential sources of concerns and develop a design that best addresses these concerns and the risk of bias. Target trial emulation is described in Chapter 4.2.6. The increasing ability to use electronical data from routine healthcare systems has opened up new opportunities for investigators to conduct studies. Many investigators use the data source(s) they have access to and are familiar with in terms of potential bias, confounding and missing data.

4.2.1. Cohort studies

In a cohort study, the investigator identifies a population from which the study subjects will be identified, defines two or more groups of subjects (referred to as study cohorts) who are at risk for the outcome of interest and differ according to their exposure, and follows them over time to observe the occurrence of the outcome of interest in the exposed and unexposed cohorts. A cohort study may also include a single cohort that is heterogeneous with respect to exposure history, and occurrence of the outcome is measured and compared between exposure groups within the cohort. The amount of follow-up of each subject in the cohorts is counted and the total person-time experience serves as the denominator for the calculation of the incidence rate of the outcome of interest. Cohorts are called *fixed* when individuals may not move from one exposure group to the other. They are called *closed* when entry is not allowed after the cohort's inception. The population of a cohort may also be called *dynamic* (or open) if it can gain and lose members who contribute to the person-time experience for the duration of their presence in the cohort. The main advantages of a cohort study are the possibility to calculate directly interpretable incidence rates of an outcome and to investigate multiple outcomes for a given exposure. The cohort design is also well suited to studies using large electronic records (such as electronic healthcare records and administrative claim data) where individual data are collected over long periods of time, allowing to study the effect of drug exposures to outcomes

occurring later. Disadvantages are the need for a large sample size and possibly a long study duration to study rare outcomes, although use of existing electronic healthcare databases allow to retrospectively observe and analyse large cohorts (see Chapters 8.2 and 9).

Cohort studies are commonly used in pharmacoepidemiology to study the utilisation and effects of medicinal products. At the beginning of the COVID-19 pandemic, it was the design of choice to compare the risk and severity of SARS-CoV-2 infection in persons using or not certain types of medicines. An example is Renin-angiotensin system blockers and susceptibility to COVID-19: an international, open science, cohort analysis (Lancet Digit Health 2021;3(2):e98-e114) where electronic health records were used to identify and follow patients aged 18 years or older with at least one prescription for RAS blockers, calcium channel blockers, thiazide or thiazide-like diuretics. Four outcomes were assessed: COVID-19 diagnosis, hospital admission with COVID-19, hospital admission with pneumonia, and hospital admission with pneumonia, acute respiratory distress syndrome, acute kidney injury, or sepsis.

4.2.2. Case-control studies

In a case-control study, the investigator first identifies cases of the outcome of interest and establishes their exposure status, but the denominators (person-time of observation) to calculate their incidence rates are not measured. A referent (traditionally called "control") group without the outcome of interest is then sampled to estimate the relative distribution of the exposed and unexposed denominators in the source population from which the cases originate. Only the relative size of the incidence rates can therefore be calculated. Advantages of a case-control study include a computational efficiency far superior to the cohort design, the possibility to initiate a study based on a set of cases already identified (e.g., in a hospital) and the possibility to study rare outcomes and their association with multiple exposures or risk factors. One of the main difficulties of case-control studies is the appropriate selection of controls independently of exposure or other relevant risk factors in order to ensure that the distribution of exposure categories among controls is a valid representation of the distribution in the source population. Another disadvantage is the difficulty to study rare exposures, as a large sample of cases and controls would be needed to identify exposed groups large enough for the planned statistical analysis.

In order to increase the efficiency of exposure assessment in case-control studies, an alternative approach is a design in which the source population is a cohort. The nested case-control design includes all cases occurring in the cohort and a pre-specified number of controls randomly chosen from the population at risk at each time a case (or other relevant event) occurs. A case-cohort study includes all cases and a randomly selected sub-cohort from the population at risk. Advantages of such designs is to allow the conduct of a set of case-control studies from a single cohort and use efficiently electronic healthcare records databases where data on exposures and outcomes are already available.

The study Impact of vaccination on household transmission of SARS-COV-2 in England (N Engl J Med. 2021;385(8):759-60) is a nested case-control study where the cohort was defined by occurrence of a laboratory-confirmed COVID-19 case occurring in a household between 4 January 2021 to 28 February 2021. A 'cases' was defined as a secondary case occurring in the same household as a COVID-19 case and a 'control' was identified as a person without infection. Exposure was defined by the presence of a vaccinated COVID-19 case vs. an unvaccinated COVID-19 case in the same household with the restriction that the vaccinated COVID-19 case had to be vaccinated 21 days prior to being diagnosed. The statistical analysis calculated the odds ratios and 95% confidence intervals for household members becoming 'cases' if the COVID-19 case was vaccinated with 21 days or more before testing positive, vs. household members where the COVID-19 case was not vaccinated.

4.2.3. Case-only designs

4.2.3.1 General considerations

Although case-only designs are not considered as traditional study designs, they are increasingly used, and have been the topic of a large amount of methodological research. Case-only designs are designs in which cases are the only subjects; they reduce confounding by using the exposure and outcome history of each case as its own control, thereby eliminating confounding by characteristics that are constant over time, such as sex, socio-economic factors, genetic factors or chronic diseases. They are also best suited to studying transient exposures in relation to acute outcomes. Control yourself: ISPE-endorsed guidance in the application of self-controlled study designs in pharmacoepidemiology (Pharmacoepidemiol Drug Saf. 2021;30(6):671–84) proposes a common terminology to facilitate critical thinking in the design, analysis and review of studies, called by the authors 'Self-controlled Crossover Observational PharmacoEpidemiologic (SCOPE)' studies. These are split into outcome-anchored (case-crossover, case-time-control and case-case-time control) and exposure-anchored (self-controlled case series and self-controlled risk interval) that are suitable for slightly different research questions.

A simple form of a self-controlled design is the sequence symmetry analysis (initially described as prescription sequence symmetry analysis), introduced as a screening tool in Evidence of depression provoked by cardiovascular medication: a prescription sequence symmetry analysis (Epidemiology 1996;7(5):478-84). Hypothesis-free screening of large administrative databases for unsuspected drug-outcome associations (Eur J Epidemiol 2018;33(6):545-55) demonstrates how the sequence symmetry analysis can screen across a very wide range of exposures and outcomes.

4.2.3.2 Case-crossover design

The case-crossover (CCO) design compares the risk of exposure in a time period prior to an outcome, with that in an earlier reference time-period, or set of time periods, to examine the effect of transient exposures on acute events (see The Case-Crossover Design: A Method for Studying Transient Effects on the Risk of Acute Events, Am J Epidemiol 1991;133(2):144-53). The case-time-control design is a modification of the case-crossover design which use exposure history data from a traditional control group to estimate and adjust for the bias from temporal changes in prescribing (The case-time-control design, Epidemiology 1995;6(3):248-53). However, if not well matched, the case-time-control group may reintroduce selection bias (see Confounding and exposure trends in case-crossover and case-time-control designs, Epidemiology 1996;7(3):231-9). Methods have been suggested to overcome the exposure-trend bias while controlling for time-invariant confounders (see Future cases as present controls to adjust for exposure trend bias in case-only studies, Epidemiology 2011;22(4):568-74). Persistent User Bias in Case-Crossover Studies in Pharmacoepidemiology (Am J Epidemiol. 2016;184(10):761-9) demonstrates that case-crossover studies of medicines that may be used indefinitely are biased upward. This bias is alleviated, but not removed completely, by using a control group. Evaluation of the Case-Crossover (CCO) Study Design for Adverse Drug Event Detection (Drug Saf. 2017;40(9):789-98) showed that the CCO design adequately performs in studies of acute outcomes with abrupt onsets and exposures characterised as transient with immediate effects.

The self-controlled case-series design (SCCS) and the self-controlled risk interval (SCRI) method were initially developed more specifically for vaccine studies and include only cases with an exposure history, with the observation period for each case and each exposure divided into risk window(s) (e.g., number of days immediately following each exposure) and a control window (observed time outside this risk window).

4.2.3.3 Self-controlled case series

A good overview of the self-controlled case series (SCCS) is provided in Tutorial in biostatistics: the self-controlled case series method (Stat Med. 2006;25(10):1768-97), Self-controlled case series methods: an alternative to standard epidemiological study designs (BMJ. 2016; 354) and Investigating the assumptions of the self-controlled case series method (Stat Med. 2018;37(4):643-58).

SCCS estimate a relative incidence, that is, incidence rates within the risk window(s) after exposure relative to incidence rates within the control window(s). The SCCS design inherently controls for time-invariant and between-individual confounding, but potential confounders that vary over time e.g., confounding by indication, within the same persons still need to be controlled for.

Three assumptions of the SCCS are that 1) events arise independently within individuals (e.g., fractures do not affect the occurrence of a subsequent fracture), 2) events do not influence subsequent follow-up, and 3) the event itself does not affect the chance of being exposed. However, SCCS studies can be adapted to circumvent these assumptions in specific situations. The third assumption is generally the most limiting one, but where the event only temporarily affects the chance of exposure, additional 'pre-exposure' windows can be included; otherwise Cases series analysis for censored, perturbed, or curtailed post-event exposures (Biostatistics 2009;10(1):3-16) describes an extended SCCS method that can address permanent changes to the chance of exposure post-event where exposure windows are short, and is suitable where the event of interest is death.

Tutorial in biostatistics: the self-controlled case series method (Stat Med. 2006;25(10):1768-97) details how to fit SCCS models using standard statistical packages. The book Self-Controlled Case Series Studies: A Modelling Guide with R (P. Farrington, H. Whitaker, Y. G. Weldeselassie, 1st Edition, Chapman and Hall/CRC, 2021) provides a more detailed account. Examples from the tutorial and book are available from http://sccs-studies.info/.

An illustrative example of an SCCS study is Opioids and the Risk of Fracture: a Self-Controlled Case Series Study in the Clinical Practice Research Datalink (Am J Epidemiol. 2021;190(7):1324-31) where the relative incidence of fracture was estimated by comparing time windows when cases were exposed following an opioid prescription and unexposed to opioids. Multiple contiguous risk windows were included to capture changes in risk from new use through to long-term use. A washout window was included after prescriptions stopped, and a pre-exposure window was included to address potential bias from event-dependent exposure. Age, season and exposure to fracture risk–increasing drugs were adjusted for. SCCS assumptions were checked using sensitivity analyses, including taking first fractures only to address independence of events, and excluding individuals who died to address events influencing follow-up.

Use of the self-controlled case-series method in vaccine safety studies: review and recommendations for best practice (Epidemiol Infect. 2011;139(12):1805-17) assesses how the SCCS method has been used across 40 vaccine studies, highlights good practices, and provides guidance on how the method should be used and reported. Using several analytical approaches is recommended, as it can reinforce conclusions or shed light on possible sources of bias when these differ for different study designs. When should case-only designs be used for safety monitoring of medical products? (Pharmacoepidemiol Drug Saf 2012;21(Suppl. 1):50-61) compares the SCCS and case-crossover methods as to their use, strengths, and major differences (directionality). It concludes that case-only analyses of intermittent users complement the cohort analyses of prolonged users because their different biases compensate for one another. It also provides recommendations on when case-only designs should, and should not, be used for drug safety monitoring. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system (Drug Saf. 2013;36(Suppl. 1):S83-S93) evaluates the performance of the SCCS design using 399 drug-health outcome pairs in 5 observational databases and 6 simulated datasets to assess four outcomes and five design choices. The Use of active Comparators in self-controlled Designs (Am J Epidemiol.

2021;190(10):2181-7) showed that presence of confounding by indication can be mitigated by using an active comparator, using an empirical example of a study of the association between penicillin and venous thromboembolism (VTE), with roxithromycin, a macrolide antibiotic, as the comparator, and upper respiratory infection, a transient risk factor for VTE, representing time-dependent confounding by indication.

4.2.3.4. Self-controlled risk interval design

The self-controlled risk interval (SCRI) design is a restricted SCCS design suitable when exposure risk windows are short. Rather than using all follow-up time available, short control windows before and/or after risk windows are selected; gaps between risk and control windows may be included e.g., to allow for a washout period. Power may be reduced as compared with the SCCS, but will often suffice for use with large databases where events are not very rare. Since each individual's observation period is short, age and time effects often do not require control. In Use of FDA's Sentinel System to Quantify Seizure Risk Immediately Following New Ranolazine Exposure (Drug Saf. 2019;42(7):897-906), new users were restricted to patients with 32 days of continuous exposure to ranolazine (i.e., capturing individuals that typically would have a 30-day dispensing). The observation period began the day after the start of the incident ranolazine dispensing and ended on the 32nd day after the index date, with two risk windows covering days 1-10 and 11-20, and the control window days 21-32. The relative incidence is calculated as a ratio of the number of events in the risk interval to the number of events in the control interval multiplied by the ratio of the length of control interval to length of risk interval from only cases.

According to the Master Protocol: Assessment of Risk of Safety Outcomes Following COVID-19 Vaccination (bestinitiative.org) (2021), the standard SCCS design is more adaptable and is thus preferred when risk or control windows may be less well-defined, when there is a need to increase statistical power, or when unmeasured time-varying confounding is a lesser concern. The SCCS design can also be more easily used to assess multiple occurrences of independent events within an individual. The SCRI design is preferred when it is feasible to have strictly defined risk and control windows for outcomes of interest, or when time varying confounding is a concern. Despite the short observation periods, SCRI may be vulnerable to time-varying confounders; a means of adjustment in SCRI studies, e.g., for steep age effects sometimes seen in studies of childhood vaccine safety, is provided in Quantifying the impact of time-varying baseline risk adjustment in the self-controlled risk interval design (Pharmacoepidemiol Drug Saf. 2015;24(12):1304-12).

4.2.4. Cross-sectional studies

Cross-sectional studies are studies that seek to collect information on a study population at a specified time point without considering the relative timing of putative outcomes and exposures. Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations (Chest 2020;158(1S):S65-S71) provides recommendations for the conduct of such studies, as well as use cases.

The data collected at the time point may include both exposure and outcome data. In studies looking at the association between drug use and a clinical outcome, use of prevalent drug users (i.e., patients already treated for some time before study follow-up begins) can introduce two types of bias. Firstly, prevalent drug users are "survivors" of the early period of treatment, which can introduce substantial (selection) bias if the risk varies with time. Secondly, covariates relevant for drug use at the time of the entry (e.g., disease severity) may be affected by previous drug utilisation, or patients may differ regarding health-related behaviours (healthy user effect). No firm inference on a causal relationship can therefore be made from the results.

The study The incidence of cerebral venous thrombosis: a cross-sectional study (Stroke 2012;43(12):3375-7) was used to provide an estimate of the background incidence of cerebral sinus venous thrombosis (CSVT) in the context of the safety assessment of COVID-19 vaccines. Patients were identified from all 19 hospitals from two Dutch provinces using specific code lists. Review of medical records and case ascertainment were conducted to include only confirmed cases. Incidence was calculated using population figures from census data as the denominator.

4.2.5. Ecological studies and case-population studies

In ecological studies, populations are the unit of analysis, for example, comparing measures of a drug's utilisation across countries and correlating it with these countries' aggregate incidence rate of an outcome. Fundamentals of the ecological design are described in Ecologic studies in epidemiology: concepts, principles, and methods (Annu Rev Public Health 1995;16:61-81) and a 'tool box' is presented in Study design VI - Ecological studies (Evid Based Dent. 2006;7(4):108).

As illustrated in Control without separate controls: evaluation of vaccine safety using case-only methods (Vaccine 2004;22(15-16):2064-70), ecological analyses assume that a strong correlation between the trend in an indicator of an exposure (vaccine coverage in this example) and the trend in incidence of a disease (trends calculated over time or across geographical regions) is consistent with a causal relationship. Such comparisons at the population level may only generate hypotheses as they do not allow controlling for time-related confounding variables, such as age and seasonal factors. Moreover, they do not establish whether the outcome primarily occurred in the exposed individuals.

Case-population studies are a form of ecological studies where cases are compared to an aggregated comparator consisting of population data. The case-population study design: an analysis of its application in pharmacovigilance (Drug Saf. 2011;34(10):861-8) explains this design and its application in pharmacovigilance for signal generation and drug surveillance. The design is also explained in Chapter 2: Study designs in drug utilization research of the textbook *Drug Utilization Research - Methods and Applications* (M Elseviers, B Wettermark, AB Almarsdóttir, et al. Editors. Wiley Blackwell, 2016). An example is a multinational case-population study aiming to estimate population rates of a suspected adverse event using national sales data in Transplantation for Acute Liver Failure in Patients Exposed to NSAIDs or Paracetamol, Drug Saf. 2013;36(2):135–44. Based on the same study, Choice of the denominator in case population studies: event rates for registration for liver transplantation after exposure to NSAIDs in the SALT study in France (Pharmacoepidemiol Drug Saf. 2013;22(2):160-7) compared sales data and healthcare insurance data as denominators to estimate population exposure and found large differences in the event rates. Choosing the wrong denominator in case-population studies might generate erroneous results. The choice of the right denominator depends not only on a valid data source but will also depend on the hazard function of the adverse event.

The case-population approach has also been adapted for vaccine safety surveillance, in particular for prospective investigation of urgent vaccine safety concerns or for the prospective generation of vaccine safety signals (see Vaccine Case-Population: A New Method for Vaccine Safety Surveillance, Drug Saf. 2016 Dec;39(12):1197-1209).

A pragmatic approach towards case-population studies is recommended: in situations where nation-wide or region-wide electronic health records (EHRs) are available and allow assessing the outcomes and confounders with sufficient validity, a case-population approach is neither necessary nor desirable, as one can perform a population-based cohort or case-control study with adequate control for confounding. In situations where outcomes are difficult to ascertain in EHRs, or where such databases do not exist, the case-population design might give an approximation of the absolute and relative risk

when both events and exposures are rare. This is limited by the ecological nature of the reference data that restricts the ability to control for confounding.

Other forms of ecological studies include interrupted time-series analyses (see Chapter 4.3.3) and the case-coverage (ecological) design mainly used for vaccine monitoring (see Chapter 16.2).

4.2.6. Target trial emulation

4.2.6.1. General principles

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available (Am J Epidemiol. 2016;183(8):758-64) introduced target trial emulation in pharmacoepidemiology as a conceptual framework helping researchers to identify and avoid potential biases in observational studies. Target trial emulation is a strategy that uses existing tools and methods to formalise the design and analysis of such studies. It stimulates investigators to identify potential sources of concerns and develop a design that best addresses these concerns and minimises the risk of bias. The first step of the strategy is to design a hypothetical ideal randomised trial ("target trial") that would answer the research question. The target trial is described with regards to all design elements: the eligibility criteria, the treatment strategies, the assignment procedure, the follow-up, the outcome, the causal contrasts, and the analysis plan. In the second step, the researcher specifies how best to emulate the design elements of the target trial using the available observational data and considering analytic approaches given the trade-offs in an observational setting.

The target trial paradigm has been shown to prevent some common biases, such as immortal time bias or prevalent user bias while also identifying situations where adequate emulation may not be possible using the data at hand (see Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses, J Clin Epidemiol. 2016;79:70-5). Target Trial Emulation: A Framework for Causal Inference From Observational Data (JAMA. 2022;328(24):2446-7) stresses, however, that the lack of randomisation and blinding still requires high attention to the prevention and/or control of selection bias, information bias and confounding, as described in Chapter 6. Successful emulation of a target trial also requires proper definition of time zero, i.e., the start of follow-up in the observational data. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available (Am J Epidemiol. 2016;183(8) 758-64) describes two unbiased choices of time zero when eligibility criteria can be met at multiple times.

The need to explicitly describe the design elements that emulate the clinical trial provides transparency on the study design, the assumptions needed to emulate the trial, and the definition of causal effects, which also increases replicability of the study. The design of both the target trial and its emulation should be compared in a table, following the example of Emulating a Target Trial of Interventions Initiated During Pregnancy with Healthcare Databases: The Example of COVID-19 Vaccination (Epidemiology 2023;34(2):238-46).

Statistical aspects of target trials are discussed in Chapters 3.6 (The target trial) and 22 (Target trial emulation) of the Causal Inference Book (Hernán MA, Robins JM (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC).

4.2.6.2. Extensions of the approach

Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses (J Clin Epidemiol. 2016;79:70-5) gives recommendations on how to deal with more complex scenarios in target trial emulation. The problem of multiple eligible points zero for patients can either be resolved by random selection or by using them all by emulating a sequence of nested trials with

increasing time zero. Inverse probability weighting is proposed to estimate the per protocol effect of sustained treatment accounting for potential selection bias due to informative censoring.

A three-step method (cloning, censoring, weighting) has been proposed in How to estimate the effect of treatment duration on survival outcomes using observational data (BMJ. 2018;360: k182) to overcome bias in studies on the effect of treatment duration (and cumulative dose), that are often impaired by selection bias and to achieve better comparability with the treatment assignment performed in clinical trials. A clone-censor-weight approach is also recommended to deal with situations where individuals' data are consistent with several strategies.

Emulating a target trial in case-control designs: an application to statins and colorectal cancer (Int J Epidemiol. 2020;49(5):1637–46) describes how to emulate a target trial using case-control data and demonstrates that better emulation reduces the discrepancies between observational and randomised trial evidence.

4.2.6.3. Target trial emulation in causal inference studies

A causal inference study is a study designed to investigate, at the individual patient level, the causal effect of an exposure in comparison to non-exposure or to another exposure. In the context of pharmacoepidemiology, the exposure is generally a medical treatment, and the outcome of interest is generally a measure of its safety or effectiveness.

ENCePP recommends that, unless an alternative strategy is justified, target trial emulation should be considered for non-interventional causal inference studies to improve internal validity and increase transparency on definitions and assumptions.

Consideration of the estimand framework (as described in the ICH Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical principles for Clinical Trial, 2019) for the design of the hypothetical trial may provide additional coherence and transparency on definitions of exposures, endpoints, intercurrent events (ICEs), strategies to manage ICEs, approach to missing data and sensitivity analyses to be emulated in the observational study. In particular, the observational study may benefit from the formalised identification of the ICEs in the hypothetical trial.

4.2.6.4. Examples

In the context of the COVID-19 pandemic, several observational studies on vaccine effectiveness used target trial emulation. The observational study BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting (N Engl J Med. 2021;384(15):1412-23) emulated a target trial of the effect of the BNT162b2 vaccine on COVID-19 outcomes by matching vaccine recipients and controls on a daily basis on a wide range of potential confounding factors. The large population size of four large healthcare organisations led to a nearly perfect matching leading to a consistent pattern of similarity between the groups in the days just before day 12 after the first dose, the anticipated onset of the vaccine effect. A similar target trial emulation design was used in Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines in U.S. Veterans (N Engl J Med. 2022;386(2):105-15).

In the field of pregnancy epidemiology, Emulating a Target Trial of Interventions Initiated During Pregnancy with Healthcare Databases: The Example of COVID-19 Vaccination (Epidemiology 2023;34(2):238-46) describes a step-by-step specification of the protocol components of a target trial and their emulation including sensitivity analyses using negative controls to evaluate the presence of confounding and, alternatively to a cohort design, a case-crossover or case-time-control design to eliminate confounding by unmeasured time-fixed factors.

In oncology, The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening (Eur J Epidemiol. 2017;32(6):495-500) compared an observational analysis that explicitly emulated a target trial of screening colonoscopy with simpler

observational analyses that do not synchronise treatment assignment and eligibility determination at time zero and/or do not allow for repeated eligibility. This comparison suggests that the lack of an explicit emulation of the target trial leads to biased estimates and shows that allowing for repeated eligibility increases the statistical efficiency of the estimates.

4.2.6.5. Target trial emulation vs. replication of an existing RCT

It is important to distinguish between target trial emulation, i.e., the emulation of a hypothetical ideal RCT, and the replication of existing RCTs, which is sometimes also called emulation. The aim of target trial emulation is to use a framework to conduct a study that avoids common biases and to transparently describe its underlying assumptions and limitations. Replication studies of existing RCTs, however, try to come as close as possible to the results of the existing, non-ideal RCT, to prove the validity of the data and the study design.

Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials (JAMA 2023;329(16):1376-85) concludes that real-world evidence studies can reach similar conclusions as RCTs when design and measurements can be closely emulated, but this may be difficult to achieve. Concordance in results varied depending on the agreement metric. Emulation differences, chance, and residual confounding can contribute to divergence in results and are difficult to disentangle Several studies have compared the results of randomised clinical trials and of observational target trial emulations designed to ask similar questions. Comparing Effect Estimates in Randomized Trials and Observational Studies From the Same Population: An Application to Percutaneous Coronary Intervention (J Am Hear Assoc. 2021;10(11):e020357) highlighted differences between the two study designs that may affect the results and be generalisable to other types of interventions: the observational study conducted in the same registry as the registry used to recruit clinical trial patients needed to be performed in a period that preceded the clinical trial; eligibility criteria differed as not all the necessary data were available for the observational study and no exclusion was based on informed consent; some outcomes could not be defined similarly and some potential confounding factors could not be measured in the observational study.

Emulation differences versus biases when calibrating RWE findings against RCTs(Clin Pharmacol Ther. 2020;107(4):735-7) provides guidance on how to investigate and interpret differences in the estimates of treatment effect in the two study types. It is also emphasised that observational effectiveness studies should not aim at emulating RCTs but at investigating questions that cannot be answered by RCTs, as in cases where randomisation would be difficult or unethical.

4.2.7. Pragmatic trials and large simple trials

4.2.7.1 Pragmatic trials

Randomised controlled trials (RCTs) are considered the gold standard for demonstrating the efficacy of medicinal products and for obtaining an initial estimate of the risk of adverse outcomes. However, they are not necessarily indicative of the benefits, risks or comparative effectiveness of an intervention when used in clinical practice. The ADAPT-SMART Glossary defines pragmatic clinical trials (PCTs) as '*trials [that] examine interventions under circumstances that approach real-world practice, with more heterogeneous patient populations, possibly less-standardised treatment protocols, and delivery in routine clinical settings as opposed to a research environment*'.

The GetReal Trial Tool: design, assess and discuss clinical drug trials in light of Real World Evidence generation (J Clin Epidemiol. 2022;149:244-253) more broadly defines PCTs as '*methodologies which incorporate real-world elements into clinical trial design, maintaining randomisation*' and describes the GetReal Trial Tool, designed to assess the impact of design choices on generalisability to routine

clinical practice, while taking into account risk of bias, precision, acceptability and operational feasibility.

The book *Pragmatic Randomized Clinical Trials Using Primary Data Collection and Electronic Health Records* (1st Edition - April 8, 2021, Eds: Cynthia Girman, Mary Ritchey) addresses practical aspects and challenges of the design, implementation, and dissemination of PCTs. The publication Series: Pragmatic trials and real world evidence: Paper 1. Introduction (J Clin Epidemiol. 2017;88:7-13) describes the main characteristics of this design and the complex interplay between design options, feasibility, acceptability, validity, precision, and generalisability of the results, and the review Pragmatic Trials (N Engl J Med. 2016;375(5):454-63) discusses the context in which a pragmatic design is relevant, and its strengths and limitations based on examples. Pragmatic trials revisited: applicability is about individualization (J. Clin. Epidemiol. 2018;99:164-166) advocates for more patient-oriented pragmatic trials and suggests to 1) develop new study designs that focus on a single person, 2) incorporate patients' perspectives on their care, and 3) integrate clinical research and medical care.

PCTs are focused on evaluating benefits and risks of treatments in patient populations and settings that are more representative of routine clinical practice. To ensure generalisability, PCTs should represent the patients to whom the treatment will be applied, for instance, inclusion criteria may be broader (e.g., allowing co-morbidity, co-medication, wider age range) and the follow-up may be minimised and allow for treatment switching. Real-World Data and Randomised Controlled Trials: The Salford Lung Study (Adv Ther. 2020;37(3):977-997) and Monitoring safety in a phase III real-world effectiveness trial: use of novel methodology in the Salford Lung Study (Pharmacoepidemiol Drug Saf. 2017;26(3):344-352) describes the model of a phase III PCT where patients were enrolled through primary care practices using minimal exclusion criteria and without extensive diagnostic testing, and where potential safety events were captured through patients' electronic health records and triggered review by the specialist safety team.

Pragmatic explanatory continuum summary (PRECIS): a tool to help trial designers (CMAJ. 2009;180(10): E45-E57) is a tool to support pragmatic trial designs and help define and evaluate the degree of pragmatism. The Pragmatic–Explanatory Continuum Indicator Summary (PRECIS) tool has been further refined and now comprises nine domains each scored on a 5 point Likert scale ranging from very explanatory to very pragmatic with an exclusive focus on the issue of applicability (The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015;350: h2147). A checklist and additional guidance is provided in Improving the reporting of pragmatic trials: an extension of the CONSORT statement (BMJ. 2008;337 (a2390):1-8), and Good Clinical Practice Guidance and Pragmatic Clinical Trials: Balancing the Best of Both Worlds (Circulation 2016;133(9):872-80) discusses the application of Good Clinical Practice to pragmatic trials, and the use of additional data sources such as registries and electronic health records for "EHR-facilitated" PCTs.

Based on the evidence that costs and complexity of conducting randomised trials lead to more restrictive eligibility criteria and shorter durations of trials, and therefore reduce the generalisability and reliability of the evidence about the efficacy and safety of interventions, the article The Magic of Randomization versus the Myth of Real-World Evidence (N Engl J Med. 2020;382(7):674-678) proposes measures to remove practical obstacles to the conduct of randomised trials of appropriate size.

The BRACE CORONA study (Effect of Discontinuing vs Continuing Angiotensin-Converting Enzyme Inhibitors and Angiotensin II Receptor Blockers on Days Alive and Out of the Hospital in Patients Admitted With COVID-19: A Randomized Clinical Trial, JAMA. 2021;325(3):254-64) is a registry-based pragmatic trial that included patients hospitalised with COVID-19 who were taking ACEIs or ARBs prior to hospital admission, to determine whether discontinuation vs. continuation of these drugs affects the

number of days alive and out of the hospital. Patients with a suspected COVID-19 diagnosis were included in the registry and followed up until diagnosis confirmation and randomised to either discontinue or continue ACEI or ARB therapy for 30 days. There was no specific treatment modification beyond discontinuing or continuing use of ACEIs or ARBs, the study team provided oversight on drug replacement based on current treatment guidelines. Treatment adherence was assessed based on medical prescriptions recorded in electronic health records after discharge.

4.2.7.2 Large simple trials

Large simple trials are pragmatic clinical trials with minimal data collection narrowly focused on clearly defined outcomes important to patients as well as clinicians. Their large sample size provides adequate statistical power to detect even small differences in effects, the clinical relevance of which can subsequently be assessed. Additionally, large simple trials include a follow-up time that mimics routine clinical practice.

Large simple trials are particularly suited when an adverse event is very rare or has a delayed latency (with a large expected attrition rate), when the population exposed to the risk is heterogeneous (e.g., different indications and age groups), when several risks need to be assessed in the same trial or when many confounding factors need to be balanced between treatment groups. In these circumstances, the cost and complexity of a traditional RCT may outweigh its advantages and large simple trials can help keep the volume and complexity of data collection to a minimum.

Outcomes that are simple and objective can also be measured from the routine process of care using epidemiological follow-up methods, for example by using questionnaires or hospital discharge records. Classical examples of published large simple trials are An assessment of the safety of paediatric ibuprofen: a practitioner based randomised clinical trial (JAMA. 1995;279:929-33) and Comparative mortality associated with ziprasidone and olanzapine in real-world use among 18,154 patients with schizophrenia: The Zodiac Observational Study of Cardiac Outcomes (ZODIAC) (Am J Psychiatry 2011;168(2):193-201).

Note that the use of the term 'simple' in the expression 'Large simple trials' refers to data structure and not to data collection. It is used in relation to situations in which a small number of outcomes are measured. The term may therefore not adequately reflect the complexity of the studies undertaken, see for example Methods for the Watch the Spot Trial. A Pragmatic Trial of More- versus Less-Intensive Strategies for Active Surveillance of Small Pulmonary Nodules (Ann Am Thorac Soc 2019;16(12): 1567–1576)

4.2.7.3 Randomised database studies

Randomised database studies can be considered a special form of a large simple trial where patients included in the trial are enrolled from a healthcare system with electronic records. Eligible patients may be identified and flagged automatically by the software, with the opportunity of allowing comparison of included and non-included patients with respect to demographic characteristics and clinical history. Database screening or record linkage can be used to collect outcomes of interest otherwise assessed through the normal process of care. Patient recruitment, informed consent and proper documentation of patient information are hurdles that still need to be addressed in accordance with the applicable legislation for RCTs.

Randomised database studies attempt to combine the advantages of randomisation and observational database studies. These and other aspects of randomised database studies are discussed in The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials (Health Technol Assess. 2014;18(43):1-146) which illustrates the practical implementation of randomised studies in general practice databases. More recent work has been conducted to extend quality standards in the Consolidated Standards of

Reporting Trials (CONSORT) to also include database studies: CONSORT extension for the reporting of randomised controlled trials conducted using cohorts and routinely collected data (CONSORT-ROUTINE): checklist with explanation and elaboration (BMJ. 2021;373:n857). These quality standards for reporting also have implications on trial design and conduct.

Published examples of randomised database studies are still scarce, however, this design is becoming more common with the increasing use of electronic health records. Pragmatic randomised trials using routine electronic health records: putting them to the test (BMJ. 2012;344:e55) describes a project to implement randomised trials in the everyday clinical work of general practitioners, comparing treatments that are already in common use, and using routinely collected electronic healthcare records both to identify participants and to gather results. The above-mentioned Salford Lung Study, and the study described in Design of a pragmatic clinical trial embedded in the Electronic Health Record: The VA's Diuretic Comparison Project (Contemp Clin Trials 2022, 116:106754) belong to this category.

A particular form of randomised database studies is the registry-based randomised trial, which uses an existing registry as a source for the identification of cases, their randomisation and their follow-up. The editorial The randomized registry trial - the next disruptive technology in clinical research? (N Engl J Med. 2013;369(17):1579-81) introduces this concept. This hybrid design aims at achieving both internal and external validity by performing a RCT in a data source with higher generalisability (such as registries). Other examples are the TASTE trial that followed patients in the long-term using data from a Scandinavian registry (Thrombus aspiration during ST-segment elevation myocardial infarction (N Engl J Med. 2013;369:1587-97) and A registry-based randomized trial comparing radial and femoral approaches in women undergoing percutaneous coronary intervention: the SAFE-PCI for Women (Study of Access Site for Enhancement of PCI for Women) trial (JACC Cardiovasc Interv. 2014;7:857-67).

The importance of large simple trials has been highlighted by their role in evaluating well-established products that were repurposed for the treatment of COVID-19. The PRINCIPLE Trial platform (for trials in primary care) and the RECOVERY Trial platform (for trials in hospitals) have been recruiting large numbers of study participants and sites within short periods of time. In addition to brief case report forms, important clinical outcomes such as death, intensive care admission and ventilation were ascertained through data linkage to existing data streams. The study Lopinavir-ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial (Lancet 2020;396:1345–52) found that in patients admitted to hospital with COVID-19, lopinavir–ritonavir was not associated with reductions in 28-day mortality, duration of hospital stay, or risk of progressing to invasive mechanical ventilation or death. On the other hand, in Dexamethasone in Hospitalized Patients with Covid-19 (N Engl J Med. 2021;384(8):693-704), the RECOVERY trial also reported that the use of dexamethasone resulted in lower 28-day mortality in patients who were receiving either invasive mechanical ventilation or oxygen alone at randomisation. Inhaled budesonide for COVID-19 in people at high risk of complications in the community in the UK (PRINCIPLE): a randomised, controlled, open-label, adaptive platform trial (Lancet 2021;398:843-55) reported on the effectiveness of an inhaled corticosteroid for COVID-19 community patients. The streamlined and reusable approaches in data collection in these still recruiting platform trials clearly were essential in the achievements to enrol larger numbers of trial participants and evaluate multiple treatments rapidly.

4.3. Specific aspects of study design

4.3.1. Positive and negative control exposures and outcomes

The validity of causal associations may be tested by using control exposures or outcomes. A negative control outcome is a variable known not to be causally affected by the treatment of interest. Likewise, a negative control exposure is a variable known not to causally affect the outcome of interest. Conversely, a positive control outcome is a variable that is understood to be positively associated with the exposure of interest and a positive control exposure is one which is known to increase the risk of the outcome of interest.

Well-selected positive and negative controls support decision-making on whether the data at hand correctly support the study results for known associations or correctly demonstrate lack of association. Positive controls with negative findings and negative controls with positive findings may signal the presence of bias, as illustrated in Utilization of Positive and Negative Controls to Examine Comorbid Associations in Observational Database Studies (Med Care 2017;55(3):244-51). This general principle, with additional examples, is described in Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies (Epidemiology 2010 May; 21(3): 383–388.) and Control Outcomes and Exposures for Improving Internal Validity of Nonrandomized Studies (Health Serv Res. 2015;50(5):1432-51). Negative controls have also been used to identify other sources of bias including selection bias and measurement bias in Brief Report: Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies (Epidemiology. 2016 Sep; 27(5): 637–641) and in Negative control exposure studies in the presence of measurement error: implications for attempted effect estimate calibration (Int J Epidemiol. 2018 Apr; 47(2): 587–596). The use of negative and positive controls has therefore been recommended as a diagnostic test to evaluate whether the study design produced valid results. Practical considerations for their selection are provided in Chapter 18. Method Validity of *The Book of OHDSI* (2021).

Selecting drug-event combinations as reliable controls nevertheless poses important challenges: it is difficult to establish for negative controls proof of absence of an association, and it is still more problematic to select positive controls because it is desirable not only to measure an association but also an accurate estimate of the effect size. This has led to attempts to establish libraries of controls that can be used to characterise the performance of different observational datasets in detecting various types of associations using a number of different study designs. Although the methods used to identify negative and positive controls may be questioned according to Evidence of Misclassification of Drug-Event Associations Classified as Gold Standard 'Negative Controls' by the Observational Medical Outcomes Partnership (OMOP) (Drug Saf. 2016;39(5):421-32), this approach may allow to separate random and systematic errors in epidemiological studies, providing a context for evaluating uncertainty surrounding effect estimates.

Beyond the detection of bias, positive and negative controls can be used to correct unmeasured confounding, such as through empirical calibration on p-values or confidence intervals, as described in Interpreting observational studies: Why empirical calibration is needed to correct p-values (Stat Med. 2014;33(2):209-18), Robust empirical calibration of p-values using observational data (Stat Med. 2016;35(22):3883-8), Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data (Proc Natl Acad Sci. USA 2018;115(11): 571-7).

The empirical calibration approach has been used in both the case-based study design (Empirical assessment of case-based methods for identification of drugs associated with acute liver injury in the French National Healthcare System database (SNDS), Pharmacoepidemiol Drug Saf. 2021;30(3):320-33) and the cohort design (Risk of depression, suicide and psychosis with hydroxychloroquine treatment for rheumatoid arthritis: a multinational network cohort study, Rheumatology (Oxford)

2021;60:3222-34). While this method may reduce the number of false positive results, it may also reduce the ability to detect a true safety or efficacy signal and is computationally expensive, as suggested in Limitations of empirical calibration of p-values using observational data (Stat Med. 2016;35(22):3869-82) and Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data (Proc Natl Acad Sci. USA 2018;115(11): 571-7).

An Overview of key negative controls techniques has been published by the Duke-Margolis Center for Health Policy, providing a brief description of key assumptions, strengths and limitations of using negative controls (Duke-Margolis/ FDA workshop on Understanding the Use of Negative Controls to Assess the Validity of Non-Interventional Studies of Treatment Using Real-World Evidence, March 8, 2023).

4.3.2. Use of an active comparator

The main purpose of using an active comparator is to reduce confounding by indication and by disease severity. Its use is optimal in the context of the new user design (see Chapter 6.1.1), where patients with the same indication initiating different treatments are compared, as described in The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application (Curr Epidemiol Rep. 2015;2(4):221-8). Active comparators implicitly restrict comparisons to patients with an indication for treatment who are actually receiving treatment. Therefore, use of an active comparator not only reduces confounding by indication, but also confounding by frailty and healthy user bias.

Active-comparator design and new-user design in observational studies (Nat Rev Rheumatol. 2015;11:437-41) points out that active comparator studies give insight how safe/effective a certain therapy is, compared to a therapeutic alternative, which is usually the more meaningful research question. Ideally, an active comparator should be interchangeable with the therapy of interest, and represent the counterfactual risk of a given event with the therapeutic alternative. This means that the active comparator should be indicated for the same disease and disease severity and have the same absolute or relative exclusion criteria. The active comparator represents the background risk in the diseased and should be known to have no effect on the event(s) of interest or competing events. If the effect of the active comparator is unknown, multiple comparators, including non-users, should be used.

Identification of an active comparator should be based on clinician input and respective guidelines, acceptability of its use within the chosen data source should be verified and the balance in patient characteristics should be reviewed as described in Core concepts in pharmacoepidemiology: Confounding by indication and the role of active comparators (Pharmacoepidemiol Drug Saf. 2022;31(3):261-269). In situations where an acceptable active comparator is lacking, such as due to unavailability of a therapeutic alternative, extensive channeling or reimbursement restrictions, the validity of the planned study needs to be assessed. Alternative methods to reduce confounding by indication, such as use of inactive comparators and alternative approaches such as methods based on propensity scores and instrumental variable analysis to balance patients' characteristics, should be considered.

4.3.3. Interrupted time series analyses and Difference-in-Differences method

In evaluating the effectiveness of population-level interventions that are implemented at a specific point in time (with clearly defined before-after periods, such as policy effect date, regulatory action date) interrupted time series (ITS) studies are becoming the standard approach. ITS, a quasi-experimental design to evaluate the longitudinal effects of interventions through regression modelling, establishes the expected pre-intervention trend for an outcome of interest. The counterfactual scenario

in the absence of the intervention serves as the comparator, the expected trend that provides a comparison for the evaluation of the impact of the intervention by examining any change occurring following the intervention period (Interrupted time series regression for the evaluation of public health interventions: a tutorial, Int J Epidemiol. 2017;46:348-55).

ITS analysis requires that several assumptions are met and its implementation is technically sophisticated, as explained in Regression based quasi-experimental approach when randomisation is not an option: Interrupted time series analysis (BMJ. 2015; 350:h2750). The use of ITS regression in pharmacovigilance impact research is illustrated in Chapter 16.4.

When data on exposed and control populations are available, Difference-in-Differences (DiD) methods are sometimes preferable. These methods compare the outcome mean or trend for exposed and control groups before and after a certain time point (usually indicating a treatment or intervention point), providing insight into the changes of the variable for the exposed population relative to the change in the negative outcome group. This approach can be a more robust approach to causal inference than ITS, by comparing the exposed group to a control group subject to the same time-varying factors. First, DiD takes the difference for both groups before and after the intervention; then it subtracts the difference of the control group from the exposed group to control for time-varying factors, thus estimating the clean impact of the intervention.

A basic introduction can be found in Impact evaluation using Difference-in-Differences (RAUSP Management Journal 2019;54:519-532) and further extensions, for example assessment of variation in treatment timing, in Difference-in-differences with variation in treatment timing (Journal of Econometrics 2021;225:254-77). A good overview applied to public health policy research is available in Designing Difference in Difference Studies: Best Practices for Public Health Policy Research (Annu Rev Public Health 2018;39:53-469). A recent review from the econometrics perspective discusses possible avenues when some core assumptions are violated and models with relaxed hypotheses are needed, and provides recommendations which can be applied to pharmacoepidemiology (What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature, J Econom. 2023;235(2); 2218-2244).

5. Definition and validation of drug exposure, outcomes and covariates

Note: except for minor text edits, Chapter 5 (formerly 4.3) has not been updated for Revision 11 of the Guide, as contents remain up-to-date.

Historically, pharmacoepidemiological studies relied on patient-reported information or paper-based health records. The rapid increase in access to electronic healthcare records and large administrative databases has changed the way exposures and outcomes are defined, measured and validated. All variables in secondary data sources should be defined with care taking into account the fact that information is often recorded for purposes other than pharmacoepidemiology. Secondary data originate mainly from four types of data sources: prescription data (e.g., UK CPRD primary care data), data on dispensing (e.g., PHARMO outpatient pharmacy database), data on payment for medication (namely claims data, e.g., IMS LifeLink PharMetrics Plus), data collected in surveys, and data from specific means of data collection (e.g., pregnancy registries, vaccine registries). Misclassification of exposure, outcome or any covariate, or incorrect categorisation of these variables, may lead to information bias (see Chapter 6).

5.1. Assessment of exposure

Exposure definitions can include simple dichotomous variables (e.g., ever vs. never exposed) or be more granular, including estimates of duration, exposure windows (e.g., current vs. past exposure) also referred to as risk periods, or dosage (e.g., current dosage, cumulative dosage over time). Consideration should be given to both the requirements of the study design and the availability of variables. Assumptions made when preparing drug exposure data for analysis have an impact on results: an unreported step in pharmacoepidemiological studies (Pharmacoepidemiol Drug Saf. 2018;27(7):781-8) demonstrates the effect of certain exposure assumptions on findings and provides a framework to report preparation of exposure data. The Methodology chapter of the book *Drug Utilization Research. Methods and Applications* (M. Elseviers, B. Wettermark, A.B. Almarsdottir et al. Ed. Wiley Blackwell, 2016) discusses different methods for data collection on drug utilisation.

The population included in these data sources follows a process of attrition: drugs that are prescribed are not necessarily dispensed, and drugs that are dispensed are not necessarily ingested. In Primary non-adherence in general practice: a Danish register study (Eur J Clin Pharmacol 2014;70(6):757-63), 9.3% of all prescriptions for new therapies were never redeemed at the pharmacy, with different percentages per therapeutic and patient groups. The attrition from dispensing to ingestion is even more difficult to measure, as it is compounded by uncertainties about which dispensed drugs are actually taken by the patients and the patients' ability to provide an accurate account of their intake.

5.2. Assessment of outcomes

A case definition compatible with the data source should be developed for each outcome of a study at the design stage. This description should include how events will be identified and classified as cases, whether cases will include prevalent as well as incident cases, exacerbations and second episodes (as differentiated from repeat codes) and all other inclusion or exclusion criteria. If feasible, prevalent cases should not be included. The reason for the data collection and the nature of the healthcare system that generated the data should also be described as they can impact on the quality of the available information and the presence of potential biases. Published case definitions of outcomes, such as those developed by the Brighton Collaboration in the context of vaccine studies, are useful but not necessarily compatible with the information available in observational data sources. For example, information on the onset or duration of symptoms, or clinical diagnostic procedures, may not be available.

Search criteria to identify outcomes should be defined and the list of codes and any used case finding algorithm should be provided. Generation of code lists requires expertise in both the coding system and the disease area. Researchers should consult clinicians who are familiar with the coding practice within the studied field. Suggested methodologies are available for some coding systems, as described in Creating medical and drug code lists to identify cases in primary care databases (Pharmacoepidemiol Drug Saf. 2009;18(8):704-7). Advances in Electronic Phenotyping: From Rule-Based Definitions to Machine Learning Models (Annu Rev Biomed Data Sci. 2018;1:53-68) reports on methods for phenotyping (finding subjects with specific conditions or outcomes) which are becoming more commonly used, particularly in multi-database studies (see Chapters 9 and 16.6). Care should be given when re-using a code list from another study as code lists depend on the study objective and methods. Public repository of codes such as Clinicalcodes.org are available and researchers are also encouraged to make their own set of coding available.

