12

Encounter Databases

Tobias Gerhard[1], Yola Moride[1], Anton Pottegård[2], and Nicole Pratt[3]

[1] *Rutgers Center for Pharmacoepidemiology and Treatment Science, Rutgers Ernest Mario School of Pharmacy, New Brunswick, NJ, USA*
[2] *Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, Odense, Denmark*
[3] *Quality Use of Medicines and Pharmacy Research Centre, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia, Australia*

Encounter databases contain electronic records of healthcare encounters for large, defined populations. They capture information on patient characteristics, prescription fills, and medical services, as part of the routine administration or reimbursement of healthcare. This is in contrast to electronic health record (EHR) databases, described in detail in Chapter 13, which are primarily intended and maintained to support patient care. Encounter data may contain records at various levels of granularity, ranging from records of individual services (fee-for-service claims) to aggregate records of care episodes (hospital discharge records). Encounter databases exist in many countries and within a number of vastly different healthcare systems. An increasing number are available for research and consequently, encounter databases have become a cornerstone of pharmacoepidemiologic research. Although they vary markedly in their specific characteristics, encounter databases share a number of defining features that warrant their discussion as a group.

While previous editions of *Pharmacoepidemiology* provided detailed information on a few select encounter databases in several dedicated chapters, this sixth edition presents this information in more general terms in a single chapter. This change in approach reflects the continued growth in the number of encounter databases used for pharmacoepidemiologic research and the often significant changes in the characteristics of individual databases over time. Rather than attempting to provide an encyclopedic description of available databases, this chapter focuses on the description of key commonalities and distinctions across encounter databases, illustrated with selected examples and supplemented by references to more comprehensive resources in the literature. The chapter now also includes a dedicated discussion of the considerations faced by researchers when evaluating the appropriateness of a specific encounter database or deciding among multiple encounter databases for their research question. The use of encounter databases for multi-database studies within distributed data networks is discussed in Chapter 25.

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.
© 2020 John Wiley & Sons Ltd. Published 2020 by John Wiley & Sons Ltd.

Description

Encounter data arise as part of the routine administration of a person's interactions with various sectors of the healthcare system. When combined, these data can be used to infer a longitudinal picture of a person's medical and treatment history. The quality of that picture, that is, its usefulness for pharmacoepidemiologic and other research, depends on the completeness and validity of the information available.

The essential attribute of all encounter databases useful for pharmacoepidemiologic research is a defined population for which healthcare services are recorded regardless of the provider or location where care is received [1]. Such databases are considered *population based* (see Chapter 17). Precise definition of the database population avoids various forms of selection bias common in nonpopulation-based studies (e.g., biased control selection in hospital-based case–control studies). Complete capture of all relevant healthcare services avoids bias from incomplete and potentially differential measurement of healthcare services (e.g., incomplete ascertainment of hospitalizations occurring in a nonparticipating healthcare system). Although representativeness of a geographic region or the general population is often desirable, it is not necessary as long as the database population is accurately defined.

While encounter databases ideally capture all healthcare services, in practice specific service types may not be captured due to the nature of the data collection process (most often due to lack of reimbursement). However, accurate qualitative description of the specific service types with lack of coverage or incomplete capture is critically important to allow evaluation of the appropriateness of the database for a given research question.

Encounter databases are maintained by a number of different entities including government agencies, insurance companies, health plans, and information services companies. The primary purpose of encounter databases is often the reimbursement of fee-for-service payment claims, and such encounter data are often referred to as claims data. In some instances, for example in US health plans with staff model delivery systems or capitated payment models, the purpose is purely administrative with no processing of payments for individual services. This distinction can be important as the accuracy and validity of data correspond to the purpose of the record. For example, claims records are routinely audited to prevent fraud and thus assure high accuracy of the data in instances where the information is directly relevant to the processing of the correct payment amount (e.g., quantity and dose of medications dispensed by a community pharmacy or type of procedure performed during an outpatient physician visit). In contrast, data elements that are not directly tied to the payment, for example the specific diagnoses associated with an outpatient visit or procedure, may be recorded with lower accuracy. In purely administrative databases, data characteristics depend on the specific data collection and quality assurance processes in place for each of the data elements.

While an ideal encounter database would capture all types of healthcare services, in practice, individual databases often lack coverage of certain service types, depending on the purpose of the database and the nature of the data collection process. The completeness of information captured in a database is a function of the types of healthcare services (data domains) included, as well as of the comprehensiveness of data capture within each domain. Encounter databases useful for pharmacoepidemiologic research typically contain the following core data domains: (1) eligibility and basic demographic information, (2) outpatient pharmacy dispensations, and (3) medical services (typically including hospitalizations; commonly also including outpatient health services).