In some circumstances, chart review or free text entries in electronic format linked to coded entries can be useful for outcome identification or confirmation. Such identification may involve an algorithm with use of multiple code lists (for example disease plus therapy codes) or an endpoint committee to adjudicate available information against a case definition. In some cases, initial plausibility checks or

subsequent medical chart review will be necessary. When databases contain prescription data only, drug exposure may be used as a proxy for an outcome, or linkage to different databases is required. The accurate date of onset is particularly important for studies relying upon timing of exposure and outcome such as in the self-controlled designs (see Chapter 4.2.3).

5.3. Assessment of covariates

In pharmacoepidemiological studies, covariates use includes selecting and matching study subjects, comparing characteristics of the cohorts, developing propensity scores, creating stratification variables, evaluating effect modifiers and adjusting for confounders. Reliable assessment of covariates is therefore essential for the validity of results. A given database may or may not be suitable for studying a research question depending on the availability of information on these covariates.

Some patient characteristics and covariates vary with time and accurate assessment is therefore time dependent. The timing of assessment of the covariates is an important factor for the correct classification of the subjects and should be clearly reported. Capturing covariates can be done at one or multiple points during the study period. In the latter scenario, the variable will be modelled as time-dependent variable (See Chapter 4.3.3).

Assessment of covariates can be performed using different periods of time (look-back periods or run-in periods). Fixed look-back periods (for example 6 months or 1 year) can be appropriate when there are changes in coding methods or in practices or when using the entire medical history of a patient is not feasible. Estimation using all available covariates information versus a fixed look-back window for dichotomous covariates (Pharmacoepidemiol Drug Saf. 2013; 22(5):542-50) establishes that defining covariates based on all available historical data, rather than on data observed over a commonly shared fixed historical window will result in estimates with less bias. However, this approach may not always be applicable, for example when data from paediatric and adult periods are combined because covariates may significantly differ between paediatric and adult populations (e.g., height and weight).

5.4. Misclassification and validation

5.4.1. Misclassification

The validity of pharmacoepidemiological studies depends on the correct assessment of exposure, outcomes and confounders. Measurement errors, i.e., *misclassification* of binary or categorical variables or *mismeasurement* of continuous variables result in *information bias*. The effect of misclassification in the presence of covariates (Am J Epidemiol. 1980;112(4):564–9) shows that misclassification of a confounder results in incomplete control for confounding.

Misclassification of exposure is *non-differential* if the assessment of exposure does not depend on the true outcome status and misclassification of outcome is non-differential if the assessment of the outcome does not depend on exposure status. Misclassification of exposure and outcome is considered *dependent* if the factors that predict misclassification of exposure are expected to also predict misclassification of outcome.

Misconceptions About Misclassification: Non-Differential Misclassification Does Not Always Bias Results Toward the Null (Am J Epidemiol. 2022; kwac03) emphasises that bias towards the null is not always "conservative" but might mask important safety signals and discusses seven exceptions to the epidemiologic 'mantra' about non-differential misclassification bias resulting in estimates towards the null. One important exception is outcome measurement with perfect specificity which results in unbiased estimates of the risk ratio.

The influence of misclassification on the point estimate should be quantified or, if this is not possible, its impact on the interpretation of the results should be discussed. FDA's Quantitative Bias Analysis Methodology Development: Sequential Bias Adjustment for Outcome Misclassification (2017) proposes a method of adjustment when validation of the variable is complete. Use of the Positive Predictive Value to Correct for Disease Misclassification in Epidemiologic Studies (Am J Epidemiol. 1993;138(11):1007–15) proposes a method based on estimates of the positive predictive value which requires validation of a sample of patients with the outcome only, while assuming that sensitivity is non-differential and has been used in a web application (Outcome misclassification: Impact, usual practice in pharmacoepidemiological database studies and an online aid to correct biased estimates of risk ratio or cumulative incidence; Pharmacoepidemiol Drug Saf. 2020;29(11):1450-5) which allows correction of risk ratio or cumulative incidence point estimates and confidence intervals for bias due to outcome misclassification based on this methodology. The article Basic methods for sensitivity analysis of biases (Int J Epidemiol. 1996;25(6):1107-16) provides different examples of methods for examining the sensitivity of study results to biases, with a focus on methods that can be implemented without computer programming. Good practices for quantitative bias analysis (Int J Epidemiol. 2014;43(6):1969-85) advocates explicit and quantitative assessment of misclassification bias, including guidance on which biases to assess in each situation, what level of sophistication to use, and how to present the results.

5.4.2. Validation

Common misconceptions about validation studies (Int J Epidemiol. 2020;49(4): 1392-6) discusses important aspects on the design of validation studies. It stresses the importance of stratification on key variables (e.g., exposure in outcome validation) and shows that by sampling conditionally on the imperfectly classified measure (e.g., case as identified by the study algorithm), only the positive and negative predictive values can be validly estimated.

Most database studies will be subject to outcome misclassification to some degree, although case adjudication against an established case definition or a reference standard can remove false positives, while false negatives can be mitigated if a broad search algorithm is used. Validity of diagnostic coding within the General Practice Research Database: a systematic review (Br J Gen Pract. 2010:60:e128 36), the book *Pharmacoepidemiology* (B. Strom, S.E. Kimmel, S. Hennessy. 6th Edition, Wiley, 2012) and Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned (Pharmacoepidemiol Drug Saf. 2012;supp1:82 9) provide examples of validation. External validation against chart review or physician/patient questionnaire is possible in some instances but the questionnaires cannot always be considered as 'gold standard'. Misclassification of exposure should also be measured based on validation, as feasible.

Linkage validation can be used when another database is used for the validation through linkage methods (see Using linked electronic data to validate algorithms for health outcomes in administrative databases, J Comp Eff Res. 2015;4:359-66). In some situations, there is no access to a resource to provide data for comparison. In this case, indirect validation may be an option, as explained in the textbook *Applying quantitative bias analysis to epidemiologic data* (Lash T, Fox MP, Fink AK. Springer-Verlag, New-York, 2009).

Structural validation of the database with internal logic checks should also be performed to verify the completeness and accuracy of variables. For example, one can investigate whether an outcome was followed by (or proceeded from) appropriate exposure or procedures or if a certain variable has values within a known reasonable range.

While the positive predictive value is more easily measured than the negative predictive value, a low specificity is more damaging than a low sensitivity when considering bias in relative risk estimates (see A review of uses of health care utilization databases for epidemiologic research on therapeutics; J Clin Epidemiol. 2005;58(4):323-37).

For databases routinely used in research, documented validation of key variables may have been done previously by the data provider or other researchers. Any extrapolation of a previous validation study should however consider the effect of any differences in prevalence and inclusion and exclusion criteria, the distribution and analysis of risk factors as well as subsequent changes to health care, procedures and coding, as illustrated in Basic Methods for Sensitivity Analysis of Biases, (Int J Epidemiol. 1996;25(6):1107-16).

6. Methods to address bias and confounding

6.1. Bias

6.1.1. Selection bias

"*Selection biases are distortions that result from procedures used to select subject and from factors that determine study participation. The common element of such biases is that the relation between exposure and disease is different for those who participate and for all those who should have been theoretically eligible for study, including those who do not participate. Because estimates of effect are conditioned on participation, the associations observed in a study represent a mix of forces that determine participation*" (Greenland, Lash. Modern Epidemiology. 3rd edition). Lack of representativeness of exposure or outcome pattern alone is not sufficient to cause selection bias. Examples of common selection biases are prevalence bias, self-selection bias, and referral bias.

Prevalence bias may occur when prevalent drug users are included in an observational study, i.e., patients already taking a therapy for some time before study follow-up began. This can cause two types of bias. Firstly, prevalent users are 'survivors' (healthy-users) of the early period of pharmacotherapy, which can introduce substantial selection bias if the risk varies with time, as seen in safety studies with unwanted exclusion from a safety assessment of persons discontinuing treatments following early adverse reactions ('depletion of susceptibles'). An illustrative example is the comparison between users of third and older generations of oral contraceptives regarding the risk of venous thrombosis where the association for the third generation was initially overestimated due to the heathy user bias in persons taking older generation contraceptives (see The Transnational Study on Oral Contraceptives and the Health of Young Women. Methods, results, new analyses and the healthy user effect, Hum Reprod Update 1999;5(6):707-20). Secondly, covariates for drug use at study entry are often influenced by the previous intake of the drug.

Self-selection in epidemiological studies may introduce selection bias and influence the validity of study results.

Referral bias occurs when patients with an abnormal test result are referred to a medical specialist at a higher rate than are patients with normal test results. In Clinical implications of referral bias in the diagnostic performance of exercise testing for coronary artery disease (J Am Heart Assoc. 2013;2(6):e000505), it was shown that exercise echocardiography and myocardial perfusion imaging are considerably less sensitive and more specific for coronary artery disease after adjustment for referral bias.

The article Collider bias undermines our understanding of COVID-19 disease risk and severity (Nat Commun. 2020;11(1):5749) describes a selection bias where a variable (a collider) is influenced by

two other variables, for example when an exposure (being a healthcare worker) and an outcome (severity of COVID-19 infection) both affect the variable determining the likelihood of being sampled (presence of PCR testing or hospitalisation). A bias would arise when the analysis includes only those people who have experienced an event such as hospitalisation with COVID-19, been tested for active infection or who have volunteered their participation. Among hospitalised patients, the relationship between any exposure that relate to hospitalisation and the severity of infection would be distorted compared to the general population. The article proposes methods for detecting and minimising the effects of collider bias. Vaccine side-effects and SARS-CoV-2 infection after vaccination in users of the COVID Symptom Study app in the UK: a prospective observational study (Lancet Infect Dis. 2021;S1473) discusses that collider bias would occur in the study if both vaccination status and COVID-19 positivity influenced the probability of participation in the study. However, it is believed that collider bias was unlikely to underlie the reduction in infections following vaccination seen in the data given that strong reductions in COVID-19 hospitalisations after vaccination were observed in other nationwide studies.

Biases in evaluating the safety and effectiveness of drugs for covid-19: designing real-world evidence studies.(Am J Epidemiol. 2021;kwab028) illustrates selection bias present in several studies evaluating the effects of drugs on SARS-CoV-2 infection and how to address them at the analysis and design stages.

Mitigating selection bias at the analysis stage

Once they have occurred, selection biases cannot be removed at the analysis stage if the factors responsible for the selection are not known or not measured. In some circumstances, it may be possible to restrict the study population by including only groups where the selection did not operate. For example, a prevalence bias may be removed by restricting the analysis to incident drug users, i.e., patients enter the study cohort only at the start of the first course of the treatment of interest (or of different treatment groups) during the study period. Consequences may include reduced precision of estimates due to lower sample size and likely reduction in the number of patients with long-term exposure. In circumstances where the factors influencing the selection are known and have been accurately measured, they can be treated as confounding factors and adjusted for at the analysis stage.

Mitigating selection bias at the design stage

The impact of selection biases should therefore be best avoided or minimised with proper consideration at study design. The new user (incident user) design helps mitigate selection bias by alleviate healthy user bias for preventive treatments in some circumstances (see Healthy User and Related Biases in Observational Studies of Preventive Interventions: A Primer for Physicians. J Gen Intern Med 2011;26(5):546-50). The article Evaluating medication effects outside of clinical trials: new-user designs (Am J Epidemiol. 2003;158 (9):915–20) defines new user designs in cohort and case-control settings. The articles The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application (Curr Epidemiol Rep. 2015;2(4):221-28) and New-user designs with conditional propensity scores: a unified complement to the traditional active comparator new-user approach (Pharmacoepidemiol Drug Saf. 2017;26(4):469-7) extend the discussion to studies with active comparators. One should be aware of the difference between a new user (which requires absence of prior use of a given drug/drug class during a prespecified washout period) and a treatment-naïve user (which requires absence of prior treatment for a given indication). A treatment-naïve status may not be ascertainable in left-truncated data.

The active comparator new user design (see Chapter 4.3.2) would ideally compare two treatments that are marketed contemporaneously. However, a more common situation is where a recently marketed drug is compared with an older established alternative. For such situations, the article Prevalent new-

user cohort designs for comparative drug effect studies by time-conditional propensity scores (Pharmacoepidemiol Drug Saf. 2017;26(4):459-68) introduces a cohort design allowing identification of matched subjects using the comparator drug at the same point in the course of disease as the (newly marketed) drug of interest. The design utilises time-based and prescription-based exposure sets to compute time-dependent propensity scores of initiating the new drug.

Observational studies of treatment effectiveness: worthwhile or worthless? (Clin Epidemiol. 2018;11:35-42) discusses how researchers can mitigate the risk of bias in the cohort design by presenting a case of the comparative effectiveness of two antidiabetic treatments using data collected during routine clinical practice.

The use of case-only designs can also reduce selection bias if the statistical assumptions of the method are fulfilled (see Chapter 4.2.3).

6.1.2. Information bias

Information bias (misclassification) arises when incorrect information about either exposure or outcome or any covariates is collected in the study or if variables are incorrectly categorised. Different factors may cause information bias. Chapter 4.3. discusses errors in definition, measurement and classification of variables and how to address them. Errors may also occur in the study design and method for data collection. Examples are the recall bias occurring in case-control studies where cases and controls can have different recall of their past exposures (see Recall bias in epidemiologic studies (J Clin Epidemiol. 1990;43(1):87-9), as well as the protopathic bias and surveillance or detection bias which are described below.

Protopathic bias

Protopathic bias arises when the initiation of a drug (exposure) occurs in response to a symptom of the (at this point undiagnosed) disease under study (outcome). For example, use of analgesics in response to pain caused by an undiagnosed tumour might lead to the erroneous conclusion that the analgesic caused the tumour. Protopathic bias, also called reverse causation, thus reflects a reversal of cause and effect (see Bias: Considerations for research practice. Am J Health Syst Pharm 2008;65(22):2159-68). This is particularly a problem in studies of drug-cancer associations and other outcomes with long latencies (see Cancer Incidence after Initiation of Antimuscarinic Medications for Overactive Bladder in the United Kingdom: Evidence for Protopathic Bias, Pharmacotherapy. 2017;37(6):673-83.)

Protopathic bias has also been described as a selection bias and it should not be confused with confounding by indication, i.e., when a variable is a risk factor for a disease among non-exposed subjects and is associated with the exposure of interest in the population from which the cases derive, without being an intermediate step in the causal pathway between the exposure and the disease (see Confounding by Indication: An Example of Variation in the Use of Epidemiologic Terminology, Am J Epidemiol. 1999;149(11):981-3).

Mitigating protopathic bias at the analysis stage

Protopathic bias may be handled by including a time-lag, (i.e., by disregarding all exposure during a specified period of time before the index date) or by restricting the analysis to cases in which the absence of relation of start of treatment to the symptoms of the outcomes is documented. Both of these methods are used in Long-Term Risk of Skin Cancer and Lymphoma in Users of Topical Tacrolimus and Pimecrolimus: Final Results from the Extension of the Cohort Study Protopic Joint European Longitudinal Lymphoma and Skin Cancer Evaluation (JOELLE) (Clin Epidemiol. 2021;13:1141-53).

Surveillance bias (or detection bias)

Surveillance or detection bias arises when patients in one exposure group have a higher probability of having the study outcome detected, due to increased surveillance, screening or testing of the outcome itself, or because of an associated symptom. For example, post-menopausal exposure to oestrogen is associated with an increased risk of bleeding that can trigger screening for endometrial cancers, leading to a higher probability of early stage endometrial cancers being detected. Any association between oestrogen exposure and endometrial cancer potentially overestimates risk, because unexposed patients with sub-clinical cancers would have a lower probability of their cancer being diagnosed or recorded. This is discussed in Alternative analytic methods for case-control studies of estrogens and endometrial cancer (N Engl J Med 1978;299(20):1089-94).

Mitigating surveillance bias at the design stage

This non-random type of misclassification bias can be reduced by selecting an unexposed comparator group with a similar likelihood of screening or testing, selecting outcomes that are likely to be diagnosed equally in both exposure groups, or by adjusting for the surveillance rate in the analysis. These issues and recommendations are outlined in Surveillance Bias in Outcomes Reporting (JAMA 2011;305(23):2462-3).

6.1.3. Time-related bias

Immortal time bias

Immortal time bias refers to a period of cohort follow-up time during which death (or an outcome that determines end of follow-up) cannot occur (K. Rothman, S. Greenland, T. Lash. *Modern Epidemiology*, 3rd Edition, Lippincott Williams & Wilkins, 2008).

Immortal time bias can arise when the period between cohort entry and date of first exposure to a drug, during which the event of interest has not occurred, is either misclassified or simply excluded and not accounted for in the analysis. Immortal time bias in observational studies of drug effects (Pharmacoepidemiol Drug Saf. 2007;16(3):241-9) demonstrates how several observational studies used a flawed approach to design and data analysis, leading to immortal time bias, which can generate an illusion of treatment effectiveness. This is frequently found in studies that compare groups of 'users' against 'non-users'. Observational studies with surprisingly beneficial drug effects should therefore be re-assessed to account for this type of bias.

Immortal Time Bias in Pharmacoepidemiology (Am J Epidemiol 2008;167(4):492-9) describes various cohort study designs leading to this bias, quantifies its magnitude under different survival distributions, illustrated with data from a cohort of lung cancer patients. For time-based, event-based and exposure-based cohort definitions, the bias in the rate ratio resulting from misclassified or excluded immortal time increases to the duration of immortal time. It is asserted that immortal time bias arises by conditioning on future exposure and that it can be avoided by analysing the data as if the exposures and outcomes were included as they developed, without ever looking into the future. Biases in evaluating the safety and effectiveness of drugs for covid-19: designing real-world evidence studies.(Am J Epidemiol. 2021; 190(8):1452-6) illustrates immortal time bias present in several studies evaluating the effects of drugs on SARS-CoV-2 infection.

Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods (Am J Epidemiol 2005;162(10):1016-23) describes five different approaches to deal with immortal time bias. The use of a time-dependent approach had several advantages: no subjects are excluded from the analysis and the study allows effect estimation at any point in time after discharge. However, changes of exposure might be predictive of the study endpoint and need

adjustment for time-varying confounders using complex methods. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes (BMJ. 2010; 340:b5087) describes how immortal time in observational studies can bias the results in favour of the treatment group and how it can be identified and avoided. It is recommended that all cohort studies should be assessed for the presence of immortal time bias using appropriate validity criteria. However, Re. 'Immortal time bias in pharmacoepidemiology' (Am J Epidemiol 2009;170(5):667-8) argues that sound efforts at minimising the influence of more common biases should not be sacrificed to that of avoiding immortal time bias.

Emulating Target Trials to Avoid Immortal Time Bias - An Application to Antibiotic Initiation and Preterm Delivery - PubMed (nih.gov) (Epidemiology. 2023;34(3):430-438) describes how a sequence of target trial emulations (see Chapter 4.2.6) can be used to avoid immortal time bias in situations where only few individuals start treatment at a certain time point.

Other forms of time-related bias

In many database studies, drugs administered during hospitalisations are unknown. Exposure misclassification bias may occur with a direction depending on whether exposure to drugs prescribed preceding hospitalisations are continued or discontinued and if days of hospitalisation are considered as gaps of exposure, especially when several exposure categories are assigned, such as current, recent and past. The differential bias arising from the lack of information on (or lack of consideration of) hospitalisations that occur during the observation period (called 'immeasurable time bias' in Immeasurable time bias in observational studies of drug effects on mortality. Am J Epidemiol. 2008;168(3):329-35) can be particularly problematic when studying serious chronic diseases that require extensive medication use and multiple hospitalisations.

In case-control studies assessing chronic diseases with multiple hospitalisations and in-patient treatment (such as the use of inhaled corticosteroids and death in chronic obstructive pulmonary disease patients), no clearly valid approach to data analysis can fully circumvent this bias. However, sensitivity analyses such as restricting the analysis to non-hospitalised patients or providing estimates weighted by exposable time may provide additional information on the potential impact of this bias, as also shown in Immeasurable time bias in observational studies of drug effects on mortality. (Am J Epidemiol. 2008;168(3):329-35).

In cohort studies where a first-line therapy (such as metformin) has been compared with second- or third-line therapies, patients are unlikely to be at the same stage of the disease (e.g., diabetes), which can induce confounding of the association with an outcome (e.g., cancer incidence) by disease duration. An outcome related to the first-line therapy may also be attributed to the second-line therapy if it occurs after a long period of exposure. Such situation requires matching on disease duration and consideration of latency time windows in the analysis (example drawn from Metformin and the Risk of Cancer. Time-related biases in observational studies. Diabetes Care 2012;35(12):2665-73).

Time-related biases in pharmacoepidemiology (Drug Saf. 2020;29(9):1101-10) further discusses several time-related biases and illustrates their impact on the effects of different COPD treatments on lung cancer, acute myocardial infarction and mortality outcomes, in studies using electronic healthcare databases. Protopathic, latency, immortal time, time-window, depletion of susceptibles, and immeasurable time biases were shown to significantly impact the effects of the study drugs on the outcomes.

Mitigating time-related bias at the design stage

Immortal time bias and other time-related biases such as prevalent bias can be avoided by emulation of a target trial, as this approach aligns assessment of eligibility and baseline information with start of follow-up (see Chapter 4.2.6).

6.2. Confounding

Confounding occurs when the estimate of measure of association is distorted by the presence of another risk factor. For a variable to be a confounder, it must be associated with both the exposure and the outcome, without being in the causal pathway.

6.2.1. Confounding by indication

Confounding by indication refers to a determinant of the outcome parameter that is present in people at perceived high risk or poor prognosis and is an indication for intervention. This means that differences in care between the exposed and non-exposed, for example, may partly originate from differences in indication for medical intervention such as the presence of specific risk factors for health problems. Another name for this type of confounding is 'channeling'. Confounding by severity is a type of confounding by indication, where not only the disease but its severity acts as confounder (see Confounding by Indication: An Example of Variation in the Use of Epidemiologic Terminology, Am J Epidemiol. 1999;149(11):981-3).

This type of confounding has frequently been reported in studies evaluating the efficacy of pharmaceutical interventions and is almost always encountered in various extents in pharmacoepidemiological studies. A good example can be found in Confounding and indication for treatment in evaluation of drug treatment for hypertension (BMJ. 1997;315:1151-4).

With the more recent application of pharmacoepidemiological methods to assess effectiveness, confounding by indication is a greater challenge and the article Approaches to combat with confounding by indication in observational studies of intended drug effects (Pharmacoepidemiol Drug Saf. 2003;12(7):551-8) focusses on its possible reduction in studies of intended effects. An extensive review of these and other methodological approaches discussing their strengths and limitations is discussed in Methods to assess intended effects of drug treatment in observational studies are reviewed (J Clin Epidemiol. 2004;57(12):1223-31).

An example of how results from a sensitivity analysis can differ from the main analysis and point towards confounding by indication is presented in First-dose ChAdOx1 and BNT162b2 COVID-19 vaccines and thrombocytopenic, thromboembolic and hemorrhagic events in Scotland (Nat Med. 2021; 27(7):1290-7), where the authors highlight the possibility of residual confounding by indication and perform a *post-hoc* self-controlled case series to adjust for time-invariant confounders.

6.2.2. Unmeasured confounding

Complete adjustment for confounders would require detailed information on clinical parameters, lifestyle or over-the-counter medications, which are often not measured in electronic healthcare records, causing residual confounding bias. Using directed acyclic graphs to detect limitations of traditional regression in longitudinal studies (Int J Public Health 2010;55(6):701-3) reviews confounding and intermediate effects in longitudinal data and introduces causal graphs to understand the relationships between the variables in an epidemiological study.

Unmeasured confounding can be adjusted for only through randomisation. When this is not possible, as most often in pharmacoepidemiological studies, the potential impact of residual confounding on the results should be estimated and considered in the discussion.

Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics (Pharmacoepidemiol Drug Saf. 2006;15(5):291-303) provides a systematic approach to sensitivity analyses to investigate the impact of residual confounding in pharmacoepidemiological studies that use healthcare utilisation databases. In this article, four basic

approaches to sensitivity analysis were identified: (1) sensitivity analyses based on an array of informed assumptions; (2) analyses to identify the strength of residual confounding that would be necessary to explain an observed drug-outcome association; (3) external adjustment of a drug-outcome association given additional information on single binary confounders from survey data using algebraic solutions; (4) external adjustment considering the joint distribution of multiple confounders of any distribution from external sources of information using propensity score calibration. The paper concludes that sensitivity analyses and external adjustments can improve our understanding of the effects of drugs in epidemiological database studies. With the availability of easy-to-apply spreadsheets (e.g., at https://www.drugepi.org/dope/software#Sensitivity), sensitivity analyses should be used more frequently, substituting qualitative discussions of residual confounding.

The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study (Am J Epidemiol. 2007;166(6):646–55) considers the extent and patterns of bias in estimates of exposure-outcome associations that can result from residual or unmeasured confounding, when there is no true association between the exposure and the outcome. Another important finding of this study was that when confounding factors (measured or unmeasured) are interrelated (e.g., in situations of confounding by indication), adjustment for a few factors can almost completely eliminate confounding.

6.2.3. Methods to address confounding

Methods to address confounding include case-only designs (see Chapter 4.2.3) and use of an active comparator (see Chapter 4.3.2). Other methods are detailed hereafter.

6.2.3.1. Disease risk scores

An approach to controlling for a large number of confounding variables is to summarise them in a single multivariable confounder score. Stratification by a multivariate confounder score (Am J Epidemiol. 1976;104(6):609-20) shows how control for confounding may be based on stratification by the score. An example is a disease risk score (DRS) that estimates the probability or rate of disease occurrence conditional on being unexposed. The association between exposure and disease is then estimated with adjustment for the disease risk score in place of the individual covariates.

DRSs are however difficult to estimate if outcomes are rare. Use of disease risk scores in pharmacoepidemiologic studies (Stat Methods Med Res. 2009;18(1):67-80) includes a detailed description of their construction and use, a summary of simulation studies comparing their performance to traditional models, a comparison of their utility with that of propensity scores, and some further topics for future research. Disease risk score as a confounder summary method: systematic review and recommendations (Pharmacoepidemiol Drug Saf. 2013;22(2);122-29), examines trends in the use and application of DRS as a confounder summary method and shows that large variation exists with differences in terminology and methods used.

In Role of disease risk scores in comparative effectiveness research with emerging therapies (Pharmacoepidemiol Drug Saf. 2012;21 Suppl 2:138–47), it is argued that DRS may have a place when studying drugs that are recently introduced to the market. In such situations, as characteristics of users change rapidly, exposure propensity scores may prove highly unstable. DRSs based mostly on biological associations would be more stable. However, DRS models are still sensitive to misspecification as discussed in Adjusting for Confounding in Early Postlaunch Settings: Going Beyond Logistic Regression Models (Epidemiology 2016;27(1):133-42).

6.2.3.2. Propensity scores

Databases used in pharmacoepidemiological studies often include records of prescribed medications and encounters with medical care providers, from which one can construct surrogate measures for

both drug exposure and covariates that are potential confounders. It is often possible to track day-by-day changes in these variables. However, while this information can be critical for study success, its volume can pose challenges for statistical analysis.

A propensity score (PS) is analogous to the disease risk score in that it combines a large number of possible confounders into a single variable (the score). The exposure propensity score (EPS) is the conditional probability of exposure to a treatment given observed covariates. In a cohort study, matching or stratifying treated and comparison subjects on EPS tends to balance all of the observed covariates. However, unlike random assignment of treatments, the propensity score may not balance unobserved covariates. Invited Commentary: Propensity Scores (Am J Epidemiol. 1999;150(4):327–33) reviews the uses and limitations of propensity scores and provide a brief outline of the associated statistical theory. The authors present results of adjustment by matching or stratification on the propensity score.

The estimated EPS summarises all measured confounders in a single variable and thus can be used in the analysis, as any other confounder, for matching, stratification, weighting or as a covariate in a regression model to adjust for the measured confounding. A description of these methods can be found in the following articles: An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies (Multivariate Behav Res. 2011;46(3):399-424), Tutorial and Case Study in Propensity Score Analysis: An Application to Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality (Multivariate Behav Res. 2011;46(1):119-51) and 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(28):3661-79).

Propensity score matching in cohort studies is frequently done 1:1, which, while allowing for selection of the best match for each member of the exposed cohort, may lead to severe depletion of the study population and the associated lower precision, especially when coupled with trimming. Increasing the matching ratio may increase precision but also negatively affect confounding control. One-to-many propensity score matching in cohort studies (Pharmacoepidemiol Drug Saf. 2012;21(S2):69-80) tests several methods for 1:n propensity score matching in simulation and empirical studies and recommends using a variable ratio that increases precision at a small cost of bias. Matching by propensity score in cohort studies with three treatment groups (Epidemiology 2013;24(3):401-9) develops and tests a 1:1:1 propensity score matching approach offering a way to compare three treatment options.

Use of EPS for stratification or weighing overcomes the precision-related limitation of matching-based methods, allowing use of a larger proportion of the study population in the analysis. Fine stratification approach is based on defining large number (50 or 100) number of EPS strata, as described in A Propensity-score-based Fine Stratification Approach for Confounding Adjustment When Exposure Is Infrequent (Epidemiology 2017;28(2):249-57).

High-dimensional Propensity Score Adjustment in Studies of Treatment Effects Using Healthcare Claims Data (Epidemiology 2009;20(4):512-22) discusses the high dimensional propensity score (hd-PS) model approach. It attempts to empirically identify large numbers of potential confounders in healthcare databases and, by doing so, to extract more information on confounders and proxies. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples (Am J Epidemiol. 2011;173(12):1404-13) evaluates the relative performance of hd-PS in smaller samples. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records (Pharmacoepidemiol Drug Saf. 2012;20(8):849-57) evaluates the use of hd-PS in a primary care electronic medical record database. In addition, the article Using high-dimensional propensity scores to automate confounding control in a distributed

medical product safety surveillance system (Pharmacoepidemiol Drug Saf. 2012;21(S1):41-9) summarises the application of this method for automating confounding control in sequential cohort studies as applied to safety monitoring systems using healthcare databases and also discusses the strengths and limitations of hd-PS. High-dimensional propensity scores for empirical covariate selection in secondary database studies: Planning, implementation, and reporting (Pharmacoepidemiol Drug Saf. 2023;32(2):93-106) provides an ISPE-endorsed overview of the hd-PS approach and recommendations on the planning, implementation, and reporting of hd-PS used for causal treatment-effect estimations in longitudinal healthcare databases. It contains a checklist with key considerations as a supportive decision tool to aid investigators in the implementation and transparent reporting of hd-PS techniques, and to aid decision-makers unfamiliar with hd-PS in the understanding and interpretation of studies using this approach.

The use of several measures of balance for developing an optimal propensity score model is described in Measuring balance and model selection in propensity score methods (Pharmacoepidemiol Drug Saf. 2011;20(11):1115-29) and further evaluated in Propensity score balance measures in pharmacoepidemiology: a simulation study (Pharmacoepidemiol Drug Saf. 2014;23(8):802-11). In most situations, the standardised difference performs best and is easy to calculate (see Balance measures for propensity score methods: a clinical example on beta-agonist use and the risk of myocardial infarction (Pharmacoepidemiol Drug Saf. 2011;20(11):1130-7) and Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review (J Clin Epidemiol 2015;68(2):112-21)). Metrics for covariate balance in cohort studies of causal effects (Stat Med 2013;33:1685-99) shows in a simulation study that the c-statistics of the PS model after matching and the general weighted difference perform as well as the standardized difference and are preferred when an overall summary measure of balance is requested. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution--a simulation study (Am J Epidemiol. 2010;172(7):843-54) demonstrates how 'trimming' of the propensity score eliminates subjects who are treated contrary to prediction and their exposed/unexposed counterparts, thereby reducing bias by unmeasured confounders.

Performance of propensity score calibration--a simulation study (Am J Epidemiol. 2007;165(10):1110-8) introduces 'propensity score calibration' (PSC). This technique combines propensity score matching methods with measurement error regression models to address confounding by variables unobserved in the main study. This is done by using additional covariate measurements observed in a validation study, which is often a subset of the main study.

Principles of variable selection for inclusion in EPS are described, for example, in Variable selection for propensity score models (Am J Epidemiol. 2006;163(12):1149-56) and in Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study (Pharmacoepidemiol Drug Saf. 2013;22(1):77-85).

Although in most situations, propensity score models, with the possible exception of hd-PS, do not have any advantages over conventional multivariate modelling in terms of adjustment for identified confounders, several other benefits may be derived. Propensity score methods may help to gain insight into determinants of treatment including age, frailty and comorbidity and to identify individuals treated against expectation. A statistical advantage of PS analyses is that if exposure is not infrequent it is possible to adjust for a large number of covariates even if outcomes are rare, a situation often encountered in drug safety research.

An important limitation of PS is that it is not directly amenable for case-control studies. A critical assessment of propensity scores is provided in Propensity scores: from naive enthusiasm to intuitive understanding (Stat Methods Med Res. 2012;21(3):273-93). Semiautomated and machine-learning based approaches to propensity score methods are currently being developed (Automated data-

adaptive analytics for electronic healthcare data to study causal treatment effects (Clin Epidemiol 2018;10:771-88).

6.2.3.3. Instrumental variables

An instrumental variable (IV) is defined in Instrumental variable methods in comparative safety and effectiveness research (Pharmacoepidemiol Drug Saf. 2010; 19(6):537-54) as a factor that is assumed to be related to treatment but is neither directly nor indirectly related to the study outcome. An IV should fulfil three assumptions: (1) it should affect treatment or be associated with treatment by sharing a common cause; (2) it should be a factor that is as good as randomly assigned so that it is unrelated to patient characteristics, and (3) it should be related to the outcome only through its association with treatment. This article also presents a practical guidance on IV analyses in pharmacoepidemiology. The article Instrumental variable methods for causal inference (Stat Med. 2014;33(13):2297-340) is a tutorial, including statistical code for performing IV analysis.

IV analysis is an approach to address uncontrolled confounding in comparative studies. An introduction to instrumental variables for epidemiologists (Int J Epidemiol. 2000;29(4):722-9) presents those developments, illustrated by an application of IV methods to non-parametric adjustment for non-compliance in randomised trials. The author mentions a number of caveats but concludes that IV corrections can be valuable in many situations. A review of IV analysis for observational comparative effectiveness studies suggested that in the large majority of studies, in which IV analysis was applied, one of the assumptions could be violated (Potential bias of instrumental variable analyses for observational comparative effectiveness research, Ann Intern Med. 2014;161(2):131-8).

The complexity of the issues associated with confounding by indication, channeling and selective prescribing is explored in Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable (Epidemiology 2006;17(3):268-75). A conventional, adjusted multivariable analysis showed a higher risk of gastrointestinal toxicity for selective COX-2-inhibitors than for traditional NSAIDs, which was at odds with results from clinical trials. However, a physician-level instrumental variable approach (a time-varying estimate of a physician's relative preference for a given drug, where at least two therapeutic alternatives exist) yielded evidence of a protective effect due to COX-2 exposure, particularly for shorter term drug exposures. Despite the potential benefits of physician-level IVs their performance can vary across databases and strongly depends on the definition of IV used as discussed in Evaluating different physician's prescribing preference based instrumental variables in two primary care databases: a study of inhaled long-acting beta2-agonist use and the risk of myocardial infarction (Pharmacoepidemiol Drug Saf. 2016;25 Suppl 1:132-41).

An important limitation of IV analysis is that weak instruments (small association between IV and exposure) lead to decreased statistical efficiency and biased IV estimates as detailed in Instrumental variables: application and limitations (Epidemiology 2006;17:260-7). For example, in the above mentioned study on non-selective NSAIDs and COX-2-inhibitors, the confidence intervals for IV estimates were in the order of five times wider than with conventional analysis. Performance of instrumental variable methods in cohort and nested case-control studies: a simulation study (Pharmacoepidemiol Drug Saf. 2014;23(2):165-77) demonstrates that a stronger IV-exposure association is needed in nested case-control studies compared to cohort studies in order to achieve the same bias reduction. Increasing the number of controls reduces this bias from IV analysis with relatively weak instruments.

Selecting on treatment: a pervasive form of bias in instrumental variable analyses (Am J Epidemiol. 2015;181(3):191-7) warns against bias in IV analysis by including only a subset of possible treatment options.

6.2.3.4. Prior event rate ratios

Another method proposed to control for unmeasured confounding is the Prior Event Rate Ratio (PERR) adjustment method, in which the effect of exposure is estimated using the ratio of rate ratios (RRs) between the exposed and unexposed from periods before and after initiation of a drug exposure, as discussed in Replicated studies of two randomized trials of angiotensin converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication (Pharmacoepidemiol Drug Saf. 2008;17(7):671-685). For example, when a new drug is launched, direct estimation of the drugs effect observed in the period after launch is potentially confounded. Differences in event rates in the period before the launch between future users and future non-users may provide a measure of the amount of confounding present. By dividing the effect estimate from the period after launch by the effect obtained in the period before launch, the confounding in the second period can be adjusted for. This method requires that confounding effects are constant over time, that there is no confounder-by-treatment interaction, and outcomes are non-lethal events.

Performance of prior event rate ratio adjustment method in pharmacoepidemiology: a simulation study (Pharmacoepidemiol Drug Saf. 2015(5);24:468-477) discusses that the PERR adjustment method can help to reduce bias as a result of unmeasured confounding in certain situations but that theoretical justification of assumptions should be provided.

6.2.3.5. Handling time-dependent confounding in the analysis

In longitudinal studies, the value of covariates may change and be measured over time. These covariates are time-dependent confounders if they are affected by prior treatment and predict the future treatment decision and future outcome conditional on the past treatment exposure (see Comparison of Statistical Approaches Dealing with Time-dependent Confounding in Drug Effectiveness Studies, Stat Methods Med Res. 2016). Methods for dealing with time-dependent confounding (Stat Med. 2013;32(9):1584-618) provides an overview of how time-dependent confounding can be handled in the analysis of a study. It provides an in-depth discussion of marginal structural models and g-computation.

G-estimation is a method for estimating the joint effects of time-varying treatments using ideas from instrumental variables methods. G-estimation of Causal Effects: Isolated Systolic Hypertension and Cardiovascular Death in the Framingham Heart Study (Am J Epidemiol. 1998;148(4):390-401) demonstrates how the G-estimation procedure allows for appropriate adjustment of the effect of a time-varying exposure in the presence of time-dependent confounders that are themselves influenced by the exposure.

The use of Marginal Structural Models can be an alternative to G-estimation. Marginal Structural Models and Causal Inference in Epidemiology (Epidemiology 2000;11(5):550-60) introduces a class of causal models that allow for improved adjustment for confounding in situations of time-dependent confounding. MSMs have two major advantages over G-estimation. Even if it is useful for survival time outcomes, continuous measured outcomes and Poisson count outcomes, logistic G-estimation cannot be conveniently used to estimate the effect of treatment on dichotomous outcomes unless the outcome is rare. The second major advantage of MSMs is that they resemble standard models, whereas G-estimation does not (see Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology 2000;11(5):561-70).

Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models (Am J Epidemiol. 2003;158(7):687-94) provides a clear example in which standard Cox analysis failed to detect a clinically meaningful net benefit of treatment because it does not appropriately adjust for time-dependent covariates that are simultaneously confounders and

intermediate variables. This net benefit was shown using a marginal structural survival model. In Time-dependent propensity score and collider-stratification bias: an example of beta2-agonist use and the risk of coronary heart disease (Eur J Epidemiol. 2013;28(4):291-9), various methods to control for time-dependent confounding are compared in an empirical study on the association between inhaled beta-2-agonists and the risk of coronary heart disease. MSMs resulted in slightly reduced associations compared to standard Cox-regression.

6.2.3.6. The trend-in-trend design

The Trend-in-trend Research Design for Causal Inference (Epidemiology 2017;28(4):529-36) presents a semi-ecological design, whereby trends in exposure and in outcome rates are compared in subsets of the population that have different rates of uptake for the drug in question. These subsets are identified through PS modelling. There is a formal framework for transforming the observed trends into an effect estimate. Simulation and empirical studies showed the design to be less statistically efficient than a cohort study, but more resistant to confounding. The trend-in-trend method may be useful in settings where there is a strong time trend in exposure, such as a newly approved drug.

6.3. Missing data

Missing data (or missing values) are defined as data value(s) that are not available for a variable in the data source of interest for a given analysis, hence are not observed. Missing data may also arise from attrition bias, non-response or poorly designed protocols. Missing data is an error as the data does not represent the true value of what is set out to be measured.

6.3.1. Impact of missing data

Missing data are a common issue in both clinical trial and observational data, and can have significant consequences on the conclusions that can be drawn from the results of an analysis for the following reasons: 1) the absence of data reduces statistical power, which refers to the probability that the test will reject the null hypothesis when it is false; 2) the unobservable data can introduce bias and increase uncertainty in the estimation of the model parameters; 3) it can reduce the representativeness of the sample; 4) it may complicate the analyses as it may render the completeness of data different between variables. Each of these elements can lead to invalid conclusions. Whether these issues are applicable to the dataset under study depends on the type of missing data (i.e., missing data mechanism).

6.3.2. Missing data mechanisms

When missing data is present, choosing the right statistical methods and making inferences are more complex, as assumptions about the processes that create missing data need to be made explicitly.

Missing data assumptions are classified into 3 categories, depending on the relationship between the unobserved values and the probability of missingness:

- Missing completely at random (MCAR): there are no systematic differences between the distribution of the missing values and the observed values. Missingness is unrelated to any variable in the analysis, including the variable with missing data itself. This is the most restrictive mechanism, but rather unrealistic.
- Missing at random (MAR): any systematic difference between the missing and observed values for a given variable can be explained by differences in other variables of the observed data. Missingness is associated with those variables, but not with the variable with missing data itself. This mechanism may be more realistic in some real-world settings.

- Missing not at random (MNAR): even after the observed data are taken into account, systematic differences remain between the missing values and the observed values. Missingness depends on the unobserved values of the variable with missing data itself.

Assumptions on missing data mechanisms determine the type of analysis that would be possible. In general, it is not possible to distinguish between these 3 mechanisms based on the observed data alone. In order words, missing data assumptions in general cannot be tested or verified. The distinction between MCAR and MAR could be made based on the observed data, but subject matter expertise and knowledge about the data collection process are needed to justify the assumption of data being MCAR or MAR. It is however not feasible to assess MAR versus MNAR based on the observed data.

6.3.3. Methods for handling missing data

Some simple solutions exist, but they generally lead to misleading inferences if the underlying assumptions on mechanisms of missingness are not valid, and they should be avoided. Examples include single imputation methods such as carrying forward the last observation in longitudinal analyses or mean substitution. Complete case analysis (CCA), i.e., removing all records with missing data, is only valid in certain circumstances, e.g., if the missing data is MCAR. Even in these circumstances, CCA will result in loss of power and increased uncertainty in the estimated parameters.

Therefore, it is advised to use other statistical methods to handle missing data, such as multiple imputation (Multiple Imputation and its Application, Wiley 2013, ISBN:9780470740521) or inverse probability weighting (Review of inverse probability weighting for dealing with missing data, Statistical Methods in Medical Research 2013;22:278-95).The choice of such statistical methods will depend on the assumed missing data mechanism.

If the missing data can be assumed to be MCAR or MAR, the Fully Conditional Specification (FCS), described in *Flexible Imputation of Missing Data* (Van Buuren S. 2nd ed. Chapman and Hall/CRC 2018, 10.1201/9780429492259), is a commonly used approach. MI utilises observed data to predict the value of missing data points, generating multiple complete data sets, performing analyses on each imputed data set, and then averaging the results.

If the missing data are assumed to be MNAR, most common statistical analysis methods are not appropriate, and would lead to biased results. There are methods to handle MNAR data, which depend on different assumptions or incorporate more specific knowledge about the missingness mechanism. One example is the not-at-random fully conditional specification (NARFCS) as described in *On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice* (Stat Med. 2018, 37(15): 2338–53, 10.1002/sim.7643).

Multiple imputation (MI) methods such as Pattern Mixture Models can be used to implement any missing data assumption (Multiple Imputation and its Application, Wiley 2013, ISBN:9780470740521).

It is important, as explained in The proportion of missing data should not be used to guide decisions on multiple imputation (J Clin Epidemiol. 2019;110:63-73), that the amount of missing data does not decide on the right MI method. In general, it is desirable to understand how sensitive to missing data assumptions are the conclusions drawn from the data, as well as to the particular method used to handle missing values. To investigate this, it is helpful to perform sensitivity analyses exploring how inferences vary under various mechanism assumptions and under various approaches.

A practice sometimes used is to create a category of the variable, or an indicator, for the missing values; however, this should be avoided. This practice can be invalid even if the data are missing completely at random, see Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression (J Am Stat Assoc. 1996;91(433):222-30) and *Missing data in*

epidemiological studies (In Armitage P, Colton T, eds. *Encyclopedia of biostatistics*. Wiley, 1998: 2641-2654.).

A concise review of methods to handle missing data is provided in the book *Statistical analysis with missing data* (Little RJA, Rubin DB. 3rd ed., Wiley 2019). The section 'Handling of missing values' in *Modern Epidemiology*, 4th ed. (T. Lash, T. VanderWeele, S. Haneuse, K.Rothman. Wolters Kluwer, 2020) is a summary of the state of the art, focused on practical issues for epidemiologists.

Other useful references on handling missing data include the books *Multiple Imputation for Nonresponse in Surveys* (Rubin DB, Wiley, 2004) and *Analysis of Incomplete Multivariate Data* (Schafer JL, Chapman & Hall/CRC, 1997), and the articles A comparison of multiple imputation methods for missing data in longitudinal studies (BMC Med Res Methodol. 2018;18(1):168), Using the outcome for imputation of missing predictor values was preferred (J Clin Epi. 2006;59(10):1092-101), and Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data (Stat Med. 2014;33(21):3725-37).

The article Framework for the treatment and reporting of missing data in observational studies: The Treatment and Reporting of Missing data in Observational Studies framework (J Clin Epi. 2021;134:79-88) focuses on missing data in non-interventional studies and provides a framework on both analysis and reporting of study results relying on incomplete data.

6.3.4. Statistical software

Many statistical procedures in standard software automatically eliminate subjects with missing data. However, a wide range of statistical software is currently available to impute missing data, mainly focusing on Multiple Imputation (MI) methods when missing data is assumed to be MAR, such as The MI Procedure of the SAS Institute. Multiple imputation of missing values (Stata J. 2004;4:227-41), and mice: Multivariate Imputation by Chained Equations in R (J Stat Soft. 2011;45(3)). A good overview of available software packages is provided in Missing data: A statistical framework for practice (Biom J. 2021;63(5): 915-47). Software tools in SAS and R for multiple imputation of missing data under MAR and MNAR have also been made available by the Drug Information Association Scientific Working Group on Estimands and Missing Data.

6.4. Triangulation

Triangulation is not a separate methodological approach, but rather a research paradigm aiming to enhance the confidence in inferred causal relationships. Triangulation in aetiological epidemiology (Int J Epidemiol. 2016;45(6):1866-86) defines triangulation as "*the practice of obtaining more reliable answers to research questions through integrating results from several different approaches, where each approach has different key sources of potential bias that are unrelated to each other*." Triangulation differs from replication by explicitly choosing data sources/data collection approaches, study designs and/or analytical approaches with different bias structures.

In Triangulation of pharmacoepidemiology and laboratory science to tackle otic quinolone safety (Basic Clin Pharmacol Toxicol. 2022;Suppl 1:75-80), laboratory studies using cell culture and rodent models were complemented with real-world data from pharmacoepidemiological studies to translate mechanistic findings and corroborate real-world evidence. In Identifying Antidepressants Less Likely to Cause Hyponatremia: Triangulation of Retrospective Cohort, Disproportionality, and Pharmacodynamic Studies (Clin Pharmacol Ther. 2022; 111(6):1258-67), analyses of three different types of data with their respective analyses are presented: a retrospective cohort study, a disproportionality analysis of patients in the Japanese Adverse Drug Event Report database, and a pharmacodynamic study examining the binding affinity for serotonin transporter.

Triangulation does not require the use of different data sources and can readily be employed in studies using electronic healthcare data, which allow investigators to use a multitude of study designs and analytical approaches. For example, in Prenatal Antidepressant Exposure and the Risk of Attention-deficit/Hyperactivity Disorder in Childhood: A Cohort Study With Triangulation (Epidemiology. 2022;33(4):581-592), a negative control analysis, a sibling analysis, and a former-user analysis were used to triangulate results.

In recent years, the use of genetic tools has become popular for the investigation of drug effects. The complementary application of drug target mendelian randomisation and colocalisation analyses can provide another layer of genetic evidence for causality, as demonstrated by Genetically proxied therapeutic inhibition of antihypertensive drug targets and risk of common cancers: A mendelian randomization analysis (PLoS Med. 2022 Feb 3;19(2):e1003897). It is recommended to use triangulation methods and formalise sensitivity analyses using *a priori* specification of potential biases and their (assumed) directions in the main analysis and by performing sensitivity/triangulation analyses explicitly addressing these biases.

7. Effect modification and interaction

Effect modification and interaction are often encountered in epidemiological research, and it is important to recognise their occurrence. The difference between these terms is rather subtle and has been described in Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators (Clin Epidemiol 2017;9:331-8) and in Tutorial: A nontechnical explanation of the counterfactual definition of effect modification and interaction (J. Clin. Epidemiol. 2021;134:113-24), which provides a more lay definition of these concepts.

Effect modification occurs when the effect of a single exposure on an outcome depends on the values of another variable, i.e., the effect modifier, which does not necessarily need to be involved in the causal pathway. Interaction occurs when there is interest in the causal effect of two exposures on an outcome and how the effect of either exposure depends upon the value of the other exposure. In Suicide and death by other causes among patients with a severe mental illness: cohort study comparing risks among patients discharged from inpatient care *v.* those treated in the community (Epidemiol Psychiatr Sci. 2022;31:e32), a test for interaction by age group indicated that younger adults with severe illness had a higher relative risk of all-cause mortality than middle-aged adults. Contrary to effect modification, interaction is a symmetric concept as two independent exposures have an equal status in the definition of interaction.

Interaction is generally studied in order to clarify aetiology, i.e., an interaction between environmental and inhered factors, while effect modification is used to identify subpopulations that are particularly susceptible to the exposure of interest. The key distinction between interaction and effect modification is that with effect modification, interest is in the effect of one single exposure on an outcome and this relationship does not have to be causal, whereas with interaction interest is in the causal effect of two exposures on an outcome. Assessment of effect modification is to identify whether the outcome of a treatment (or exposure) differs across patient population subgroups. To check the presence of an effect modifier, one can stratify the study population by a certain covariate, e.g., gender, and compare the effects in these subgroups. These subgroups can be constructed based on a priori knowledge regarding the effect modifier or derived from analysing the observed data and covariates itself.

It is recommended to perform a formal statistical test to assess if there are statistically significant differences for the effects (i.e., the measures) between subgroups (see CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials, J Clin Epidemiol. 2010;63(8):e1-37 and Interaction revisited: the difference between two estimates, BMJ. 2003;326(7382):219). The study report should explain which measure was used to examine these

differences and specify which subgroup analyses were predefined in the study protocol and which ones were performed at analysis stage (see Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology 2007;18(6):805-35).

The presence of effect modification depends on which measure is used in the study (absolute or relative) and can be measured in two ways: on an additive scale (based on risk differences [RD]), or on a multiplicative scale (based on relative risks [RR]). An example of potential effect modifier in studies assessing the risk of occurrence of events associated with drug use is the presence or severity of the underlying illness. Novel antihyperglycaemic drugs and prevention of chronic obstructive pulmonary disease exacerbations among patients with type 2 diabetes: population based cohort study (BMJ. 2022;379:e071380) observed effect measure modification on a multiplicative scale with history of asthma, but not for age, sex or severity when studying the association between GLP-1 receptor agonists with risk of severe and moderate COPD exacerbations among patients with type 2 diabetes and COPD, compared with sulfonylureas.

Evidence derived from studies considering effect modification provides more information and may lead to stronger conclusions about treatment effects. In the absence of prior knowledge about which covariates to consider as potential effect modifiers, one may test the data to investigate their presence. In False discovery rate control for effect modification in observational studies (Electron J Statist. 2018;12(2):3232-53), several analyses are proposed to test the presence of effect modification using the observed data itself.

Interaction is often considered in regression models whose design have variables with one or more levels; binary, categorical or dichotomised. For the evaluation of the interaction variable, the standard measure is the relative excess risk due to interaction (RERI), as explained in the textbook *Modern Epidemiology* (T. Lash, T.J. VanderWeele, S. Haneuse, K. Rothman. 4th Edition, Wolters Kluwer, 2020). Other measures of interaction include the attributable proportion (A) and the synergy index (S). Most measures, such as the S measure, are limited to binary variables.

Using statistical measures only, it is often difficult to understand the direction and size of an interaction effect. Therefore, visually inspecting the data using bar graphs (i.e., categorical variables) or line graphs (i.e., continuous variables) is another way of assessing and interpreting the marginal effects of interaction terms.

For further recommendations regarding reporting, Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration (Epidemiology 2007;18(6):805-35), Recommendations for presenting analyses of effect modification and interaction (Int J Epidemiol. 2012;41(2):514-20), Confidence interval estimation of interaction (Epidemiology. 1992;3(5):452-6) and The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE) (BMJ. 2018;363:k3532) and Causal inference and effect estimation using observational data (J Epidemiol Community Health 2022;76:960–6) are useful resources. They provide recommendations on how to describe methods used to examine interactions and present the results:

- Separate effects (rate ratios, odds ratios or risk differences, with confidence intervals and p-values) of the exposure of interest (e.g. drug) and of the effect modifier (e.g. gender) and of their joint effect using one single reference category (preferably the stratum with the lowest risk of the outcome), as suggested in Estimating measures of interaction on an additive scale for preventive exposures (Eur J Epidemiol. 2011;26(6):433-8), as this provides enough information to calculate effect modification on an additive or multiplicative scale (see Modeling Ratios or Differences? Let the Data Tell Us, AJPH Methods 2017;107(7):1087-91);
- Effects of the exposure (e.g., drug) within strata of the potential effect modifier (e.g., gender);

- Measures of effect modification on both additive (e.g., RERI) and multiplicative (e.g., S) scales including confidence intervals;
- List of the confounders for which the models assessing the association between exposure and outcome were adjusted for.

The article Evaluating sources of bias in observational studies of angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker use during COVID-19: beyond confounding (J Hypertens. 2021;39(4):795-805) highlights that factors associated with differences in hypertension phenotype, and the renin-angiotensin system (and by extension ACEi/ARB use), may modify the strength of the effect size between ACEi/ARB use and the COVID-19 outcomes. These factors should be assessed as potential effect-modifying factors rather than confounding factors, as treating these factors as confounders can induce bias. It further emphasises the above recommendations that if present, effect size estimates should be presented across strata (including 95% confidence intervals) along with measures of interaction on both the additive and multiplicative scales.

IL-6 inhibition in the treatment of COVID-19: A meta-analysis and meta-regression (J Infect. 2021;82(5):178-85) estimates the relative risk of mortality between arms of RCTs comparing IL-6 inhibitors (tocilizumab and sarilumab) to placebo or standard of care in adults with COVID-19. Meta-regression was used to investigate treatment effect modification and showed no evidence of such effect by patient characteristics.

Nonsteroidal Antiinflammatory Drugs and Susceptibility to COVID-19 (Arthritis Rheumatol. 2021;73(5):731-39) investigated whether active use of NSAIDs increases susceptibility to developing suspected or confirmed COVID-19 compared to the use of other common analgesics. There was no evidence of effect modification by age or sex.

8. Approaches to data collection

There are two main approaches for data collection: collection of data to address the specific research question under study ('primary data collection') or use of data already collected for another purpose, e.g., for clinical management of patients, for reimbursement purposes ('secondary use of data'), or for another research question. The distinction between primary data collection and secondary use of data is important for marketing authorisation holders as it implies different regulatory requirements for the collection and reporting of suspected adverse reactions, as described in Module VI of the Guideline on good pharmacovigilance practice (GVP) - Management and reporting of adverse reactions to medicinal products. It has also implications in terms of data privacy requirements.

Secondary use of data has become a common approach in pharmacoepidemiological studies due to the large availability of electronic healthcare records, administrative claims and other existing data sources (see Chapter 8.2), and its increased efficiency and lower costs. In addition, networking between centres active in the pharmacoepidemiology and pharmacovigilance fields has been rapidly changing the landscape of medicines' safety and effectiveness research in Europe and worldwide, both in terms of availability of networks of data sources, and networks of organisations and researchers able to contribute to a particular study with a particular data source (see Chapter 9).

8.1. Primary data collection

8.1.1. General considerations

The methodological aspects of studies using primary data collection (also sometimes referred to as field studies or prospective studies) are well covered in the textbooks and guidelines referred to in the

Introduction. Annex 1 of Module VIII of the Good pharmacovigilance practice provides examples of study designs based on prospective/primary data collection, such as cross-sectional study, prospective cohort study, and active surveillance. For completeness, surveys and randomised controlled trials are also presented below as examples of primary data collection.

Studies using primary data collection in clinical care or community-based settings have allowed the evaluation of drug-disease associations for rare complex conditions that require very large source populations and in-depth case assessment by clinical experts. Classic historical examples are Appetite-Suppressant Drugs and the Risk of Primary Pulmonary Hypertension (N Engl J Med. 1996;335:609-16), The design of a study of the drug etiology of agranulocytosis and aplastic anemia (Eur J Clin Pharmacol. 1983;24:833-6) and Medication Use and the Risk of Stevens–Johnson Syndrome or Toxic Epidermal Necrolysis (N Engl J Med. 1995;333:1600-8). For some conditions, case-control surveillance networks have been developed and used for selected studies and for signal generation and evaluation, e.g., Signal generation and clarification: use of case-control data (Pharmacoepidemiol Drug Saf 2001;10:197-203).

Data can be collected using paper, electronic case report forms or, increasingly, study-specific smartphone or web applications provided to patients. This approach has been used during the COVID-19 pandemic, as illustrated, for example, in COVID-19 vaccine waning and effectiveness and side-effects of boosters: a prospective community study from the ZOE COVID Study (Lancet Infect Dis. 2022:S1473-3099(22)00146-3): in this longitudinal, prospective, community-based study, data on demographic characteristics, comorbidities, symptoms, SARS-CoV-2 tests and results, and vaccinations, were self-reported through an app, with participants prompted to daily reporting through app notifications. Possibilities, Problems, and Perspectives of Data Collection by Mobile Apps in Longitudinal Epidemiological Studies: Scoping Review (J Med Internet Res. 2021;23(1):e17691) concludes that using mobile technologies can help to overcome challenges linked to data collection in epidemiological research, but the applicability and acceptance of these mobile apps in various subpopulations vary and need to be further studied. In addition, self-reported data may introduce information bias or selection bias, and since participants are self-selected, they might not be fully representative of the general population.

8.1.2. Surveys

The book *Research Methods in Education* (J. Check, RK. Schutt, Sage Publications, 2011) defines survey research as "*the collection of information from a sample of individuals through their responses to questions*" (p. 160). This type of research allows for a variety of methods to recruit participants, collects data and utilises various instruments.

A survey is the collection of data on specific health and quality of life aspects, knowledge, attitudes, behaviour, practices, opinions, beliefs, or feelings of selected groups of individuals from a specific sampling frame, by asking them questions in person or by post, phone or online. They generally have a cross-sectional design, but repeated measures over time may be performed for the assessment of trends.

Surveys have long been used in fields such as market research, social sciences and epidemiology. General guidance on constructing and testing the survey questionnaire, modes of data collection, sampling frames and ways to achieve representativeness can be found in general texts (*Survey Sampling* (L. Kish, Wiley, 1995) and *Survey Methodology* (R.M. Groves, F.J. Fowler, M.P. Couper et al., 2nd Edition, Wiley 2009). The book *Quality of Life: the assessment, analysis and interpretation of patient-related outcomes* (P.M. Fayers, D. Machin, 3rd Edition, Wiley, 2016) offers a comprehensive review of the theory and practice of developing, testing and analysing health-related quality of life questionnaires in different settings.

Surveys have an important role in the evaluation of the effectiveness of risk minimisation measures (RMM) or of a risk evaluation and mitigation strategy (REMS) (see Chapter 16.4). The application of methods described in the aforementioned textbooks needs adaptation for surveys to evaluate the effectiveness of RMM or REMS. For example, the extensive methods for questionnaire development of quality of life scales (construct, criterion and content validity, inter-rater and test-retest reliability, sensitivity and responsiveness) are not appropriate to questionnaires for RMM which are often used only once. The EMA and FDA issued guidance documents on the conduct of surveys for risk minimisation (RM) which, together, encompass the selection of risk minimisation measures, study design, instrument development, data collection, processing and data analysis and presentation of results. This guidance include the draft EMA Guideline on good pharmacovigilance practices (GVP) Module XVI – Risk minimisation measures: selection of tools and effectiveness indicators (Rev 3) (2021), the FDA draft guidance for industry REMS Assessment: Planning and Reporting on REMS (2019) and the FDA Guidance on Survey Methodologies to Assess REMS Goals That Relate to Knowledge (2019). A checklist to assess the quality of studies evaluating risk management programs is provided in The RIMES Statement: A Checklist to Assess the Quality of Studies Evaluating Risk Minimization Programs for Medicinal Products (Drug Saf. 2018;41(4): 389-401). The article Are Risk Minimization Measures for Approved Drugs in Europe Effective? A Systematic Review (Expert Opin Drug Saf. 2019;18(5):443-54) highlights the need for improvement in the methods and presentation of results and for more hybrid designs that link survey data with health and safety outcomes as requested by regulators. This article also reports on low response rates found in many studies, allowing for the possibility of important bias. The response rate should therefore be reported in a standardised way in surveys to allow comparisons. Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys (2016) of the American Association for Public Opinion Research provides standard definitions which can be adapted to RM surveys and the FDA Guidance on Survey Methodologies to Assess REMS Goals That Relate to Knowledge (2019) provides guidance for RM surveys.

An important aspect of surveys is sampling, often using a clustered random sample. However, attention shall be paid to the selection of the original list of subjects in the target population. For example, if the evaluation of the awareness about an educational material is part of the objectives, the same lists which were used to distribute the educational material cannot be used for sampling the survey, otherwise a selection bias cannot be excluded.

The increasing use of online RMM require that survey methods adapt but should not sacrifice representativeness by accessing only populations which visit these websites. They should provide evidence that the results using these sampling methods are not biased. Similarly, the increasing use of healthcare professional and patient panels needs to ensure that survey methods do not sacrifice representativeness by accessing only self-selected participants in these panels and should provide evidence that the results are not biased by using these convenient sampling frames. The influence of information given to survey subjects about the survey prior to its completion should attempt to minimise the influence of this information to reduce bias.

The issue of thresholds to assess the effectiveness of RMM remains a topic of debate. This topic is discussed in the aforementioned EMA and FDA documents and the article Are Risk Minimization Measures for Approved Drugs in Europe Effective? A Systematic Review (Expert Opin Drug Saf. 2019;18(5):443-54). The thresholds need to be viewed in the context of their potential impact on the benefit-risk balance. Composite thresholds for all of three aspects (awareness, knowledge, and behaviour) of RM effectiveness are hardly achieved.

The draft EMA Guideline on good pharmacovigilance practices (GVP) Module XVI – Risk minimisation measures: selection of tools and effectiveness indicators (Rev 3) (2021) encourages the evaluation of process indicators being linked to health outcomes. A holistic evaluation of non-targeted effects as well

as product-specific targeted effects has so far been performed in only a minority of studies, as shown in Risk Minimisation Evaluation with Process Indicators and Behavioural or Health Outcomes in Europe: Systematic Review (Expert Opin Drug Saf. 2019;18(5):443-54).

8.1.3. Randomised controlled trials

Randomised controlled trials are an experimental design that involves primary data collection. There are numerous textbooks and publications on methodological and operational aspects of clinical trials which are not covered here. An essential guideline on clinical trials is the European Medicines Agency (EMA) Guideline for good clinical practice E6(R2), which specifies obligations for the conduct of clinical trials to ensure that the data generated in the trial are valid. From a legal perspective, the Volume 10 of the Rules Governing Medicinal Products in the European Union contains all guidance and legislation relevant for conduct of clinical trials. A number of documents are under revision.

The way clinical trials are conducted in the European Union (EU) has undergone a major change when the Clinical Trial Regulation (Regulation (EU) No 536/2014) came into effect and replaced the existing Directive 2001/20/EC.

Hybrid data collection as used in pragmatic trials, large simple trials and randomised database studies are described in Chapter 4.2.7.

8.2. Secondary use of data

Secondary use of data refers to the utilisation of data already collected for other purposes. These data can be further linked to prospectively collected medical and non-medical data. Electronic healthcare databases (e.g., claims databases, electronic health records) and patient registries are examples of data sources that can be leveraged as secondary data for pharmacoepidemiological studies.

The last decades have witnessed the development of key data resources, expertise and methodology that have allowed use of such data for pharmacoepidemiology. The ENCePP Inventory of Data Sources contains information on existing European and worldwide databases that may be used for pharmacoepidemiological research. However, this field is continuously evolving.

A description of the main features, applications and limitations of frequently used electronic healthcare databases for pharmacoepidemiology research in the United States and in Europe is presented in the textbook *Pharmacoepidemiology* (B. Strom, S.E. Kimmel, S. Hennessy. 6th Edition, Wiley, 2019, Chapters 11-14).

In order to assist in the selection and appropriate use, including the assessment of strengths and limitations, of data sources for pharmacoepidemiological research, the ISPE-endorsed Guidelines for Good Database Selection and use in Pharmacoepidemiology Research (Pharmacoepidemiol Drug Saf. 2012;21(1):1-10) highlights potential limitations of data sources for secondary use containing routinely collected healthcare information, such as electronic health records (from either primary or secondary care) and claims databases, and recommends procedures for data analysis and interpretation. A section of the guideline is dedicated to multi-database studies which may be defined as "*studies using at least two healthcare databases, which are not linked with each other at an individual person level, either because they insist on different populations, or because, even if populations overlap, local regulations forbid record linkage*" (see Chapter 9). References to data quality and validation procedures, data processing/transformation, and data privacy and security (see Chapter 12.2) are also provided. In Different Strategies to Execute Multi-Database Studies for Medicines Surveillance in Real-World Setting: A Reflection on the European Model (Clin Pharmacol Ther. 2020;108(2):228-235), four strategies to conduct multi-database studies are discussed (see also Chapter 9). Specific processes have also been proposed to identify fit-for-purpose data sources to

address research questions. For example, The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework (Clin Pharmacol Ther. 2022;111(1):122-34) provides a structured and detailed stepwise approach for the identification and feasibility assessment of candidate data sources for a specific study. In order to help signpost regulators, researchers, industry and evidence reviewers to the relevant data sources to address a research question, the joint EMA-EU Heads of Medicines Agency Big Data Steering Group has also published a list of metadata (2022) describing data sources and studies and defined following extensive consultation of interested parties. This list will be used in the rebuilding and enhancement of the ENCePP Inventory of Data sources. The experience will show how such initiatives can support the validity and transparency of study results and ultimately the level of confidence in the evidence provided. It should also be acknowledged that many investigators naturally use the data source(s) they can directly access and are familiar with.

The FDA Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Health Care Data Sets (2013) provides criteria for best practice that apply to the study design, analysis, conduct and documentation. It emphasizes that investigators should understand the potential limitations of electronic healthcare data systems, make provisions for their appropriate use and refer to validation studies of outcomes of interest in the proposed study and captured in the database. This is also covered in the UK MHRA guidance on the use of real-world data in clinical studies to support regulatory decisions (2021). Guidance for conducting studies within electronic healthcare databases can also be found in the International Society for Pharmacoepidemiology Guidelines for Good Pharmacoepidemiology Practices (ISPE GPP, 2015), in particular sections IV-B (Study conduct, Data collection). This guidance emphasises the importance of patient data protection.

The use of real-world data (RWD) for the generation of real-world evidence (RWE) for regulatory decision-making has been addressed by guidelines issued by regulatory agencies. The article Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe (Clin Pharmacol Ther. 2019;106(1):36-9) describes the operational, technical and methodological challenges for the acceptability of RWD for regulatory purposes and presents possible solutions to address these challenges. The draft FDA guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (2021) provides recommendations focused on regulatory studies using electronic health records and claims databases, and a more general draft guidance provides Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products (December 2021). More information on RWD and RWE are available in Chapter 16.7, Real-world evidence and pharmacoepidemiology.

The Joint ISPE-ISPOR Special Task Force Report on Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness (2017) recommends good research practices for designing and analysing retrospective databases for comparative effectiveness research (CER) and reviews methodological issues and possible solutions for CER studies based on secondary data analysis (see also Chapter 16.1). Many of the principles are applicable to studies with other objectives than CER, but some aspects of pharmacoepidemiological studies based on secondary use of data, such as data quality, ethical issues, data ownership and privacy, are not covered.

Most of the examples and methods covered in Chapter 4 are based on studies and methodologic developments concerning secondary use of healthcare databases, since this is one of the most frequent approaches used in pharmacoepidemiology.

8.3. Patient registries

8.3.1. Definitions

A patient registry is defined in the EMA's Guideline on registry-based studies (2021) as an organised system that collects uniform data (clinical and other) to identify specified outcomes for a population defined by a particular disease, condition or exposure. It should be considered as an infrastructure for the standardised recording of data from routine clinical practice on individuals identified by a characteristic or an event.

A registry-based study is the investigation of a research question using the data collection infrastructure or patient population of one or several patient registries. A registry-based study may be a non-interventional study or a clinical trial.

8.3.2. General guidance on patient registries

The EMA's Guideline on registry-based studies (2021) includes an Annex discussing several aspects of good practice considered relevant for the use of registries for registry-based studies and other possible regulatory purposes. It addresses the registry population, data elements, quality management, governance and data sharing.

The EMA's Scientific Advice Working Party issued a Qualification Opinion for several registry platforms, including the ECFSPR for cystic fibrosis, the EBMT for blood & marrow transplantation and the Enroll HD for Huntington disease, with an evaluation of their potential use as data sources for registry-based studies in specific regulatory contexts. These opinions provide an indication of the methodological components expected by regulators for using a disease registry for such studies.

The US Agency for Health Care Research and Quality (AHRQ) published 'good registry practices' under the title Registries for Evaluating Patient Outcomes: A User's Guide, 4th Edition (2020), which provide comprehensive methodological guidance on planning, design, implementation, analysis, interpretation and evaluation of the quality of a registry.

The FDA issued the draft guidance Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry (2021). This guidance provides sponsors (marketing authorisation applicants and holders) and other relevant stakeholders with considerations when proposing to design a registry, or when using an existing registry to support regulatory decision-making about a drug's effectiveness or safety.

8.3.3. Types of patient registries

The characteristic or event defining entry into a patient registry may be the diagnosis of a disease (*disease registry)*, the occurrence of a condition or event (e.g., *pregnancy registry*), a birth defect (e.g., *birth defect registry*), a molecular or a genomic feature, or any other patient characteristics.

The term *product registry* has been used for a data collection system where data are collected on patients exposed to a particular medicinal product, single substance or therapeutic class in order to evaluate their use or their effects. Such system should rather be considered a clinical trial or non-interventional study, as data is collected for the purpose of a specific pre-planned analysis in line with performing a trial/study. Moreover, it does not include specific aspects related to the use of patient registries as source population or existing data collection system.

The terms *population registry* or *register* have been used to describe the type of registries that exist in European Nordic countries. In these countries, a comprehensive registration of data covering the entire

population allows linkage between different patient registries that may include hospital encounters, diagnoses and procedures, such as the Norwegian Patient Registry, the Danish National Patient Registry or the Swedish National Patient Register. Review of 103 Swedish Healthcare Quality Registries (J Intern Med. 2015; 277(1): 94–136) describes healthcare 'quality' registries initiated mostly by Swedish physicians that focus on specific disorders. Data recorded may include aspects of disease management, self-reported quality of life, lifestyle and general health status and provide an important data source for research.

8.3.4. Registry-based studies

As outlined in Imposed registries within the European postmarketing surveillance system (Pharmacoepidemiol Drug Saf. 2018;27(7):823-26) and the EMA's Guideline on registry-based studies (2021), there are important methodological differences between the registries and the conduct of registry-based studies. Patient registries are often integrated into routine clinical practice with systematic and sometimes automated data capture in electronic healthcare records. A registry-based study may only use the data relevant for the specific study objectives, is often limited in time and may need to be enriched with additional information on outcomes, lifestyle data, immunisation or mortality information. Such information may be obtained from linkage to existing databases such as national cancer registries, prescription databases or mortality records.

Results obtained from analyses of registry data may be affected by the same biases as those of studies described in Chapter 5 of this Guide. Factors that may influence the enrolment of patients in a registry may be numerous (including clinical, demographic and socio-economic factors) and difficult to predict and identify. This will potentially result in a biased sample of the patient population in case the recruitment has not been exhaustive. Bias may also be introduced by differential completeness of follow-up and data collection.

As illustrated in The randomized registry trial--the next disruptive technology in clinical research? (N Engl J Med. 2013; 369(17): 1579-81) and Registry-based randomized controlled trials: what are the advantages, challenges and areas for future research? (J Clin Epidemiol. 2016;80:16-24), and more recently in Registry randomised trials: a methodological perspective (BMJ Open 2023; 13(3)), randomised registry-based trials may support enhanced generalisability of findings, rapid consecutive enrolment, and the potential completeness of follow-up for the reference population, when compared with conventional randomized effectiveness trials. Defining key design elements of registry-based randomised controlled trials: a scoping review (Trials 2020;21(1):552) concludes that the low cost, reduced administrative burden and enhanced external validity make registries an attractive research methodology to be used to address questions of public health importance. However, the issues of data integrity, completeness, timeliness, validation and adjudication of endpoints need to be carefully addressed.

8.3.5. Interoperability between registries

A complexity of using registry data for regulatory purposes and analyses is the need for interoperability between different registries covering a same disease or condition. In most cases, there is no global alignment on how to collect data (data format, expression of a variable) in registries and often no mandatory standards to be applied for the data collected (content/variables). Interoperability of disease registries has been addressed in several workshops on disease-specific registries organised by EMA. The reports of these workshops are available on the EMA Patient registries initiative website. They describe the expectations from different stakeholders on common data elements to be collected and the best practices on topics such as governance, data quality control, data sharing or reporting of safety data.

One way to approach the challenge of heterogeneity between registries is the adaptation of globally common data structures in preparing registry data for joint analyses. One example is the Observational Medical Outcomes Partnership common data model (OMOP CDM) of the Observational Health Data Sciences and Informatics - OHDSI group. The OMOP CDM was originally designed for electronic healthcare records and claims data representing the majority of the 331 data sources from 34 countries. Data mapped to the OMOP CDM in January 2022, as stated by OHDSI in Our Journey (p. 36), resulted in 810 million unique patient records. Registry data is only slowly getting introduced to the OMOP CDM.

8.3.6. Registries which capture special populations

Special populations can be identified based on age (e.g., birth, paediatric or elderly), pregnancy status, renal or hepatic function, race, or genetic differences. Some registries are focused on these particular populations.

For paediatric populations, specific and detailed information as neonatal age (e.g., in days), pharmacokinetic parameters and organ maturation need to be considered and is usually missing from traditional data sources, therefore paediatric-specific registries are important. The Guideline on good pharmacovigilance practices (GVP) Product- or Population-Specific Considerations IV: Paediatric population (2018) provides further relevant information. An example of registry which focuses on paediatric patients is Pharmachild, which captures children with juvenile idiopathic arthritis undergoing treatment with methotrexate or biologic agents.

Pregnancy registries include pregnant women followed until the end of pregnancy and provide information on pregnancy outcomes. Use of pregnancy registries for observational studies on adverse effects of medicinal products administered during pregnancy are often faced with multiple challenges, which may vary from registry to registry. They include not only the recruitment and retention of pregnant women, but also the identification of relevant control groups for comparisons and the complete recording of information on pregnancy outcomes. Embryonic and early foetal loss are often not recognised or recorded and data on the gestational age at which these events occur are often missing. Non-interventional studies may therefore require linkage with data captured in birth defects registries, teratology information services or electronic health care records where mother-child linkage is possible. The EMA Draft Guideline on good pharmacovigilance practices. Product- or Population-Specific Considerations III – Pregnancy prevention programme and other pregnancy-specific risk minimisation measures (2022) provides methodological recommendations for use of a pregnancy registry for data collection in additional pharmacovigilance activities. The FDA Draft Postapproval Pregnancy Safety Studies Guidance for Industry (2019) include recommendations for designing a pregnancy registry with a description of research methods and elements to be addressed. The Systematic overview of data sources for drug safety in pregnancy research (2016) provides an inventory of pregnancy exposure registries and alternative data sources on safety of prenatal drug exposure and discusses their strengths and limitations. Examples of population-based registries allowing to assess outcome of drug exposure during pregnancy are the European network of registries for the epidemiologic surveillance of congenital anomalies EUROCAT, the EUROmediSAFE inventory of data sources, and the pan-Nordic registries which record drug use during pregnancy as illustrated in Selective serotonin reuptake inhibitors and venlafaxine in early pregnancy and risk of birth defects: population based cohort study and sibling design (BMJ 2015;350:h1798).

In the context of rare diseases, the European Reference Networks (ERNs), consisting of 24 virtual networks composed of healthcare providers across Europe, aim to facilitate discussion on such complex diseases and conditions that require highly specialised treatment, and concentrated knowledge and resources. One of the purposes of the European Rare Disease Research Coordination and Support

Action consortium (ERICA), in which all 24 networks take part, is to build on the strengths of the individual ERNs and create a platform that integrates all ERNs research and innovation capacity. Various activities intend to advance the development and integration of ERN-wide rare disease registries and their utilisation for joint research initiatives. The Network supports the creation of biorepositories within and across ERNs, and promotes the use of the European Platform on Rare Diseases Registration (EU RD Platform) for research.

Other registries that focus on special populations can be found in the ENCePP Inventory of data sources, in the European Platform on Rare Diseases Registration (EU RD Platform), and in Orphanet.

8.3.7. Disease registries in regulatory practice and health technology assessment

Use of real-world data (RWD), including registry data, to support regulatory decision-making is a topic of high interest. Several studies have evaluated the frequency and usefulness of information based on RWD in marketing authorisation applications, but did not present results stratified by data source. The article Marketing Authorization Applications Made to the European Medicines Agency in 2018-2019: What was the Contribution of Real-World Evidence? (Clin Pharmacol Ther. 2022;111(1):90-7) shows that registries were the most common type of RWD sources referred to in marketing authorisation applications and extensions of indications submitted to the EMA in 2018 and 2019 (60.3% and 46.4% respectively of the medicinal products presented with RWE). The follow-up study described in Contribution of Real- World Evidence in European Medicines Agency's Regulatory Decision Making (Clin Pharmacol Ther. 2023;113(1):135-151) provides an in-depth review of real-world evidence (RWE) submitted in recent centralised applications in the EU, illustrated by examples of RWE contribution to regulatory decision-making.

The article Patient Registries: An Underused Resource for Medicines Evaluation: Operational proposals for increasing the use of patient registries in regulatory assessments (Drug Saf. 2019;42(11):1343-51) proposes sets of measures to improve use of registries in relation to: (1) nature of the data collected and registry quality assurance processes; (2) registry governance, informed consent, data protection and sharing; and (3) stakeholder communication and planning of benefit-risk assessments. The EMA's Guideline on registry-based studies (2021) discusses methodological aspects for the use of registries for conducting registry-based studies and recommends performing a feasibility assessment of the suitability of a registry for a specific research question to facilitate early discussions with regulators. The use of registries to support the post-authorisation collection of data on safety and effectiveness of medicinal products in the routine treatment of diseases is also discussed in the EMA Guideline on good pharmacovigilance practices (GVP) – Module VIII -Post-authorisation safety studies (2017) and the EMA Scientific guidance on post-authorisation efficacy studies (2016).

As outlined in Real World Data in Health Technology Assessment of Complex Health Technologies - PMC (Front Pharmacol. 2022; 13: 837302), incorporating data from clinical practice into the drug development process is of growing interest for Health Technology Assessment (HTA) bodies and payers since reimbursement decisions can benefit from better estimation and prediction of effectiveness of treatments at the time of product launch. An example where registries can provide clinical practice data is the building of predictive models that incorporate data from both randomised clinical trials (RCTs) and registries to generalise results observed in RCTs to a real-world setting. In this context, the EUnetHTA Joint Action 3 project has issued the Registry Evaluation and Quality Standards Tool (REQueST) aiming to guide the evaluation of registries for effective use in HTA.

Patient experience data collected through patient registries can inform medicine development, enhance regulatory decision-making, and result in more patient-relevant outcomes to study, however, the generation of these data remains challenging as highlighted for example in A review of patient-

reported outcomes used for regulatory approval of oncology medicinal products in the European Union between 2017 and 2020 (Front Med (Lausanne). 2022 Aug 12;9:968272).

8.3.8. Registry catalogues

Several data source catalogues provide different levels of access to different amounts of information on disease registries, such as the ENCePP Resource database of data sources, the EHDEN data partners listing or the EMIF Catalogue. The European Platform on Rare Diseases Registration (EU RD Platform) serves as a platform for information on registries for rare diseases and has developed a harmonised set of common data elements for rare disease registration.

In the context of the EMA/HMA Big Data Initiative, the ENCePP Resource database of data sources and the EU PAS Register will be enhanced and replaced in 2024 by two new catalogues of RWD sources and non-interventional studies in view of facilitating the identification by regulators, researchers and pharmaceutical companies of data sources and studies suitable to address research questions, based on the FAIR (findable, accessible, interoperable and reusable) data principles.

8.4. Spontaneous reports

Note: Chapter 8.4. (formerly 7.4.) has not been updated for Revision 11 of the Guide, as contents remain up-to-date.

Spontaneous reports of suspected adverse drug reactions remain a cornerstone of pharmacovigilance and are collected from a variety of sources, including healthcare providers, national authorities, pharmaceutical companies, medical literature, and directly from patients.

EudraVigilance is the European Union data processing network and management system for reporting and evaluating suspected adverse drug reactions (ADRs). Other major systems for collections of spontaneous reports are the FDA's Adverse Event Reporting System (FAERS), the FDA's Vaccine Adverse Event Reporting System (VAERS) and the WHO global database of individual case safety reports, VigiBase, that pools reports from the members of the WHO programme for international drug monitoring. These systems deal with the electronic exchange of Individual Case Safety Reports (ICSRs), the early detection of possible safety signals and the continuous monitoring and evaluation of potential safety issues in relation to reported ADRs. Spontaneous case reports represent the first line of evidence and the majority of safety signals is based on them, as described in A description of signals during the first 18 months of the EMA pharmacovigilance risk assessment committee (Drug Saf. 2014;37(12):1059-66).

The main strengths of spontaneous reporting systems are:

i) they cover *all types of authorised medicines used* in any setting (primary, secondary and specialised healthcare) and *all reasons for use* including authorised indications, off-label, misuse and abuse;

ii) they are built to obtain information specifically to *evaluate the likelihood that a particular treatment is the cause of an observed adverse event*. The data collection concentrates on variables relevant to this objective directing reporters towards careful coding and communication of the main aspects of an ADR (e.g., event dates, medical history and co-morbidities, concomitant treatments, etc.);

iii) they are designed to collect *and make the information on suspected ADRs rapidly available for analysis*.

The application of knowledge discovery in databases to post-marketing drug safety: example of the WHO database (Fundam Clin Pharmacol. 2008;22(2):127-40) describes known limitations of spontaneous ADR reporting systems, which can be grouped into four main categories:

i) *factors influencing reporting dynamics*, whereby known or unknown factors, such as workload of healthcare professionals or increased media coverage and public awareness, may influence the reporting rate, leading respectively to under-reporting or to a comparative increase in the reporting rate affecting the reliability of estimates of signals of disproportionate reporting;

ii) *insufficient clinical information* reported, not allowing a satisfactory case evaluation and/or the identification of possible risk factors, which is crucial to establish the likely causal relationship between exposure to the product and occurrence of the adverse drug reaction;

iii) *misclassification of diagnosis* is closely related to the factors influencing reporting dynamics, where extensive media coverage and public awareness not only stimulates reporting, but may influence the interpretation of symptoms, such that symptoms similar to the ones of the disorder in the media coverage, are likely to be reported as suspected cases of that disorder to the detriment of other disorders with similar symptoms, potentially leading to a misclassification of diagnosis;

iv) *lack of collection of control information*, as these databases are case-only databases and thus cannot provide actual medicinal product exposure information nor information on the disease incidence.

Another challenge of spontaneous reporting databases is the quality of the information provided and adherence to reporting rules; for this reason, comprehensive and multi-faceted quality activities are often an integral part of these systems (see Detailed guide regarding the EudraVigilance data management activities by the European Medicines Agency Rev 1 for an example). One aspect of the data quality activities regards report duplication. Duplicates are separate and unlinked records that refer to one and the same case of a suspected ADR and may mislead signal assessment or distort statistical screening. They are generally detected by individual case review of all reports or by computerised duplicate detection algorithms. In Performance of probabilistic method to detect duplicate individual case safety reports (Drug Saf. 2014;37(4):249-58) a probabilistic method applied to VigiBase highlighted duplicates that had been missed by a rule-based method and also improved the accuracy of manual review. In the study, however, a demonstration of the performance of de-duplication methods to improve signal detection is lacking. The EMA and FDA have also implemented probabilistic duplicate detection in their databases.

More recently, there have been attempts to boost the computerised detection of duplicates using Natural Language Processing (NLP) techniques to identify similarities on the narrative of reports, as demonstrated in Using Probabilistic Record Linkage of Structured and Unstructured Data to Identify Duplicate Cases in Spontaneous Adverse Event Reporting Systems (Drug Saf. 2017;40(7):571–82).

For the above reasons, it is advised that the cases underlying a potential safety signal from spontaneous reports should be verified from a clinical perspective and preferably supported by pharmacological information before further investigation. Anecdotes that provide definitive evidence (BMJ. 2006;333(7581):1267-9) describes uncommon examples where this is not necessary, where strong and well documented spontaneous reports can be convincing to support the existence of a signal.

Patient reporting is an important source of suspected adverse drug reactions. Factors affecting patient reporting of adverse drug reactions: a systematic review (Br J Clin Pharmacol. 2017;83(4):875-83) describes the practical difficulties with patient reporting and highlights the patients' motivation to make their ADRs known to prevent similar suffering in other patients. The value of patient reporting to the pharmacovigilance system: a systematic review (Br J Clin Pharmacol. 2017;83(2):227-46) concludes that patient reporting adds new information and perspective about ADRs in a way otherwise unavailable, and this can contribute to better regulatory decision-making. Patient Reporting in the EU: Analysis of EudraVigilance Data (Drug Saf. 2017;40(7):629-45) also concludes that patient reporting

complements reporting by health care professionals and that patients are motivated to report especially those ADRs that affect their quality of life.

The information collected in spontaneous reports is a reflection of a clinical event that has been attributed to the use of one or more suspected medicinal products. Although the majority of information provided in the ICSRs is coded, the description of the clinical event, as well as the interpretation of the reporter, contains valuable information for signal detection purposes. Examples are the description of timing and course of the reactions, of the presence or absence of additional risk factors and of the medical history of the patient. Knowledge of the local healthcare system, its corresponding guidelines and the possibilities to follow-up for more detailed information are considered important during this review.

Since only part of this information is coded and can be used in statistical analyses, it remains important to review the underlying cases for signal detection purposes.

The increase in systematic collection of ICSRs in large electronic databases has allowed the application of data mining and statistical techniques for the detection of safety signals (see Chapter 11). Validation of statistical signal detection procedures in EudraVigilance post-authorisation data: a retrospective evaluation of the potential for earlier signalling (Drug Saf. 2010;33(6): 475-87) shows that the statistical methods applied in EudraVigilance can provide significantly early warning in a large proportion of drug safety concerns. Nonetheless, this approach should supplement, rather than replace, other pharmacovigilance methods.

The report 'Characterisation of Databases (DBs) Used for Signal Detection (SD): Results of a Survey of IMI PROTECT Work Package (WP) 3 Participants' (Pharmacoepidemiol Drug Saf. 2012;21(Suppl.3): abstract no. 496 pp 233) shows the heterogeneity of spontaneous databases and the lack of comparability of signal detection methods employed.

Chapters IV and V of the Report of the CIOMS Working Group VIII 'Practical aspects of Signal detection in Pharmacovigilance' present sources and limitations of spontaneously-reported drug-safety information and databases that support signal detection. Appendix 3 of the report provides a list of international and national spontaneous reporting system databases.

Finally, in EudraVigilance Medicines Safety Database: Publicly Accessible Data for Research and Public Health Protection (Drug Saf. 2018;41(7):665-75), the authors describe how these databases, focusing on EudraVigilance, have been made more easily accessible for external stakeholders. This has allowed to provide better access to information on suspected adverse reactions for healthcare professionals and patients, and opportunities for health research for academic institutions.

8.5. Social media

Note: Chapter 8.5. (formerly 7.5.) has not been updated for Revisions 10 and 11 of the Guide.

8.5.1. Definition

Technological advances have dramatically increased the range of data sources that can be used to complement traditional ones and may provide compelling insights into or relevant to effectiveness and safety of health interventions such as medicines and their risk minimisation measures, benefit-risk communications and related stakeholder engagement. Such data include those from digital media that exist in a computer-readable format and can be extracted from websites, web pages, blogs, vlogs, social networking sites, internet forums, chat rooms and health portals. A recent addition to the digital media data is biomedical data collected through wearable technology (e.g., heart rate, physical activity and sleep pattern, dietary patterns). These data are unsolicited and generated in real time.

A subset of digital media data are social media data. The European Commission's Digital Single Market Glossary defines social media as "*of Web 2.0 and that allow the creation and exchange of user-generated content. It employs mobile and web-based technologies to create highly interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content.*"

8.5.2. Use in pharmacovigilance

Social media content analyses have been used to provide insights into patients' perceptions of the effectiveness and safety of medicines and for the collection of patient reported outcomes, as discussed in Web-based patient-reported outcomes in drug safety and risk management: challenges and opportunities? (Drug Saf. 2012;35(6):437-46).

The IMI WEB-RADR European collaborative project explored different aspects related to the use of social media data for pharmacovigilance and summarised its recommendations in Recommendations for the Use of Social Media in Pharmacovigilance: Lessons From IMI WEB-RADR (Drug Saf 2019;42(12):1393-407). The French Vigi4Med project, which evaluated the use of social media, mainly web forums, for pharmacovigilance activities, published a set of recommendations in Use of Social Media for Pharmacovigilance Activities: Key Findings and Recommendations from the Vigi4Med Project (Drug Saf. 2020;43(9):835-51).

A further possible use of social media data would be as a source of information for signal detection or assessment. Studies including Using Social Media Data in Routine Pharmacovigilance: A Pilot Study to Identify Safety Signals and Patient Perspectives (Pharm Med. 2017;31(3): 167-74) and Assessment of the Utility of Social Media for Broad-Ranging Statistical Signal Detection in Pharmacovigilance: Results from the WEB-RADR Project (Drug Saf. 2018;41(12):1355–69) evaluated whether analysis of social media data (specifically Facebook and Twitter posts) could identify pharmacovigilance signals early, but in their respective settings, found that this was not the case.

The study Using Social Media Data in Routine Pharmacovigilance: A Pilot Study to Identify Safety Signals and Patient Perspectives (Pharm Med. 2017;31(3): 167-74) also tried to determine the quantity of posts with resemblance to adverse events and the types and characteristics of products that would benefit from social media content analysis. It concludes that, although analysis of data from social media did not identify new safety signals, it can provide unique insight into the patient perspective.

From a regulatory perspective, social media is a source of potential reports of suspected adverse reactions and marketing authorisation holders are legally obliged to screen websites under their management and assess whether reports of adverse reactions qualify for spontaneous reporting (see Good Pharmacovigilance Practice Module VI, section VI.B.1.1.4.). Principles for continuous monitoring of the safety of medicines without overburdening established pharmacovigilance systems and a regulatory framework on the use of social media in pharmacovigilance have been proposed in Establishing a Framework for the Use of Social Media in Pharmacovigilance in Europe (Drug Saf. 2019;42(8):921-30).

Sentiment analyses of social media content may offer future opportunities for regulators into public perceptions about the safety of medicines and trustworthiness of regulatory bodies. This can inform and evaluate specific safety communication strategies aiming at effective and safe use of medicines. For example, a recent study provided insight into public sentiments about vaccination of pregnant women by stance, discourse and topic analysis of social media posts in "Vaccines for pregnant women?! Absurd" – Mapping maternal vaccination discourse and stance on social media over six months (Vaccine 2020;38(42): 6627-38).

8.5.3. Challenges

While offering the promise of new research models and approaches, the rapidly evolving social media environment presents many challenges including the need for strong and systematic processes for data selection and validation, and study implementation. Articles which detail associated challenges are: Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies (Drug Saf. 2015;38(10): 921-30.) and Social media and pharmacovigilance: A review of the opportunities and challenges (Br J Clin Pharmacol. 2015;80(4): 910-20).

There is currently no defined strategy or framework in place in order to meet the standards around data selection and validity and methods for data analysis, and their regulatory acceptance may therefore be lower than for traditional sources. However, more tools and methods for analysing unstructured data are becoming available, especially for pharmacoepidemiology and pharmacovigilance research, as in Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts (J Am Med Inform Assoc. 2017 Feb 22), Social Media Listening for Routine Post-Marketing Safety Surveillance (Drug Saf. 2016;39(5):443-54) and Social Media Research (Chapter 11 in Communicating about Risks and Safe Use of Medicines, Adis Singapore, 2020, pp 307-332). However, the recognition and disambiguation of references to medicines and adverse events in free text remains a challenge and performance evaluations need to be critically assessed as discussed in Prospective Evaluation of Adverse Event Recognition Systems in Twitter: Results from the Web-RADR Project (Drug Saf. 2020;43(8):797-808).

8.5.4. Data protection

The EU General Data Protection Regulation (GDPR) introduces EU-wide legislation on personal data and security. It specifies that the impact of data protection at the time of study design concept should be assessed and reviewed periodically. Other technical documents may also be applicable such as Smartphone Secure Development Guidelines (2017) published by the European Network and Information Security Agency (ENISA), which advises on design and technical solutions. The principles of these security measures are found in the European Data Protection Supervisor (EDPS) opinion on mobile health (Opinion 1/2015 Mobile Health-Reconciling technological innovation with data protection.