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Data domains may be maintained in separate files within a single *integrated* database (e.g., US private and governmental databases), or in multiple autonomous databases that together function as a *federated* virtual database (e.g., Nordic healthcare databases), depending on whether the data are collected and maintained by a single or multiple entities. Both integrated and federated databases require reliable linkage of an individual's records over time and between data domains. Box 12.1 summarizes commonly available data elements within the core data domains. The content of the core data domains often varies across individual databases in terms of which types of healthcare services are captured. While some databases are limited to hospital discharge data, many also capture data on outpatient office-based physician visits, outpatient clinic visits, long-term care facilities, dental, and vision. Another example of incomplete data capture within a data domain is incomplete or lack of recording of over-the-counter medication fills in prescription databases. Similar variability across databases exists in terms of access to nonencounter data, such as electronic health records, laboratory test results, diagnostic examinations, provider specialty/characteristics, vital statistics, or disease registries. Lastly, profound differences also exist in data structure and coding systems.

Because the primary purpose of encounter data is administrative, any inferences about a patient's medical history made from these data have to be carefully evaluated. Validation of encounter data, ranging from the validation of individual data elements to the validation of complex encounter data-based algorithms, is critical for rigorous pharmacoepidemiologic research with encounter databases (see Chapter 37). Validation necessitates the ability to reliably link an individual's encounter data to nonencounter data sources that serve as the external gold standard, such as electronic or paper medical records, disease registries, or survey data. Furthermore, linkage with

Box 12.1 Core data domains in encounter databases

Membership	Patient identifier, sex, age/date of birth, race/ethnicity (not universally available), zip code, dates of enrollment and disenrollment, benefits package/eligibility category (if applicable)
Medical	
Outpatient services	Patient identifier, encounter date, service location (physician office, hospital outpatient, etc.), procedure codes (e.g., CPT, HCPCS), primary and secondary diagnosis codes (e.g., ICD-10-CM), provider identifier, provider profession/specialty
Inpatient services	Patient identifier, primary diagnosis, secondary diagnoses, admission and discharge dates, length of stay, patient destination, hospital identifier. Inpatient data generally do not include information on in-hospital medication use and typically represent summaries for an entire hospital stay, resulting in some lack of detail
Pharmacy	Patient identifier, unique drug identifier (e.g., US-NDC, Nordic article number) which identifies generic name, brand name, dosage form, and strength (crosswalks may be needed for some databases while others include the individual data elements coded by the unique identifier), date dispensed, quantity dispensed, prescription duration/days supply
	Typically not recorded: indication for the prescription, inpatient drug use, over-the-counter drugs

CPT, Current Procedural Terminology; HCPCS, Healthcare Common Procedure Coding System; ICD, International Classification of Diseases; NDC, National Drug Code.

10.1002/9781119431431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119431431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

complementary nonencounter data resources or *ad hoc* data collection (see Chapter 16) is also commonly implemented in order to supplement an encounter database with variables that are required to answer a specific research question but are not available in the database, such as lifestyle factors or disease severity.

Because of their size, population-based nature, comprehensive capture of the full spectrum of healthcare encounters, and ability to rapidly assemble cohorts and identify outcomes among them, encounter databases represent a tremendous resource for pharmacoepidemiologic studies. For some research questions, encounter data may be sufficient on their own, particularly when the outcome of interest has been previously validated and data on all important confounders are available within the database. In many instances, however, validation of outcomes and supplementation with external data is necessary. In these cases, the encounter databases provide the study foundation (population base and comprehensive capture of healthcare interactions) with certain data elements critical to the study question fleshed out through linkage with additional data resources.

Attributes of Encounter Databases

Although encounter databases share a basic set of defining characteristics, they differ in numerous attributes that deserve consideration when evaluating the fit of a database to address a specific research question [2,3]. Importantly, in some databases, such as US commercial insurance databases, these attributes can be heterogeneous across individual people, as availability of supplemental data (e.g., laboratory results or ability to retrieve medical records) or even core data domains (e.g., pharmacy data) may be restricted to subsets of the full database population. In these instances, suitability of the database (e.g., in terms of sample size and representativeness) should be evaluated based on the subset of the population in a given database for which the attributes required to address the question under study (i.e., key study variables) are available rather than the database population as a whole.

Population and Coverage Period

The population captured is a critically important consideration when examining the suitability of an encounter database for the study of a specific research question. The *size of the database* is typically one of the key criteria when considering an encounter database for a specific research question, in comparison to both electronic medical record databases and alternative encounter databases. A large study population is generally necessary to ensure adequate statistical power when exposures or outcomes are rare (particularly when both are rare), effect sizes are small, and when subgroup effects or treatment effect heterogeneity are of interest. In addition, some common study designs and analytic methods may further increase the size of the database necessary to achieve adequate statistical power. For example, the new-user active comparator design results in study populations that often represent only a small fraction of the total number of users of a drug of interest during the study period [4]; restriction, a common approach to reduce confounding, can substantially decrease the size of study cohorts [5]; and instrumental variable methods are statistically inefficient compared to standard regression approaches (see Chapter 44) [6].

In addition to the size of the database, the *characteristics of the database population* have to be carefully considered. As a general rule, the population covered by an encounter database is a function of the underlying healthcare system in the respective country during the study period. Knowledge of these systems is a prerequisite for informed consideration and use of databases for pharmacoepidemiologic research. Databases in countries or regions with universal single-payer coverage, such as Taiwan, South Korea, Canadian provinces (with variations in drug benefits between provinces), and the northern European countries, generally include

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the entire population and do not impose eligibility restrictions. All individuals are included and membership is maintained throughout a person's life regardless of qualifying factors such as age, employment or financial situation. As such, the characteristics of the population included in these databases are stable over time and closely track the characteristics of the population of the respective country or region as a whole. In contrast, database populations in countries or regions with less complete or more fragmented coverage, first and foremost in the US, are heterogeneous and far more complicated. The fragmentation of the US healthcare system, in particular, leads to a complex landscape for encounter databases, with different databases covering distinctly different subsets of the US population (discussed in more detail below).

Furthermore, individuals may be included in different databases at different points in time based on their personal situation (e.g., employment and state of residence), resulting in short average enrollment periods (dwell times) in any specific database environment. *Dwell time* is an important consideration particularly when the research question involves studying a long-term effect of a medicine. Similarly, when dwell time is short, it becomes increasing difficult to study new users of medicines as a lag time at the start of an individual's data capture is required to differentiate incident from prevalent medication exposure.

Lastly, the *time period covered by a database* often determines its usefulness for a given study question, depending on the start of data collection and recency of the latest available data. Studies examining trends in drug utilization over time or studies on the long-term effects of drugs, such as those with cancer as an outcome, are best served by databases with long coverage periods and a stable population. Studies of newly approved medications primarily require the most current data available. The US Medicaid Analytic Extracts (MAX) data, for example, are generally not appropriate for studies of recently approved drugs, due to an approximately three-year lag in data availability. Importantly, when studying long-term utilization trends or long-term drug effects, it is important to be aware of any changes over time in health service reimbursement and administration and appreciate their impact on drug utilization.

Services Covered and Data Completeness

For obvious reasons, medication data are a prerequisite for all encounter databases used for pharmacoepidemiologic research. Generally, these data are limited to information on medications dispensed by community pharmacies. Drugs administered during hospital stays or in long-term care units, in the emergency room, or in outpatient physician office settings are typically not included. The latter, however, can in some instances be captured through drug-specific outpatient procedure codes (e.g., drug-specific procedure codes for injection administration). In-hospital databases are discussed in Chapter 14. Over-the-counter (OTC) drug use is generally not recorded, unless OTC drugs are prescribed and specifically covered by the insurance or health system [7]. In databases for which data capture depends on a reimbursement mechanism, drug dispensings may also be missing in cases where drugs are paid for entirely out of pocket (i.e., because the cash price is lower than the required co-payment) [8], or for nonreimbursable drugs (benzodiazepines, for example, were excluded from reimbursement by Medicare Part D prior to 2013) [9].

Lastly, drug formularies, stepped therapy requirements, and prior authorization programs may impose restrictions on availability and co-payments and thus have a significant impact on use rates of individual medications and medication classes. Individual formularies may apply to an entire database population or vary widely across individuals, depending on the underlying healthcare system.

Encounter databases also vary substantially in terms of which medical services are included and, importantly, what information is captured

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

about these services. Most widely used encounter databases capture hospital services, including emergency departments. Hospital services are generally recorded as hospital discharge data that summarize information for an entire hospital or emergency department stay rather than provide documentation of individual services. Differences, however, exist in the granularity of these data, such as number of diagnosis fields and availability of procedure codes.