9. Research networks for multi-database studies

9.1. General considerations

A growing number of pharmacoepidemiological studies use data from networks of databases, often from different countries. Pooling data across different databases affords insight into the generalisability of the results and may improve precision. Some of these networks are based on long-term contracts with selected partners and are very well structured (such as Sentinel, the Vaccine Safety Datalink (VSD), the Canadian Network for Observational Drug Effect Studies (CNODES)) and the recently set-up Data Analysis and Real World Interrogation Network (DARWIN EU®). Others are collaborations based on open science principles such as the Observational Health Data Sciences and Informatics (OHDSI) program.

In Europe, collaborations for multi-database studies have been strongly encouraged as part of the drug safety research funded by the European Commission (EC) as well as public-private partnerships such as the Innovative Medicines Initiative (IMI). This funding resulted in the conduct of groundwork necessary to overcome the hurdles of data sharing across countries for specific projects (e.g. PROTECT, ADVANCE, EMIF, EHDEN, ConcePTION) and specific post-authorisation studies. The European Commission is currently establishing an European Health Data Space (EHDS) and major

breakthroughs in this field are expected, with the Joint Action Towards the European Health Data Space – TEHDAS developing joint European principles for the secondary use of health data.

The 2009 H1N1 influenza pandemic (see Safety monitoring of Influenza A/H1N1 pandemic vaccines in EudraVigilance, Vaccine 2011;29(26):4378-87) and the 2020 COVID-19 pandemic showed the value of an operational infrastructure to rapidly and effectively monitor the safety of therapeutics and vaccines. In this context, EMA established contracts with academic and private partners to support readiness of research networks to perform observational research. Three dedicated projects started in 2020: ACCESS (vACcine Covid-19 monitoring readinESS), CONSIGN (COVID-19 infectiOn aNd medicineS In preGNancy) and E-CORE (Evidence for COVID-19 Observational Research Europe). Other initiatives have emerged to address specific COVID-19 related research questions, such as the CVD-COVID-UK consortium (see Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource, BMJ. 2021;373:n826), providing a secure access to linked health data from primary and secondary care, registered deaths, COVID-19 laboratory data, vaccination data and cardiovascular specialist audits. Similarly, linked data have been made available in trusted research environments in Scotland and Wales.

EMA funded several studies to address research questions on the monitoring of COVID-19 vaccines using federated analytics with a common data model (CDM), which resulted in publications on background rates of adverse events of special interest (Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study, BMJ. 2021;373:n1435; Background rates of 41 adverse events of special interest for COVID-19 vaccines in 10 European healthcare databases - an ACCESS cohort study, Vaccine. 2023;41(1):251-262); thrombosis and risk of coagulopathy post-COVID-19 (Venous or arterial thrombosis and deaths among COVID-19 cases: a European network cohort study, Lancet Infectious Diseases 2022;22(8):1142-52); comparative risk of thrombosis and thrombocytopenia following COVID-19 vaccines (Comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with different covid-19 vaccines: international network cohort study from five European countries and the US, BMJ. 2022;379:e071594); and myocarditis (Myocarditis and pericarditis associated with SARS-CoV-2 vaccines: A population-based descriptive cohort and a nested self-controlled risk interval study using electronic health care data from four European countries, Front Pharmacol. 2022;13:1038043).

In this Chapter, the term networking is used to reflect collaboration between researchers for sharing expertise and resources. The ENCePP Database of Research Resources, which provides an inventory of research centres and data sources collaborating on specific pharmacoepidemiology and pharmacovigilance studies in Europe, may facilitate such networking by allowing the identification of research centres and data sources by country, study, type of research, and other relevant fields.

The use of research networks in medicines safety and utilisation, and in disease epidemiology, is well established, with a significant body of practical experience. Their use in effectiveness research is now increasing (see Assessing strength of evidence for regulatory decision making in licensing: What proof do we need for observational studies of effectiveness?, Pharmacoepidemiol Drug Saf. 2020;29(10):1336-40).

From a methodological point of view, studies adopting a multi-database design have many advantages over single database studies:

- It increases the *size* of the study population. This especially facilitates research on rare events, on medicines used in specialised settings (see Ability of primary care health databases to assess medicinal products discussed by the European Union Pharmacovigilance Risk Assessment Committee, Clin Pharmacol Ther. 2020;107(4):957-65), or when the interest is in subgroup effects.

- It exploits the *heterogeneity* of treatment options across countries, which allows studying the effect of different medicines used for the same indication, or specific patterns of utilisation.
- It exploits differences in outcome/event rates across countries/regions.
- It provides additional knowledge on the *generalisability* of results and on the *consistency of associations*, for instance whether a safety issue can be identified in several countries. Possible inconsistencies might be caused by different biases or truly different effects in the databases, revealing causes of differential effects, and these might be investigated.
- It involves experts from various countries addressing case definitions, terminologies, coding in databases, and research practices. This provides opportunities to increase *consistency of results* of observational studies.
- For primary data collection from multiple data sources, it *shortens the time* needed for obtaining the desired sample size and therefore accelerates the investigation of safety issues or other outcomes.

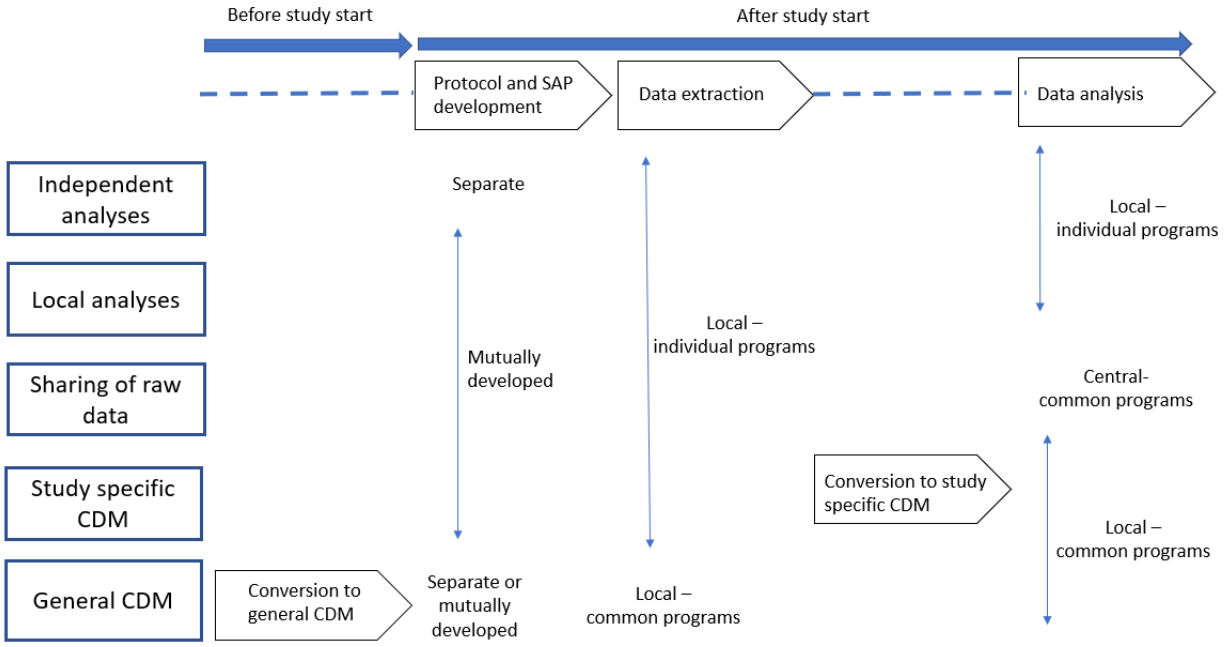
The articles Approaches for combining primary care electronic health record data from multiple sources: a systematic review of observational studies (BMJ Open 2020;10(10): e037405) and Different strategies to execute multi-database studies for medicines surveillance in real world setting: a reflection on the European model (Clin Pharmacol Ther. 2020;108(2):228-35) describe key characteristics of studies using multiple data sources and different models applied for combining data or results from multiple databases. A common characteristic of all models is the fact that data partners maintain physical and operational control over electronic data in their existing environment, and therefore, the data extraction is always performed locally. Differences, however, exist in the following areas: use of a common protocol; use of a CDM; and where and how the data analysis is conducted.

Use of a CDM implies that local formats are translated into a predefined, common data structure, which allows launching a similar data extraction and analysis script across several databases. Sometimes the CDM also imposes a common terminology, such as for the OMOP CDM. The CDM can be systematically applied on the entire database (generalised CDM) or on the subset of data needed for a specific study (study-specific CDM). While transforming the database in a CDM, comparisons between source and target data across all variables and dimensions is strongly recommended as part of the quality control of the process, in order to make sure that the transformation faithfully represents the source data, both in terms of completeness and accuracy. A number of tools exist for checking the resulting data, including the OHDSI DataQualityDashboard, which involves thousands of checks for conformance, completeness, and plausibility, based on the harmonised framework for data quality assessment developed by Khan et al. (EGEMS 2016;4(1):1244).

In the European Union, study specific CDMs have generated results for several projects, and several databases have been converted to a generalised CDM version that exists alongside the native version. This conversion was accelerated as a result of the observational research needed to respond to the COVID-19 pandemic. An example of application of generalised CDMs are studies conducted in the OHDSI community, such as Association of angiotensin converting enzyme (ACE) inhibitors and angiotensin 2 receptor blockers (ARB) on COVID-19 incidence and complications or the ConcePTION study From Inception to ConcePTION: Genesis of a Network to Support Better Monitoring and Communication of Medication Safety During Pregnancy and Breastfeeding (Clin Pharmacol Ther. 2022;111(1):321-31). More recently, DARWIN EU® has galvanised the use of the OMOP CDM for regulatory purposes, with the completion of the first studies, and the planned commissioning of many additional studies in the coming years (see list of completed DARWIN EU® studies).

9.2. Models of studies using multiple data sources

Studies may be classified into five categories according to specific choices in the steps needed for their execution, i.e., protocol and statistical analysis plan (SAP) development, location of data extraction and analysis (locally or centrally), methods for data extraction and analysis (using individual or common programs, use of a CDM, and which type of CDM: study-specific or general CDM). The key steps needed to execute each study model are presented in the following Figure and explained in this section.



9.2.1. Independent analyses: separate protocols, local and individual data extraction and analysis, no CDM

The traditional model to combine data from multiple data sources consists in data extraction and analysis performed independently at each centre, based on separate protocols. This is usually followed by a meta-analysis of the different estimates obtained (see Chapter 10 and Annex 1).

This type of model, when viewed as a means to combine results from multiple data sources on the same research questions, may be considered as a baseline situation which a research network should try to improve on for the study design. Meta-analyses also facilitate the evaluation of heterogeneity of results across different independent studies and they could be performed retrospectively regardless of the model of studies used, in line with the recommendations from the Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project (Pharmacoepidemiol Drug Saf. 2016;25(S1):156-65). Investigating heterogeneity may provide useful information on the issue under investigation, and explaining such variation should also be attempted if the data sources can be accessed. An example of such an investigation is Assessing heterogeneity of electronic health-care databases: A case study of background incidence rates of venous thromboembolism (Pharmacoepidemiol Drug Saf. 2023 Apr 17. doi: 10.1002/pds.5631). This approach increases consistency in findings from observational drug effect studies or reveals causes of differential drug effects.

9.2.2. Local analysis: common protocol, local and individual data extraction and analysis, no CDM

In this model, data are extracted and analysed locally, with site-specific programs developed by each centre, on the basis of a common protocol and a common SAP agreed by all study partners. The common SAP defines and standardises exposures, outcomes and covariates, analytical programmes and reporting formats. The results of each analysis, either at the subject level or in an aggregated format depending on the governance of the network, are shared and can be pooled together using meta-analysis.

This approach allows the assessment of database or population characteristics and their impact on estimates, but it reduces the variability of results determined by differences in design. Examples of research networks that use the common protocol approach are PROTECT (as described in Improving Consistency and Understanding of Discrepancies of Findings from Pharmacoepidemiological Studies: the IMI PROTECT Project, Pharmacoepidemiol Drug Saf. 2016;25(S1): 1-165), which has implemented this approach in collaboration with CNODES (see Major bleeding in users of direct oral anticoagulants in atrial fibrillation: A pooled analysis of results from multiple population-based cohort studies, Pharmacoepidemiol Drug Saf. 2021;30(10):1339-52).

This approach requires very detailed common protocols and data specifications that reduce variability in interpretation by researchers.

9.2.3. Sharing of data: common protocol, local and individual data extraction, central analysis

In this approach, a common protocol is agreed by the study partners. Data intended to be used for the study are locally extracted with site-specific programs, transferred without analysis and conversion to a CDM, and pooled and analysed at the central partner receiving them. Data received at the central partner can be reformatted to a common structure to facilitate the analysis.

This approach applies when databases are very similar in structure and content, as for some Nordic registries and the Italian regional databases. Examples of such models are Protocol: Methodology of the brodalumab assessment of hazards: a multicentre observational safety (BRAHMS) study (BMJ. Open 2023;13(2):e066057) and All-cause mortality and antipsychotic use among elderly persons with high baseline cardiovascular and cerebrovascular risk: a multi-center retrospective cohort study in Italy (Expert Opin. Drug Metab. Toxicol. 2019;15(2):179-88).

The central analysis allows for assessment of pooled data adjusting for covariates on an individual patient level and removing an additional source of variability linked to the statistical programming and analysis. However, this model becomes more difficult to implement due to the stronger privacy requirements for sharing patient level data.

9.2.4. Study specific CDM: common protocol, local and individual data extraction, local and common analysis, study specific CDM

In this approach, a common protocol is agreed by the study partners. Data intended to be used for the study are locally extracted and transformed into an agreed CDM. The data in the CDM are then processed locally in every site with one common program. The output of the common program is transferred to a specific partner. The output to be shared may be an analytical dataset or study estimates, depending on the governance of the network. This model is explained in From Inception to ConcePTION: Genesis of a Network to Support Better Monitoring and Communication of Medication Safety During Pregnancy and Breastfeeding (Clin Pharmacol Ther. 2022;111(1):321-31).

Examples of research networks that used this approach by employing a study-specific CDM with transmission of anonymised patient-level data (allowing a detailed characterisation of each database) are EU-ADR (as explained in Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how?, J Intern Med 2014;275(6):551-61), SOS, ARITMO, SAFEGUARD, GRIP, EMIF, EUROmediCAT, ADVANCE, VAC4EU and ConcePTION. In all these projects, a CDM was utilised, and R, SAS, STATA or Jerboa scripts used to create and share common analytics. Diagnosis codes for case finding can be mapped across terminologies by using the Codemapper developed in ADVANCE (see CodeMapper: semiautomatic coding of case definitions, Pharmacoepidemiol Drug Saf. 2017;26(8):998-1005). An example of a study performed using this model is Background rates of 41 adverse events of special interest for COVID-19 vaccines in 10 European healthcare databases - an ACCESS cohort study, Vaccine. 2023;41(1):251-262).

9.2.5. General CDM: common protocol, local and common data extraction and analysis, general CDM

In this approach, the local databases are transformed into a CDM prior to, and are agnostic to, any study protocol. When a study is required, a common protocol is developed and a centrally created analysis program is created that runs locally on each database to extract and analyse the data. The output of the common programs shared may be an analytical dataset or study estimates, depending on the governance of the network.

Examples of research networks which use a generalised CDM are the Sentinel Initiative (as described in The US Food and Drug Administration Sentinel System: a national resource for a learning health system, Journal of the American Medical Informatics Association, Volume 29, December 2022, Pages 2191–2200) OHDSI – Observational Health Data Sciences and Informatics, the Canadian Network for Observational Drug Effect Studies (CNODES), and EMA's Data Analysis and Real World Interrogation Network (DARWIN EU®). The latter uses the same CDM as OHDSI, and combines previously existing analytical pipelines with bespoke newly developed ones, based on an EMA-endorsed catalogue of Standardised Analytics.

The main advantage of a general CDM is that it can be used for nearly any study involving the same database converted into the CDM. OHDSI and DARWIN EU® are based on the Observational Medical Outcomes Partnership (OMOP) CDM which is now used by many organisations and has been tested for its suitability for safety studies (see, for example, Validation of a common data model for active safety surveillance research, J Am Med Inform Assoc. 2012;19(1):54–60; and Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?: A Validation Study Based on Prescribing Codeine in Children, Clin Pharmacol Ther. 2020;107(4):915-25). Conversion into the OMOP CDM requires formal mapping of database items to standardised concepts, which is a resource intensive and iterative process. Iterations on the same databases usually lead to gains in efficiency. Mapping expertise and software are also constantly developed to support and accelerate the conversion process. Examples of studies performed with the OMOP CDM in Europe are Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study (J Am Med Inform Assoc. 2020;27(8):1268-77); Safety of hydroxychloroquine, alone and in combination with azithromycin, in light of rapid wide-spread use for COVID-19: a multinational, network cohort and self-controlled case series study (Lancet Rheumatol. 2020;11(2):e698–711); Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study (BMJ. 2021;373:n1435); Venous or arterial thrombosis and deaths among COVID-19 cases: a European network cohort study (Lancet Infectious Diseases 2022;22(8):P1142-52); and Comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events

associated with different covid-19 vaccines: international network cohort study from five European countries and the US (BMJ. 2022;379:e071594).

In A Comparative Assessment of Observational Medical Outcomes Partnership and Mini-Sentinel Common Data Models and Analytics: Implications for Active Drug Safety Surveillance (Drug Saf. 2015;38(8):749-65), it is suggested that slight conceptual differences between the Sentinel and the OMOP models do not significantly impact on identifying known safety associations. Differences in risk estimations can be primarily attributed to the choices and implementation of the analytic approach.

A review of IT-architecture, legal considerations, and statistical methods for federated analyses is presented in Federated analyses of multiple data sources in drug safety studies (Pharmacoepidemiol Drug Saf. 2023 Mar;32(3):279-286).

9.3. Challenges of different models

The different models described above present several challenges, as detailed below.

Related to the database content:

- Differences in the underlying health care systems,
- Different mechanisms of data generation and collection as well as data availability,
- Mapping of different drugs and disease dictionaries (e.g., SNOMED, the International Classification of Disease, 10th Revision (ICD-10), Read codes),
- Free text medical notes in different languages,
- Differences in the validation of study variables and access to source documents for validation,
- Differences in the type and quality of information contained within each database.

Related to the organisation of the network:

- Different ethical and governance requirements in each country regarding processing of anonymised or pseudo-anonymised healthcare data,
- Issues linked to intellectual property and authorship,
- Implementing quality controls procedures at each partner and across the entire network,
- Sustainability and funding mechanisms,
- The networks tend to become very topic specific over time and to become isolated in 'silos'.

Each model has strengths and weaknesses in facing the above challenges, as illustrated in Data Extraction and Management in Networks of Observational Health Care Databases for Scientific Research: A Comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies (eGEMs 2016;4(1):2). In particular, a central analysis or a CDM provide protection from problems related to variation in how protocols are implemented by individual analysts (as described in Quantifying how small variations in design elements affect risk in an incident cohort study in claims; Pharmacoepidemiol Drug Saf. 2020;29(1):84-93). Several of the networks have made their codes, common data models and analytics software publicly available, such as OHDSI, DARWIN EU®, Sentinel, VAC4EU. This is one of the potential solutions to minimise reproducibility issues in multi-database studies.

Timeliness or speed of running studies is important in order to meet short regulatory timelines, in circumstances where prompt decision-making is needed. Solutions need therefore to be further developed and introduced to be able to run multi-database studies with shorter timelines.

Independently from the model used, a critical factor that should be considered in speeding up studies relates to having tasks completed that are independent of any particular study. This includes all activities associated with governance, such as having prespecified agreements on data access, processes for protocol development and study management, and identification and characterisation of a large set of databases. This also includes some activities related to the analysis, such as creating common definitions for frequently used variables, and creating common analytical systems for the most typical and routine analyses. This latter point is made easier with the use of CDMs with standardised analytics and tools that can be re-used to support faster analysis, as demonstrated in DARWIN EU®, where analytical pipelines are being developed to fulfil the needs of EMA-commissioned studies based on pre-specified analysis plans (see Catalogue of Standard Analyses).

10. Systematic reviews and meta-analysis

Identification and integration of evidence derived from results from several studies with the same or similar research objective can extend our understanding of the research question. A systematic literature review aims to collect in a systematic and explicit manner all empirical evidence that fits pre-specified eligibility criteria to answer a specific research question and to critically appraise relevant results. A meta-analysis involves the use of statistical techniques to integrate and summarise the results of identified studies. The focus of this activity may be to learn from the diversity of designs, results and associated gaps in knowledge as well as to obtain overall risk estimates. An example of a systematic review and meta-analysis of results of individual studies with potentially different designs is given in Variability in risk of gastrointestinal complications with individual NSAIDs: results of a collaborative meta-analysis (BMJ. 1996;312(7046):1563-6), which compared the relative risk of serious gastrointestinal complications reported with individual NSAIDs by conducting a systematic review of twelve hospital and community-based case-control and cohort studies and found a relation between use of the drugs and admission to hospital for haemorrhage or perforation.

Systematic review and meta-analysis of observational studies and other epidemiological sources are becoming as common as those of randomised clinical trials (RCTs). Challenges in systematic reviews that assess treatment harms (Ann Intern Med. 2005;142:1090-9) explains the different reasons why both are important in providing relevant information and knowledge for pharmacovigilance. However, the method of analysis differs when meta-analyses pool evidence from observational studies compared to RCTs. This also applies to network meta-analyses that are now being performed using observational studies (see NMA incorporating RWE, Overview of evidence synthesis and network meta-analysis – RWE Navigator, and Chapter 16.1 of this Guide on comparative effectiveness research). It is also increasingly common to perform a meta-analysis of individual patient data which allows for additional opportunities but has challenges.

A detailed guidance on the methodological conduct and reporting of systematic reviews and meta-analysis is reported in Annex 1 of this Guide and includes links to several relevant resources.

It should be noted that meta-analyses, even of RCTs, shares characteristics with observational research as subjective criteria are often involved in the selection of studies to include. Careful planning in design of a meta-analysis and pre-specification of selection criteria, outcomes and analytical methods before review of any study results may thus contribute to the confidence placed in the results. A further useful reference is the CIOMS Working Group X Guideline on Evidence Synthesis and Meta-Analysis for Drug Safety (Geneva 2016).

Framework for the synthesis of non-randomised studies and randomised controlled trials: a guidance on conducting a systematic review and meta-analysis for healthcare decision making (BMJ Evid Based Med. 2020 Dec 9:bmjebm-2020-111493) provides guidance on when and how to combine evidence from observational studies and RCTs.

Renin-angiotensin-system inhibitors and all-cause mortality in patients with COVID-19: a systematic review and meta-analysis of observational studies (J Hypertens. 2021;39(4):784-794) included 31 cohort and three case-control studies which were combined and analysed separately due to the inherent differences in study designs. The meta-analysis Lopinavir/Ritonavir for COVID-19: a Systematic Review and Meta-Analysis (J Pharm Pharm Sci. 2021;24:246-57) compared the efficacy and safety of lopinavir/ritonavir to other treatment options using data up to April 2021 from 14 studies, including data from clinical registry databases. Safety of Tenofovir Disoproxil Fumarate (TDF) for Pregnant Women facing the COVID-19 Pandemic (Am J Epidemiol. 2021:kwab109) assessed the teratogenicity of tenofovir using a claims-based pregnancy cohort of women with HIV and combined the results with prior data by conducting a systematic literature review and a meta-analysis.

11. Signal detection methodology and application

Note: Chapter 11 (formerly 10) has not been updated for Revision 11 of the Guide, as contents remain up-to-date.

11.1. General aspects of signal detection

A general overview of methods for signal detection and recommendations for their application are provided in the report of the CIOMS Working Group VIII Practical aspects of signal detection in pharmacovigilance; empirical results on various aspects of signal detection obtained from the IMI PROTECT project have been summarised in Good Signal Detection Practices: Evidence from IMI PROTECT. (Drug Saf. 2016;39(6):469-90).

The EU Guideline on good pharmacovigilance practices (GVP) Module IX (Rev 1)- Signal Management defines signal management as the set of activities performed to determine whether, based on an examination of individual case safety reports (ICSRs), aggregated data from active surveillance systems, studies, literature information or other data sources, there are new risks associated with an active substance or a medicinal product or whether risks have changed. Signal management covers all steps from detecting signals (signal detection), through their validation and confirmation, analysis, prioritisation and assessment to recommending action, as well as the tracking of the steps taken and of any recommendations made. The Guideline on good pharmacovigilance practices (GVP) - Module IX Addendum I – Methodological aspects of signal detection from spontaneous reports of suspected adverse reactions describes the components of an effective signal detection system and lists some of the methodological aspects that have been proved to be effective and that should be considered. Implementation details of such a system are not provided as they may be database dependent.

The FDA's Guidance for Industry-Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment provides best practice for documenting, assessing and reporting individual case safety reports and case series and for identifying, evaluating, investigating and interpreting safety signals, including recommendations on data mining techniques and use of pharmacoepidemiological studies.

11.2. Methods of statistical signal detection

Quantitative analysis of spontaneous adverse drug reaction reports is routinely used in drug safety research. Several articles have been published on statistical signal detection. Quantitative signal detection using spontaneous ADR reporting (Pharmacoepidemiol Drug Saf. 2009;18(6):427-36) describes the core concepts behind the most common methods, the proportional reporting ratio (PRR), reporting odds ratio (ROR), information component (IC) and empirical Bayes geometric mean (EBGM). The authors also discuss the role of Bayesian shrinkage in screening spontaneous reports and the importance of changes over time in screening the properties of the measures.

Additionally, they discuss major areas of controversy (such as stratification and evaluation and implementation of methods) and give some suggestions as to where emerging research is likely to lead. Data mining for signals in spontaneous reporting databases: proceed with caution (Pharmacoepidemiol Drug Saf. 2007;16(4):359–65) reviews data mining methodologies and their limitations and provides useful points to consider before incorporating data mining as a routine component of any pharmacovigilance program. Disproportionality Analysis for Pharmacovigilance Signal Detection in Small Databases or Subsets: Recommendations for Limiting False-Positive Associations (Drug Saf. 2020;43(5):479-87) evaluates the impact of database size on the performance of disproportionality analysis, with regards to limiting spurious associations.

Methods such as multiple logistic regression have the theoretical capability to reduce masking and confounding by co-medication and underlying disease. Regression-Adjusted GPS Algorithm describes the use of regression to increase the discriminatory power of the Gamma Poisson Shrinkage (GPS) algorithm. Data-Driven Prediction of Drug Effects and Interactions (Sci Transl Med. 2012 Mar 14; 4(125): 125ra31) describes the application of regression methods to correct for synthetic associations caused by hidden, or unmeasured, covariates as well as those from indication and concomitant drug use. The letter Logistic regression in signal detection: another piece added to the puzzle (Clin Pharmacol Ther. 2013;94(3):312) highlights the variability of results obtained in different studies based on this method and the daunting computational task it requires. More work is needed on its value for pharmacovigilance in the real-world setting.

A more recent proposal involves a broadening of the basis for computational screening of individual case safety reports, by considering multiple aspects of the strength of evidence in a predictive model. This approach combines disproportionality analysis with features such as the number of well-documented reports, the number of recent reports and geographical spread of the case series (Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank, Drug Saf. 2014;37(8):617–28). In a similar spirit, logistic regression has been proposed to combine a disproportionality measure with a measure of unexpectedness for the time-to-onset distribution (Use of logistic regression to combine two causality criteria for signal detection in vaccine spontaneous report data, Drug Saf. 2014;37(12):1047-57). In A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands (Pharmacoepidemiol Drug Saf. 2018;27(2):199-205), five relevant characteristics (number of reports, disproportionality, Naranjo score, proportion of MAH reports, proportion of HCP reports) were chosen as potential predictors in the model and tested against the presence in the Summary of Product Characteristics (SmPC) of each unique drug-ADR association at the time of the analysis. All candidate predictors were included into the final model with an increased screening efficiency. The authors comment that the choice of candidate predictors may depend on each spontaneous report databases but that the method of generating a prediction model-based priority list of signals could be useful in other databases.

Methods for statistical signal detection tend to classify reports based on reported adverse event terms considered one at time. Broader categories such as High-Level Terms or Standardized MedDRA Queries are sometimes used to group similar adverse events and improve sensitivity. However, this may be at the expense of specificity. Consensus clustering for case series identification and adverse event profiles in pharmacovigilance (Artif Intell Med, 2021; 122:1-9) proposes a different approach where cluster analysis attempts to identify case series describing similar clinical conditions, accounting for the complete sets of signs, symptoms, and diagnoses on each report.

Disproportionality methods are usually calculated on the cumulative data and therefore do not provide a direct insight into temporal changes in frequency of reports. Methodologies to monitor changes in the frequency of reporting over time have been developed with the focus to enhance pharmacovigilance when databases are small, when drugs have established safety profiles and/or when product quality

defects, medication errors and cases of abuse or misuse are of concern. Automated method for detecting increases in frequency of spontaneous adverse event reports over time (J Biopharm Stat. 2013; 23(1):161-77) presents a regression method with both smooth trend and seasonal components, while An algorithm to detect unexpected increases in frequency of reports of adverse events in EudraVigilance (Pharmacoepidemiol Drug Saf. 2018;27(1):38-45) presents the testing of a model based on a negative binomial time-series regression model on thirteen historical concerns. Additionally, a modification of the Information Component to screen for spatial-temporal disproportionality is described in Using VigiBase to Identify Substandard Medicines: Detection Capacity and Key Prerequisites (Drug Saf. 2015; 38(4): 373–82). Despite the promising results of these methods, and even if theoretically they seem appealing, limited work has been performed to assess their effectiveness.

11.3. Triage of statistical safety signals

The revised guidance on Screening for adverse reactions in EudraVigilance describes methods for screening adverse drug reactions used by the European Medicines Agency and national competent authorities. The proposed methods complement the classical disproportionality analysis with additional data summaries, based on both statistical and clinical considerations. This approach is based on the fact that, although disproportionality methods have demonstrated to detect many adverse reactions before other currently used methods of signal detection, this is not true for all types of adverse reactions.

Hence a comprehensive and efficient routine signal detection system will seek to integrate a number of different methods to prioritise the drug event combinations for further evaluation. For the methods recommended, the guidance addresses elements of their interpretation, their potential advantages and limitations and the evidence behind. Areas of uncertainty that require resolution before firm recommendations can be made are also mentioned.

As understanding increases regarding the mechanisms at a molecular level that are involved in adverse effects of drugs it would be expected that this information will inform efforts to predict and detect drug safety problems. Such modelling is presented in Data-driven prediction of drug effects and interactions (Sci Transl Med. 2012 14;4(125):125ra31) and should be a major focus of drug safety research activities. An example of an application of this concept is illustrated in the paper Cheminformatics-aided pharmacovigilance: application to Stevens-Johnson Syndrome (J Am Med Inform Assoc. 2016; 23(5): 968–78) where the authors apply a Quantitative Structure-Activity Relationship (QSAR) model to predict the drugs associated with Stevens Johnson syndrome in a pharmacovigilance database. In Target Adverse Event Profiles for Predictive Safety in the Postmarket Setting (Clin Pharmacol Ther. 2021;109(5):1232-43), the authors identify drugs that share pharmacological targets with the drug of interest and use information from these drugs to predict post-marketing adverse drug reactions of the drug of interest. Machine learning on data from the FDA Adverse Event Reporting System, peer-reviewed literature and FDA drug labels is used for the prediction. In Role of serotonin and norepinephrine transporters in antidepressant-induced arterial hypertension: a pharmacoepidemiological-pharmacodynamic study (Eur J Clin Pharmacol. 2020 Sep;76(9):1321-1327.) disproportionality analysis on Vigibase was combined with a pharmacodynamic study to study the relationship between SRIs ands SNRIs and arterial hypertension, taking in to account the affinity for noradrenergic and serotonergic receptors.

With pharmacovigilance databases increasing in size, manual review of all cases becomes a non-scalable process both because the increasing number of cases to review in each potential signal and because it is difficult to summarise hundreds of case reports in a narrative format. To address some of these issues there has been recent experimentation with machine learning and natural language

processing techniques. Towards Automating Adverse Event Review: A Prediction Model for Case Report Utility (Drug Saf. 2020 Apr;43(4):329-338) notes the need to develop modernised pharmacovigilance practices and shows the feasibility of developing a tool predictive of ICSR utility. Feature engineering and machine learning for causality assessment in pharmacovigilance: Lessons learned from application to the FDA Adverse Event Reporting System (Comput Biol Med. 2021 Aug;135:104517) describes the use of machine learning techniques to quickly eliminate non-assessable reports.

11.4. Performance comparison of signal detection methods

The role of data mining in pharmacovigilance (Expert Opin Drug Saf. 2005;4(5):929-48) explains how signal detection algorithms work and addresses questions regarding their validation, comparative performance, limitations and potential for use and misuse in pharmacovigilance.

An empirical evaluation of several disproportionality methods in a number of different spontaneous reporting databases is given in Comparison of statistical detection methods within and across spontaneous reporting databases (Drug Saf. 2015;38(6);577-87).

Performance of pharmacovigilance signal detection algorithms for the FDA adverse event reporting system (Clin Pharmacol Ther. 2013;93(6):539-46) describes the performance of signal-detection algorithms for spontaneous reports in the US FDA adverse event reporting system against a benchmark constructed by the Observational Health Data Sciences and Informatics community. It concludes that logistic regression performs better than traditional disproportionality analysis. Other studies have addressed similar or related questions, for examples Large-scale regression-based pattern discovery: The example of screening the WHO global Drug Safety database (Stat Anal. Data Min. 2010;3(4):197–208), Are all quantitative postmarketing signal detection methods equal? Performance characteristics of logistic regression and Multi-item Gamma Poisson Shrinker (Pharmacoepidemiol Drug Saf. 2012; 21(6):622–30 and Data-driven prediction of drug effects and interactions (Sci Transl Med. 2012;4(125):125ra31).

11.5. Stratification and sub-group analyses

Many statistical signal detection algorithms disregard the underlying diversity and give equal weight to reports on all patients when computing the expected number of reports for a drug-event pair. This may give them vulnerability to confounding and distortions due to effect modification, and could result in true signals being masked or false associations being flagged as potential signals. Stratification and/or subgroup analyses might address these issues, and whereas stratification is implemented in some standard software packages, routine use of subgroup analyses is less common. Performance of stratified and subgrouped disproportionality analyses in spontaneous databases (Drug Saf. 2016; 39(4):355-64) performed a comparison across a range of spontaneous report databases and covariates and found subgroup analyses to improve first pass signal detection, whereas stratification did not; subgroup analyses by patient age and country of origin were found to bring the greatest value.

11.6. Masking

Masking is a statistical issue by which true signals of disproportionate reporting are hidden by the presence of other products in the database. It is a phenomenon often observed when external factors, such as solicited schemes of reporting of adverse drug reactions or media attention, affect the reporting dynamics leading to a relative increase in the reporting rate for a specific medicinal product. As the change in reporting dynamics can be restricted in time and location, masking is not fully

understood, but can be highly impactful if the reporting dynamics change dramatically over a long period and across multiple countries, such as seen in the COVID-19 world-wide vaccination campaigns.

Publications have described methods assessing the extent and impact of the masking effect of measures of disproportionality. They include A conceptual approach to the masking effect of measures of disproportionality (Pharmacoepidemiol Drug Saf. 2014;23(2):208-17), with an application described in Assessing the extent and impact of the masking effect of disproportionality analyses on two spontaneous reporting systems databases (Pharmacoepidemiol Drug Saf. 2014;23(2):195-207), Outlier removal to uncover patterns in adverse drug reaction surveillance - a simple unmasking strategy (Pharmacoepidemiol Drug Saf. 2013;22(10):1119-29) and A potential event-competition bias in safety signal detection: results from a spontaneous reporting research database in France (Drug Saf. 2013;36(7):565-72). The value of these methods in practice needs to be further investigated.

11.7. Complementary role of databases

A time-consuming step in signal detection of adverse reactions is the determination of whether an effect is already recorded in the product information. A database which can be searched for this information allows filtering or flagging reaction monitoring reports for signals related to unlisted reactions, thus improving considerably the efficiency of the signal detection process by restricting attention to those drugs and adverse event not already considered causally related. In research, it permits an evaluation of the effect of background restriction on the performance of statistical signal detection. An example of such database is the PROTECT Database of adverse drug reactions (EU SPC ADR database), a structured Excel database of all adverse drug reactions (ADRs) listed in Chapter 4.8 of the SmPC of medicinal products authorised in the European Union (EU) according to the centralised procedure, based exclusively on the Medical Dictionary for Regulatory Activities (MedDRA) terminology. Efforts to identify adverse drug reactions in regulatory documents using natural language processing are being explored and could help build and maintain such databases in the future. ADE Eval: An Evaluation of Text Processing Systems for Adverse Event Extraction from Drug Labels for Pharmacovigilance (Drug Saf. 2021;44(1):83-94) presents a systematic evaluation of different such approaches.

Other large observational databases such as claims and electronic health records databases are potentially useful as part of a larger signal detection and refinement strategy. Modern methods of pharmacovigilance: detecting adverse effects of drugs (Clin Med 2009;9(5):486-9) describes the strengths and weaknesses of different data sources for signal detection (spontaneous reports, electronic patient records and cohort-event monitoring). A number of studies have considered the use of observational data in electronic systems that complement existing methods of safety surveillance e.g. the PROTECT, OHDSI and Sentinel projects. Useful Interplay Between Spontaneous ADR Reports and Electronic Healthcare Records in Signal Detection (Drug Saf. 2015;38(12):1201-10) investigates the potential of using electronic health records alongside spontaneous reporting systems to improve signal detection, concluding that the former may have additional value for adverse events with a high background incidence. Toward multimodal signal detection of adverse drug reactions (J Biomed Inform. 2017;76:41-9) concludes that utilising and jointly analysing multiple data sources may lead to improved signal detection but development of this approach requires a deeper understanding the data sources used, additional benchmarks and further research on methods to generate and synthesise signals.

12. Statistical analyses

12.1. General considerations

Compared to the protocol that includes a section outlining the analyses, the SAP is a more technical, stand-alone document describing in detail the planned analyses, population definitions and methodology.

Given the influence of statistical decisions on study conclusions, a well-documented and transparent statistical plan is essential. Developing a SAP forces researchers to think about which data to collect, in which format. This may then guide decisions on e.g., measurement instruments and timing of (repeated) measurements.

Further guidance on general principles and justification for the need for a SAP are provided in *Design of Observational Studies* (P.R. Rosenbaum, Springer Series in Statistics, 2020).

The following objectives of a SAP apply to most studies, including observational studies:

- *Transparency* as to how the analysis will proceed, by specifying in advance the methodology that will be applied. A SAP should always be completed prior to start of data analysis. Revisions after the start of the analysis might be possible, provided these changes are noted and justified in a revised SAP.
- *Communication* to the study team, especially statisticians, involved in the study. It promotes good planning and efficiency for other stakeholders such as reviewers and the target audience of the study. Readers of observational research might dismiss important findings if they were not pre-specified.
- *Reproducibility* so that in the future, for similar studies, the same analytical steps can be performed. The SAP should be sufficiently detailed so that it can be followed and reproduced by any statistician. Thus, it should provide clear and complete templates for each analysis.
- *Validity of study outcomes,* with the SAP enabling the researcher to separate the pre-planned analyses to address the research question from data-driven analyses, to understand and interpret the data.

Pre-specification of statistical and epidemiological analyses can be challenging for data that are not collected specifically to answer the research question. This is often the case in observational studies where secondary use of data is frequent (see Chapter 8.2). Nevertheless, The Value of Statistical Analysis Plans in Observational Research: Defining High-Quality Research From the Start (JAMA 2012;308(8):773-4) provides arguments to produce a SAP for observational research which is more vulnerable to issues of reproducibility. A main component of an observational study is an initial raw dataset including a set of variables that do not usually provide a direct answer to the research question. The SAP details the statistical calculations that will be performed on these observed data and the patterns of results that will in turn be interpreted.

Specific to observational studies, strong emphasis needs to be given to measures applied to control and possibly quantify bias. Avoiding bias in observational studies: part 8 in a series of articles on evaluation of scientific publications (Dtsch Arztebl Int. 2009;106(41):664-8) explains how these methodological issues can be avoided by careful planning. Factors that may bias the results of observational studies are described in Chapter 6.1. In this context, thoughtful specification of the way missing values will be handled and the use of a small part of the data as a pilot set to guide the analysis can be useful approaches. Handling of missing data is discussed in Chapter 6.3.

In some studies, analyses that are not pre-specified will be performed in response to observations in the data, in order to support interpretation of the results. It is important to distinguish between such data-driven analyses and pre-specified findings. Post-hoc modifications to the analytical strategy should be duly noted and justified in the revision history of the SAP.

12.2. Timing of the statistical analysis plan

The SAP is crucial for guiding data analysis and therefore, it is useful to formulate it at an early stage. In particular, the SAP should be developed before any informal inspection of aspects of the data or results that might influence opinions regarding the study hypotheses. Ideally, the SAP will be developed as soon as the protocol is finalised.

The SAP may be submitted to regulatory authorities as part of a submission package, e.g., as an appendix to the protocol and/or study report. The SAP is stored in the study master file, and is used during audits to check if statistical analyses are performed as planned. The role of the SAP is explained in the International Council for Harmonisation (ICH) E9 guideline 'Statistical principles for clinical trials, and these recommendations may also apply to observational studies.

A particular concern in retrospective studies is that decisions about the analysis should be made blinded to any knowledge of the results. This should be a consideration in the study design, particularly when feasibility assessments are to be performed to inform the design phase. Such feasibility assessments should be independent of the study results (see Chapter 2).

At any cost, a SAP should always be completed before the data have been unblinded for the statistician. This contributes to transparency of the study process and confirms that the set of analyses have not been influenced by the data. Making alterations to a planned statistical analysis after seeing the data increases the risk of bias, inflates the probability of type I errors, and might jeopardise the validity of the study findings, and acceptance of the study.

12.3. Information in the statistical analysis plan

A SAP is usually structured to reflect the protocol but will provide more granularity regarding the statistical methodology and population definition. Ideally, it includes and addresses the following elements in detail:

- Statistics on who wrote the SAP, its version number, when it was approved, and who signed it.
- Testable hypotheses to answer the study objectives (see Chapter 2). Defining primary and secondary objectives is important to avoid 'data dredging' and must correspond to the research question. A hypothesis is the product of deductive reasoning, going from general premises to specific results one would expect if those general premises were indeed true. This usually involves a set of possible relationships between a set of variables. It should be clearly stated how each outcome will be measured. Negative findings may be equally important as positive findings. Another reason to carefully choose the primary outcome is to minimise multiplicity effects. These occur when there are multiple statistical tests needed to assess the primary outcome, which increases the likelihood of false positives.
- Definitions of study variables. Outcome variables based on historical data may involve complex transformations to approximate clinical variables not explicitly measured in the dataset used. These transformations should be discriminated from those made to improve the fit of a statistical model. In either case, the rationale should be provided. In the latter case, this will include which tests of fit will be used and under what conditions a transformation will be used. Next to the outcomes, the other variables used in the study also need to be further formalised. The formatting

(e.g., categorisation, dichotomisation), modifications or derivations need to be described, with a special attention given to time-dependent variables (e.g., age, BMI).

- Study design (see Chapter 4) and sample size considerations. It should be noted that in observational studies using data that already exist and where no additional data is collected, sample size is not preclusive and the ethical injunction against 'underpowered' studies has no obvious force provided the results, in particular the 'absence of effect' and 'insufficient evidence', which should be properly presented and interpreted. The anticipated overall number of study subjects, as well as the minimum number per strata for stratified analyses, can be provided as an indicative guide.
- Interim analyses. If considered, interim analyses can be beneficial. In observational studies, there may be incentives to perform such analyses for early stopping of continued (costly) data collection due to already clear observed associations or futility. An interim analysis requires careful advanced planning and appropriate adjustments to the statistical approach.
- Data sources, study population and analysis set. This section of the SAP includes a description of the data sources and linkage methods, inclusion and exclusion criteria, withdrawal/follow-up, study entry/index date, baseline subject characteristics and potential confounding variables, and full analysis set. In addition, the section identifies which of the subjects are to be included in the different analyses. Attention should also be given to the creation and description of the study groups/cohorts and definition of subgroups, as applicable.
- Analytical methods. This section should describe effect measures and statistical methods used to address each study objective, such as:
  - Methods for dealing with missing data.
  - Methods for dealing with outliers.
  - Procedures for dealing with protocol variations, non-compliance, and withdrawals.
  - Methods for estimating points and intervals.
  - Rules for calculating composite or derived variables, including data-driven definitions and any additional details required to minimise ambiguity.
  - Baseline and covariate data used.
  - Definition of study period (study entry/index date, follow-up period, study exit)
  - Inclusion of randomisation factors (if applicable).
  - Methods for dealing with data from several locations/sources.
  - Methods for dealing with treatment interactions.
  - Methods for multiple comparisons and subgroup analysis.
  - Computer systems and statistical software packages used to analyse data.
- Statistical principles including confidence intervals and level of significance. The level of statistical significance to be employed, as well as whether one-tailed or two-tailed tests will be used, should be specified. Observational studies may be subject to repeated testing of accumulating data, which needs adjustment of significance levels to reduce inflated type-I errors (false positive findings). When false positives are a greater concern, a smaller confidence interval should be considered. Any planned adjustment of the significance level to control for type 1 error that can arise from comparisons across multiple subgroups or analysis of multiple predictors or outcomes (secondary

analyses) should be presented. However, different objectives of the study may require a lower or higher strength of evidence – for instance, policy recommendations regarding drug licensing may require a lower chance of false positive decisions when deciding whether further investigation is needed for a product safety issue. It should be noted that statistical packages often employ standard procedures – for instance default p-values (i.e., 5%) or confidence intervals (i.e., 95%).

- Sensitivity analyses. Sensitivity analyses allow to study the effect of potential violations of assumptions and/or results depending on specific observations (subjects) and are used to support the conclusions of the main analysis. Analyses to merely explore the data are considered exploratory analyses and should be described as such.
- Decision criteria. If decisions are drawn from the study results, a section of the SAP should explain the different outcomes that might be selected for each decision, which statistics influence the decision-making process, and which values of the statistics will be considered to support each outcome.
- Tables and figures for presentation of the study results. Skeleton tables should include a title, row labels and column entries be clearly spelled out, with only figures/numbers in the cells lacking. The analysis will produce the contents of the cells in a targeted manner, that is, hardly any other numbers will need to be generated. The same principle applies to graphs.

Consideration of the estimand framework is recommended to help informing choices regarding study design and data analysis and clarify how to interpret study findings. Tell me what you want, what you really really want: estimands in observational pharmacoepidemiologic comparative effectiveness and safety studies (Pharmacoepidemiol Drug Saf. 2023 Mar 22) discusses how defining an estimand is instrumental to the process of designing and analysing pharmacoepidemiological comparative effectiveness or safety studies. It applies the ICH Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical principles for Clinical Trial (2019) on estimands to three case studies and shows how defining an estimand ensures that the study targets a treatment effect that aligns with the treatment decision the study aims to inform.

For further reading on how to draft a SAP tailored to observational studies, see DEBATE-statistical analysis plans for observational studies (BMC Med Res Methodol. 2019;19(1):233); Guide to the statistical analysis plan (Pediatric Anesthesia 2019;29:237-42) which provides an exhaustive list of SAP items applicable to both prospective and retrospective observational studies; and The value of statistical analysis plans in observational research: defining high-quality research from the start (JAMA. 2012;308(8):773-4). A good example of a SAP where the main components are included can be found in Necrotizing soft tissue infections - a multicentre, prospective observational study (INFECT): protocol and statistical analysis plan (Acta Anaesthesiol Scand. 2018;62(2);272-79). Modern Epidemiology, 4th ed. (T. Lash, T.J. VanderWeele, S. Haneuse, K. Rothman. Wolters Kluwer, 2020) summarises the phases in a statistical analysis that should all be thought out and described beforehand.

13. Quality management

13.1. General principles of quality management

Quality in research is a measure of excellence that impacts medicines development and public health. What is quality management system (QMS)? (American Society for Quality, 2022) defines a QSM as a formalised system that documents processes, procedures, and responsibilities for achieving quality policies and objectives. Quality management principles as described in ISO Quality management principles are applicable to pharmacoepidemiological research. ISO 9000:2015 describes the fundamental concepts and principles of quality management which are universally applicable to

organisations and specifies the terms and definitions that apply to quality management and quality management system standards. The book *Total Quality Management-Key Concepts and Case Studies* (D.R. Kiran, BSP Books, Elsevier, 2016) deals with the management principles and practices that govern the quality function and presents all the aspects of quality control and management in practice.

The Commission Implementing Regulation (EU) No 520/2012 and the Good pharmacovigilance practices (GVP) Module I provide a framework for the quality management of pharmacovigilance and safety studies of authorised medicinal products.

Measurable quality can be achieved by:

- Quality planning: establishing structures (including validated computerised systems) and planning integrated and consistent processes;
- Quality assurance and control: monitoring and evaluating how effectively the structures and processes have been established and how effectively the processes are being carried out;
- Quality improvement: correcting and improving the structures and processes where necessary.

Pharmacoepidemiological research may be based on primary data collection or secondary use of data which is collected for other purposes (see Chapter 8). Primary data collection is a controlled process to which all steps of quality management should apply. Secondary use of data requires quality control addressing *a posteriori* data quality irrespective of its use (also part of the concept of *reliability* mentioned in the next section, e.g., detection of missing information, errors made during a transfer or conversion, outliers, duplicates, implausible values), as well as data quality in the context of its use for a specific study (also named *relevance*).

Pharmacoepidemiological research is also becoming more complex and may use a very large amount of data. In such situation, managing quality implies a risk-based approach. Risk-based quality management is incorporated as Good Clinical Practice expectation in ICH E8 (R1) and addressed in the European Commission's Risk proportionate approaches in clinical trials (2017), EMA's Reflection paper on risk-based quality management in clinical trials (2013) and GVP Module III on Pharmacovigilance inspections (2014).

The considerations and recommendations in Chapter 5. regarding the definition and validation of exposure, outcomes and covariates are essential aspects to be addressed for quality management. Adequate information on data sources is needed in order to identify real-world data sources and to assess their suitability for specific research questions. The HMA-EMA Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources (2022) provides recommendations on the identifiability of data sources based on the data elements described in the List of metadata for Real World Data catalogues (2022). This catalogue will also help assessing the quality of data sources proposed to be used in a study protocol or referred to in a study report.

13.2. Data quality frameworks

Large electronic data sources such as electronic healthcare records, insurance claims and other administrative data have opened up new opportunities for investigators to rapidly conduct pharmacoepidemiological studies and clinical trials in real-world settings, with a large number of subjects. A concern is that these data have not been collected systematically for research on the utilisation, safety or effectiveness of medicinal products, which could affect the validity, reliability and reproducibility of the investigation. Several data quality frameworks have been developed to understand the strengths and limitations of the data to answer a research question, the impact they may have on the study results, and the decisions to be made to complement available data. The

dimensions covered by these frameworks overlap, with different levels of details. Quality Control Systems for Secondary Use Data (2022) lists the domains addressed in several of them.

The following non-exhaustive list provides links to published data quality frameworks generally applicable to data sources, with a short description of their content.

- The draft HMA-EMA Data Quality Framework for EU medicines regulation (2022) provides general considerations on data quality that are relevant for regulatory decision-making, definitions for data dimensions and sub-dimensions, as well as ideas for their characterisation and related metrics. It also provides an analysis of what data quality actions and metrics can be put in place in different scenarios and introduces a maturity model to drive the evolution of automation to support data-driven regulatory decision making. The proposed data dimensions include Reliability (with sub-dimensions of Precision, Accuracy and Plausibility), Extensiveness (with sub-dimensions of Completeness and Coverage), Coherence (with the sub-dimensions of formal, structural and semantic coherence, Uniqueness, Conformance and Validity), Timeliness and Relevance.
- The European Health Data Space Data Quality Framework (2022) of the Joint Action Towards the European Health Data Space (TEHDAS) project has defined six dimensions deemed the most important ones at data source level: reliability, relevance, timeliness, coherence, coverage and completeness.
- Kahn's A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data (eGEMs. 2016;4(1):1244) describes a framework with three data quality categories: Conformance (with sub-categories of Value, Relational Conformance and Computational Conformance), Completeness, and Plausibility (with sub-categories of Uniqueness, Atemporal Plausibility and Temporal Plausibility). These categories are applied in two contexts: Verification and Validation. This framework is used by the US National Patient-Centered Clinical Research Network (PCORnet), with an additional component, Persistence, and by the Observational Health Data Science and Informatics (OHDSI) network. Based on this framework, the Data Analytics chapter of the Book of OHDSI (2021) provides an automated tool performing the data quality checks in databases conforming to the OMOP common data model. Increasing Trust in Real-World Evidence Through Evaluation of Observational Data Quality (J Am Med Inform Assoc. 2021;28(10):2251-7) describes an open source R package that executes and summarises over 3,300 data quality checks in databases available in OMOP.
- Duke-Margolis Center for Health Policy's Characterizing RWD Quality and Relevancy for Regulatory Purposes (2018) and Determining Real-World Data's Fitness for Use and the Role of Reliability (2019) specify that determining if a real-world dataset is fit-for-regulatory-purpose is a contextual exercise, as a data source that is appropriate for one purpose may not be suitable for other evaluations. A real-world dataset should be evaluated as fit-for-purpose if, within the given clinical and regulatory context, it fulfils two dimensions: Data Relevancy (including Availability of key data elements, Representativeness, Sufficient subjects and Longitudinality) and Data Reliability with two aspects: Data Quality (Validity, Plausibility, Consistency, Conformance and Completeness) and Accrual.

Data quality frameworks have been described for specific types of data sources and specific objectives. For example, the EMA's Guideline on Registry-based studies (2021) describes four quality components for use of patient registries (mainly disease registries) for regulatory purposes: Consistency, Completeness, Accuracy and Timeliness. A roadmap to using historical controls in clinical trials – by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG) (Orphanet J Rare Dis. 2020;15:69) describes the main sources of RWD to be used as historical controls, with an Appendix providing guidance on factors to be evaluated in the assessment of the relevance of RWD sources and resultant analyses.

Algorithms have been proposed to identify fit-for-purpose data to address research questions. For example, The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework (Clin Pharmacol Ther. 2022;111(1):122-34) and its update, A Structured Process to Identify Fit-for Purpose Study Design and Data to Generate Valid and Transparent Real-World Evidence for Regulatory uses (Clin Pharmacol Ther. 2023;113(6):1235-1239), aim to complement FDA's framework for RWE with a structured and detailed stepwise approach for the identification and feasibility assessment of candidate data sources for a specific study. The update emphasises the importance of initial study design, including designing a hypothetical target trial as a benchmark for the real-world study design before proceeding to data feasibility assessment. Whilst the approach of data feasibility assessment should be recommended, the complexity of some of the algorithms may discourage their use in practice. The experience will show to which extent they can support the validity and transparency of study results and ultimately the level of confidence in the evidence provided. It is also acknowledged that many investigators simply use the data source(s) they have access to and are familiar with in terms of potential bias, confounding and missing data.

13.3. Quality management in clinical trials

Rules, procedures, roles and responsibilities of quality assurance and quality control for clinical trials and biomedical research are well defined and described in many documents, such as the ICH E6 (R2) Good clinical practice, the European Forum for Good Clinical Practice (EFCGP) Guidelines, the Imperial College Academic Health Science Centre (AHSC)'s Quality Control and Quality Assurance SOP or the article Quality by Design in Clinical Trials: A Collaborative Pilot With FDA (Ther Innov Regul Sci. 2013; 47(2):161-6).

13.4. Quality management in observational studies

Quality management principles applicable to observational studies with primary data collection or secondary use of data are described in the Commission Implementing Regulation (EU) No 520/2012, GVP Module I, FDA's Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Health Care Data Sets, in recommendations from scientific societies such as the ISPE Guidelines for Good Pharmacoepidemiology Practices or the Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP): a guideline developed by the German Society for Epidemiology (Eur J Epidemiol. 2019;34(3):301-17), and in general epidemiology textbooks cited in the Introduction of this Guide. The Strengthening the Reporting of Observational studies in Epidemiology (STROBE) Statement Guidelines for reporting observational studies has established recommendations for improving the quality of reporting of observational studies and seeks to ensure a clear presentation of what was planned, done, and found.

The following articles are practical examples of quality aspects implementation or assessment in different settings:

- Poor reporting quality of observational clinical studies comparing treatments of COVID-19 - a retrospective cross-sectional study (BMC Med Res Methodol. 2022;22(1):2) found a poor reporting quality of observational studies on the treatment of COVID-19 throughout the year 2020 with a mean adherence of 45.6% to the STROBE checklist items in 147 observational studies.
- Quality of observational studies in prestigious journals of occupational medicine and health based on Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: a cross-sectional study (BMC Res Notes 2018;11:266) found that all sub-items of the STROBE statement were reported in 63.7%, not reported in 29.7% and not applicable in 6.6% of the 60 studies evaluated.

- Chapter 11 '*Data Collection and Quality Assurance*' of the Agency for Healthcare Research and Quality (AHRQ)'s Registries for Evaluating Patient Outcomes: A User's Guide, 4th Edition (2020) reviews key areas of data collection, cleaning, storage, and quality assurance for registries, with practical examples.
- Validation and validity of diagnoses in the General Practice Research Database (GPRD): a systematic review (Br J Clin Pharmacol. 2010;69:4-14) assesses the quality of the methods used to validate diagnoses in the former GPRD.
- Quality assurance in non-interventional studies (Ger Med Sci. 2009;7:Doc 29: 1-14) proposes measures of quality assurance that can be applied at different stages of non-interventional studies without compromising the character of non-intervention.

14. Dissemination and communication of study results

Aspects of dissemination and communication of study results include, but are not limited to, reports to health authorities and study sponsors, presentations in scientific fora, scientific publications, patient-focused communications, and websites dedicated to publishing study reports.

Important specific points relating to reporting of study results that are common to the various guidelines cited in this Chapter are that:

- Sources of research funding should always be disclosed whether in oral or written presentation of results.
- A dissemination and communication strategy should be pre-defined as part of the funding contract for a given study.
- All results with a scientific or public health impact must be reported to relevant authorities and made publicly available without undue delay.
- Quantitative measures of association should be reported rather than just results of statistical testing.

The Declaration of Helsinki provides overarching guidance on the registration, publication and dissemination of research results. Every research study involving human subjects must be registered in a publicly accessible database before recruitment of the first subject. A means to achieve this with pharmacoepidemiology and pharmacovigilance studies is through registration of protocols and reports of studies in the European Union electronic Register of Post-Authorisation Studies (EU PAS Register), ideally before they start. Making protocols and study results public in this Register is compulsory only for study imposed by regulators, however, transparent communication of research outside of regulatory requirements is encouraged. In addition to the EU PAS Register, study protocols and reports can also be registered and posted on other platforms: ClinicalTrials.gov includes specific guidelines for the posting of non-interventional research, while since 2020, the Open Science Forum has a specific registration portal for observational studies. The prospective register PROSPERO can be used for systematic reviews and meta-analyses.

Transparency of real-world evidence (RWE) studies starts with transparency of the study protocol. The European Medicines Agency (EMA) Guidance for the format and content of the protocol of non-interventional post-authorisation safety studies has been available since 2012 and provides the template required for study protocols submitted to EU regulators. Derived from an international consensus, the HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force (Pharmacoepidemiol Drug Saf. 2023;32(1):44-55) became available in 2023 to guide structure

and content of RWE study protocols with a focus on providing best possible detail of the operational study parameters used to create analytic datasets from the data collected for the study. A key component for design transparency, also included in the HARPER template, is the Graphical Depiction of Longitudinal Study Designs in Health Care Databases (Ann Intern Med. 2019;19;170(6):398-406), a framework of graphical representations that clarifies critical design choices. It helps researchers and reviewers to think systematically about time-related aspects in the context of typical study designs when designing studies or preparing manuscripts, which may increase confidence in evidence generated from non-randomised database studies.

Authors of publications should conform to the guidelines, including authorship criteria, established by the International Committee of Medical Journal Editors (ICJME) 'Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals'.

The ENCePP Code of Conduct states that "*the (primary) lead investigator shall be ultimately responsible for the design of the protocol, the conduct of the study, the analysis and interpretation of the study results and the preparation and publication of the study outcome.*"

The ISPE GPP contain a section on communication (section V) which includes a statement that there is an ethical obligation to disseminate findings of potential scientific or public health importance and that research sponsors (government agencies, private sector, etc.) shall be informed of study results in a manner that complies with local regulatory requirements. The European Medicines Agency (EMA) Guidance for the format and content of the protocol of non-interventional post-authorisation safety studies states that plans for disseminating and communicating study results are to be described as part of study planning activities.

The EMA Guidance for the format and content of the final study report of non-interventional post-authorisation safety studies (PASS) provides a template for final study reports that may be applied to any non-interventional study, including meta-analyses and systematic reviews. The FDA's Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Health Care Data Sets includes a description of all the elements that should be addressed and included in the final study report of such studies.

The Strengthening the Reporting of Observational studies in Epidemiology (STROBE) Statement Guidelines for reporting observational studies has established recommendations for improving the quality of reporting of observational studies and seeks to ensure a clear presentation of what was planned, done, and found. Of note, the aim of these guidelines was not to require the reporting of observational research in a rigid format, but to address what should be the essential information contained in a publication on an observational study. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement (PLoS Med. 2015;12(10):e1001885) was created as an extension to the STROBE statement to address reporting items specific to observational studies using routinely collected health data. RECORD makes additional recommendations on the reporting of methods of selection of study populations, exposures, outcomes and covariates (including codes or algorithms used), whether validation has been conducted, the level of access to databases used, and data linkages that were required to conduct the study. The RECORD-PE statement (BMJ. 2018;363:k3532) aims to extend existing STROBE and RECORD guidelines providing guidance for the reporting of pharmacoepidemiological studies using routinely collected data. Retraction of COVID-19 Pharmacoepidemiology Research Could Have Been Avoided by Effective Use of Reporting Guidelines (Clin Epidemiol. 2020;12:1403-1420) evaluated two retracted articles on the effectiveness and risk of hydroxychloroquine in COVID-19 patients, and demonstrated that transparent and complete reporting would have provided peer-reviewers and editors with sufficient information to question the methods used and the validity of the results.

The Good ReseArch for Comparative Effectiveness (GRACE) guidance includes recommendations on reporting comparative effectiveness studies. The STARD guidelines (BMJ Open 2016;14;6(11):e012799) focus on reporting diagnostic accuracy studies.

The Meta-analysis of Observational Studies in Epidemiology (MOOSE) group (JAMA. 2000;283(15):2008-15) has developed a consensus statement and recommendations for reporting meta-analyses of observational studies. It is equivalent to the STROBE Statement and the Consolidated Standards of Reporting Trials Consolidated Standards for Reporting Trials (CONSORT) 2010 Statement for RCTs, in focusing primarily on communication and list the minimum requirements for adequate reporting. The authors recommend a broad inclusion of studies and conduct of post-hoc sensitivity testing on the dependence of the results on factors such as quality of underlying papers, design, accounting for confounders, etc. The authors comment on the particular problems in merging observational studies with highly variable sets of confounders that were or were not controlled for, but they do not suggest any solution or give any references to possible ways to address it. As pragmatic trials increase in our field, another CONSORT extension focused on this type of studies, Improving the reporting of pragmatic trials: an extension of the CONSORT Statement (BMJ. 2008;337:a2390) might be also relevant.

The Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement (BMJ. 2009;339:b2535) is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. While focused on randomised trials, PRISMA can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions. PRISMA may also be useful for critical appraisal of published systematic reviews, although it is not designed as a quality assessment instrument.

Module VI of the Guideline on good pharmacovigilance practices (GVP) addresses the legal requirements which are applicable regards submission of individual reports of suspected adverse reactions associated with medicinal products authorised in the European Union. The Guidelines for Submitting Adverse Event Reports for Publication (Pharmacoepidemiol Drug Saf. 2007;16(5): 581–7) also list key elements that have to be included when publishing a report of one or more adverse events. These guidelines have been endorsed by the International Society for Pharmacoepidemiology (ISPE) and the International Society of Pharmacovigilance (ISoP) and are available on their websites.

Additional guidance on reporting of study results is provided in the ENCePP Checklist for Study Protocols and Code of Conduct and the IEA GEP guideline.

A preprints is an author's research manuscript prior to formal peer review which is deposited on a public server. Preprint publication can speed the sharing of work with important and immediate public health implications, and stimulate pre-publication comment from a wide research audience. Guidance on preprints can be found at journal and institution sites, including, for example, The Lancet and Nature.

15. Data protection and ethical aspects

Note: Chapter 15 (formerly 14) has not been updated for Revision 11 of the Guide, as contents remain up-to-date.

15.1. Personal data protection in the European Union

In the European Union, the conduct of pharmacoepidemiological studies needs to respect applicable Union data protection rules, namely the General Data Protection Regulation (EU) 2016/679 (GDPR) and Member State laws adopted in line with the GDPR (for example further conditions or limitations with

regard to the processing of genetic data, biometric data or data concerning health), which apply to processing carried out by organisations and bodies operating within the EU (for more details regarding the territorial scope of the GDPR, see EDPB Guidelines 3/2018 on the territorial scope of the GDPR, Article 3). Regulation (EU) 2018/1725 (EUDPR) apply to the personal data processing by Union institutions, bodies, offices and agencies.

Personal data is information that relates to an identified or identifiable individual. An identifiable individual is one who can be identified, directly or indirectly. Where it is possible to identify an individual directly from the information being processed, then that information is personal data. Where an individual cannot be directly identified from that information, it is still important to consider whether the individual is identifiable. For this, all the information being processed should be taken into account together with all the means reasonably likely to be used to identify that individual.

Special categories of personal data need more protection because they concern sensitive information. They include amongst others information revealing racial or ethnic origin, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. Special categories of personal data can only be processed if specific conditions set out in Article 9 of GDPR and Article 10 of EUDPR are met.

EudraLex - EU pharmaceutical legislation – the regulatory information for human medicines on the EMA website, the Good pharmacovigilance practices and ENCePP provide for methodological and ethical standards and ensure that private interests do not prevail over the general interest of public health. In this context, the Union data protection legislation is an enabler that promotes high data protection standards whilst providing the foundation for scientific research for the purpose of development, authorisation and supervision of medicinal products.

For interventional research, Clinical Trial Regulation (EU) 536/2014 and the Guidelines for Good Clinical Practice (Commission Directive 2005/28/EC) apply. It also applies to trials authorised under the previous legislation if they are still ongoing three years after the Regulation has come into operation. In addition, marketing authorisation holders (MAHs) and investigators must follow relevant national guidance of those Member States where the study is being conducted. To explain the interplay between the Clinical Trials Regulation and the GDPR the European Commission has published dedicated Questions and Answers.

Post-Authorisation Safety Studies (PASS) may be interventional or non-interventional. They may be conducted voluntarily or imposed on the marketing authorisation holder (MAH). Article 36 of the Commission Implementing Regulation (EU) No 520/2012 specifies that for post-authorisation safety studies (PASS) imposed as an obligation, MAHs shall ensure that all study information is handled and stored in a way that ensure the confidentiality of the study records of the study subjects. Section VIII.B.6. of the GVP Module VIII - Post-authorisation safety studies (Rev. 3) recommends that these provisions should also be applied to PASS that are voluntarily initiated, managed or financed by a MAH.

The ISPE Good pharmacoepidemiology practice provides recommendations on the protection of human subjects and refers to the ISPE guidelines on Data Privacy, Medical Record Confidentiality, and Research in the Interest of Public Health. It also recommends that the plans for protecting human subjects should be described in a stand-alone section of the study protocol.

The Data Protection Authorities (DPAs) of the Member States are competent for monitoring and enforcing the application of the GDPR. They are the natural interlocutors and first point of contact for the public, businesses and public administrations for questions regarding the GDPR. The Data Protection Authorities' role includes informing controllers and processors of their obligations and raising the general public's awareness and understanding of the risks, rules, safeguards and rights in relation to data processing.

The European Data Protection Board (EDPB) is an independent European body which is composed of representatives of the national DPAs (of all Union and EEA Member States) and the EDPS. The EDPB is established by Art 68 of the GDPR and is empowered to make binding decisions towards national DPAs to ensure the consistent application of Union data protection law. The EDPB may also issue general guidance (including guidelines, recommendations and best practice). Certain guidance adopted by the predecessor of the EDPB, the Article 29 Working Party (WP) are still applicable and provide interpretation of data protection principles under Union law.

15.2. Scientific integrity and ethical conduct

Principles of scientific integrity and ethical conduct are paramount in any medical research. The Declaration of Helsinki (2013) provides ethical principles addressed primarily to physicians participating in medical research involving human subjects, including research on identifiable human material and data and is the main document on human research ethics. The ENCePP Code of Conduct (Revision 4, 2018) offers standards for scientific independence and transparency of research in pharmacoepidemiology and pharmacovigilance and promotes best practice for the interactions between investigators and study funders in critical areas such as planning, conduct and reporting of studies. As a core transparency measure, it recommends that the protocols of all pharmacoepidemiology and pharmacovigilance studies should be registered in the European Union electronic Register of Post-Authorisation Studies (EU PAS Register), ideally before they start. The Code also recommends that study findings should be published irrespective of positive or negative results.

Guided by three core values (best science, strengthening public health and transparency), the ADVANCE Code of Conduct for Collaborative Vaccine Studies (Vaccine 2017;35(15):1844-55) includes recommendations about 10 topics: Scientific integrity, Scientific independence, Transparency, Conflicts of interest, Study protocol, Study report, Publication, Subject privacy, Sharing of study data, Research contract, and be used for research on any type of medicinal product. Each topic includes a definition, a set of recommendations and a list of additional reading. The concept of the study team is introduced as a key component of the ADVANCE Code of Conduct with a core set of roles and responsibilities. It also provides direct access to a comprehensive list of relevant guidelines.

The Good Pharmacoepidemiology Practices (GPP) (2015) of the International Society for Pharmacoepidemiology (ISPE) proposes practices and procedures that should be considered to help ensure the quality and integrity of pharmacoepidemiological research, including detailed guidance for protocol development, roles and responsibilities, study conduct, communication, reporting of adverse events and archiving. The Good Epidemiology Practice (GEP) (2007) of the International Epidemiological Association addresses four general ethical principles for research (Autonomy, Beneficence, Non-maleficence and Justice) and proposes rules for good research behaviour in relation to working with personal data, data documentation, publication, the exercise of judgment and scientific misconduct.

The CIOMS International Ethical Guidelines for Health-related Research Involving Humans (Geneva: 2016) provides detailed commentary on how universal ethical principles should be applied, with particular attention to conducting research in low-resource settings. It includes 25 guidelines addressing different topics, settings and population groups concerned by health-related research.

The Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals (2021) by the International Committee of Medical Journal Editors (ICJME) include clear statements on ethical principles related to publication in biomedical journals. Authorship and contributorship, editorship, peer review, conflicts of interest, privacy and confidentiality and protection of human subjects and animals in research are addressed.

The Agency for Healthcare Research and Quality (AHRQ) published Registries to Evaluate Patient Outcomes: a User's guide, 4th Edition, 2020, which is a reference for establishing, maintaining and evaluating the success of registries created to collect data about patient outcomes. Section II: 'Legal and Ethical Considerations for Registries' is a specific chapter dedicated to ethics, data ownership, and privacy. The concepts within are focused on US law.

More specifically on data used for the purpose of pharmacoepidemiology and pharmacovigilance studies, the HMA-EMA Joint Big Data Taskforce Phase II report: 'Evolving Data-Driven Regulation' (2019) acknowledges (in section 5.7) that data sharing and secondary use of data for research raise ethical issues which require identification, examination and guidance. The report uses Floridi and Taddeo's definition of data ethics: a new branch of ethics which "*studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing, and use), algorithms (including artificial intelligence, artificial agents, machine learning, and robots), and corresponding practices (including responsible innovation, programming, hacking, and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values)*". The Task Force report provides a set of recommendations for secure and ethical use of data ensuring that personal data are protected and that ethical challenges are addressed.

16. Specific topics

16.1. Comparative effectiveness research

16.1.1. Introduction

Comparative effectiveness research (CER) is designed to inform healthcare decisions for the prevention, the diagnosis and the treatment of a given health condition. CER therefore compares the potential benefits and harms of therapeutic strategies available in routine practice. The compared interventions may be related to similar treatments, such as competing medicines within the same class or with different mechanism of actions, or to different therapeutic approaches, such as surgical procedures and drug therapies. The comparison may focus only on the relative medical benefits and risks of the different options, or it may weigh both their costs and their benefits. The methods of comparative effectiveness research (Annu Rev Public Health 2012;33:425-45) defines the key elements of CER as a) a head-to-head comparison of active treatments, b) study population typical of the day-to-day clinical practice, and c) evidence focussed on informing healthcare and tailored to the characteristics of individual patients. CER is often discussed in the regulatory context of real-world evidence (RWE) generated by clinical trials or non-interventional (observational) studies using real-world data (RWD) (see Chapter 16.6).

The term 'Relative effectiveness assessment (REA)' is also used when comparing multiple technologies or a new technology against standard of care, while 'rapid' REA refers to performing an assessment within a limited timeframe in the case of a new marketing authorisation or a new indication granted for an approved medicine (see What is a rapid review? A methodological exploration of rapid reviews in Health Technology Assessments, Int J Evid Based Healthc. 2012;10(4):397-410).

16.1.2. Methods for comparative effectiveness research

CER may use a variety of data sources and methods. Methods to generate evidence for CER are divided below in four categories according to the data source: randomised clinical trials (RCTs), observational data, synthesis of published RCTs and cross-design synthesis.

16.1.2.1. CER based on randomised clinical trials

RCTs are considered the gold standard for demonstrating the efficacy of medicinal products but they rarely measure the benefits, risks or comparative effectiveness of an intervention in post-authorisation clinical practice. Moreover, relatively few RCTs are designed with an alternative therapeutic strategy as a comparator, which limits the utility of the resulting data in establishing recommendations for treatment choices. For these reasons, other methodologies such as pragmatic trials and large simple trials may be used to complement traditional confirmatory RCTs in CER. These trials are discussed in Chapter 4.2.7. The estimand framework described in the ICH E9-R1 Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials (2019) should be considered in the planning of comparative effectiveness trials as it provides coherence and transparency on important elements of CER, namely definitions of exposures, endpoints, intercurrent events (ICEs), strategies to manage ICEs, approach to missing data and sensitivity analyses.

In order to facilitate comparison of results of CER between clinical trials, the COMET (Core Outcome Measures in Effectiveness Trials) Initiative aims at developing agreed minimum standardized sets of outcomes ('core outcome sets', COS) to be assessed and reported in effectiveness trials of a specific condition. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey (PLoS One 2016;11(1):e0146444) provides an updated review of studies that have addressed the development of COS for measurement and reporting in clinical trials. It is also worth noting that regulatory disease guidelines also establish outcomes of clinical interest to assess if a new therapeutic intervention works. Use of the same endpoint across RCTs thus facilitate comparisons.

16.1.2.2. CER using observational data

Use of observational data in CER

Although observational data from Phase IV trials, post-authorisation safety studies (PASS), or other RWD sources can be used to assess comparative effectiveness (and safety), it is generally inappropriate to use such data as a *replacement* for randomised evidence, especially in a confirmatory setting. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials (JAMA 2023;329(16):1376-85) concludes that RWE studies can reach similar conclusions as RCTs when design and measurements can be closely emulated, but this may be difficult to achieve. Concordance in results varied depending on the agreement metric. Emulation differences, chance, and residual confounding can contribute to divergence in results and are difficult to disentangle. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? (Clin Pharmacol. Ther. 2017;102(6):924-33) suggests that RWE may be preferred over randomised evidence when studying a highly promising treatment for a disease with no other available treatment and where ethical considerations may preclude randomising patients to placebo, particularly if the disease is likely to result in severely compromised quality of life or mortality. In these cases, RWE could support medicines regulation by providing evidence on the safety and effectiveness of the therapy against the typical disease progression observed in the absence of treatment. This comparator disease trajectory may be assessed from historical controls that were diagnosed prior to the availability of the new treatment, or other sources.

When Can We Rely on Real-World Evidence to Evaluate New Medical Treatments? (Clin Pharmacol Ther. 2021; 111(1): 30–4) recommends that decisions regarding use of RWE in the evaluation of new treatments should depend on the specific research question, characteristics of the potential study settings and characteristics of the settings where study results would be applied, and take into account three dimensions in which RWE studies might differ from traditional clinical trials: use of RWD, delivery of real-world treatment and real-world treatment assignment. Observational data have, for instance, been used in proof-of-concept studies on anaplastic lymphoma kinase-positive non-small cell lung cancer, in pivotal trials on acute lymphoblastic leukaemia, thalassemia syndrome and haemophilia A, and in studies aimed at label expansion for epilepsy (see Characteristics of non-randomised studies

using comparisons with external controls submitted for regulatory approval in the USA and Europe: a systematic review, BMJ Open. 2019;1;9(2):e024895; The Use of External Controls in FDA Regulatory Decision Making, Ther Innov Regul Sci. 2021;55(5):1019–35; and Application of Real-World Data to External Control Groups in Oncology Clinical Trial Drug Development, Front Oncol. 2022;11:695936).

Outside of specific circumstances, observational data and clinical trials are considered *complementary* to generate comprehensive evidence. For example, clinical trials may include historical controls from observational studies, or identify eligible study participants from disease registries. In defense of pharmacoepidemiology--embracing the yin and yang of drug research (N Engl J Med 2007;357(22):2219-21) shows that strengths and weaknesses of RCTs and observational studies may make both designs necessary in the study of drug effects. Hybrid approaches for CER allow to enrich clinical trials with observational data, for example:

- Use of historical data to partially replace concurrent controls in randomised trials (see A roadmap to using historical controls in clinical trials - by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG), Orphanet J Rare Dis. 2020;15:69);
- Use of historical data as prior evidence for relative treatment effects (see Prior Elicitation for Use in Clinical Trial Design and Analysis: A Literature Review, Int J Environ Res Public Health 2021;18(4):1833);
- Construction of external control groups in single arm studies and Phase IV trials (see the draft FDA guidance Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products (2023), A Review of Causal Inference for External Comparator Arm Studies (Drug Saf. 2022;45(8):815-37) and Methods for external control groups for single arm trials or long-term uncontrolled extensions to randomized clinical trials, Pharmacoepidemiol Drug Saf. 2020; 29(11):1382–92).

Methods for CER using observational data

The use of non-randomised data for causal inference is notoriously prone to various sources of bias. For this reason, it is strongly recommended to carefully design or select the source of RWD and to adopt statistical methods that acknowledge and adjust for major risks of bias (e.g. confounding, missing data).

A framework to address these challenges adopts counterfactual theory to treat the observational study as an emulation of a randomised trial. Target trial emulation (described in Chapter 4.2.6.) is a strategy that uses existing tools and methods to formalise the design and analysis of observational studies. It stimulates investigators to identify potential sources of concerns and develop a design that best addresses these concerns and the risk of bias.

Target trial emulation consists in designing first a hypothetical ideal randomised trial ("target trial") that would answer the research question. A second step identifies how to best emulate the design elements of the target trial (including its eligibility criteria, treatment strategies, assignment procedure, follow-up, outcome, causal contrasts and pre-specified analysis plan) using the available observational data source and the analytic approaches to apply, given the trade-offs in an observational setting. This approach may prevent some common biases, such as immortal time bias or prevalent user bias while also identifying situations where adequate emulation may not be possible using the data at hand. Emulating a Target Trial of Interventions Initiated During Pregnancy with Healthcare Databases: The Example of COVID-19 Vaccination (Epidemiology 2023;34(2):238-46) describes a step-by-step specification of the protocol components of a target trial and their emulation including sensitivity analyses using negative controls to evaluate the presence of confounding and, alternatively to a cohort design, a case-crossover or case-time-control design to eliminate confounding by unmeasured time-fixed factors. Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines

in U.S. Veterans (N Engl J Med. 2022;386(2):105-15) used target trial emulation to design a study where recipients of each vaccine were matched in a 1:1 ratio according to their baseline risk factors. This design could not be applied where baseline measurements are not collected at treatment start, which may be the case in some patient registries. Use of the estimand framework of the ICH E9 (R1) Addendum to design the target trial may increase transparency on the choices and assumptions needed in the observational study to emulate key trial protocol components, such as the estimand, exposure, intercurrent events (and the strategies to manage them), the missing data and the sensitivity analyses, and therefore may help evaluate the extent to which the observational study addresses the same question as the target trial. Studies on the effect of treatment duration are also often impaired by selection bias: How to estimate the effect of treatment duration on survival outcomes using observational data (BMJ. 2018;360: k182) proposes a 3-step approach (cloning, censoring, weighting) that could be used with target trial simulation to achieve better comparability with the treatment assignment performed in the trial and overcome bias in the observational study.

Statistical inference methods that can be used for conducting causal inference in non-interventional studies are described in Chapter 6.2.3 and include multivariable regression (to adjust for confounding, missing data, measurement error, and other sources of bias), propensity score methods (to adjust for confounding bias), prognostic or disease risk score methods (to adjust for confounding), G-methods and marginal structure models (to adjust for time-dependent confounding), and imputation methods (to adjust for missing data). In some situations, these methods can also be used to adjust for instrumental variables or to estimate prior event rate ratios. Causal Inference in Oncology Comparative Effectiveness Research Using Observational Data: Are Instrumental Variables Underutilized? (J Clin Oncol. 2023;41(13):2319-2322) summarises the key assumption, advantages and disadvantages of methods of causal inference in CER to adjust for confounding, including regression adjustment, propensity scores, difference-in differences, regression discontinuity and instrumental variable, highlighting that different methods can be combined. In some cases, observational studies may substantially benefit from collecting instrumental variables, and this should be considered early on when designing the study. For example, Dealing with missing data using the Heckman selection model: methods primer for epidemiologists (Int J Epidemiol. 2023;52(1):5-13) illustrates the use of instrumental variables to address data that are missing not at random. Another example is discussed in Association of Osteoporosis Medication Use After Hip Fracture With Prevention of Subsequent Nonvertebral Fractures: An Instrumental Variable Analysis (JAMA Netw Open. 2018;1(3):e180826.), where instrumental variables are used to adjust for unobserved confounders.

The Agency for Healthcare Research and Quality (AHRQ)'s Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide (2013) identifies minimal standards and best practices for observational CER. It provides principles on a wide range of topics for designing research and developing protocols, with relevant questions to be addressed and checklists of key elements to be considered. The RWE Navigator website discusses methods using observational RWD with a focus on effectiveness research, such as the source of RWD, study designs, approaches to summarising and synthesising the evidence, modelling of effectiveness and methods to adjust for bias and governance aspects. It also presents a glossary of terms and case studies.

A roadmap to using historical controls in clinical trials - by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG) (Orphanet J Rare Dis. 2020;15:69) describes methods to minimise disadvantages of using historical controls in clinical trials, i.e. frequentist methods (e.g. propensity score methods and meta-analytical approach) or Bayesian methods (e.g. power prior method, adaptive designs and the meta-analytic combined [MAC] and meta-analytic predictive [MAP] approaches for meta-analysis). It also provides recommendations on approaches to apply historical controls when they are needed while maximising scientific validity to the extent feasible.

In the context of hybrid studies, key methodological issues to be considered when combining RWD and RCT data include:

- Differences between the RWD and RCT in terms of data quality and applicability,
- Differences between available RWD sources (e.g., due to heterogeneity in studied populations, differences in study design, etc.),
- Risk of bias (particularly for RWD),
- Generalisability (especially for RCT findings beyond the overall treatment effect).

Methods for systematic reviews and meta-analyses of observational studies are presented in Chapter 10 and Annex 1 of this Guide. They are also addressed in the Cochrane Handbook for Systematic Reviews of Interventions and the Methods Guide for Effectiveness and Comparative Effectiveness Reviews presented in section 16.1.2.3 of this Chapter.

Assessment of observational studies used in CER

Given the potential for bias and confounding in CER based on observational non-randomised studies, the design and results of such studies need to be adequately assessed. The Good ReseArch for Comparative Effectiveness (GRACE) (IQVIA, 2016) provides guidance to enhance the quality of observational CER studies and support their evaluation for decision-making using the provided checklist. How well can we assess the validity of non-randomised studies of medications? A systematic review of assessment tools (BMJ Open 2021;11:e043961) examined whether assessment tools for non-randomised studies address critical elements that influence the validity of findings from non-randomised studies for CER. It concludes that major design-specific sources of bias (e.g., lack of new-user design, lack of active comparator design, time-related bias, depletion of susceptibles, reverse causation) and statistical assessment of internal and external validity are not sufficiently addressed in most of the tools evaluated, although these critical elements should be integrated to systematically investigate the validity of non-randomised studies on comparative safety and effectiveness of medications. The article also provides a glossary of terms, a description of the characteristics the tools and a description of methodological challenges they address.

Comparison of results of observational studies and RCTs

Even if observational studies are not appropriate to replace RCTs for many CER topics and cannot answer exactly the same research question, comparison of their results for a same objective is currently a domain of interest. The underlying assumption is that if observational studies consistently match the results of published trials and predict the results of ongoing trials, this may increase the confidence in the validity of future RWD analyses performed in the absence of randomised trial evidence. In a review of five interventions, Randomized, controlled trials, observational studies, and the hierarchy of research designs (N Engl J Med 2000;342(25):1887-92) found that the results of well-designed observational studies (with either a cohort or case-control design) did not systematically overestimate the magnitude of treatment effects. Interim results from the 10 first emulations reported in Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies: First Results From the RCT DUPLICATE Initiative (Circulation 2021;143(10):1002-13) found that differences between the RCT and corresponding RWE study populations remained but the RWE emulations achieved a hazard ratio estimate that was within the 95% CI from the corresponding RCT in 8 of 10 studies. Selection of active comparator therapies with similar indications and use patterns enhanced the validity of RWE. Final results of this project are discussed in the presentation Lessons Learned from Trial Replication Analyses: Findings from the DUPLICATE Demonstration Project (Duke-Margolis Center for Health Policy Workshop, 10 May 2022). Emulation Differences vs. Biases When Calibrating Real-World Evidence Findings Against Randomized Controlled Trials (Clin Pharmacol Ther. 2020;107(4):735-

7) provides guidance on how to investigate and interpret differences in treatment effect estimates from the two study types.

An important source of selection bias leading to discrepancies between results of observational studies and RCTs may be the use of prevalent drug users in the former. Evaluating medication effects outside of clinical trials: new-user designs (Am J Epidemiol 2003;158(9):915-20) explains the biases introduced by use of prevalent drug users and how a new-user (or incident user) design eliminate these biases by restricting analyses to persons under observation at the start of the current course of treatment. The incident user design in comparative effectiveness research (Pharmacoepidemiol Drug Saf. 2013; 22(1):1–6) reviews published CER case studies in which investigators had used the incident user design and discusses its strengths (reduced bias) and weaknesses (reduced precision of comparative effectiveness estimates). Unless otherwise justified, the incident user design should always be used.

16.1.2.3. CER based on evidence synthesis of published RCTs

The Cochrane Handbook for Systematic Reviews of Interventions (version 6.2, 2022) describes in detail the process of preparing and maintaining systematic reviews on the effects of healthcare interventions. Although its scope is focused on Cochrane reviews, it has a much wider applicability. It includes guidance on the standard methods applicable to every review (planning a review, searching and selecting studies, data collection, risk of bias assessment, statistical analysis, GRADE and interpreting results), as well as more specialised topics. The (GRADE) working group (Grading of Recommendations Assessment, Development, and Evaluation) offers a structured process for rating quality of evidence and grading strength of recommendations in systematic reviews, health technology assessment and clinical practice guidelines. The Methods Guide for Effectiveness and Comparative Effectiveness Reviews (AHRQ, 2018) provides resources supporting comparative effectiveness reviews. They are focused on the US Effective Health Care (EHC) programme and may therefore have limitations as regards their generalisability.

A pairwise meta-analysis of RCT results is used when the primary aim is to estimate the relative effect of two interventions. Network meta-analysis for indirect treatment comparisons (Statist Med. 2002;21:2313–24) introduces methods for assessing the relative effectiveness of two treatments when they have not been compared directly in a randomised trial but have each been compared to other treatments. Overview of evidence synthesis and network meta-analysis – RWE Navigator discussed methods and best practices and gives access to published articles on this topic. A prominent issue that has been overlooked by some systematic literature reviews and network meta-analyses is the fact that RCTs included in a network meta-analysis are usually not comparable with each other even though they all compared to placebo. Different screening and inclusion/exclusion criteria often create different patient groups, and these differences are rarely discussed in indirect comparisons. Before indirect comparison are performed, researchers should therefore check the similarity/differences between the RCTs.

16.1.2.4. CER based on cross-design synthesis

Decision-making should ideally be based on all available evidence, including both randomised and non-randomised studies, and on both individual patient data and published aggregated data. Clinical trials are highly suitable to investigate efficacy but less practical to study long-term outcomes or rare diseases. On the other hand, observational data offer important insights about treatment populations, long-term outcomes (e.g., safety), patient-reported outcomes, prescription patterns, active comparators, etc. Combining evidence from these two sources could therefore be helpful to reach certain effectiveness/safety conclusions earlier or to address more complex questions. Several methods have been proposed but are still experimental. The article Framework for the synthesis of non-randomised studies and randomised controlled trials: a guidance on conducting a systematic

review and meta-analysis for healthcare decision making (BMJ Evid Based Med. 2022;27(2):109-19) uses a 7-step mixed methods approach to develop guidance on when and how to best combine evidence from non-randomised studies and RCTs to improve transparency and build confidence in summary effect estimates. It provides recommendations on the most appropriate statistical approaches based on analytical scenarios in healthcare decision making and highlights potential challenges for the implementation of this approach.

16.1.3. Methods for Relative Effectiveness Assessment (REA)

Methodological Guidelines for Rapid REA of Pharmaceuticals (EUnetHTA, 2013) cover a broad spectrum of issues on REA. They address methodological challenges that are encountered by health technology assessors while performing rapid REA and provide and discuss practical recommendations on definitions to be used and how to extract, assess and present relevant information in assessment reports. Specific topics covered include the choice of comparators, strengths and limitations of various data sources and methods, internal and external validity of studies, the selection and assessment of endpoints and the evaluation of relative safety.

16.1.4. Specific aspects

16.1.4.1. Secondary use of data for CER

Electronic healthcare records, patient registries and other data sources are increasingly used in clinical effectiveness studies as they capture real clinical encounters and may document reasons for treatment decisions that are relevant for the general patient population. As they are primarily designed for clinical care and not research, information on relevant covariates and in particular on confounding factors may not be available or adequately measured. These aspects are presented in other chapters of this Guide (see Chapter 6, Methods to address bias and confounding; Chapter 8.2, Secondary use of data, and other chapters for secondary use of data in other contexts) but they need to be specifically considered in the context of CER. For example, the Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG) Roadmap to using historical controls in clinical trials (Orphanet J Rare Dis. 2020;15:69) describes the main sources of RWD to be used as historical controls, with an Appendix providing guidance on factors to be evaluated in the assessment of the relevance of RWD sources and resultant analyses.

16.1.4.2 Data quality

Data quality is essential to ensure the rigor of CER and secondary use of data requires special attention. Comparative Effectiveness Research Using Electronic Health Records Data: Ensure Data Quality (SAGE Research Methods, 2020*)* discusses challenges and share experiences encountered during the process of transforming electronic health record data into a research quality dataset for CER. This aspect and other quality issues are also discussed in Chapter 13 on Quality management.

In order to address missing information, some CER studies have attempted to integrate information from healthcare databases with information collected *ad hoc* from study subjects. Enhancing electronic health record measurement of depression severity and suicide ideation: a Distributed Ambulatory Research in Therapeutics Network (DARTNet) study (J Am Board Fam Med. 2012;25(5):582-93) shows the value of linking direct measurements and pharmacy claims data to data from electronic healthcare records. Assessing medication exposures and outcomes in the frail elderly: assessing research challenges in nursing home pharmacotherapy (Med Care 2010;48(6 Suppl):S23-31) describes how merging longitudinal electronic clinical and functional data from nursing home sources with Medicare and Medicaid claims data can support unique study designs in CER but pose many challenging design and analytic issues.

16.1.4.3. Transparency and reproducibility

Clear and transparent study protocols for observational CER should be used to support the evaluation, interpretation and reproducibility of results. Use of the HARPER protocol template (HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force, Pharmacoepidemiol Drug Saf. 2023;32(1):44-55) is recommended to facilitate protocol development and addressing important design components. Public registration and posting of the protocol, disease and drug code lists, and statistical programming is strongly recommended to ensure that results from comparative effectiveness studies can be replicated using the same data and/or design, as emphasised in *Journal of Comparative Effectiveness Research* welcoming the submission of study design protocols to foster transparency and trust in real-world evidence (J Comp Eff Res. 2023;12(1):e220197). The EU PAS Register and ClinicalTrials.gov should be used for this purpose.

16.2. Vaccine safety and effectiveness

16.2.1. Vaccine safety

16.2.1.1. General considerations

The book *Vaccination Programmes | Epidemiology, Monitoring, Evaluation* (Hahné, S., Bollaerts, K., & Farrington, P., Routledge, 2021) is a comprehensive textbook addressing most of the concepts presented in this Chapter. For contents related to safety monitoring of vaccines, it further builds on the 2014 ADVANCE Report on appraisal of vaccine safety methods that described a wide range of direct and indirect methods for vaccine safety assessment. Specific aspects related to vaccine safety and effectiveness are discussed in several documents:

- The Report of the CIOMS/WHO Working Group on Definition and Application of Terms for Vaccine Pharmacovigilance (2012) provides definitions and explanatory notes for the terms 'vaccine pharmacovigilance', 'vaccination failure' and 'adverse event following immunisation (AEFI)'.
- The Guide to active vaccine safety surveillance: Report of CIOMS working group on vaccine safety – executive summary (Vaccine 2017;35(32):3917-21) describes the process for selecting the best approach to active surveillance considering key implementation issues, including in resource-limited countries.
- The CIOMS Guide to Vaccine Safety Communication (2018) addresses vaccine safety communication aspects for regulators, vaccination policy-makers, and other stakeholders, when introducing vaccines in populations, based on selected examples.
- The Brighton Collaboration provide a resource to facilitate and harmonise collection, analysis, and presentation of vaccine safety data, including case definitions for outcomes of interest, including adverse events of special interest (AESIs).
- Module 4 (Surveillance) of the e-learning training course Vaccine Safety Basics of the World Health Organization (WHO) describes pharmacovigilance principles, causality assessment procedures, surveillance systems, and places safety in the context of the vaccine benefit/risk profile.
- Recommendations on vaccine-specific aspects of the EU Pharmacovigilance System, including on risk management, signal detection and post-authorisation safety studies (PASS) are presented in Module P.I: Vaccines for prophylaxis against infectious diseases (EMA, 2013) of the Good Pharmacovigilance Practices (GVP).