Even greater variation between databases exists in the capture of outpatient services. For example, in contrast to databases in the US, Canada, Taiwan, and South Korea, the Nordic countries do not maintain a database of outpatient office-based physician visits, though visits to outpatient hospital/specialty clinics are captured. As such, Nordic database studies of outcomes that do not result in hospitalizations or require outpatient office-based diagnoses for adjustment of confounding have to rely on medication use as a proxy for outpatient office-based diagnoses [10]. Capture of other service types, such as dental, vision, or long-term care, also depends on the database and the patient's specific insurance coverage. Lastly, particularly in the US, specific benefits such as mental health or other specialty services may be excluded ("carved out") in certain benefits packages and thus are not captured for individuals covered under these plans. For many databases, it is thus important to evaluate the availability of data on specific service types not only at the level of the database but at the individual level and over time using information on each person's benefit package.

Finally, databases differ in the information available about the patient and service provider. For example, data on the patient's race and ethnicity are generally not available in US administrative claims databases but are available in US governmental databases. Similarly, databases differ in the availability of provider specialty and identity for physician medical services as well as prescriber specialty and identity for dispensing data.

Linkage to Nonencounter Data

Many pharmacoepidemiologic research questions cannot be answered with encounter data alone. Some questions will require randomized trials (see Chapter 32) or prospective primary data collection (see Chapter 16). However, linkages to complementary sources of data may help to overcome inherent limitations of encounter data. Commonly used sources for nonencounter data include electronic or paper medical records, laboratory results, cause of death registries or autopsy records, disease or immunization registries, census data, biobanks, or survey data.

Linkage of encounter data to complementary data sources serves two distinct purposes: (1) validation of encounter-based information against an external gold standard, and (2) provision of supplementary data not available in the encounter database. Linkage to an external gold standard, ideally the medical record, for a sample of cases is particularly critical in order to facilitate outcome validation and calculation of positive predictive values (PPVs) of encounter data-based algorithms. In the absence of the medical record, validation may be performed against disease registries or patient self-report/survey. The validity of pharmacoepidemiologic drug and diagnosis data as well as approaches to the conduct of validation studies are discussed in detail in Chapter 37. The ability to retrieve medical records for outcome validation varies between databases and is often a critical factor in database selection.

Linkage to nonencounter data may also be necessary to provide supplemental information on variables that are unmeasured or poorly measured in the encounter data but necessary to adjust for confounding or appropriate restriction of the study population (e.g., indication for drug prescribing, lifestyle factors, measures of disease severity). Supplemental information such as laboratory test results or autopsy records may also be required for outcome ascertainment (e.g., Hba1c level as an outcome for a study

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

on the comparative effectiveness of various hypoglycemic agents).

Due to privacy restrictions that prevent the sharing or use of personal identifiers, retrieval of medical records or information obtained through direct contact with physicians or patients is generally not performed by investigators, but rather facilitated through third parties (e.g., retrieval of redacted medical records for US Medicaid and Medicare) or handled internally by employees of the participating health plans (e.g., in several US commercial insurance databases). Depending on the database, encounter and nonencounter data may be available under the same umbrella organization (e.g., linkage to EHRs in many US health plans) or require linkage to outside entities (e.g., retrieval of hospital medical records for US Medicaid beneficiaries for the purpose of outcome validation), which greatly affects the feasibility, efficiency, cost, and success rates of the linkage.

Healthcare data linkages are governed by both privacy restrictions and the availability of common linkage variables in the respective databases. Privacy regulations governing the ability to link personal health information are complex and vary between countries and database owners, and over time. When these regulations do not preclude linkage, health information databases can be linked using either deterministic or probabilistic methods [11]. Briefly, in deterministic linkage, a unique identifier or a combination of several nonunique variables available in both databases must match exactly (though the match can be implemented based on transformed versions of the variables, e.g., phonetic codes instead of names to minimize the impact of spelling errors). Deterministic linkage is most useful if reliable unique identifiers are available (e.g., US social security number) but is also achievable with combinations of multiple nonunique variables (e.g., birth dates, admission dates, and names). However, use of variables with low discriminative power and errors or missingness in the matching variable(s) will lead to a high number of overlooked (false-negative) matches.

Probabilistic linkage methods can reduce the number of overlooked (false-negative) matches by allowing imperfect matches due to partially inaccurate or missing data but in turn may produce false-positive matches. Choice of matching method thus involves a trade-off between false-negative matches (i.e., missed matches) and false-positive matches (i.e., incorrectly matched records). Simulation studies have suggested that deterministic linkage is an equally valid but less computationally intensive method for databases with low rates of missingness and error in the linkage variables [12]. However, probabilistic linkage is more accurate in error-