- The WHO Covid-19 vaccines: safety surveillance manual (WHO, 2020) was developed upon recommendation of the WHO Global Advisory Committee on Vaccine Safety (GACVS) and describes categories of surveillance strategies: passive, active, cohort event monitoring, and sentinel surveillance. While developed for COVID-19 vaccines, it can be used to guide pandemic preparedness activities for the monitoring of novel vaccines.
- A vaccine study design selection framework for the postlicensure rapid immunization safety monitoring program (Am J Epidemiol. 2015;181(8):608-18) describes in a tabular form strengths and weaknesses of study designs and can be broadly applied to vaccine research questions beyond safety assessment.

16.2.1.2. Signal detection and validation

Besides a qualitative analysis of spontaneous case reports or case series, quantitative methods such as disproportionality analyses (described in Chapter 11) and observed-to-expected (O/E) analyses are routinely employed in signal detection and validation for vaccines. Several documents discuss the merits and review the methods of these approaches for vaccines.

Disproportionality analyses

GVP Module P.I: Vaccines for prophylaxis against infectious diseases describes aspects to be considered when applying methods for vaccine disproportionality analyses, including choice of the comparator group and use of stratification. Effects of stratification on data mining in the US Vaccine Adverse Event Reporting System (VAERS) (Drug Saf. 2008;31(8):667-74) demonstrates that stratification can reveal and reduce confounding and unmask some vaccine-event pairs not found by crude analyses. However, Stratification for Spontaneous Report Databases (Drug Saf. 2008;31(11):1049-52) highlights that extensive use of stratification in signal detection algorithms should be avoided, as it can mask true signals. Vaccine-Based Subgroup Analysis in VigiBase: Effect on Sensitivity in Paediatric Signal Detection (Drug Saf. 2012;35(4):335-46) further examines the effects of subgroup analyses based on the relative distribution of vaccine/non-vaccine reports in paediatric adverse drug reaction data (ADR) data. In Performance of Stratified and Subgrouped Disproportionality Analyses in Spontaneous Databases (Drug Saf. 2016;39(4):355-64), subgrouping by vaccines/non-vaccines resulted in a decrease in both precision and sensitivity in all spontaneous report databases that contributed data.

Optimization of a quantitative signal detection algorithm for spontaneous reports of adverse events post immunization (Pharmacoepidemiol Drug Saf. 2013;22(5): 477–87) explores various ways of improving performance of signal detection algorithms for vaccines.

Adverse events associated with pandemic influenza vaccines: comparison of the results of a follow-up study with those coming from spontaneous reporting (Vaccine 2011;29(3):519-22) reported a more complete pattern of reactions when using two complementary methods for first characterisation of the post-marketing safety profile of a new vaccine, which may impact on signal detection.

In Review of the initial post-marketing safety surveillance for the recombinant zoster vaccine (Vaccine 2020;38(18):3489-500), the time-to-onset distribution of zoster vaccine-adverse event pairs was used to generate a quantitative signal of unexpected temporal relationship.

Bayesian disproportionality methods have also been developed to generate disproportionality signals. In Association of Facial Paralysis With mRNA COVID-19 Vaccines: A Disproportionality Analysis Using the World Health Organization Pharmacovigilance Database (JAMA Intern Med. 2021;e212219), a potential safety signal for facial paralysis was explored using the Bayesian neural network method.

In Disproportionality analysis of anaphylactic reactions after vaccination with messenger RNA coronavirus disease 2019 vaccines in the United States (Ann Allergy Asthma Immunol. 2021; S1081-

1206(21)00267-2) the CDC Wide-ranging Online Data for Epidemiologic Research (CDC WONDER) system was used in conjunction with proportional reporting ratios to evaluate whether rates of anaphylaxis cases reported in the VAERS database following administration of mRNA COVID-19 vaccines is disproportionately different from all other vaccines.

Signaling COVID-19 Vaccine Adverse Events (Drug Saf. 2022 Jun 23:1–16) discusses the extent, direction, impact, and causes of masking, an issue associated with signal detection methodologies, in which signals for a product of interest are hidden by the presence of other reported products, which may limit the understanding of the risks associated with COVID-19 and other vaccines, and delay their identification.

Observed-to-expected analyses and background incidence rates

In vaccine vigilance, O/E analyses compare the 'observed' number of cases of an adverse event occurring in vaccinated individuals and recorded in a data collection system (e.g. a spontaneous reporting system or an electronic healthcare database) and the 'expected' number of cases that would have naturally occurred in the same population without vaccination, estimated from available incidence rates in a non-vaccinated population. O/E analyses constitute a first step in the continuum of safety signal evaluation, and can guide further steps such as a formal pharmacoepidemiological study. GVP Module P.I: Vaccines for prophylaxis against infectious diseases (EMA, 2013) suggests conducting O/E analyses for signal validation and preliminary signal evaluation when prompt decision-making is required, and there is insufficient time to review a large number of individual cases. It discusses key requirements of O/E analyses: an observed number of cases detected in a passive or active surveillance system, near real-time exposure data, appropriately stratified background incidence rates calculated on a population similar to the vaccinated population (for the expected number of cases), the definition of appropriate risk periods (where there is suspicion and/or biological plausibility that there is a vaccine-associated increased risk of the event) and sensitivity analyses around these measures. O/E analyses may require some adjustments for continuous monitoring due to inflation of type 1 error rates when multiple tests are performed. The method is further discussed in Pharmacoepidemiological considerations in observed-to-expected analyses for vaccines (Pharmacoepidemiol Drug Saf. 2016;25(2):215-22) and the review Near real-time vaccine safety surveillance using electronic health records - a systematic review of the application of statistical methods (Pharmacoepidemiol Drug Saf. 2016;25(3):225-37).

O/E analyses require several pre-defined assumptions based on the requirements listed above. Each of these assumptions can be associated with uncertainties. How to manage these uncertainties is also addressed in Pharmacoepidemiological considerations in observed-to-expected analyses for vaccines (Pharmacoepidemiol Drug Saf. 2016;25(2):215–22). Observed-over-Expected analysis as additional method for pharmacovigilance signal detection in large-scaled spontaneous adverse event reporting (Pharmacoepidemiol Drug Saf. 2023;32(7):783-794) uses two examples of events of interest (idiopathic peripheral facial paralysis and Bell's palsy) in the context of the COVID-19 immunisation campaigns, when very large numbers of case safety reports (ICSRs) had to be timely handled.

Use of population based background rates of disease to assess vaccine safety in childhood and mass immunisation in Denmark: nationwide population based cohort study (BMJ. 2012;345:e5823) illustrates the importance of collecting background rates by estimating risks of coincident associations of emergency consultations, hospitalisations and outpatients consultations, with vaccination. Rates of selected disease events for several countries may vary by age, sex, method of ascertainment, and geography, as shown in Incidence Rates of Autoimmune Diseases in European Healthcare Databases: A Contribution of the ADVANCE Project (Drug Saf. 2021;44(3):383-95), where age-, gender-, and calendar-year stratified incidence rates of nine autoimmune diseases in seven European healthcare databases from four countries were generated to support O/E analyses. Guillain-Barré syndrome and

influenza vaccines: A meta-analysis (Vaccine 2015; 33(31):3773-8) suggests that a trend observed between different geographical areas would be consistent with a different susceptibility of developing a particular adverse reaction among different populations. In addition, comparisons with background rates may be invalid if conditions are unmasked at vaccination visits (see Human papillomavirus vaccination of adult women and risk of autoimmune and neurological diseases, J Intern Med. 2018;283:154-65)).

Several studies have generated background incidence rates of AESIs for COVID-19 vaccines and discuss methodological challenges related to identifying AESIs in electronic health records (EHRs) (see The critical role of background rates of possible adverse events in the assessment of COVID-19 vaccine safety, Vaccine 2021;39(19):2712-18).

In Arterial events, venous thromboembolism, thrombocytopenia, and bleeding after vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: population based cohort study (BMJ. 2021;373:n1114), observed age- and sex-specific rates of events among vaccinated people were compared with expected rates in the general population calculated from the same databases, thereby removing a source of variability between observed and expected rates. Where this is not possible, rates from multiple data sources have shown to be heterogeneous, and the choice of relevant data should take into account differences in database and population characteristics related to different diagnoses, recording and coding practices, source populations (e.g., inclusion of subjects from general practitioners and/or hospitals), healthcare systems, and linkage ability (e.g., to hospital records). This is further discussed in Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study (BMJ. 2021;373:n1435) and Background rates of five thrombosis with thrombocytopenia syndromes of special interest for COVID-19 vaccine safety surveillance: Incidence between 2017 and 2019 and patient profiles from 38.6 million people in six European countries (Pharmacoepidemiol Drug Saf. 2022;31(5):495-510).

Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study (EClinicalMedicine. 2023;58:101932) used data from primary care, electronic health records, and insurance claims mapped to a common data model to characterise the incidence rates of AESIs, also following SARS-CoV-2 infection (considered a confounder), compared them to historical rates in the general population, and addressed issues of heterogeneity.

Historical comparator designs may generate false positives, as discussed in Bias, Precision and Timeliness of Historical (Background) Rate Comparison Methods for Vaccine Safety Monitoring: An Empirical Multi-Database Analysis (Front Pharmacol. 2021;12:773875), which explores the effect of empirical calibration on type 1 and 2 errors using outcomes presumed to be unrelated to vaccines (negative control outcomes) as well as positive controls (outcomes simulated to be caused by the vaccines).

Factors Influencing Background Incidence Rate Calculation: Systematic Empirical Evaluation Across an International Network of Observational Databases (Front Pharmacol. 2022;13:814198) examined the sensitivity of rates to the choice of design parameters using 12 data sources to systematically examine their influence on incidence rates using 15 AESIs for COVID-19 vaccines. Rates were highly influenced by the choice of the database (varying by up to a factor of 100), the choice of anchoring (e.g., health visit, vaccination, or arbitrary date) for the time-at-risk start, the choice of clean window and time-at-risk duration, but less so by secular or seasonal trends. It concluded that results should be interpreted in the context of study parameter choices.

Sequential methods

Sequential methods, as described in Early detection of adverse drug events within population-based health networks: application of sequential methods (Pharmacoepidemiol Drug Saf. 2007;16(12):1275-84), allow O/E analyses to be performed on a routine (e.g., weekly) basis using cumulative data with adjustment for multiplicity. Such methods are routinely used for near-real time surveillance in the Vaccine Safety Datalink (VSD) (see Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project, Am J Epidemiol 2010;171(2):177-88). Potential issues are described in Challenges in the design and analysis of sequentially monitored postmarket safety surveillance evaluations using electronic observational health care data (Pharmacoepidemiol Drug Saf. 2012;21(S1):62-71). A review of signals detected over 3 years in the VSD concluded that care with data quality, outcome definitions, comparator groups, and duration of surveillance, is required to enable detection of true safety issues while controlling for error (Active surveillance for adverse events: the experience of the Vaccine Safety Datalink Project, Pediatrics 2011;127(S1):S54-S64).

A new self-controlled case series method for analyzing spontaneous reports of adverse events after vaccination (Am J Epidemiol. 2013;178(9):1496-504) extends the self-controlled case series approach (see Chapter 4.2.3, and 16.2.2.2 in this Chapter) to explore and quantify vaccine safety signals from spontaneous reports using different assumptions (e.g., considering large amount of underreporting, and variation of reporting with time since vaccination). The method should be seen as a signal strengthening approach for quickly exploring a signal prior to a pharmacoepidemiological study (see for example, Kawasaki disease and 13-valent pneumococcal conjugate vaccination among young children: A self-controlled risk interval and cohort study with null results, PLoS Med. 2019;16(7):e100284).

The tree-based scan statistic (TreeScan) is a statistical data mining method that can be used for the detection of vaccine safety signals from large health insurance claims and electronic health records (Drug safety data mining with a tree-based scan statistic, Pharmacoepidemiol Drug Saf. 2013;22(5):517-23). A Broad Safety Assessment of the 9-Valent Human Papillomavirus Vaccine (Am J Epidemiol. 2021;kwab022) and A broad assessment of covid-19 vaccine safety using tree-based data-mining in the vaccine safety datalink (Vaccine. 2023;41(3):826-835) used the self-controlled tree-temporal scan statistic which does not require pre-specified outcomes or specific post-exposure risk periods. The method requires further evaluation of its utility for routine vaccine surveillance in terms of requirements for large databases and computer resources, as well as predictive value of the signals detected.

16.2.1.3. Study designs for vaccine safety assessment

A complete review of vaccine safety study designs and methods for hypothesis-testing studies is included in the ADVANCE Report on appraisal of vaccine safety methods (2014) and in Part IV of the book *Vaccination Programmes | Epidemiology, Monitoring, Evaluation* (Hahné, S., Bollaerts, K., & Farrington, P., Routledge, 2021).

Current Approaches to Vaccine Safety Using Observational Data: A Rationale for the EUMAEUS (Evaluating Use of Methods for Adverse Events Under Surveillance-for Vaccines) Study Design (Front Pharmacol. 2022;13:837632) provides an overview of strengths and limitations of study designs for vaccine safety monitoring and discusses the assumptions made to mitigate bias in such studies.

Methodological frontiers in vaccine safety: qualifying available evidence for rare events, use of distributed data networks to monitor vaccine safety issues, and monitoring the safety of pregnancy interventions (BMJ Glob Health. 2021;6(Suppl 2):e003540) addresses multiple aspects of pharmacoepidemiological vaccine safety studies, including study designs.

Cohort and case-control studies

There is a large body of published literature reporting on the use of the cohort design (and to a lesser extent, the case-control design) for the assessment of vaccine safety. Aspects of these designs presented in Chapters 4.2.1 and 4.2.2 are applicable to vaccine studies (for the cohort design, see also the examples of studies on background incidence rates in paragraph 16.2.2.1 of this Chapter). A recent illustration of the cohort design is provided in Clinical outcomes of myocarditis after SARS-CoV-2 mRNA vaccination in four Nordic countries: population based cohort study (BMJ Med. 2023 Feb 1;2(1):e000373) which used nationwide register data to compare clinical outcomes of myocarditis associated with vaccination, with COVID-19 disease, and with conventional myocarditis, with respect to readmission to hospital, heart failure, and death, using the Kaplan-Meier estimator approach.

Cohort-event monitoring

Prospective cohort-event monitoring (CEM) including active surveillance of vaccinated subjects using smartphone applications and/or web-based tools has been extensively used to monitor the safety of COVID-19 vaccines, as primary data collection was the only means to rapidly identify safety concerns when the vaccines started to be used at large scale. A definition of cohort-event monitoring is provided in The safety of medicines in public health programmes : pharmacovigilance, an essential tool (WHO, 2006, Chapter 6.5, Cohort event monitoring, pp 40-41). Specialist Cohort Event Monitoring studies: a new study method for risk management in pharmacovigilance (Drug Saf. 2015;38(2):153-63) discusses the rationale and features to address possible bias, and some applications of this design. COVID-19 vaccine waning and effectiveness and side-effects of boosters: a prospective community study from the ZOE COVID Study (Lancet Infect Dis. 2022:S1473-3099(22)00146-3) is a longitudinal, prospective, community-based study to assess self-reported systemic and localised adverse reactions of COVID-19 booster doses, in addition to effectiveness against infection (a confounder). Self-reported data may introduce information bias, as some participants might be more likely to report symptoms and some may drop out; however, multi-country CEM studies allow to include large populations, as shown in Cohort Event Monitoring of Adverse Reactions to COVID-19 Vaccines in Seven European Countries: Pooled Results on First Dose (Drug Saf. 2023;46(4):391-404).

Case-only designs

Traditional designs such as the cohort and case-control designs (see Chapters 4.2.1 and 4.2.2) may be difficult to implement in circumstances of high vaccine coverage (for example, in mass immunisation campaigns such as for COVID-19), a lack of an appropriate comparator group (e.g., unvaccinated), or a lack of adequate covariate information at the individual level. Frequent sources of confounding are underlying health status and factors influencing the likelihood of being vaccinated, such as access to healthcare or belonging to a high-risk group (see paragraph 16.2.4.1 on Studies in special populations in this Chapter). In such situations, case-only designs may provide stronger evidence than large cohort studies as they control for fixed individual-level confounders (such as demographics, genetics, or social deprivation) and have similar, sometimes higher, power (see Control without separate controls: evaluation of vaccine safety using case-only methods, Vaccine 2004;22(15-16):2064-70). Case-only designs are discussed in Chapter 4.2.3.

Several publications have compared traditional and case-only study designs for vaccine studies:

- Epidemiological designs for vaccine safety assessment: methods and pitfalls (Biologicals 2012;40(5):389-92) used three designs (cohort, case-control, and self-controlled case-series (SCCS)) to illustrate aspects such as case definition, limitations of data sources, uncontrolled confounding, and interpretation of findings.
- Comparison of epidemiologic methods for active surveillance of vaccine safety (Vaccine 2008; 26(26):3341-45) performed simulations to compare four designs (matched cohort, vaccinated-only

(risk interval) cohort, case-control, and SCCS). The cohort design allowed for the most rapid signal detection, less false-positive error and highest statistical power in sequential analyses. However, one limitation of this simulation was the lack of case validation.

- The simulation study Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations (J Clin Epidemiol. 2006; 59(8):808-18) compared four designs (cohort, case-control, risk-interval and SCCS) and concluded that all were valid, however, with contrasting strengths and weaknesses. The SCCS, in particular, proved to be an efficient and valid alternative to the cohort design.
- Hepatitis B vaccination and first central nervous system demyelinating events: Reanalysis of a case-control study using the self-controlled case series method (Vaccine 2007;25(31):5938-43) describes how the SCCS found similar results as the case-control design but with greater precision, based on the assumption that exposures are independent of earlier events, and recommended that case-series analyses should be conducted in parallel to case-control analyses.

It is increasingly considered good practice to use combined approaches, such as a cohort design and sensitivity analyses using a self-controlled method, as this provides an opportunity for minimising some biases that cannot be taken into account in the primary design (see for example, Myocarditis and pericarditis associated with SARS-CoV-2 vaccines: A population-based descriptive cohort and a nested self-controlled risk interval study using electronic health care data from four European countries; Front Pharmacol. 2022;13:1038043).

While the SCCS is suited to secondary use of data, it may not always be appropriate in situations where rapid evidence generation is needed, since follow-up time needs to be accrued. In such instances, design approaches include the SCRI method that can be used to shorten observation time (see The risk of Guillain-Barre Syndrome associated with influenza A (H1N1) 2009 monovalent vaccine and 2009-2010 seasonal influenza vaccines: Results from self-controlled analyses, Pharmacoepidemiol. Drug Saf 2012;21(5):546-52; and Chapter 4.2.3); O/E analyses using historical background rates (see Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project, Am J Epidemiol 2010;171(2):177-88); or traditional case-control studies (see Guillain-Barré syndrome and adjuvanted pandemic influenza A (H1N1) 2009 vaccine: multinational case-control study in Europe, BMJ 2011;343:d3908).

Nevertheless, the SCCS design is an adequate method to study vaccine safety, provided the main requirements of the method are taken into account (see Chapter 4.2.3). An illustrative example is shown in Bell's palsy and influenza(H1N1)pdm09 containing vaccines: A self-controlled case series (PLoS One. 2017;12(5):e0175539). In First dose ChAdOx1 and BNT162b2 COVID-19 vaccinations and cerebral venous sinus thrombosis: A pooled self-controlled case series study of 11.6 million individuals in England, Scotland, and Wales (PLoS Med. 2022;19(2):e1003927), pooled primary care, secondary care, mortality, and virological data were used. The authors discuss the possibility that the SCCS assumption of event-independent exposure may not have been satisfied in the case of cerebral venous sinus thrombosis (CVST) since vaccination prioritised risk groups, which may have caused a selection effect where individuals more likely to have an event were less likely to be vaccinated and thus less likely to be included in the analyses. In First-dose ChAdOx1 and BNT162b2 COVID-19 vaccines and thrombocytopenic, thromboembolic and hemorrhagic events in Scotland (Nat Med. 2021; 27(7):1290-7), potential residual confounding by indication in the primary analysis (a nested case-control design) was addressed by a SCCS to adjust for time-invariant confounders. Risk of acute myocardial infarction and ischaemic stroke following COVID-19 in Sweden: a self-controlled case series and matched cohort study (Lancet 2021;398(10300):599-607) showed that a COVID-19 diagnosis is an independent risk factor for the events, using two complementary designs in Swedish healthcare data: a SCCS to calculate incidence rate ratios in temporal risk periods following COVID-19 onset, and a matched

cohort study to compare risks within 2 weeks following COVID-19 to the risk in the background population.

A modified self-controlled case series method for event-dependent exposures and high event-related mortality, with application to COVID-19 vaccine safety (Stat Med. 2022;41(10):1735-50) used data from a study of the risk of cardiovascular events, together with simulated data, to illustrate how to handle event-dependent exposures and high event-related mortality, and proposes a newly developed test to determine whether a vaccine has the same effect (or lack of effect) at different doses.

Estimating the attributable risk

The attributable risk of a given safety outcome (assuming a causal effect attributable to vaccination) is an important estimate to support public health decision-making in the context of vaccination campaigns. In the population-based cohort study Investigation of an association between onset of narcolepsy and vaccination with pandemic influenza vaccine, Ireland April 2009-December 2010 (Euro Surveill. 2014;19(17):15-25), the relative risk was calculated as the ratio of the incidence rates for vaccinated and unvaccinated subjects, while the absolute attributable risk was calculated as the difference in incidence rates. Safety of COVID-19 vaccination and acute neurological events: A self-controlled case series in England using the OpenSAFELY platform (Vaccine. 2022;40(32):4479-4487) used primary care, hospital admission, emergency care, mortality, vaccination, and infection surveillance data linked through a dedicated data analytics platform, and calculated the absolute risk of selected AESIs.

Case-coverage design

The case-coverage design is a type of ecological design using exposure information on cases, and population data on vaccination coverage to serve as control. It compares odds of exposure in cases to odds of exposure in the general population, similar to the screening method used in vaccine effectiveness studies (see below paragraph16.2.3.3 in this Chapter). However, it does not control for residual confounding and is prone to selection bias introduced by propensity to seek care (and vaccination) and by awareness of possible occurrence of a specific outcome, and does not consider underlying medical conditions, with limited comparability between cases and controls. In addition, it requires reliable and granular vaccine coverage data corresponding to the population from which cases are drawn, to allow control of confounding by stratified analyses (see for example, Risk of narcolepsy in children and young people receiving AS03 adjuvanted pandemic A/H1N1 2009 influenza vaccine: retrospective analysis, BMJ. 2013; 346:f794).

16.2.2. Vaccine effectiveness

16.2.2.1. General considerations

The book *Vaccination Programmes | Epidemiology, Monitoring, Evaluation* (Hahné, S., Bollaerts, K., & Farrington, P., Routledge, 2021) discusses the concept of vaccine effectiveness and provides further insight into the methods discussed in this section. The book *Design and Analysis of Vaccine Studies* (ME Halloran, IM Longini Jr., CJ Struchiner, Ed., Springer, 2010) presents methods and a conceptual framework of the different effects of vaccination at the individual and population level, and includes methods for evaluating indirect, total and overall effects of vaccination in populations.

A key reference is Vaccine effects and impact of vaccination programmes in post-licensure studies (Vaccine 2013;31(48):5634-42), which reviews methods for the evaluation of the effectiveness of vaccines and vaccination programmes and discusses design assumptions and biases to consider. A framework for research on vaccine effectiveness (Vaccine 2018;36(48):7286-93) proposes

standardised definitions, considers models of vaccine failure, and provides methodological considerations for different designs.

Evaluation of influenza vaccine effectiveness: a guide to the design and interpretation of observational studies (WHO, 2017) provides an overview of methods to study influenza vaccine effectiveness, also relevant for other vaccines. Evaluation of COVID-19 vaccine effectiveness (WHO, 2021) provides guidance on how to monitor COVID-19 vaccine effectiveness using observational study designs, including considerations relevant to low- and middle-income countries. Methods for measuring vaccine effectiveness and a discussion of strengths and limitations are presented in Exploring the Feasibility of Conducting Vaccine Effectiveness Studies in Sentinel's PRISM Program (CBER, 2018). Although focusing on the planning, evaluation, and modelling of vaccine efficacy trials, Challenges of evaluating and modelling vaccination in emerging infectious diseases (Epidemics 2021:100506) includes a useful summary of references for the estimation of indirect, total, and overall effects of vaccines.

16.2.2.2. Sources of exposure and outcome data

Data sources for vaccine studies largely rely on vaccine-preventable infectious disease surveillance (for effectiveness studies) and vaccine registries or vaccination data recorded in healthcare databases (for both safety and effectiveness studies). Considerations on validation of exposure and outcome data are provided in Chapter 5.

Infectious disease surveillance is a population-based, routine public health activity involving systematic data collection to monitor epidemiological trends over time in a defined catchment population, and can use various indicators. Data can be obtained from reference laboratories, outbreak reports, hospital records or sentinel systems, and use consistent case definitions and reporting methods. There is usually no known population denominator, thus surveillance data cannot be used to measure disease incidence. Limitations include under-detection/under-reporting (if passive surveillance) or over-reporting (e.g., due to improvements in case detection or introduction of new vaccines with increased disease awareness). Changes/delays in case counting or reporting can artificially reduce the number of reported cases, thus artificially increasing vaccine effectiveness. Infectious Disease Surveillance (International Encyclopedia of Public Health 2017;222-9) is a comprehensive review including definitions, methods, and considerations on use of surveillance data in vaccine studies. The chapter on Routine Surveillance of Infectious Diseases in *Modern Infectious Disease Epidemiology* (J. Giesecke. 3rd Ed. CRC Press, 2017) discusses how surveillance data are collected and interpreted, and identifies sources of potential bias. Chapter 8 of *Vaccination Programmes | Epidemiology, Monitoring, Evaluation* outlines the main methods of vaccine-preventable disease surveillance, considering data sources, case definitions, biases and methods for descriptive analyses.

Granular epidemiological surveillance data (e.g., by age, gender, pathogen strain) are of particular importance for vaccine effectiveness studies. Such data were available from the European Centre for Disease Prevention and Control and the WHO Coronavirus (COVID-19) Dashboard during the COVID-19 pandemic and, importantly, also included vaccine coverage data.

EHRs and claims-based databases constitute an alternative to epidemiological surveillance data held by national public health bodies, as illustrated in Using EHR data to identify coronavirus infections in hospitalized patients: Impact of case definitions on disease surveillance (Int J Med Inform. 2022;166:104842), which also recommends using sensitivity analyses to assess the impact of variations in case definitions.

Examples of vaccination registries, and challenges in developing such registries, are discussed in Vaccine registers-experiences from Europe and elsewhere (Euro Surveill. 2012;17(17):20159), Validation of the new Swedish vaccination register - Accuracy and completeness of register data (Vaccine 2020; 38(25):4104-10), and Establishing and maintaining the National Vaccination Register

in Finland (Euro Surveill. 2017;22(17):30520). Developed by WHO, Public health surveillance for COVID-19: interim guidance describes key aspects of the implementation of SARS-CoV-2 surveillance, including a section on vaccine effectiveness monitoring in relation to surveillance systems.

16.2.2.3. Study designs for vaccine effectiveness assessment

Traditional cohort and case-control designs

The case-control design has been used to evaluate vaccine effectiveness, but the likelihood of bias and confounding is a potential important limitation. The articles Case-control vaccine effectiveness studies: Preparation, design, and enrollment of cases and controls (Vaccine 2017; 35(25):3295-302) and Case-control vaccine effectiveness studies: Data collection, analysis and reporting results (Vaccine 2017; 35(25):3303-8) provide recommendations on best practices for their design, analysis and reporting. Based on a meta-analysis of 49 cohort studies and 10 case-control studies, Efficacy and effectiveness of influenza vaccines in elderly people: a systematic review (Lancet 2005;366(9492):1165-74) highlights the heterogeneity of outcomes and study populations included in such studies and the high likelihood of selection bias. In A Dynamic Model for Evaluation of the Bias of Influenza Vaccine Effectiveness Estimates From Observational Studies (Am J Epidemiol. 2019;188(2):451-60), a dynamic probability model was developed to evaluate biases in passive surveillance cohort, test-negative, and traditional case-control studies.

Non-specific effects of vaccines, such as a decrease of mortality, have been claimed in observational studies but can be affected by bias and confounding. Epidemiological studies of the 'non-specific effects' of vaccines: I--data collection in observational studies (Trop Med Int Health 2009;14(9):969-76.) and Epidemiological studies of the non-specific effects of vaccines: II--methodological issues in the design and analysis of cohort studies (Trop Med Int Health 2009;14(9):977-85) provide recommendations for observational studies conducted in high mortality settings; however, these recommendations have wider relevance.

The cohort design has been widely used to monitor the effectiveness of COVID-19 vaccines; the following two examples reflect early times of the pandemic, and its later phase when several vaccines were used, reaching wider population groups and used according to different types of vaccination schedule depending on national policies: BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting (N Engl J Med. 2021;384(15):1412-23) used data from a nationwide healthcare organisation to match vaccinated and unvaccinated subjects according to demographic and clinical characteristics, to assess effectiveness against infection, COVID-19 related hospitalisation, severe illness, and death. Vaccine effectiveness against SARS-CoV-2 infection, hospitalization, and death when combining a first dose ChAdOx1 vaccine with a subsequent mRNA vaccine in Denmark: A nationwide population-based cohort study (PLoS Med. 2021;18(12):e1003874) used nationwide linked registries to estimate VE against several outcomes of interest of a heterologous vaccination schedule, compared to unvaccinated individuals. As vaccination coverage increased, using a non-vaccinated comparator group became no longer feasible or suitable, and alternative comparators were needed (see paragraph below on comparative effectiveness).

More recently, pharmacoepidemiological studies have assessed the effectiveness of COVID-19 booster vaccination, which uncovered new methodological challenges, such as the need to account for time-varying confounding. Challenges in Estimating the Effectiveness of COVID-19 Vaccination Using Observational Data (Ann Intern Med. 2023;176(5):685-693) describes two approaches to target trial emulation to overcome limitations due to confounding or designs not considering the evolution of the pandemic over time and the rapid uptake of vaccination. Comparative effectiveness of different primary vaccination courses on mRNA-based booster vaccines against SARs-COV-2 infections: a time-varying cohort analysis using trial emulation in the Virus Watch community cohort (Int J Epidemiol.

2023 Apr 19;52(2):342-354) conducted trial emulation by meta-analysing eight cohort results to reduce time-varying confounding-by-indication.

Test-negative case-control design

The test-negative case-control design aims to reduce bias associated with misclassification of infection and confounding by healthcare-seeking behaviour, at the cost of sometimes difficult-to-test assumptions. The test-negative design for estimating influenza vaccine effectiveness (Vaccine 2013;31(17):2165-8) explains the rationale, assumptions and analysis of this design, originally developed for influenza vaccines. Study subjects were all persons seeking care for an acute respiratory illness, and influenza VE was estimated from the ratio of the odds of vaccination among subjects testing positive for influenza to the odds of vaccination among subject testing negative. Test-Negative Designs: Differences and Commonalities with Other Case-Control Studies with "Other Patient" Controls (Epidemiology. 2019 Nov;30(6):838-44) discusses advantages and disadvantages of the design in various circumstances. The use of test-negative controls to monitor vaccine effectiveness: a systematic review of methodology (Epidemiology 2020;31(1):43-64) discusses challenges of this design for various vaccines and pathogens, also providing a list of recommendations.

In Effectiveness of rotavirus vaccines in preventing cases and hospitalizations due to rotavirus gastroenteritis in Navarre, Spain (Vaccine 2012;30(3):539-43), electronic clinical reports were used to select cases (children with confirmed rotavirus infection) and test-negative controls (children who tested negative for rotavirus in all samples), under the assumption that the rate of gastroenteritis caused by pathogens other than rotavirus is the same in both vaccinated and unvaccinated subjects. A limitation is sensitivity of the laboratory test, which may underestimate vaccine effectiveness. In addition, if the viral type is not available, it is not possible to study the association between vaccine failure and a possible mismatch between vaccine strains and circulating strains. These learnings still apply today in the context of COVId-19 vaccines.

The article Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness (Am J Epidemiol. 2016;184(5):345-53; see also the related Comments) uses directed acyclic graphs to characterise potential biases and shows how they can be avoided or minimised. In Estimands and Estimation of COVID-19 Vaccine Effectiveness Under the Test-Negative Design: Connections to Causal Inference (Epidemiology 2022;33(3):325-33), an unbiased estimator for vaccine effectiveness using the test-negative design is proposed under the scenario of different vaccine effectiveness estimates across patient subgroups.

In the multicentre study in 18 hospitals 2012/13 influenza vaccine effectiveness against hospitalised influenza A(H1N1)pdm09, A(H3N2) and B: estimates from a European network of hospitals (EuroSurveill 2015;20(2):pii=21011), influenza VE was estimated based on the assumption that confounding due to health-seeking behaviour is minimised since all individuals needing hospitalisation are likely to be hospitalised.

Postlicensure Evaluation of COVID-19 Vaccines (JAMA. 2020;324(19):1939-40) describes methodological challenges of the test-negative design applied to COVID-19 vaccines and discusses solutions to minimise bias. Covid-19 Vaccine Effectiveness and the Test-Negative Design (N Engl J Med. 2021;385(15):1431-33) uses the example of a published study in a large hospital network to provide considerations on how to report findings and assess their sensitivity to biases specific to the test-negative design. The study Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study (BMJ 2021;373:n1088) linked routine community testing and vaccination data to estimate effectiveness against confirmed symptomatic infection, COVID-19 related hospital admissions and case fatality, and estimated the odds ratios for testing positive to SARS-CoV-2 in

vaccinated compared to unvaccinated subjects with compatible symptoms. The study also provides considerations on strengths and limitations of the test-negative design.

Case-population, case-coverage, and screening methods

These methods are related, and all include, to some extent, an ecological component such as vaccine coverage or epidemiological surveillance data at population level. Terms to refer to these designs are sometimes used interchangeably. The case-coverage design is discussed above in paragraph 16.2.2.2. Case-population studies are described in Chapter 4.2.5 and in Vaccine Case-Population: A New Method for Vaccine Safety Surveillance (Drug Saf. 2016;39(12):1197-209).

The screening method estimates vaccine effectiveness by comparing vaccination coverage in positive (usually laboratory confirmed) cases of a disease (e.g., influenza) with the vaccination coverage in the population from which the cases are derived (e.g., in the same age group). If representative data on cases and vaccination coverage are available, it can provide an inexpensive and rapid method to provide early estimates or identify changes in effectiveness over time. However, Application of the screening method to monitor influenza vaccine effectiveness among the elderly in Germany (BMC Infect Dis. 2015;15(1):137) emphasises that accurate and age-specific vaccine coverage data are crucial to provide valid estimates. Since adjusting for important confounders and assessing product-specific effectiveness is generally challenging, this method should be considered mainly as a supplementary tool to assess crude effectiveness. COVID-19 vaccine effectiveness estimation using the screening method – operational tool for countries (2022) also provides a good introduction to the method and its strengths and limitations.

Indirect cohort (Broome) method

The indirect cohort method is a case-control type design which uses cases caused by non-vaccine serotypes as controls, and uses surveillance data, instead of vaccination coverage data. Use of surveillance data to estimate the effectiveness of the 7-valent conjugate pneumococcal vaccine in children less than 5 years of age over a 9 year period (Vaccine 2012;30(27):4067-72) evaluated the effectiveness of a pneumococcal conjugate vaccine against invasive pneumococcal disease and compared to the results of a standard case-control design conducted during the same time period. The authors consider the method most useful shortly after vaccine introduction, and less useful in a setting of very high vaccine coverage and fewer cases. Using the indirect cohort design to estimate the effectiveness of the seven valent pneumococcal conjugate vaccine in England and Wales (PLoS One 2011;6(12):e28435) and Effectiveness of the seven-valent and thirteen-valent pneumococcal conjugate vaccines in England: The indirect cohort design, 2006-2018 (Vaccine 2019;37(32):4491-98) describe how the method was used to estimate effectiveness of various vaccine schedules as well as for each vaccine serotype.

Density case-control design

Effectiveness of live-attenuated Japanese encephalitis vaccine (SA14-14-2): a case-control study (Lancet 1996;347(9015):1583-6) describes a case-control study of incident cases in which the control group consisted of all village-matched children of a given age who were at risk of developing disease at the time that the case occurred (density sampling). The effect measured is an incidence density rate ratio. In Vaccine Effectiveness of Polysaccharide Vaccines Against Clinical Meningitis - Niamey, Niger, June 2015 (PLoS Curr. 2016;8), a case-control study compared the odds of vaccination among suspected meningitis cases to controls enrolled in a vaccine coverage survey performed at the end of the epidemic. A simulated density case-control design randomly attributing recruitment dates to controls based on case dates of onset was used to compute vaccine effectiveness. In Surveillance of COVID-19 vaccine effectiveness: a real-time case–control study in southern Sweden (Epidemiol Infect. 2022;150:1-15) a continuous density case-control sampling was performed, with the control group

randomly selected from the complete study cohort as individuals without a positive test the same week as the case or 12 weeks prior.

Waning immunity

Studying how immunity conferred by vaccination wanes over time requires consideration of within-host dynamics of the pathogen and immune system, as well as the associated population-level transmission dynamics. Implications of vaccination and waning immunity (Proc Biol Sci. 2009; 276(1664):2071-80) combined immunological and epidemiological models of measles infection to examine the interplay between disease incidence, waning immunity and boosting.

Global Varicella Vaccine Effectiveness: A Meta-analysis (Pediatrics 2016; 137(3):e20153741) highlights the challenges to reliably measure effectiveness when some confounders cannot be controlled for, force of infection may be high, degree of exposure in study participants may be variable, and data may originate from settings where there is evidence of vaccine failure. Several estimates or studies may therefore be needed to accurately conclude in waning immunity. Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression (Lancet 2022;399(10328):924-944) reviews evidence of changes in efficacy or effectiveness with time since full vaccination for various clinical outcomes; biases in evaluating changes in effectiveness over time, and how to minimise them, are presented in a tabular format. Effectiveness of Covid-19 Vaccines over a 9-Month Period in North Carolina (N Engl J Med. 2022;386(10):933-941) linked COVID-19 surveillance and vaccination data to estimate reduction in current risks of infection, hospitalisation and death as a function of time elapsed since vaccination, and demonstrated durable effectiveness against hospitalisation and death while waning protection against infection over time was shown to be due to both declining immunity and emergence of the delta variant.

Vaccine effectiveness estimates over time are subject to bias from differential depletion of susceptibles (persons at risk of infection) between vaccinated and unvaccinated groups, which can lead to biased estimates of waning effectiveness. Depletion-of-susceptibles bias in influenza vaccine waning studies: how to ensure robust results (Epidemiol Infect. 2019;147:e306) recommends to study only vaccinated persons, and compare for each day the incidence in persons with earlier or later dates of vaccination, to assess waning as a function of vaccination time. Identifying and Alleviating Bias Due to Differential Depletion of Susceptible People in Postmarketing Evaluations of COVID-19 Vaccines (Am J Epidemiol. 2022;191(5):800-11) outlines scenarios under which bias can arise and identifies approaches to minimise these biases.

Comparative vaccine effectiveness

Comparing vaccine benefits has traditionally been performed using head-to-head immunogenicity studies, while comparative effectiveness designs have been used mostly to compare vaccination schedules, vaccine formulations, or administration routes (e.g., for measles, mumps and rubella (MMR), influenza, or pneumococcal vaccines; see for example, Analysis of relative effectiveness of high-dose versus standard-dose influenza vaccines using an instrumental variable method (Vaccine 2019;37(11):1484-90). Methods to account for measured and unmeasured confounders in influenza relative vaccine effectiveness studies: A brief review of the literature (Influenza Other Respir. Viruses 2022;16(5):846-850) discusses methods to account for confounding in such studies. In The risk of non-specific hospitalised infections following MMR vaccination given with and without inactivated vaccines in the second year of life. Comparative self-controlled case-series study in England (Vaccine 2019;37(36):5211-17) the SCCS design was used to compare the effectiveness of the MMR vaccine alone with the MMR vaccine in combination with PCV7 or with both PCV7 and the combined Hib-MenC vaccine. Comparative effectiveness of pneumococcal vaccination with PPV23 and PCV13 in COPD patients over a 5-year follow-up cohort study, (Sci Rep 2021;11(1):15948.) used a prospective cohort

design to compare effectiveness between the 23-valent vaccine, the 13-valent vaccine, and no vaccination.

The COVID-19 vaccination campaigns increased the interest in, and triggered, comparative effectiveness studies. Postmarketing studies: can they provide a safety net for COVID-19 vaccines in the UK? (BMJ Evid Based Med. 2020:bmjebm-2020-111507) discusses methodological and operational aspects and provides considerations on head-to-head vaccine comparisons. Assessment of Effectiveness of 1 Dose of BNT162b2 Vaccine for SARS-CoV-2 Infection 13 to 24 Days After Immunization (JAMA Netw Open. 2021;4(6):e2115985) compared the effectiveness of the first vaccine dose between two post-immunisation periods. Comparative effectiveness of the BNT162b2 and ChAdOx1 vaccines against Covid-19 in people over 50 (Nat Commun. 2022;13(1):1519) used data from the UK Biobank linked to primary care, hospital admissions, and COVID-19 testing data, to compare the effectiveness of BNT162b2 vs. ChAdOx1s against COVID-19 infection and hospitalisation, using propensity score modelling. Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines in U.S. Veterans (N Engl J Med. 2022;386(2):105-15) and Comparative effectiveness of BNT162b2 versus mRNA-1273 covid-19 vaccine boosting in England: matched cohort study in OpenSAFELY-TPP used a target trial emulation design.

Comparative vaccine effectiveness studies may require larger sample sizes, as they aim to detect smaller effect sizes as opposed to effectiveness studies for a single vaccine, where an unvaccinated group is used as a comparator. Various sources of confounding (such as self-seeking testing behaviour) should be considered, and appropriate methods used, such as (propensity score) matching, instrumental variable analysis, inverse probability of treatment weighting, use of negative control outcomes, off-season outcomes (for influenza vaccines) and positive control outcomes. For some vaccines (e.g., COVID-19 vaccines), variant-specific comparative effectiveness data are important, taking into consideration the correlation between vaccine schedules and calendar periods, and therefore with variants in circulation at a given time.

Impact studies

Vaccine impact studies estimate disease reduction in a community. These studies are typically ecological or modelling analyses that compare disease outcomes pre- and post-vaccine introduction. Reductions in disease outcomes are observed through direct effects of vaccination in vaccinated people, and indirect effects due to reduced transmission within a community. Other concurrent interventions or phenomena unrelated to vaccine effects, such as changes in risk behaviours or healthcare practices, may reduce disease outcomes and confound the assessment of vaccine impact (see The value of vaccine programme impact monitoring during the COVID-19 pandemic, Lancet 2022;399(10320):119-21). For example, for a paediatric vaccine, the impact of vaccination can be quantified in the targeted age group (overall effect) or in other age groups (indirect effect). For an overview, see Vaccine effects and impact of vaccination programmes in post-licensure studies (Vaccine 2013;31(48):5634-42).

Direct and indirect effects in vaccine efficacy and effectiveness (Am J Epidemiol. 1991;133(4):323-31) describes how parameters intended to measure direct effects must be robust and interpretable in the midst of complex indirect effects of vaccine intervention programmes. Lack of impact of rotavirus vaccination on childhood seizure hospitalizations in England - An interrupted time series analysis (Vaccine 2018; 36(31):4589-92) discusses possible reasons for negative findings compared to previous studies. In a review of 65 articles, Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis (Lancet. 2019;394(10197):497–509) compared the prevalence or incidence of several HPV-related endpoints between the pre- and post-vaccination periods with stratification by sex, age, and years since introduction of HPV vaccination.

Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data (Lancet. 2021;397(10287):1819-29) evaluated the public health impact of vaccination using national surveillance and vaccine uptake data. Although such population-level data are ecological, and teasing apart the impact of the vaccination programme from the impact of non-pharmaceutical interventions is complex, declines in incident cases by age group were shown to be aligned with high vaccine coverage rather than initiation of the nationwide lockdown.

Accumulated effectiveness data has suggested the potential for a population-level effect of COVID-19 vaccination, which has been critical to control the pandemic. Community-level evidence for SARS-CoV-2 vaccine protection of unvaccinated individuals (Nat Med. 2021;27(8):1367-9) analysed vaccination records and test results in a large population, while mitigating the confounding effect of natural immunity and the spatiotemporally dynamic nature of the epidemic, and showed that vaccination provided cross-protection to unvaccinated individuals in the community.

Transmission studies

Vaccination programmes have indirect effects at the population-level, also called herd immunity, as a result of reduced transmission. Besides measuring the direct effect of vaccination in vaccine effectiveness studies, it is important to assess whether vaccination has an effect on transmission. As a high-risk setting, households can provide evidence of such impact.

Among the first studies of the impact of COVID-19 vaccination on transmission, Effect of Vaccination on Household Transmission of SARS-CoV-2 in England (N Engl J Med. 2021;385(8):759-60) was a nested case-control study estimating odds ratios for household members becoming secondary cases if the case was vaccinated within 21 days or more before testing positive, vs. household members where the case was not vaccinated. Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel (Science. 2022;375(6585):1151-54) assessed the effectiveness of BNT162b2 against susceptibility to infection and infectiousness, comparing pre- and post-Delta periods, using a chain binomial model applied to data from a large healthcare organisation. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study (Lancet Infect Dis. 2022;22(2):183-95) ascertained secondary transmission by longitudinally following index cases and their contacts (regardless of symptoms) early after exposure to the Delta variant, and highlights the importance of community studies to characterise transmission in highly vaccinated populations.

Specific limitations of transmission studies such as likelihood of information bias (misclassification) and selection bias, should be considered when interpreting findings and are discussed in the above references.

Cluster design

A cluster is a group of subjects sharing common characteristics: geographical (community, administrative area), health-related (hospital), educational (schools), or social (household). In cluster randomised trials, clusters instead of individual subjects are randomly allocated to an intervention, whereas in infectious disease epidemiology studies, clusters are sampled based on aspects of transmission (e.g., within a community) or a vaccination programme. This design is often used in low and middle income settings and can measure vaccination interventions naturally applied at the cluster level or when the study objectives require a cluster design (e.g., to estimate herd immunity).

Meningococcal B Vaccine and Meningococcal Carriage in Adolescents in Australia (N Engl J Med. 2020;382(4):318-27) used cluster randomisation to assign students, according to school, to receive

4CMenB vaccination, either at baseline or at 12 months (as a control) to measure oropharyngeal carriage.

In The ring vaccination trial: a novel cluster randomised controlled trial design to evaluate vaccine efficacy and effectiveness during outbreaks, with special reference to Ebola (BMJ. 2015;351:h3740), a newly diagnosed Ebola case served as the index case to form a "ring", which was then randomised to immediate or delayed vaccination with inclusion based on tracing cases using active surveillance instead of randomisation. In Assessing the safety, impact and effectiveness of RTS,S/AS01 E malaria vaccine following its introduction in three sub-Saharan African countries: methodological approaches and study set-up (Malar J. 2022;21(1):132), active surveillance was used to enrol large numbers of children in vaccinated and unvaccinated clusters as part of the WHO Malaria Vaccine Implementation Programme, to conduct temporal (before vs. after) and concurrent (exposed vs. unexposed) cluster comparisons. Clusters were selected based on geographically limited areas with demographic surveillance in place and infrastructure to monitor population health and vaccination programmes.

Misclassification in studies of vaccine effectiveness

Like vaccine safety studies, studies of vaccine effectiveness rely on accurate identification of vaccine exposure and of cases of the targeted vaccine-preventable disease/infection, but in practice, diagnostic tests, clinical case definitions and vaccination records often present inaccuracies. Bias due to differential and non-differential disease- and exposure misclassification in studies of vaccine effectiveness (PLoS One 2018;15;13(6):e0199180) explores through simulations the impact of non-differential and differential disease- and exposure-misclassification when estimating vaccine effectiveness using cohort, case-control, test-negative case-control and case-cohort designs. This can also be applied to safety outcomes, especially those with a complex natural history such as neurological or potential immune mediated diseases, and is particularly relevant for secondary use of data, where validation studies may be needed in a first step. Misclassification can lead to significant bias and its impact strongly depends on the vaccination scenarios. A web application designed in the ADVANCE project is publicly available to assess the potential (joint) impact of possibly differential disease- and exposure misclassification.

16.2.3. Specific aspects of pharmacoepidemiological vaccine studies

16.2.3.1. Studies in special populations

Special populations include pregnant and breastfeeding persons, immunocompromised patients (including transplanted patients), paediatric populations, older adults/the elderly, and patients with rare disorders. Post-authorisation studies are often required for these populations, which are usually not included in the clinical development of vaccines. In real-world settings, special populations are often the subject of specific vaccination recommendations, which may impact study designs and choice of an appropriate comparator. This was the case, for example, of COVID-19 vaccines which initially targeted high-risk priority groups. The article Vaccine safety in special populations (Hum Vaccin. 2011;7(2):269-71) highlights design issues when evaluating vaccine safety in these populations. Methodological challenges include defining the study population (particularly for immunocompromised populations), low sample size due to rare outcomes, accounting for comorbidities and other types of confounders, or difficulty in identifying cases or disease duration and severity in immunocompromised patients.

Influenza vaccination for immunocompromised patients: systematic review and meta-analysis by etiology (J Infect Dis. 2012;206(8):1250-9) illustrates the importance of performing stratified analyses by aetiology of immunocompromised status and limitations due to residual confounding, differences within and between etiological groups and small sample size in some subgroups. In anticipation of the

design of post-authorisation vaccine effectiveness and safety studies, the study Burden of herpes zoster in 16 selected immunocompromised populations in England: a cohort study in the Clinical Practice Research Datalink 2000–2012 (BMJ Open 2018;8(6): e020528) illustrated the challenges of defining an immunocompromised cohort and a relevant comparator cohort in a primary healthcare database. Validation of a Method to Identify Immunocompromised Patients with Severe Sepsis in Administrative Databases (Ann Am Thorac Soc. 2016;13(2):253-8) provides considerations on identifying this group of patients in large administrative databases.

Pregnant and breastfeeding persons represent an important group to be addressed when monitoring vaccine use; Annex 2 of this Guide provides guidance on methods to evaluate medicines in pregnancy and breastfeeding, including for vaccine studies. The Guidance for design and analysis of observational studies of foetal and newborn outcomes following COVID-19 vaccination during pregnancy (Vaccine 2021;39(14):1882-6) provides useful insights on study design, data collection, and analytical issues in COVID-19 vaccine safety studies in pregnant people, and can be applied to other vaccines.

16.2.3.2. Meta-analyses

The guidance on conducting meta-analyses of pharmacoepidemiological studies of safety outcomes (Annex 1 of this Guide) is also applicable to vaccines. A systematic review evaluating the potential for bias and the methodological quality of meta-analyses in vaccinology (Vaccine 2007;25(52):8794-806) provides a comprehensive overview of quality and limitations of meta-analyses. Meta-analysis of the risk of autoimmune thyroiditis, Guillain-Barré syndrome, and inflammatory bowel disease following vaccination with AS04-adjuvanted human papillomavirus 16/18 vaccine (Pharmacoepidemiol Drug Saf. 2020;29(9):1159-67) combined data from 18 randomised controlled trials, one cluster-randomised trial, two large observational retrospective cohort studies, and one case-control study, resulting in a large sample size for these rare events. The Systematic review and meta-analysis of the effectiveness and perinatal outcomes of COVID-19 vaccination in pregnancy (Nat Commun. 2022;13(1):2414) generated evidence on a large number of adverse pregnancy and perinatal outcomes.

Meta-analytical methods are increasingly used in multi-database studies (see Chapter 9) to combine data generated at country level to obtain pooled risk estimates in large populations. In SARS-CoV-2 Vaccination and Myocarditis in a Nordic Cohort Study of 23 Million Residents (JAMA Cardiol. 2022;7(6):600-12), four cohort studies were conducted in linked nationwide health registers in Denmark, Finland, Norway, and Sweden according to a common protocol; the results were combined using meta-analysis and the homogeneity of country-specific estimates was tested.

16.2.3.3. Pharmacogenetic studies

There is increasing interest in the role of genomics in pharmacoepidemiology (see Chapter 16.3), including for the study of vaccine safety outcomes (see Adversomics: a new paradigm for vaccine safety and design, Expert Rev Vaccines 2015; 14(7): 935–47). Vaccinomics and Adversomics in the Era of Precision Medicine: A Review Based on HBV, MMR, HPV, and COVID-19 Vaccines (J Clin Med. 2020;9(11):3561) highlights that knowledge of genetic factors modulating responses to vaccination could contribute to the evaluation of vaccine safety and effectiveness. In State-wide genomic epidemiology investigations of COVID-19 in healthcare workers in 2020 Victoria, Australia: Qualitative thematic analysis to provide insights for future pandemic preparedness (Lancet Reg Health West Pac. 2022;25:100487), a large SARS-CoV-2 genomic epidemiological investigation identified transmission dynamics using a newly developed set of metadata. Genetic risk and incident venous thromboembolism in middle-aged and older adults following COVID-19 vaccination (J Thromb Haemost. 2022;20(12):2887-2895) used data from the UK Biobank to estimate hazard ratios of the associations between a polygenic risk score and post-vaccination incident veinous thromboembolism.

16.2.3.4. Generic protocols

Generic protocols, also referred to as template or master protocols, provide a standardised structure to support study design and protocol development. Such protocols have supported the urgent need for COVID-19 vaccine monitoring, often based, in Europe, on the EMA Guidance for the format and content of the protocol of non-interventional post-authorisation safety studies (2012).

A protocol for generating background rates of AESIs for the monitoring of COVID-19 vaccines (2021) was developed by the vACcine Covid-19 monitoring readinESS (ACCESS) consortium, which also published Template study protocols (2021) to support the design of safety studies, based on both cohort-event monitoring and secondary use of data. The protocol Rapid assessment of COVID-19 vaccines safety concerns through electronic health records- a protocol template from the ACCESS project compares the suitability of the ecological design and the unadjusted self-controlled risk interval (SCRI) for rapid safety assessment, by type of AESI. Other published templates include FDA's Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring protocol, the COVID-19 Vaccine Safety Active Monitoring Protocol and the Master Protocol: Assessment of Risk of Safety Outcomes Following COVID-19 Vaccination (FDA BEST Initiative, 2021); and the Template for observational study protocols for sentinel surveillance of adverse events of special interest (AESIs) after vaccination with COVID-19 vaccines (WHO, 2021).

The ACCESS consortium also published template protocols (2021) for COVID-19 vaccine effectiveness studies using the cohort and test-negative case-control designs. The Core protocol for ECDC studies of COVID-19 vaccine effectiveness against hospitalisation with Severe Acute Respiratory Infection laboratory-confirmed with SARS-CoV-2 (ECDC, 2021) presents the main elements to consider to design multi-centre, multi-country hospital-based COVID-19 vaccine effectiveness studies in patients hospitalised with severe acute respiratory infections (SARI).

The DRIVE project developed a Core protocol for type/brand specific influenza vaccine effectiveness studies - Test-negative design studies and a Core protocol for population-based database cohort-studies, and the COVIDRIVE consortium a Brand-specific COVID-19 vaccine effectiveness protocol to assess effectiveness against severe COVID-19 disease.

Generic protocols for retrospective case-control studies and retrospective cohort studies to assess the effectiveness of rotavirus and influenza vaccination in EU Member States are published by ECDC and describe potential data sources to identify virological outcomes. The Protocol for Cluster Investigations to Measure Influenza Vaccine Effectiveness (ECDC, 2009) builds on the cluster design to generate rapid/early influenza season estimates in settings where investigation can take place at the same time as vaccination is carried out (e.g. schools, care homes). The generic study protocol to assess the impact of rotavirus vaccination (ECDC, 2013) lists the information to be collected to compare the incidence/proportion of rotavirus cases in the period before and after vaccine introduction.

Although developed for specific vaccines, all these protocols can be tailored to other vaccine exposures and outcomes, as they address the most important aspects to consider for the design of vaccine safety and effectiveness studies.

16.3. Design, implementation and analysis of pharmacogenomic studies

16.3.1. Introduction

Individual differences in the response to medicines encompass variation in both efficacy/effectiveness and safety, including the risk of severe adverse drug reactions. Clinical factors influencing response include disease severity, age, gender, and concomitant drug use. However, natural genetic variation that influences the expression or activity of proteins involved in drug disposition (absorption,

metabolism, distribution, and excretion) as well as the protein targets of drug action (such as enzymes and receptors) may be an important additional source of inter-individual variability in both the beneficial and adverse effects of drugs (see Pharmacogenomics: translating functional genomics into rational therapeutics. Science 1999;286(5439):487-91).

Pharmacogenetics is defined as the study of variation in the DNA sequence as a determinant of drug response. Drug response may vary as a result of differences in the DNA sequence present in the germline or, in the case of cancer treatments, due to somatic variation in the DNA arising in cancer cells (see The Roles of Common Variation and Somatic Mutation in Cancer Pharmacogenomics, Oncol Ther. 2019;7(1):1-32; Systematic pan-cancer analysis of mutation-treatment interactions using large real-world clinicogenomics data, Nat Med. 2022 Aug;28(8):1656-1661). Notably, in the case of treatment or prevention of infectious diseases, the genome of both the pathogen and the host may influence drug and vaccine responses, either independently, interactively or jointly (see Pharmacogenomics and infectious diseases: impact on drug response and applications to disease management, Am J Health Syst Pharm. 2002;59(17):1626-31; The potential of genomics for infectious disease forecasting, Nat Microbiol. 2022 Nov;7(11):1736-1743). For example, COVID-19 vaccine effectiveness changes significantly according to SARS-CoV-2 variant, likely due to vaccine-escape mutations in the virus genome (see Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries, Genomics 2021; 113(4):2158-2170 and Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant, N Engl J Med 2022; 386:1532-46). When incorporated, the study of genetic variation underlying drug response can complement information on clinical factors and disease sub-phenotypes to optimise the prediction of treatment response and reduce the risk of adverse reactions. The identification of variation in genes that modify the response to drugs provides an opportunity to optimise safety and effectiveness of the currently available drugs and to develop new drugs for paediatric and adult populations (see Drug discovery: a historical perspective, Science 2000;287(5460):1960-4).

The broader term *pharmacogenomics* has been used to describe the study changes both in the DNA and RNA, and how they may determine drug response. However, the distinction between pharmacogenetics and pharmacogenomics is arbitrary, and both terms are used interchangeably. It is important to note that pharmacogenomics is one of several approaches available to identify useful biomarkers of drug effects. Other approaches include, but are not limited to, epigenomics (the study of gene expression changes not attributable to changes in the DNA sequence), transcriptomics, proteomics (protein function and levels, see Precision medicine: from pharmacogenomics to pharmacoproteomics, Clin Proteom. 2016; 13:25), and metabolomics.

16.3.2. Identification of genetic variants influencing drug response

Approaches

Identification of genetic variation associated with important drug or therapy-related outcomes can be carried out by three main technologies. Their choice may be dictated by whether the aim is research and discovery or clinical application, and whether the genetic variants being sought occur at high or low frequency in the population or patient group(s) being evaluated. The strategy to identify genetic variants will depend on the aim and design of the pharmacogenetic study or the clinical application (see Methodological and statistical issues in pharmacogenomics, J Pharm Pharmacol. 2010;62(2):161-6). For illustration, to assess clinical applications, technologies might be used to identify genetic variants where there is already prior knowledge about the gene or the variant (*candidate gene studies*). These studies require prior information about the likelihood of the polymorphism, gene, or gene-product interacting with a drug or drug pathway, and thus, resources can be directed to several important genetic polymorphisms with a higher *a priori* chance of relevant drug-gene interactions.

Moving towards individualized medicine with pharmacogenomics (Nature 2004;429(6990):464-8) explains that lack or incompleteness of information on genes from previous studies may result in the failure in identifying every important genetic determinant in the genome.

In contrast, *genome-wide scan approaches* are discovery orientated and use technologies that identify genetic variants across the genome without previous information or gene/variant hypothesis (hypothesis-generating or hypothesis-agnostic approach). Genome-wide approaches are widely used to discover the genetic basis of common complex diseases where multiple genetic variations contribute to disease risk. The same study design is applicable to identification of genetic variants that influence treatment response. However, common variants in the genome, if functional, have generally small effect sizes, and therefore large sample sizes should be considered, for example by pooling different studies as done by the CHARGE Consortium with its focus on cardiovascular diseases (see The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium as a model of collaborative science, Epidemiology 2013;24(3):346-8). By comparing the frequency of genetic variants between drug responders and non-responders, or those with or without drug toxicity, genome-wide approaches can identify important genetic determinants. They may detect variants in genes, which were previously not considered as candidate genes, or even variants outside of the genes. However, because of the concept of linkage disequilibrium, whereby certain genetic determinants tend to be co-inherited together, it is possible that the genetic associations identified through a genome-wide approach may not be truly biologically functional polymorphisms, but instead may simply be a linkage-related marker of another genetic determinant that is the true biologically relevant genetic determinant. Thus, this approach is considered discovery in nature. Furthermore, failure to cover all relevant genetic risk factors can still be a problem, though less than with the candidate gene approach. It is therefore essential to conduct replication studies in independent cohorts and validation studies (*in vivo and in vitro*) to ascertain the generalisability of findings to populations of individuals, to characterise the mechanistic basis of the effect of these genes on drug action, and to identify true biologic genetic determinants. Importantly, allele frequencies differ across populations, and these differences should be accounted for to reduce biases when designing and analysing pharmacogenetic studies, and to ensure equity when implementing pharmacogenomics in the healthcare setting (see Preventing the exacerbation of health disparities by iatrogenic pharmacogenomic applications: lessons from warfarin, Pharmacogenomics 2018 19(11):875-81).

More recently, pharmacogenomic studies have also been performed in large national biobanks which link genetic data to healthcare data for cohorts of hundreds of thousands of subjects, such as the UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age (PLoS Med. 2015;12(3):e1001779) and the Estonian Biobank (see Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu, Int J Epidemiol. 2015;44(4):1137-47). Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions and other studies (Genet Med. 2019;21(6):1345-54) shows that these large-scale resources represent unique opportunities to discover novel and rare variants.

Technologies used for detection of genetic variants

The main technologies are:

- *Genotyping and array-based technologies* which are the most feasible and cost-effective approach for most large-scale clinical utility studies and for clinical implementation, either through commercial or customised arrays. They can identify hundreds of thousands of genetic variants within one or several genes, including a common form of variations known as single nucleotide polymorphisms (SNPs). The identification of genetic determinants is limited to the variants included in the array, and thus, it cannot be used to discover novel variants. Generally, they are chosen on

the grounds of biological plausibility, which may have been proven before in previous studies, or of knowledge of functional genes known to be involved in pharmacokinetic and pharmacodynamics pathways or related to the disease or intermediate phenotype.

- *Sanger sequencing* represents the gold standard used in clinical settings for confirming genetic variants since it was first commercialised in 1986. More recently, Sanger sequencing has been replaced by other sequencing methods to increase the speed and reduce the cost of DNA sequencing, especially for automated analysis involving large numbers of samples.
- *Next generation sequencing* (NGS) is a high-throughput sequencing technology that identifies genetic variants across the genome (whole genome sequencing; WGS) or the exome (whole exome sequencing; WES) without requiring prior knowledge on genetic biomarkers. These techniques may prove valuable in early research settings for discovery of novel or rare variants, and for the detection of structural variants and copy number variation which are common in pharmacogenes such as *CYP2D6* (see A Review of the Important Role of CYP2D6 in Pharmacogenomics, Genes (Basel) 2020;11(11):1295; Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene-drug interaction between CYP2C19 and CYP2D6 and SSRIs, Eur J Hum Genet. 2022 Oct;30(10):1114-1120; Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2C19 Genotype and Clopidogrel Therapy: 2022 Update, Clin Pharmacol Ther. 2022 Nov;112(5):959-967). As use of clinical WGS testing increases, the return of secondary pharmacogenomic findings will benefit from greater understanding of rare and novel variants.

Variant curation and annotation

Lastly, the identification of genetic variants requires careful curation and annotation to ensure that their description and allelic designation is standardised. Common pharmacogenomic variants and haplotypes (combinations of sequence variants in the same individual) are catalogued by the Pharmacogene Variation Consortium (PharmVar) using a 'star allele' nomenclature. The use of this nomenclature is historic and in human disease genetics, the reference sequence identifier (rs-id) is more commonly used as to assign a genetic variant unambiguously. Although the star allele nomenclature remains the most widely used classification in pharmacogenomic research it is recognised to have several limitations. Pharmacogenomic haplotypes and star alleles can lack accurate definition and validation, and there may be limited annotation of phenotypic effects. In addition, current classifications also exclude many rare variants which are increasingly recognised as having an important effect, as described in Pharmacogenetics at Scale: An Analysis of the UK Biobank (Clin Pharmacol Ther. 2021;109(6):1528-37). Some authors have called for an effort to standardise annotation sequence variants (see The Star-Allele Nomenclature: Retooling for Translational Genomics, Clin Pharmacol Ther. 2007;82(3):244–8).

16.3.3. Study designs

Several options are available for the design of pharmacogenetic studies to ascertain the effect and importantly the clinical relevance and utility of obtaining pharmacogenetic information to guide prescribing decisions regarding the choice and dose of agent for a particular condition (see Prognosis research strategy (PROGRESS) 4: Stratified medicine research, BMJ. 2013;346:e5793).

RCTs, both pre- and post-authorisation, provide the opportunity to address several pharmacogenetic questions. Pharmacogenetics in randomized controlled trials: considerations for trial design (Pharmacogenomics 2011;12(10):1485-92) describes three different trial designs differing in the timing of randomization and genotyping, and Promises and challenges of pharmacogenetics: an overview of study design, methodological and statistical issues (JRSM Cardiovasc Dis. 2012;1(1)) discusses outstanding methodological and statistical issues that may lead to heterogeneity among

reported pharmacogenetic studies and how they may be addressed. Pharmacogenetic trials can be designed (or *post hoc* analysed) with the intention to study whether a subgroup of patients, defined by certain genetic characteristics, respond differently to the treatment under study. Alternatively, a trial can verify whether genotype-guided treatment is beneficial over standard care. Obvious limitations with regard to the assessment of rare adverse drug events or low prevalence genetic variants are the large sample size required and its related high costs. In order to make a trial as efficient as possible in terms of time, money and/or sample size, it is possible to opt for an adaptive trial design, which allows prospectively planned modifications in design after patients have been enrolled in the study. Such a design uses accumulating data to decide how to modify aspects of the study during its progress, without undermining the validity and integrity of the trial. An additional benefit is that the expected number of patients exposed to an inferior/harmful treatment can be reduced (see Potential of adaptive clinical trial designs in pharmacogenetic research, Pharmacogenomics 2012;13(5):571-8).

Observational studies are an alternative and can be family-based (using twins or siblings) or population-based (using unrelated individuals). The main advantage of family-based studies is the avoidance of bias due to population stratification. A clear practical disadvantage for pharmacogenetic studies is the requirement to study families where patients have been treated with the same drugs (see Methodological quality of pharmacogenetic studies: issues of concern, Stat Med. 2008;27(30):6547-69).

Population-based studies may be designed to assess drug-gene interactions as cohort (including exposure-only), case-cohort and case-control studies (including case-only, as described in Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls! Am J Epidemiol. 1996;144(3):207-13). Sound pharmacoepidemiological principles as described in this Guide also apply to observational pharmacogenetic studies. A specific type of confounding due to population stratification needs to be considered in pharmacogenetic studies, and, if present, needs to be dealt with. Its presence may be obvious where the study population includes more than one immediately recognisable ethnic group; however, in other studies stratification may be more subtle. Population stratification can be detected by the Pritchard and Rosenberg's method, which involves genotyping additional SNPs in other areas of the genome and testing for association between them and outcome (see Association mapping in structured populations, Am J Hum Genet. 2000;67(1):170-81). In genome-wide association studies, the data contained within the many SNPs typed can be used to assess population stratification without the need to undertake any further genotyping. Several methods have been suggested to control for population stratification such as genomic control, structure association and EIGENSTRAT. These methods are discussed in Methodological quality of pharmacogenetic studies: issues of concern (Stat Med. 2008;27(30):6547-69), Softwares and methods for estimating genetic ancestry in human populations (Hum Genomics 2013;7(1):1) and Population Stratification in Genetic Association Studies (Curr Protoc Hum Genet. 2017;95:1.22.1–1.22.23).

The main advantage of exposure-only and case-only designs is the smaller sample size that is required, at the cost of not being able to study the main effects of drug exposure (case-only) or genetic variant (exposure-only) on the outcome. Furthermore, interaction can be assessed only on a multiplicative scale, whereas from a public health perspective, additive interactions are very relevant. Up till now GWAS with gene*interactions have not been very rewarding because of the required huge power. However, this is likely to improve as genetic data is linked to longitudinal clinical data in large biobanks, as described in Drug Response Pharmacogenetics for 200,000 UK Biobank Participants (Biocomputing 2021;184-95). An important condition that has to be fulfilled for case-only studies is that the exposure is independent of the genetic variant, e.g., prescribers are not aware of the genotype of a patient and do not take this into account, directly or indirectly (by observing clinical characteristics associated with the genetic variant). In the exposure-only design, the genetic variant

should not be associated with the outcome, for example variants of genes coding for cytochrome p-450 enzymes. When these conditions are fulfilled and the main interest is in the drug-gene interaction, these designs may be an efficient option. In practice, case-control and case-only studies usually result in the same interaction effect as empirically assessed in Bias in the case-only design applied to studies of gene-environment and gene-gene interaction: a systematic review and meta-analysis (Int J Epidemiol. 2011;40(5):1329-41). The assumption of independence of genetic and exposure factors can be verified among controls before proceeding to the case-only analysis. Further development of the case-only design for assessing gene-environment interaction: evaluation of and adjustment for bias (Int J Epidemiol. 2004;33(5):1014-24) conducted sensitivity analyses to describe the circumstances in which controls can be used as proxy for the source population when evaluating gene-environment independence. The gene-environment association in controls will be a reasonably accurate reflection of that in the source population if baseline risk of disease is small (<1%) and the interaction and independent effects are moderate (e.g., risk ratio<2), or if the disease risk is low (e.g., <5%) in all strata of genotype and exposure. Furthermore, non-independence of gene-environment can be adjusted in multivariable models if non-independence can be measured in controls. Further methodological considerations and assumptions of study designs in pharmacogenomics research are discussed in A critical appraisal of pharmacogenetic inference (Clin Genet. 2018;93(3): 498-507).

Lastly, variation in prevalence and effect of pharmacogenetic variants across different ethnicities is an important consideration for study design and ultimately clinical utility, cost-effectiveness and implementation of testing. International research collaborations, as demonstrated in several studies (see HLA-B*5701 Screening for Hypersensitivity to Abacavir, N Engl J Med. 2008;358(6):568-79; and Effect of Genotype-Guided Oral P2Y12 Inhibitor Selection vs Conventional Clopidogrel Therapy on Ischemic Outcomes After Percutaneous Coronary Intervention: The TAILOR-PCI Randomized Clinical Trial, JAMA. 2020; 25;324(8):761-71), encourage greater representation of different populations and ensure broader applicability of pharmacogenomic study results. Diverse ethnic representation in study recruitment is important to detect the range of variant alleles of importance across different ethnic groups and reduce inequity in the clinical impact of pharmacogenomic testing once implemented.

16.3.4. Data collection

The same principles and approaches to data collection as for other pharmacoepidemiological studies can be followed (see Chapter 8 of this Guide on Approaches to Data Collection). An efficient approach to data collection for pharmacogenetic studies is to combine secondary use of electronic health records with primary data collection (e.g., collection of biological samples to extract DNA).

Examples are provided in SLCO1B1 genetic variant associated with statin-induced myopathy: a proof-of-concept study using the clinical practice research datalink (Clin Pharmacol Ther. 2013;94(6):695-701), Diuretic therapy, the alpha-adducin gene variant, and the risk of myocardial infarction or stroke in persons with treated hypertension (JAMA. 2002;287(13):1680-9) and Interaction between the Gly460Trp alpha-adducin gene variant and diuretics on the risk of myocardial infarction (J Hypertens. 2009;27(1):61-8). Another approach to enrich electronic healthcare records with data from biological samples is record linkage to biobanks as illustrated in Genetic variation in the renin-angiotensin system modifies the beneficial effects of ACE inhibitors on the risk of diabetes mellitus among hypertensives (Hum Hypertens. 2008;22(11):774-80). A third approach is to use active surveillance methods to fully characterise drug effects such that a rigorous phenotype can be developed prior to genetic analysis. This approach was followed in Adverse drug reaction active surveillance: developing a national network in Canada's children's hospitals (Pharmacoepidemiol Drug Saf. 2009;18(8):713-21) and EUDRAGENE: European collaboration to establish a case-control DNA collection for studying the genetic basis of adverse drug reactions (Pharmacogenomics 2006;7(4):633-8).

16.3.5. Data analysis

The focus of data analysis should be on the measure of effect modification (see Chapter 7). Attention should be given to whether the mode of inheritance (e.g., dominant, recessive or additive) is defined *a priori* based on prior knowledge from functional studies. However, investigators are usually naïve regarding the underlying mode of inheritance. A solution might be to undertake several analyses, each under a different assumption, though the approach to analysing data raises the problem of multiple testing (see Methodological quality of pharmacogenetic studies: issues of concern, Stat Med. 2008;27(30):6547-69). The problem of multiple testing and the increased risk of type I error is in general a problem in pharmacogenetic studies evaluating multiple SNPs, multiple exposures and multiple interactions. The most common approach to correct for multiple testing is to use the Bonferroni correction. This correction may be considered too conservative and runs the risk of producing null results. Other approaches to adjust for multiple testing include permutation testing and false discovery rate (FDR) control, which are less conservative. The FDR, described in Statistical significance for genome-wide studies (Proc Natl Acad Sci. USA 2003;100(16):9440-5), estimates the expected proportion of false-positives among associations that are declared significant, which is expressed as a q-value.

Alternative innovative methods are becoming increasingly used, such as Mendelian Randomization (see Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality, Annu Rev Genomics Hum Genet. 2015;16:327-50), systems biology, Bayesian approaches, data mining (see Methodological and statistical issues in pharmacogenomics, J Pharm Pharmacol. 2010;62(2):161-6) and polygenic risk scores (see Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations, Nat Genet. 2018;50(9):1219-1224; The potential of polygenic scores to improve cost and efficiency of clinical trials, Nat Commun. 2022;13(1):2922; Polygenic heterogeneity in antidepressant treatment and placebo response, Transl Psychiatry. 2022;12(1):456; Genetic risk and incident venous thromboembolism in middle-aged and older adults following COVID-19 vaccination, J Thromb Haemost. 2022;20(12):2887-2895).

Important complementary approaches include the conduct of individual patient data meta-analyses and/or replication studies to avoid the risk of false-positive findings.

An important step in analysis of genome-wide association studies data that needs to be considered is the conduct of rigorous quality control procedures before conducting the final association analyses. This becomes particularly important when phenotypic data were originally collected for a different purpose ("secondary use of data"). Relevant guidelines include Guideline for data analysis of genomewide association studies (Cancer Genomics Proteomics 2007;4(1):27-34) and Statistical Optimization of Pharmacogenomics Association Studies: Key Considerations from Study Design to Analysis (Curr Pharmacogenomics Person Med. 2011;9(1):41-66).

To improve both reproducibility, efficiency and interoperability across multiple data sources, the use of a common data model (CDM) is increasingly used in pharmacoepidemiological studies (see Chapter 9). Some healthcare databases with genomic data have been mapped to CDMs (see Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond, J Am Med Inform Assoc. 2022;30(1):103-111). However, CDMs were developed for routinely collected healthcare and claims data, and thus, data on genetic variation are in general not yet integrated.

16.3.6. Reporting

The guideline STrengthening the Reporting Of Pharmacogenetic Studies: Development of the STROPS guideline (PLOS Medicine 2020;17(9):e1003344) should be followed for reporting findings of pharmacogenetic studies. Essential Characteristics of Pharmacogenomics Study Publications (Clin

Pharmacol Ther. 2019;105(1):86-91) also provides recommendations to ensure that all the relevant information is reported in pharmacogenetic studies. As pharmacogenetic information is increasingly found in drug labels, as described in Pharmacogenomic information in drug labels: European Medicines Agency perspective (Pharmacogenomics J. 2015;15(3):201–10), it is essential to warrant consistency across the reporting of pharmacogenetic studies. Additional efforts by regulatory agencies, international organisations or boards to standardise the reporting and utilisation of pharmacogenetic studies will be discussed in the next section.

16.3.7. Clinical implementation and resources

An important step towards the implementation of the use of genotype information to guide pharmacotherapy is the development of clinical practice guidelines. A valuable pharmacogenomics knowledge resource is PharmGKB which curates and disseminates information about the impact of human genetic variation on drug responses, including genotype-phenotype relationships, potentially clinically actionable gene-drug associations, clinical guidelines, and drug labels. The development and publication of clinical practice guidelines for pharmacogenomics has been driven by international initiatives including the Clinical Pharmacogenetics Implementation Consortium, the European Medicines Agency Pharmacogenomics Working Party, and the DPWG: Dutch Pharmacogenetics Working Group. See also Pharmacogenetics: From Bench to Byte— An Update of Guidelines (Clin Pharmacol Ther. 2011;89(5):662–73); Use of Pharmacogenetic Drugs by the Dutch Population (Front Genet. 2019;10:567); and the Canadian Pharmacogenomics Network for Drug Safety. Evidence of clinical utility and cost-effectiveness of pharmacogenomic tests is important to support the translation of clinical guidelines into policies for implementation across health services, such as pharmacogenomic testing for *DPYD* polymorphisms with fluoropyrimidine therapies (see EMA recommendations on DPD testing prior to treatment with fluorouracil, capecitabine, tegafur and flucytosine).

The clinical implementation of pharmacogenomic testing requires consideration of complex clinical pathways and the multifactorial nature of drug response. Translational research and clinical utility studies can identify issues arising from the translation of pharmacokinetic or retrospective studies into real-world implementation of pharmacogenomic testing (see Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan, N Engl J Med. 2011;364(12):1126-33). Careful consideration is required in the interpretation of gene variants which cause a spectrum of effects. Binary interpretation or thresholds for phenotypic categorisation within clinical guidelines may result in different treatment recommendations for patients who would ultimately have the same drug response. In addition, the safety, efficacy and cost-effectiveness of alternative treatments are important factors in assessing the overall health benefit to patients from pharmacogenomic testing.

Further, some groups of patients may require specific treatment guidelines. Research studies such as the NICHD-funded Optimal Medication Management for Mothers with Depression (OPTI-MOM) aim to understand how best to manage drug therapy in pregnant women and investigate the impact of pharmacogenomics with the goal of generating treatment guidelines for proactive management during pregnancy (see Rationale and design for an investigation to optimize selective serotonin reuptake inhibitor treatment for pregnant women with depression, Clin Pharmacol Ther. 2016;100(1):31-3; and Pharmacogenomics in pregnancy. Semin Perinatol. 2020;44(3):151222).

Within clinical practice, the choice of technology for testing must be mapped to the clinical pathway to ensure that test results are available at an appropriate time to guide decision-making. Other key factors for clinical implementation include workforce education in pharmacogenomics, multidisciplinary pathway design, digital integration and tools to aid shared decision making (see Attitudes of clinicians following large-scale pharmacogenomics implementation, Pharmacogenomics J. 2016;16(4):393-8; Pharmacogenomics Implementation at the National Institutes of Health Clinical Center, J Clin

Pharmacol. 2017;57 (Suppl 10):S67-S77; The implementation of pharmacogenomics into UK general practice: a qualitative study exploring barriers, challenges and opportunities, J Community Genet. 2020;11(3):269-77; Implementation of a multidisciplinary pharmacogenomics clinic in a community health system, Am J Health Syst Pharm. 2016;73(23):1956-66).

Large-scale international population studies of clinical utility in pharmacogenomics will contribute to understanding these real-world implementation factors, including studies underway with the U-PGx (see Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium, Clin Pharmacol Ther. 2017;101(3):341-58) and The IGNITE Pharmacogenetics Working Group: An Opportunity for Building Evidence with Pharmacogenetic Implementation in a Real-World Setting, Clin Transl Sci. 2017;10(3):143-6).

The clinical utility of pharmacogenetic testing before starting drug treatment is well documented for several single gene–drug pairs. To further improve the understanding of how genetic variation may increase the risk of adverse drug reactions, the Medicines and Healthcare Products Regulatory Agency (MHRA) together with Genomics England launched the Yellow Card biobank, which will contain genetic data and patient samples, and will operate alongside the MHRA's Yellow Card reporting site for suspected side effects and adverse incidents involving medicines and medical devices.

Beyond single gene genotyping, a recent study has investigated the clinical benefit of using a pharmacogenetic panel to guide prescription and showed that pharmacogenetics-guided prescribing resulted in a 30% reduction of clinically relevant adverse drug reactions (see A 12-gene pharmacogenetic panel to prevent adverse drug reactions: an open-label, multicentre, controlled, cluster-randomised crossover implementation study, Lancet. 2023;401(10374):347-356). Lastly, international networks on pharmacogenomics research provide biological insights into emerging diseases and can support public health actions. For example, the COVID-19 Host Genetics Initiative (The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic, Eur J Hum Genet. 2020; 28(6): 715–8) has enabled rapid genetic association studies on COVID-19 and advanced the global knowledge of SARS-CoV-2 infection by creating a common repository for COVID-19 genetic studies (https://www.covid19hg.org/) and performing powered meta-analyses (Mapping the human genetic architecture of COVID-19, Nature 2021; 600:472-7). Although the discovery of genetic variants associated with susceptibility and severity of COVID-19 disease is challenged by the accurate ascertainment of cases and controls (Understanding COVID-19 through genome-wide association studies, Nature Genetics 2022; 54:368–9), the COVID-19 HGI identified novel host genetic factors associated with COVID-19 and created a framework for international collaboration for future genetic discoveries in emerging pandemics.

16.4. Methods for pharmacovigilance impact research

Note: Chapter 16.4. (formerly 15.4.) has not been updated for Revision 11 of the Guide, as contents remain up-to-date.

16.4.1. Introduction

Pharmacovigilance activities aim to protect patients and promote public health. This includes implementing risk minimisation measures that lead to changes in the knowledge and behaviour of individuals (e.g., patients, consumers, caregivers and healthcare professionals) and in healthcare practice. Impact research aims to generate evidence to evaluate the outcomes of these activities which may be intended or unintended. This approach has been adopted in the EMA Guideline on good pharmacovigilance practices (GVP) - Module XVI – Risk minimisation measures: selection of tools and

effectiveness indicators (Rev 2), which is currently undergoing revision (see Guideline on good pharmacovigilance practices (GVP) - Module Risk Minimisation Measures for the draft of Rev. 3).

Pharmacovigilance activities are frequently examined for their impact on processes of healthcare delivery, such as healthcare outcomes or drug utilisation patterns following changes to the product information. In addition, measuring dissemination of risk minimisation is of importance as well as changes in knowledge, awareness and behaviour of healthcare professionals and patients.

These effects can be assessed separately, or combined in a framework, which is more challenging and therefore rarely done. An example of such a standardised framework includes evaluation of the effectiveness of risk minimisation measures through four domains: data, knowledge, behaviour and outcomes (Evaluating the effectiveness of risk minimisation measures: the application of a conceptual framework to Danish real-world dabigatran data; Pharmacoepidemiol Drug Saf. 2017;26(6):607-14). Further testing of this method is needed, however, to ascertain its usefulness in regulatory practice.

Measuring the impact of pharmacovigilance activities may be challenging as these activities may target stakeholder groups at different levels of the healthcare system, co-exist with other unrelated events that can influence healthcare, and can use several tools applied simultaneously or sequentially to deliver information and influence behaviour (Measuring the impact of pharmacovigilance activities, challenging but important; Br J Clin Pharmacol. 2019;85(10):2235-7). In addition to the intended outcomes of pharmacovigilance activities, there may be unintended outcomes which are important to be measured as they could counteract the effectiveness of risk minimisation. Another challenging aspect is separating the outcomes of individual pharmacovigilance activities from simultaneous events such as media attention, reimbursement policies, publications in scientific journals, changes in clinical guidelines and practice, or secular trends in health outcomes.

This Chapter provides a detailed guidance on the methodological conduct of impact studies.

16.4.2. Outcomes

Outcomes to be studied in impact research are closely tied to the nature and objective of the pharmacovigilance activities. Because regulatory actions are mostly tailored to individual medicinal products, there is no standard outcome that could be measured for each activity and the concepts outlined in this chapter need to be applied on a case-by-case basis (Post-approval evaluation of effectiveness of risk minimisation: methods, challenges and interpretation; Drug Saf. 2014;37(1):33-42).

Outcome measures provide an overall indication of the level of risk reduction that has been achieved with a specific risk minimisation measure in place. This may also require measuring outcomes not linked to the specific medicinal product but representing potential unintended consequences of regulatory interventions e.g., change of non-target drug use in a population leading to less favourable health outcomes. Examples are provided in Table XVI.1 of the Guideline on good pharmacovigilance practices (GVP) - Module Risk Minimisation Measures.

Relevant outcomes may include: information dissemination and risk knowledge; changes in behaviour or clinical practice; drug utilisation patterns (e.g. prescribing or dispensing rates, use of treatment alternatives); and health outcomes (Measuring the impact of medicines regulatory interventions - Systematic review and methodological considerations; Br J Clin Pharmacol. 2018;84(3):419-33).

Dissemination of information and risk knowledge can be assessed in a quantitative, qualitative or mixed-methods manner. Quantitative assessment can involve measuring the proportion of healthcare professionals and patients aware of the risk minimisation measure as well as their level of comprehension (Effectiveness of Risk Minimization Measures to Prevent Pregnancy Exposure to

Mycophenolate-Containing Medicines in Europe; Pharmaceut Med. 2019;33(5):395-406). Qualitative measures often focus on understanding of attitudes about the risk minimisation measure, impact of external factors on implementation and information update whilst mixed methods utilise both qualitative and quantitative approaches.

Assessment of behavioural changes is performed to measure if changes towards intended behaviour have been achieved, and to what extent. These measures align with those applied when measuring dissemination of information and risk knowledge. Quantitative assessment can include measuring the proportion of patients exposed to a medicinal product which is not in accordance with authorised use (off label use, contraindicated use, interactions). A qualitative assessment may allow an in-depth understanding of enablers and barriers in relation to awareness, attitudes towards use of the medicinal product and the causes why intended outcomes may not have been achieved.

Health outcomes should preferably be measured directly. They may include clinical outcomes such as all-cause mortality, congenital defects or other conditions that prompted the pharmacovigilance activity. Direct measurement of health outcomes is not always feasible or may not be necessary, for example when it can be replaced with indirect measures. Indirect surrogate measures may use data on hospitalisations, emergency department admissions or laboratory values e.g. blood pressure as a surrogate for cardiac risk, as outlined in Practical Approaches to Risk Minimisation for Medicinal Products: Report of CIOMS Working Group IX. An example of use of a surrogate measure is glycaemic outcomes (HbA1C change from baseline) in patients with diabetes mellitus using the Veterans Integrated Services Network database; the results confirmed a 45% discontinuation of thiazolidinedione use in this population and a worsening of glycaemic control following safety warning publicity in 2007, which may have driven the decline in usage of this class of medicines (Impact of thiazolidinedione safety warnings on medication use patterns and glycemic control among veterans with diabetes mellitus; J Diabetes Complications 2011;25(3):143-50).

Depending on the nature of the safety concern and the regulatory intervention, or when the assessment of patient-relevant health outcomes is unfeasible (e.g. inadequate number of exposed patients, rare adverse reaction), the dissemination of safety information, risk knowledge or behavioural changes may be alternative objectives of impact research (Guideline on good pharmacovigilance practices (GVP) - Module VIII – Post-authorisation safety studies (Rev 3)).

16.4.3. Considerations on data sources

The impact of pharmacovigilance activities can be measured using both primary and secondary data collection, although the literature shows that the latter is more commonly used (Measuring the impact of medicines regulatory interventions - Systematic review and methodological considerations; Br J Clin Pharmacol. 2018;84(3):419-33). Chapter 7 of this Guide provides a general description of the main characteristics, advantages and disadvantages of various data sources. Chapter 7.1.2. provides guidance on primary data collection through surveys.

The impact of pharmacovigilance activities should be interpreted with a view to the limitations of the data sources used for the evaluation (A General Framework for Considering Selection Bias in EHR-Based Studies: What Data Are Observed and Why?; EGEMS. (Wash DC.) 2016;4(1):1203). Researchers should have a clear understanding of the limitations of the different data sources when planning their research and assess whether these limitations could impact the results in one direction or the other in such a way that their interpretation may be significantly influenced, for example due to bias or unmeasured confounders. As for all observational studies, the evaluation of the usefulness and limitation of a given data source for the study requires a very good understanding of the research question.

Primary data collection, via interviews or surveys, can usually never cover the complete target population. Therefore, a sampling approach is often required which can involve those that prescribe, dispense or use the medicinal product. Sampling should be performed in accordance with the Guideline on good pharmacovigilance practices (GVP) - Module XVI Addendum II, ensuring target population representativeness. The following elements should be considered to minimise bias and optimise generalisability: sampling procedures (including sample size), design and administration of the data collection instrument, analytical approaches and overall feasibility (including ethics).

Different databases are unlikely to capture all impact–relevant outcomes, even when they are linked to one another. Data of good quality may be available on hard outcomes such as death, hospital admission, emergency room visit or medical contacts but claims databases rarely capture primary care diagnoses, symptoms, conditions or other events that do not lead to a claim, such as suicidal ideation, abuse or misuse. An accurate definition of the outcomes also often requires the development of algorithms that need validation in the database that will be used for impact measurement.

Nurse-Led Medicines' Monitoring for Patients with Dementia in Care Homes: A Pragmatic Cohort Stepped Wedge Cluster Randomised Trial (PLoS One 2015;10(10):e0140203) reported that only about 50% of the less serious drug-related problems listed in the product information are recorded in patient notes. If generalisable to electronic data sources, this would indicate that incomplete recording of patient-reported outcomes of low severity may reduce the likelihood of identifying some outcomes related to a pharmacovigilance activity, for example a change in the frequency of occurrence of an adverse drug reaction (ADR). Combining different approaches such as integrating a patient survey would be necessary to overcome this situation.

Missing information on vulnerable populations, such as pregnant women, and missing mother-child or father-child links is a significant barrier to measuring the impact of paternal/maternal exposure or behaviour. For example, the impact of pregnancy prevention programmes could not be accurately assessed using European databases that had been used to report prescribing in pregnancy (The limitations of some European healthcare databases for monitoring the effectiveness of pregnancy prevention programmes as risk minimisation measures; Eur J Clin Pharmacol. 2018;74(4):513-20). This was largely due to inadequate data on planned abortions and exposure to oral contraceptives.

Depending on the initial purpose of the data source used for impact research, information on potential confounders may be missing, such as indication of drug use, co-morbidities, co-medication, smoking, diet, body mass index, family history of disease or recreational drug use. Missing information may impair a valid assessment of risk factors for changes in health care practice, but this limitation should be considered in light of the research question. In some settings, record linkage between different types of data sources including different information could provide comprehensive data on the frequency of ADRs and potential confounders (Health services research and data linkages: issues, methods, and directions for the future; Health Serv Res. 2010;45(5 Pt 2):1468-88; Selective Serotonin Reuptake Inhibitor (SSRI) Antidepressants in Pregnancy and Congenital Anomalies: Analysis of Linked Databases in Wales, Norway and Funen, Denmark; PLoS One 2016;11(12):e0165122; Linking electronic health records to better understand breast cancer patient pathways within and between two health systems; EGEMS. (Wash DC.) 2015;3(1):1127).

16.4.4. Study designs

16.4.4.1. Single time point cross-sectional study

The cross-sectional study design as defined in Appendix 1.1.2.1 of the Guideline on good pharmacovigilance practices (GVP) - Module VIII – Post-authorisation safety studies (Rev 3) collects data at a single point in time after implementation of a regulatory intervention. However, cross-

sectional studies have limitations as a sole measure of the impact of interventions. Cross-sectional studies may include data collected through surveys and can be complemented with data from other studies, e.g. on patterns of drug use (Healthcare professional surveys to investigate the implementation of the isotretinoin Pregnancy Prevention Programme: a descriptive study; Expert Opin Drug Saf. 2013;12(1):29-38; Prescriptive contraceptive use among isotretinoin users in the Netherlands in comparison with non-users: a drug utilisation study; Pharmacoepidemiol Drug Saf. 2012;21(10):1060-6).

16.4.4.2. Before-and-after study

A before-and-after study is defined as an evaluation (at one point in time) before and (one point in time) after the date of the intervention and/or its implementation. When uncontrolled, before-and-after studies need to be interpreted with caution as any baseline trends are ignored, potentially leading to the intervention effect being incorrectly estimated. Including a control (e.g., a population that did not receive the intervention or a drug not targeted by the risk minimisation measure) can strengthen this design by minimising potential confounding. However, identifying a suitable control group may be challenging or unfeasible as any regulatory action aimed at reducing risk is intended to be applied to the entire target population (see Post-approval evaluation of effectiveness of risk minimisation: methods, challenges and interpretation; Drug Saf. 2014;37(1):33-42 and Measuring the impact of medicines regulatory interventions - Systematic review and methodological considerations; Br J Clin Pharmacol. 2018;84(3):419-33). When a suitable control group is available, the difference-in-differences (DiD) method can be used. The DiD method is a controlled before-and-after design whereby comparisons are made between two similar groups under different conditions. The outcome can be measured either at a single pre-intervention and post-intervention time point, or by comparing pre- and post-intervention means, but it does not incorporate time. The DiD method then takes the difference for both groups (exposed and control) before and after the intervention, thereby controlling for varying factors in estimating the impact of the intervention (see The use of controls in interrupted time series studies of public health interventions; Int J Epidemiol 2018;47:2082–93 and Difference-in-Differences Method in Comparative Effectiveness Research: Utility with Unbalanced Groups; Appl Health Econ Health Policy. 2016; 14: 419–29). The DiD method relies upon the assumption that both groups are similar and trends are parallel, hence may be susceptible to residual confounding as a result of differences between the groups.

16.4.4.3. Time series design

A time series is a sequence of data points (values) usually gathered at regularly spaced intervals over time. These data points can represent a value or a quantification of outcomes that are used for impact research. The underlying trend of a particular outcome is 'interrupted' by a regulatory intervention at a known point in time. Time series data can be analysed using various methods, including interrupted time series (ITS) and Joinpoint analysis.

16.4.4.4. Cohort study

The cohort study design as defined in Appendix 1.1.2.2 of the Guideline on good pharmacovigilance practices (GVP) - Module VIII – Post-authorisation safety studies (Rev 3) can be useful in impact research to establish the base population for the conduct of drug utilisation studies or to perform aetiological studies.

Cohort studies can be used to study exposure to the medicine targeted by regulatory interventions before and after its implementation, and indeed to perform drug utilisation studies in clinical populations targeted by these interventions. To model their impact on health outcomes, more complex study designs may be required, that are the subject of further research.

The following are examples of cohort studies being used for:

- Impact research evaluating pregnancy prevention programmes (Isotretinoin exposure during pregnancy: a population-based study in The Netherlands; BMJ. Open 2014;4(11):e005602);
- Drug utilisation in target populations (Impact of EMA regulatory label changes on systemic diclofenac initiation, discontinuation, and switching to other pain medicines in Scotland, England, Denmark, and The Netherlands; Drug Saf. 2020;29(3):296-305);
- Aetiological studies examining the impact on health outcomes (Measuring the Effectiveness of Safety Warnings on the Risk of Stroke in Older Antipsychotic Users: A Nationwide Cohort Study in Two Large Electronic Medical Records Databases in the United Kingdom and Italy; Drug Saf. 2019;42(12):1471-85).

16.4.4.5. Randomised controlled trial

The randomised controlled trial (RCT) as defined in Appendix 1.1.2.2 of the Guideline on good pharmacovigilance practices (GVP) - Module VIII – Post-authorisation safety studies (Rev 3) can be useful in evaluating the effectiveness of different interventions but it is not always possible to randomise individual participants and few examples exist (Improved therapeutic monitoring with several interventions: a randomized trial; Arch Intern Med. 2006;166(17):1848-54). Designs including cluster randomised trials or step-wedge trials may be more feasible, in which randomisation is conducted at the level of organisation, when a phased roll-out is being considered (Research designs for studies evaluating the effectiveness of change and improvement strategies; Qual Saf Health Care 2003;12(1):47-52). RCTs could be considered more often to generate evidence on the impact of pharmacovigilance interventions by evaluating interventions that potentially enhance agreed safety information and normal methods of dissemination and communication channels.

16.4.5. Analytical methods

The analytical methods to be applied in impact research depend on the study design and approach to data collection. Various types of analyses have been used to assess the impact of a regulatory guidance, as described in: Measuring the impact of medicines regulatory interventions - Systematic review and methodological considerations (Br J Clin Pharmacol. 2018;84(3):419-33); Impact of regulatory guidances and drug regulation on risk minimization interventions in drug safety: a systematic review (Drug Saf. 2012;35(7):535-46); and A descriptive review of additional risk minimisation measures applied to EU centrally authorised medicines 2006-2015 (Expert Opin Drug Saf. 2017;16(8):877-84).

16.4.5.1 Descriptive statistics

Descriptive measures are the basis of quantitative analyses in studies evaluating the impact of regulatory interventions. Whilst appropriate to describe the population to understand generalisability, simple descriptive approaches do not determine whether statistically significant changes have occurred (Measuring the impact of medicines regulatory interventions - Systematic review and methodological considerations; Br J Clin Pharmacol. 2018;84(3):419-33). When simple descriptive statistics are used, they are often insufficiently valid to determine statistical significance.

16.4.5.2 Time series analysis

Interrupted time series (ITS) analysis

ITS analysis, sometimes referred to as interrupted segmented regression analysis, can provide statistical evidence about whether observed changes in a time series represent a real decrease or increase by accounting for secular trends. ITS has commonly been used to measure the impact of regulatory interventions and is among the more robust approaches to pharmacovigilance impact

research (Measuring the impact of medicines regulatory interventions - Systematic review and methodological considerations; Br J Clin Pharmacol. 2018;84(3):419-33; Impact of EMA regulatory label changes on systemic diclofenac initiation, discontinuation, and switching to other pain medicines in Scotland, England, Denmark, and The Netherlands; Pharmacoepidemiol Drug Saf. 2020;29(3):296-305; The Effect of Safety Warnings on Antipsychotic Drug Prescribing in Elderly Persons with Dementia in the United Kingdom and Italy: A Population-Based Study; CNS Drugs 2016;30(11):1097-109).

ITS is well suited to study changes in outcomes that are expected to occur relatively quickly following an intervention, such as change in prescribing, and can consist of averages, proportions, counts or rates. ITS can be used to estimate a variety of outcomes including: the immediate change in outcome after the intervention; the change in trend in the outcome compared to before the intervention; and the effects at specific time periods following the intervention.

Common segmented regression models fit a least squares regression line to each time segment and assume a linear relationship between time and the outcome within each segment.

When the effects of interventions take time to manifest, this can be accounted for through the use of lag times in the analysis to avoid incorrect specification of the intervention effect. To model these effects, one can exclude from the analysis outcome values that occur during the lag or during the intervention period. Alternatively, with enough data points, the period may be modelled as a separate segment.

ITS regression requires that the time point of the intervention is known prior to the analysis and sufficient data points are collected before and after the intervention for adequate power. Studies with a small number of data points should be interpreted with caution as they may be underpowered.

An assumption of ITS segmented regression analysis is that time points are independent of each other. Autocorrelation is a measure of how correlated data collected closely together in time are with each other. If autocorrelation is present, it may violate the underlying model assumptions that observations are independent of each other and can lead to an over-estimation of the statistical significance of effects. Autocorrelation can be checked by examining autocorrelation and partial autocorrelation function plots and checking the Durbin-Watson statistic or performing the Breusch-Godfrey test (Testing for serial correlation in least squares regression. I; Biometrika. 1950;37(3-4):409-28; Testing for serial correlation in least squares regression. II; Biometrika. 1951;38(1-2):159-78). Factors such as autocorrelation, seasonality and non-stationarity should therefore be checked and may require more complicated modelling approaches if detected, e.g. autoregressive integrated moving average (ARIMA) models (Impact of FDA Black Box Warning on Psychotropic Drug Use in Noninstitutionalized Elderly Patients Diagnosed With Dementia: A Retrospective Study; J Pharm Pract. 2016;29(5):495-502; IMI Work Package 5: Benefit –Risk Integration and Visual Representation).

Long time periods may also be affected by historical changes in trend that can violate model assumptions. Therefore, data should always be visually inspected and reported.

Data point outliers that are explainable, such a sudden peak in drug dispensing in anticipation of a drug restriction policy can be controlled for using an indicator term. Outliers that result from random variation can be treated as regular data point.

Another caveat when conducting ITS analysis relates to possible outcome measure ceiling or floor effects. For example, when studying the impact of an intervention in improving the proportion of patients treated with a drug, the outcome has a natural ceiling of 100% and thus, depending of the initial level of measurement, minimal change in the outcome is observed.

Time-varying confounding, such as from concomitant interventions, may be addressed by use of a control outcome in the same population or a control population using the same outcome. An advantage on ITS analysis is the ease in stratifying results by different groups.

Joinpoint analysis

Accurately establishing the date of the intervention time period may be challenging (e.g. during a phased roll out of a regulatory intervention or when attempting to assess different parts of a regulatory intervention). In such instances, more complex modelling techniques and other approaches time series approaches could be considered.

Statistical analysis using joinpoint regression identifies the time point(s) where there is a marked change in trend (the 'joinpoint') in the time series data and estimates the regression function compared with previously identified joinpoints. Joinpoints can be identified by using permutation tests using Monte Carlo methods or Bayesian Information Criterion approaches (Permutation tests for joinpoint regression with applications to cancer rates; Stat Med. 2000;19(3):335-51). As the final number of joinpoints is established on the basis of a statistical criterion, their position is not fixed. Therefore, joinpoint regression does not require that the date of the regulatory intervention is pre-specified. It can be used to estimate the average percent change in an outcome, which is a summary measure of the trend over a pre-specified fixed interval. It can also be used to undertake single or pairwise comparisons.

16.4.5.3 Other statistical techniques

Different types of regression models can be applied to the time series data once it has been properly organised depending upon the exact question being asked such as Poisson regression (Interrupted time series regression for the evaluation of public health interventions: a tutorial; Int J Epidemiol. 2017;46(1):348-55. Erratum in: Int J Epidemiol. 2020;49(4):1414). These methods are based on the assumption that error terms are normally distributed. When time series analysis measurements are based at extreme values (e.g. all are near 0% or near 100% or with low cell counts near 0) alternative approaches may be required (e.g. aggregate binomial regression models) and advice from an experienced statistician is recommended.

16.4.5.4 Examples of impact research using time series analysis

Before-and-after after time series have been used to evaluate the effects of:

- Paracetamol pack size reductions introduced in the UK in 1998 on poisoning deaths and liver transplants (Long term effect of reduced pack sizes of paracetamol on poisoning deaths and liver transplant activity in England and Wales: interrupted time series analyses; BMJ. 2013;346:f403);
- Black Triangle Label on Prescribing of New Drugs in the United Kingdom (Impact of the black triangle label on prescribing of new drugs in the United Kingdom: lessons for the United States at a time of deregulation; Pharmacoepidemiol Drug Saf. 2017;26(11):1307-13);
- FDA boxed warning on the duration of use for depot medroxprogesterone acetate (The impact of the boxed warning on the duration of use for depot medroxprogesterone acetate; Pharmacoepidemiol Drug Saf. 2017;26(7):827-36);
- Withdrawal of fusafungine from the market on prescribing of antibiotics, other nasal or throat preparations and tyrothricin in Germany (Effect of withdrawal of fusafungine from the market on prescribing of antibiotics and other alternative treatments in Germany: a pharmacovigilance impact study; Eur J Clin Pharmacol. 2019;75(7):979-84);

- FDA black box warning on fluoroquinolone and alternative antibiotic use in southeastern US hospitals (Impact of FDA black box warning on fluoroquinolone and alternative antibiotic use in southeastern US hospitals; Infect Control Hosp Epidemiol. 2019;40(11):1297-1300);
- A re-analysis of published UK impact studies showed that UK regulatory risk communications were associated with significant changes in targeted prescribing and potential changes in clinical outcomes (Impact of medicines regulatory risk communications in the UK on prescribing and clinical outcomes: Systematic review, time series analysis and meta-analysis; Br J Clin Pharmacol. 2020;86(4):698-710).

Examples of the use of Joinpoint regression analysis:

- Scientific publications, FDA advisories and media exposure on glitazone use (Changes in glitazone use among office-based physicians in the U.S., 2003-2009; Diabetes Care. 2010;33(4):823-5);
- The fall of hormone replacement therapy in England following the results of the women's health initiative (What was the immediate impact on population health of the recent fall in hormone replacement therapy prescribing in England? Ecological study; J Public Health (Oxf.). 2010;32(4):555-64).

16.4.5.5 Regression modelling

Multivariable regression allows controlling for potential confounding factors or to study factors associated with the impact or non-impact of regulatory interventions.

An analysis with multivariate regression was used in Measuring the Effectiveness of Safety Warnings on the Risk of Stroke in Older Antipsychotic Users: A Nationwide Cohort Study in Two Large Electronic Medical Records Databases in the United Kingdom and Italy (Drug Saf. 2019;42(12):1471-85). The Medicines and Healthcare Regulatory Agency (MHRA) and the Italian Drug Agency (AIFA) both launched a safety warning on the risk of stroke and all-cause mortality with antipsychotics in older people with dementia. In the UK, the MHRA launched a warning in March 2004 for the use of risperidone and olanzapine which was expanded to all antipsychotics in March 2009. In Italy, AIFA restricted prescribing of antipsychotics in the elderly to specific prescribing centres in July 2005, which was followed by communication about these restrictions in May 2009. A retrospective new-user cohort study was undertaken to estimate incidence rates of stroke in elderly incident antipsychotic users. The authors showed a significant reduction of stroke after both safety warnings in the UK, while there was no impact of the warning on incidence rates of stroke in Italy. Metabolic screening in children receiving antipsychotic drug treatment (Arch Pediatr Adolesc Med. 2010;164(4):344-51) measured the impact of a class warning issued by the Food and Drug Administration (FDA) for all second-generation antipsychotics (SGAs) regarding the risk of hyperglycaemia and diabetes mellitus in 2003. This warning stated that glucose levels should be monitored in at-risk patients. A retrospective new-user cohort study was undertaken to estimate population-based rates of glucose and lipid testing in children after the availability of FDA warnings and to identify predictors of the likelihood of receiving glucose or lipid testing among SGAs-treated children after adjusting for covariates. Children without diabetes taking albuterol but no SGA drugs were used as controls. The authors showed that most included children starting treatment with SGAs did not receive recommended glucose and lipid screening.

More sophisticated methodologies, such as propensity-score matching (Chapter 5.2.3.2), instrumental variable analysis (Chapter 5.2.3.3) and time-varying exposures and covariates (Chapter 5.2.3.5) may be implemented in regression analyses if relevant.

Whichever design and method of analysis is used, consideration should be given to reporting both relative and absolute effects.

16.4.5.6 Other types of analytical methods

Metrics such as "Population Impact Number of Eliminating a Risk factor over time t" (PIN-ER-t), and "Number of Events Prevented in a Population" (NEPP) have proven valuable in assessing the impact of removing a risk factor on public health, and may be useful in assessing impact of regulatory interventions. Illustrative examples for population impact analyses include Potential population impact of changes in heroin treatment and smoking prevalence rates: using Population Impact Measures (Eur J Public Health 2009;19(1):28-31) and Assessing the population impact of low rates of vitamin D supplementation on type 1 diabetes using a new statistical method (JRSM Open 2016;7(11):2054270416653522). Further, statistical analysis using impact metrics is possible where proxy measures are used to assess the impact that one event or resource has on another, as shown in Communicating risks at the population level: application of population impact numbers (BMJ. 2003;327(7424):1162-5); the benefit-risk case study report for rimonabant in IMI Work Package 5: Benefit –Risk Integration and Visual Representation; and in Population Impact Analysis: a framework for assessing the population impact of a risk or intervention (J Public Health (Oxf.) 2012;34(1):83-9).

Predictive modelling techniques may provide an insight into future impact of regulatory actions. Modelling the risk of adverse reactions leading to product withdrawal alongside drug utilisation data can assess the number of patients at risk of experiencing the adverse reactions per year, and provide an estimate of the number of patients per year which are protected from as a result of regulatory action (Population Impact Analysis: a framework for assessing the population impact of a risk or intervention; J Public Health (Oxf.) 2012;34(1):83-9; Assessing the population impact of low rates of vitamin D supplementation on type 1 diabetes using a new statistical method; JRSM Open 2016;7(11):2054270416653522).

Chronographs, typically used for rapid signal detection in observational longitudinal databases, have been used to visualise the impact of regulatory actions. Although this is a novel method that could potentially be applied to rapidly assess impact, the method lacks ways to control for confounding. In addition, further validation may be required to understand in which situations this works well or not (A Novel Approach to Visualize Risk Minimization Effectiveness: Peeping at the 2012 UK Proton Pump Inhibitor Label Change Using a Rapid Cycle Analysis Tool; Drug Saf. 2019;42(11):1365-76).

16.4.6. Measuring unintended effects of regulatory interventions

Pharmacovigilance activities can have unintended consequences, which could in some cases counteract the effectiveness of risk minimisation measures. To determine the net attributable impact of pharmacovigilance activities, besides the intended outcomes, other outcomes associated with potential unintended consequences may need to be measured and incorporated into the design of impact research (see Table XVI.1 of the Guideline on good pharmacovigilance practices (GVP) - Module Risk Minimisation Measures). Examples of such studies include the Effect of withdrawal of fusafungine from the market on prescribing of antibiotics and other alternative treatments in Germany: a pharmacovigilance impact study (Eur J Clin Pharmacol. 2019;75(7):979-84), which was associated with an increase in prescribing of other nasal or throat preparations but no increase in alternative antibiotic prescribing. Another example concerns the unintended increased use of conventional antipsychotics in two European countries after the introduction of EU risk minimisation measures for the risk of stroke and all-cause mortality with atypical antipsychotic drug use (The Effect of Safety Warnings on Antipsychotic Drug Prescribing in Elderly Persons with Dementia in the United Kingdom and Italy: A Population-Based Study; CNS Drugs 2016;30(11):1097-109). Further, prescribers may extrapolate warnings for one group of patients to other groups (spill-over effects), although they may not share the same risk factors. In 2003, the FDA warned of an association between SSRI prescription and suicidality in paediatric patients (<18 years of age). Subsequently, the number of prescriptions of

SSRIs in newly diagnosed adult patients fell without compensation by alternative medicines or treatment (Spillover effects on treatment of adult depression in primary care after FDA advisory on risk of pediatric suicidality with SSRIs; Am J Psychiatry 2007;164(8):1198-205).

Socio-economic factors may also play an important role in implementing regulatory interventions at local level. It has been suggested that practices in affluent communities are more likely to implement regulatory interventions faster than over-stretched or under-resourced practices in more deprived communities and that permanent changes in daily practice in these communities may take longer (THE INTERNATIONAL MARCÉ SOCIETY FOR PERINATAL MENTAL HEALTH BIENNIAL SCIENTIFIC CONFERENCE; Arch Womens Ment Health 2015;18:269–408; Prescribing of antipsychotics in UK primary care: a cohort study; BMJ Open 2014;4(12):e006135).

Both health care service providers and users may circumvent or 'work round' restrictions. Where medicines are restricted or restrictions are perceived as inconvenient, patients may turn to buying medicines over the internet, self-medicating with over-the-counter medicines or using herbals or other complementary medicines. Healthcare professionals may subvert requirements for additional documentation by realigning diagnostic categories (Changes in rates of recorded depression in English primary care 2003-2013: Time trend analyses of effects of the economic recession, and the GP contract quality outcomes framework (QOF); J Affect Disord. 2015;180:68-78) or switch to medicines where patient monitoring is not mandated (Incorporating Comprehensive Management of Direct Oral Anticoagulants into Anticoagulation Clinics; Pharmacotherapy 2017;37(10):1284-97). The effects of progressive dextropropoxyphene withdrawal in the EU since 2007 on prescribing behaviour showed an increased use of same level analgesics but also an increased use of paracetamol as monotherapy. Aggregated dispensation data suggested that the choice of analgesics depended on physician speciality, healthcare setting, indication, patients' comorbidities and age, underlining the complexity and international differences in pain management (Use of analgesics in France, following dextropropoxyphene withdrawal; BMC Health Serv Res. 2018;18(1):231).

16.5. Artificial intelligence in pharmacoepidemiology

16.5.1. Introduction

Artificial intelligence (AI) is a catch-all term for a set of tools and techniques that allow machines to do activities commonly described as requiring human-level intelligence. While no consensus on a definition of AI exists, a common trend is an analogy to human intelligence, however, this is unhelpful as it suggests an idea of Artificial *General* Intelligence, whereas current techniques and tools are dedicated to assist specific tasks, i.e., Artificial *Narrow* Intelligence.

Machine Learning (ML) is considered a subset of AI and reflects the ability of computers to identify and extract rules from data rather than those rules being explicitly coded by a human. Deep Learning (DL) is a subtype of ML with increased complexity of how it parses and analyses data. The rules identified by ML or DL applications constitute an algorithm and the outputs are often said to be data-driven, as opposed to rules explicitly coded by a human that form knowledge-based algorithms.

Natural language processing (NLP) sits at the interface of linguistics, computer science and AI and is concerned with providing machines with the ability to understand text and spoken words. NLP can be subset into statistical NLP, which uses ML or DL approaches and symbolic NLP, which uses a semantic rule-based methodology. Applications of AI in pharmacoepidemiology can be broadly classified into those that extract and structure some data and those that produce some insight.

16.5.2. Applications of AI in pharmacoepidemiology

16.5.2.1. Data extraction

AI techniques can be used to extract text data from unstructured documents transforming it into information available in a structured, research-ready format to which statistical techniques can be applied. A potential application being explored is in extracting data from medical notes, usually including a named-entity recognition, i.e., discovering mentions of entities of a specific class or group such as medication or diseases, and a relation extraction, allowing to relate sets of entities, e.g., a medicine and an indication.

The 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text (J Am Med Inform Assoc. 2011;18(5):552-6) presents three tasks: a concept extraction of medical concepts from patient reports; a classification task focused on assigning assertion types for medical problem concepts; and a relation classification task focused on assigning relation types that hold between medical problems, tests, and treatments. Multiple algorithms were compared showing promising results for concept extraction. In NEAR: Named entity and attribute recognition of clinical concepts (J Biomed Inform. 2022;130:104092), three DL models were created for the same data used in the 2010 i2b2 challenge and have showed an improvement in performance.

Some of the first applications of ML and NLP to extract information from clinical notes focused on the identification of adverse drug events in medical notes, as illustrated in publications such as A method for systematic discovery of adverse drug events from clinical notes (J Am Med Inform Assoc. 2015;22(6):1196-204), Detecting Adverse Drug Events with Rapidly Trained Classification Models (Drug Saf. 2019;42(1):147-56) and MADEx: A System for Detecting Medications, Adverse Drug Events, and Their Relations from Clinical Notes (Drug Saf. 2019;42(1):123-33).

Another common application for medical concept extraction from clinical text is the identification of a relevant set of patients, often referred to as computable phenotyping as exemplified in Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications (J Am Med Inform Assoc. 2010;17(5):507-13). Combining deep learning with token selection for patient phenotyping from electronic health records (Sci Rep. 2020;10(1):1432) describes the development of DL models to construct a computable phenotype directly from the medical notes.

A large body of research has focused on extracting information from clinical notes in electronic health records. The approach can also be applied with some adjustment to other sets of unstructured data, including spontaneous reporting systems, as reflected in Identifying risks areas related to medication administrations - text mining analysis using free-text descriptions of incident reports (BMC Health Serv Res. 2019;19(1):791), product information documentation such as presented in Machine learning-based identification and rule-based normalization of adverse drug reactions in drug labels (BMC Bioinformatics. 2019;20(Suppl. 21):707) or even literature screening for systematic reviews as explored in Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool (Syst Rev. 2018 Mar 12;7(1):45).

In the systematic review Use of unstructured text in prognostic clinical prediction models: a systematic review (J Am Med Inform Assoc. 2022 Apr 27;ocac058), data extraction from unstructured text was shown to be beneficial in most studies. However, data extraction from unstructured text data does not show perfect accuracy (or related metric) and may have wide variability with respect to model performance for the same data extraction task, as shown in ADE Eval: An Evaluation of Text Processing Systems for Adverse Event Extraction from Drug Labels for Pharmacovigilance (Drug Saf. 2021;44(1):83-94). Thus, the application of these techniques should consider the objective in terms of precision or recall. For instance, a model that identifies medical concepts in a spontaneous report of an adverse drug reaction from a patient and maps it to a medical vocabulary might preferably focus on

achieving a high recall, as false positives can be picked up in the manual review of the potential signal, whereas models with high precision and low recall may introduce irretrievable loss of information. In other words, ML models to extract data are likely to introduce some error and thus the error tolerance for the specific application needs to be considered.

16.5.2.2. Data insights

In pharmacoepidemiology, data insights extracted with ML models are typically one of three categories: confounding control, clinical prediction models and probabilistic phenotyping.

Propensity score methods are a predominant technique for confounding control (see Chapter 6.2.3.2). In practice, the propensity score is most often estimated using a logistic regression model, in which treatment status is regressed on observed baseline characteristics. In Evaluating large-scale propensity score performance through real-world and synthetic data experiments (Int J Epidemiol. 2018;47(6):2005-14) and A comparison of machine learning algorithms and covariate balance measures for propensity score matching and weighting (Biom J. 2019;61(4):1049-72) ML models were explored as alternatives to traditional logistic regression with a view to improve propensity score estimation. The theoretical advantages of using ML models include an automatisation procedure, by dispensing the need for investigator-defined covariate selection, and better modelling of non-linear effects and interactions. However, most studies in this field use synthetic or plasmode data and applications in real-world data need to be further explored.

The concept of rule-based, knowledge-based algorithms and risk-based stratification is not new to medicine and healthcare, the Framingham risk score being one of the most well-known. Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review (J Am Med Inform Assoc. 2022;29(5):983-9) shows that there is a growing trend to develop data-driven clinical prediction models. However, the problem definition is often not clearly reported, and the final model is often not completely presented. This trend was exacerbated with the COVID-19 pandemic, where over 400 papers on clinical prediction models were published (see Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal, BMJ. 2020;369:m1328). The authors also suggest that prediction models are poorly reported, and at high risk of bias such that their reported predictive performance is probably optimistic, which was confirmed for several models in Clinical prediction models for mortality in patients with covid-19: external validation and individual participant data meta-analysis (BMJ. 2022;378:e069881). This is common, as has been reported in External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination (Journal of Clinical Epidemiology. 2015;68(1):25–34.). While guidelines for reporting that are specific for AI prediction models are still under development (Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence, BMJ Open 2021;11:e048008), the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement can be used (BMJ 2015;350:g7594). Further, PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies (Ann Intern Med. 2019;170:51-58) supports the evaluation of prediction models. A review of checklists for reporting AI use is reported in Time to start using checklists for reporting artificial intelligence in health care and biomedical research: a rapid review of available tools (2022 IEEE 26th International Conference on Intelligent Engineering Systems (INES), IEEE 2022. p. 000015–20). A checklist for assessing bias in a ML algorithm is provided in A clinician's guide to understanding and critically appraising machine learning studies: a checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML) (European Heart Journal - Digital Health. 2022;3(2):125–40).

Clinical prediction models have also been applied for safety signal detection with some degree of success as exemplified in A supervised adverse drug reaction signalling framework imitating Bradford

Hill's causality considerations (J Biomed Inform. 2015;56:356-68). For the evaluation of safety and utility, the Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI can be used (BMJ 2022;377:e070904).

Probabilistic phenotyping is another potential use of ML in pharmacoepidemiology. It refers to the development of a case definition using a set of labelled examples to train a model and the outputting of the probability of a phenotype as a continuous trait. It differs from ML-based computable phenotyping mentioned earlier, as probabilistic phenotyping takes a set of features and estimates a probability of a phenotype whereas for the computable phenotyping, the ML technique merely extracts information that identifies a relevant case.

Methods for diagnosis phenotyping are discussed in Methods for Clinical Evaluation of Artificial Intelligence Algorithms for Medical Diagnosis. (Radiology. 2023 Jan;306(1):20–31). Validation of phenotyping of outcomes in pharmacoepidemiology, but not specifically AI related, is discussed in Core concepts in pharmacoepidemiology: Validation of health outcomes of interest within real-world healthcare databases (Pharmacoepidemiology and Drug Safety. 2023;32(1):1–8).

Identifying who has long COVID in the USA: a machine learning approach using N3C data (Lancet Digit Health. 2022;S2589-7500(22)00048-6) describes the development of a probabilistic phenotype of patients with long COVID using ML models and showed a high accuracy. Probabilistic phenotyping can be applied in wider contexts. In An Application of Machine Learning in Pharmacovigilance: Estimating Likely Patient Genotype From Phenotypical Manifestations of Fluoropyrimidine Toxicity (Clin Pharmacol Ther. 2020; 107(4): 944–7), a ML model using clincal manifestations of adverse drug reactions is used to estimate the probability of having a specific genotype, known to be correlated with severe but varied outcomes.

As development of probabilistic phenotypes is likely to increase, tools to assess the performance characteristics such as PheValuator: Development and evaluation of a phenotype algorithm evaluator (J Biomed Inform. 2019;97:103258) become more relevant.

Another possible category of use is hypothesis generation in causal inference, but this requires further research. For instance, in Identifying Drug-Drug Interactions by Data Mining: A Pilot Study of Warfarin-Associated Drug Interactions (Circ Cardiovasc Qual Outcomes. 2016;9(6):621-628) known warfarin–drug interactions and unknown possible interactions were identified using random forests.

16.5.3. Explainable AI

As AI decisions, predictions, extractions and other output can be incorrect, and sometimes especially so for a subgroup of people, it can cause risks and ethical concerns that must be investigated. As deep learning models are not directly interpretable, methods to explain their decisions have been developed. However, these provide only an approximation that might not resemble the underlying model and the performance is rarely tested.

In The false hope of current approaches to explainable artificial intelligence in health care (Lancet Digit Health. 2021;3(11):e745-e750), the authors show that incorrect explanations from current explainability methods can cause problems for decision making for individual patients, and they explain that these explainable AI methods are unlikely to achieve their asserted goals for patient-level decision support.

In Artificial intelligence in pharmacovigilance: A regulatory perspective on explainability (Pharmacoepidemiol Drug Saf. 2022;31(12):1308-1310) the authors argue that although by default pharmacovigilance models should require explainability, model performance may outweigh

explainability in processes with high error-tolerance where, for instance, a human-in-the-loop is required, and the need for explainability should follow a risk-based approach.

16.6. Real-world evidence and pharmacoepidemiology

16.6.1. Introduction

The pharmacoepidemiology community has a long tradition of producing, evaluating, and interpreting observational data to provide evidence on the use, safety and effectiveness of medicines. The increasing ability to electronically capture and store data from routine healthcare systems and transform it into additional knowledge has opened up new opportunities for investigators to conduct studies. The terms real-world data (RWD) and real-world evidence (RWE) have been increasingly used since the early 2000's to denote evidence generated from observational data collected during routine patient-level healthcare interactions. In medicines evaluation, evidence relying on RWD is now frequently submitted across the lifecycle of a product to complement and contextualise clinical trial knowledge with information from the routine healthcare setting, but the place of RWD in regulatory decision-making is still a subject of debate (see for example Replacing RCTs with real world data for regulatory decision making: a self-fulfilling prophecy? BMJ. 2023:380:e073100). Contribution of Real-World Evidence in European Medicines Agency's Regulatory Decision Making (Clin Pharmacol Ther. 2023;113(1):136-51) reports that RWD/RWE was considered not supportive or was not further addressed in the regulatory evaluation report for 15 of 26 applications submitted to EMA in 2018-2019, where RWD/RWE was included to support efficacy pre-authorisation. Many issues discussed in the evaluation reports with respect to RWE were weaknesses related to methodological aspects, highlighting the need for adequate pharmacoepidemiological and statistical expertise in the generation of RWE.

There is currently no internationally agreed definition of RWD and RWE. Real World Evidence – Where Are We Now? (N Engl J Med. 2022;386(18):1680-2) emphasises that these terms are being used inconsistently and sometimes interchangeably across different health domains. Although evolving, a consistent terminology is yet to be established.

This chapter discusses ENCePP's views on definitions of RWD and RWE, their role in medicines approval and evaluation, their relation to evidence generated by clinical trials, and why pharmacoepidemiological methods remain essential for the generation and assessment of RWD and RWE.

16.6.2. Definitions

The FDA's Real-World Evidence website defines RWD as "*the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources"* and RWE as "*clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD*". These definitions are widely used and have been adopted by other regulatory agencies. There is however a debate about the appropriateness of including both the nature of the data and the way it is collected ("routinely") in the definition of RWD. RWD is commonly understood as observational data from various origins (e.g., electronic healthcare records, claims data, registries) but Marketing Authorization Applications Made to the European Medicines Agency in 2018-2019: What was the Contribution of Real-World Evidence? (Clin Pharmacol Ther. 2021;111(1):90-7) illustrates the difficulty of applying definitions of RWD in authorisation applications, notably when RWD is included as an element of clinical trials. Real-World Trends in the Evaluation of Medical Products (Am J Epidemiol. 2023;192(1):1-5) states that there is room for interpretation as to the data considered as RWD, for example data collected outside of health care settings for research purposes such as those collected

through patient self-report outside of clinical encounters, or data collected through new technologies such as wearable biometric devices. This comment also applies to genetic data that are often collected outside the context of routine care or clinical trials but are generally considered as RWD. It is also noted that the data quality frameworks developed for RWD (see Chapter 13.2) examine how accurately the data represent the original information and how suitable they are but not how routinely they have been collected.

The view of ENCePP is that the specificity of RWD in comparison to any other observational data lies in the requirement for a true representation of the "real-world" patient characteristics (i.e., data with a high external validity) without influence of any specific study conditions. An assessment and validation of this real-world attribute, e.g., by external validation or benchmarking, is needed to provide assurance that it applies, or at least to evaluate and understand the deviation that may exist. A simpler definition of RWD could therefore only refer to patient data in contemporary clinical practice.

RWE is information derived from the analysis of RWD using sound epidemiological and statistical practices. The term RWE does not refer to specific methodologies and overlaps with pharmacoepidemiology, although it only partially overlaps with traditional classification of clinical research such as randomised vs. observational, prospective vs. retrospective or primary data collection vs. secondary use of data. The term RWE is nevertheless useful to state that the evidence originates from RWD, in the same way as the term experimental evidence is sometimes used to state that the evidence is based on experimental data.

16.6.3. Use of real-world evidence in medicines evaluation

There are many examples where RWD and RWE can be submitted to support medicines evaluation and regulatory decision-making. Three main objectives are identified in EMA's DARWIN EU®: Multi-stakeholder information webinar (2022; slides 14-21):

- to support the planning and validity of applicant studies, for example to inform the recruitment in pre- and post-authorisation studies, to examine the impact of planned inclusion/exclusion criteria, to measure the representativeness of the CT population (treatment and control arm) vs. the real-world target population and to evaluate whether the standard of care used in the control arm of a CT is comparable with the current real-word standard of care;
- to understand the clinical context, for example to evaluate the incidence, prevalence and characteristics of diseases, to generate evidence on the actual clinical standards of care and compare them in different populations, and to characterise real-world drug use (incidence, prevalence, amount, duration, switching patterns);
- to investigate associations and impact, for example to investigate the association between treatment exposure and either effectiveness or safety outcomes (including use of RWD as external control group), and to monitor the implementation and the effectiveness of risk minimisation measures.

Several studies have recently attempted to measure the frequency of use of RWD or RWE in marketing authorisation applications and the extent to which these data were actually utilised for decision-making, see, for example:

- Use of real-world evidence in postmarketing medicines regulation in the European Union: a systematic assessment of European Medicines Agency referrals 2013-2017. BMJ Open 2019;9(10):e028133
- Marketing Authorization Applications Made to the European Medicines Agency in 2018-2019: What was the Contribution of Real-World Evidence? Clin Pharmacol Ther. 2021;111(1):90-7

- The Role of Real- World Evidence in FDA- Approved New Drug and Biologics License Applications. Clin Pharmacol Ther. 2022;111(1):133-44;
- Use of Real-World Data and Evidence in Drug Development of Medicinal Products Centrally Authorized in Europe in 2018–2019. Clin Pharmacol Ther. 2022;111(1):310-20.

Due to variability in definitions, data sources, study designs and acceptability of RWD by regulatory decision-making bodies, very different estimates were found in these studies, with percentages of authorisation applications including RWE ranging from 39.9% to 100%.

How to enhance the suitability and acceptability of RWD/RWE to support authorisation applications is a matter of discussion and several publications have made proposals:

- Contribution of Real-World Evidence in European Medicines Agency's Regulatory Decision Making (Clin Pharmacol Ther. 2023;113(1):136-51) provides an in-depth analysis of the actual contribution of RWE in the decision-making on marketing authorisation approvals of applications submitted to EMA in 2018-2019, why such information was not considered supportive in some cases and how it contributed to the approval decision in other cases. It discusses suggestions to enable broader use of RWE in medicines development, including provision of data on mechanisms of action where RWE is used to extrapolate efficacy data from adults to children, previous experience with the medicinal product outside the EU application, description of the disease population and natural course of the disease, and early interactions (such as through scientific advice) between applicants and regulators to discuss the expected value of RWD to answer a specific research question, their limitations and how they could be minimised.
- Harnessing Real-World Evidence to Advance Cancer Research (Curr. Oncol. 2023;30(2):1844-59) proposes a strategy with four steps: 1) to identify meaningful and well-defined clinical questions answerable with available RWD rather than scenarios for which RCTs are necessary and feasible; 2) to rely on high-quality RWD representative of the population of interest and contemporary clinical practice and with documented data completeness and provenance; 3) to use appropriate study designs accounting for data limitations, bias, confounding and sensitivity analyses; 4) to use clear, transparent and replicable study methodology to increase the confidence in the results.
- Assessing and Interpreting Real-World Evidence Studies: Introductory Points for New Reviewers (Clin Pharmacol. 2022;111(1):145-9) details three aspects: the research question evaluated in the RWE study must align with the question of interest, with a recommendation to break it down according to the Population, Intervention, Comparator Outcome and Timing (PICOT) framework; the study design must use valid methods minimising selection bias, information bias and confounding, with a recommendation to use the target trial framework to help plan and design the RWE study; and the data must be suitable to address the research question, with elements of reliability (incl. plausibility and missingness) and relevance.
- When Can We Rely on Real-World Evidence to Evaluate New Medical Treatments? (Clin Pharmacol Ther. 2021;111(1):30-4) recommends that decisions regarding use of RWE in the evaluation of new treatments should depend on the specific research question, characteristics of the potential study settings and characteristics of the settings where study results would be applied, and take into account three dimensions in which RWE studies might differ from traditional clinical trials: use of RWD, delivery of real-world treatment and real-world treatment assignment.
- Real-world evidence to support regulatory decision making: New or expanded medical product indications (Pharmacoepidemiol Drug Saf. 2021;30(6):685-93) reviews more specifically study designs used to generate RWE, including pragmatic trials, externally controlled trials and non-randomised healthcare database studies, among others.

- Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe (Clin Pharmacol Ther. 2019; 106(1):36-9) specifies four criteria for acceptability of RWE for regulatory purposes: it should be derived from data sources of demonstrated good quality, valid (with both internal and external validity), consistent (or heterogeneity should be explained) and adequate in terms of amount of information provided.
- When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? (Clin Pharmacol. Ther. 2017;102(6):924-33) suggests that RWE is likely to be preferred over RCTs when studying a highly promising treatment for a disease with no other available treatments, where ethical considerations may preclude randomising patients to placebo, particularly if the disease is likely to result in severely compromised quality of life or mortality. In these cases, RWE could support product regulation by providing evidence on the safety and effectiveness of the therapy against the typical disease progression observed in the absence of treatment.
- Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0 (Pharmacoepidemiol Drug Saf. 2017;26(9):1018-32) highlights that substantial improvement in reproducibility, rigor and confidence in RWE generated from healthcare databases could be achieved with greater transparency about study parameters used to create analytic datasets from longitudinal healthcare databases and provides lists of specific parameters to be reported to increase reproducibility of studies.

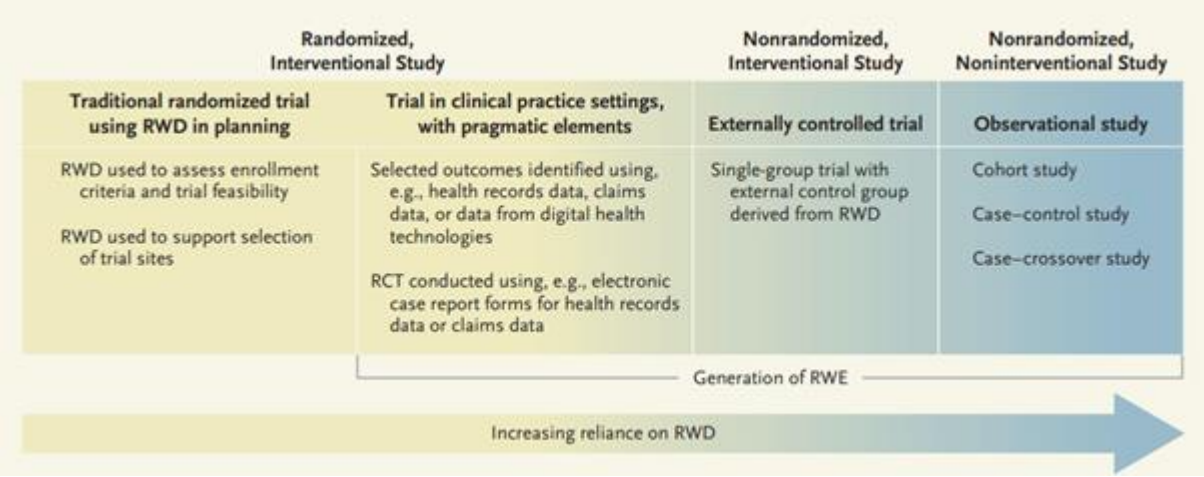
Regulatory agencies have also published methodological recommendations to medicines developers on the submission of RWD/RWE within their applications to support their evaluation and acceptability:

- Among other guidance available on the FDA's Real-World Evidence website, a draft FDA guidance for industry provides Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products(2021), the draft guidance Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (2021) provides recommendations in three domains: data sources (relevance of data source and data capture), study design elements (time period, study population, exposure, outcome, covariates) and data quality, and the draft guidance Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products (2023) provides recommendations to sponsors and investigators considering the use of externally controlled clinical trials to provide evidence of the safety and effectiveness of a drug product.
- The MHRA guidance on the use of real-world data in clinical studies to support regulatory decisions (2021) emphasises the importance of the quality of the data source, including its accuracy, validity, variability, reliability and provenance, with areas of consideration prior to submitting the study protocol. The MHRA guideline on randomised controlled trials using real-world data to support regulatory decisions (2021) provides points to consider when planning a prospective randomised trial using RWD sources with the intention of using the trial to support a regulatory decision, together with examples of scenarios, endpoints and designs.
- Health Canada's Elements of Real-World Data/Evidence Quality throughout the Prescription Drug Product Life Cycle (2019) provides overarching principles to guide the generation of RWE and an overview of some of the elements that should be addressed in protocol development and documentation of data quality within submissions containing RWE.
- The Guidance for Reporting Real-World Evidence (2023) lays the foundation for the use of RWE in regulatory approval and Health Technology Assessment (HTA) in Canada; it focusses on guiding the evaluation of a study, providing core reporting standards and prioritising transparency in reporting while accounting for practical challenges related to RWD and RWE.

- The EMA's Guideline on registry-based studies (2021) provides recommendations on key methodological aspects that are specific to the use of patient registries by marketing authorisation applicants and holders planning to conduct registry-based studies for regulatory purposes.

16.6.4. Real-world evidence vs. clinical trials

The value of RWE to provide valid evidence on medicinal products as compared to clinical trials is a frequent subject of debate in the context of regulatory assessments, especially for medicines effectiveness where a departure from traditional clinical trials has been called on to speed-up their pace, reduce their cost and increase their generalisability. While RCTs are the gold standard for demonstrating the efficacy of medicinal products, they rarely measure the benefits and risks of an intervention when used in contemporary clinical practice and the current thinking is moving away from the long-held position that RWE is always inferior due to the likelihood of bias. Randomized Controlled Trials Versus Real World Evidence: Neither Magic Nor Myth (Clin Pharmacol Ther. 2021;109(5):1212–8) illustrates that the question is not about RCTs *vs*. RWE but about RCTs *and* RWE. In other words, use of observational evidence should generally not be considered to *replace* RCT information, except in specific circumstances, but both are *complementary,* as RWE may provide additional data, such as on longer follow-up of interventions and on treatment effects in populations not included in RCTs. Real World Evidence – Where Are We Now? (N Engl J Med. 2022;386(18):1680-2) suggests that randomised, non-randomised interventional and non-randomised non-interventional studies may rely on RWD for different objectives and therefore generate RWE, as illustrated by the following diagram:



Reliance on RWD in Representative Types of Study Design. RCT denotes randomized, controlled trial; RWD real-world data; and RWE real-world evidence. Source: Concato J, Corrigan-Curay JD. Real World Evidence – Where Are We Now? (N Engl J Med. 2022;386(18):1680-2).

Statistical Considerations When Using RWD and RWE in Clinical Studies for Regulatory Purposes: A Landscape Assessment (Statistics in Biopharmaceutical Research 2023;15:1,3-13) discusses examples of when RWD can be incorporated into the design of various study types, including RCTs and purely observational studies, and reviews biostatistical challenges and methods for the use of RWE for medicinal product development.

A current domain of research is the assessment of whether non-interventional RWE studies can provide the same results as RCTs performed for the same research question. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials (JAMA 2023;329(16):1376-

85) concludes that RWE studies can reach similar conclusions as RCTs when design and measurements can be closely emulated, but this may be difficult to achieve. Concordance in results varied depending on the agreement metric. Emulation differences, chance, and residual confounding can contribute to divergence in results and are difficult to disentangle.

16.6.5. Real-world evidence and pharmacoepidemiology

All the elements cited above to generate valid and reliable RWE using RWD are related to fundamental principles of pharmacoepidemiology. The widespread use of the concept of RWD/RWE has stimulated the use, accessibility and quality control of data sources as well as methodological developments to prevent and control bias and confounding, for example confounding by indication. Pharmacoepidemiologists should therefore take a leadership role and embrace this concept as a domain of research supporting regulatory decisions on medicinal products and public health in general. The following list includes areas of pharmacoepidemiological expertise that ENCePP considers important to develop and disseminate:

- Knowledge about RW data source metadata and its characteristics
- Understanding of different data types (e.g., primary care, specialist care, hospital care, disease registries, claims data, longitudinal drug prescription, dispensing or other drug utilisation data).
- Understanding of the context in which the data are collected, which should include – but not be limited to – local diagnostic criteria, local prescribing practices, local prescribing formularies, local coding practices, reimbursement policies, etc.
- Understanding of real-world data sources, including:
  - common coding terminologies for drug exposure and clinical events,
  - common data models,
  - assessment of data quality (incl. data quality metrics, data quality frameworks, misclassification and missingness, benchmarking),
  - their limitations and the statistical approaches to address them
- Knowledge about appropriate methods to establish meaningful RW evidence
- Expertise in epidemiological study designs, including traditional designs as well as case-only and case-population designs; studies with primary data collection vs. secondary use of data; prevalent-user vs. incident-user designs, positive and negative control exposures and outcomes; use of active exposure vs. non-exposure comparator groups.
- Knowledge of mechanisms of bias in observational studies (information bias, selection bias, confounding) and methods to address them at the design and analytical stages (incl. restriction, matching, stratification, modelling, use of propensity score methods, multiple imputation); methods to address unmeasured confounding and time-dependent confounding.
- Knowledge in handling effect modification, interaction and heterogeneity in observational studies.
- Expertise in assessing and validating different exposures, outcomes and covariates in observational studies.
- Knowledge in causal inference methods (incl. missing data handling, target trial emulation and interplay with ICH E9 (R1)).
- Knowledge in evidence synthesis, meta-analysis and data pooling.

- Experience in assessing a statistical analysis plan for an RWE study.

Annex I

Available as a separate document on https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml.

Annex 2

Available as a separate document on https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml.

References

Available as a separate document on https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml.

Authors list

Available as a separate document on https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml.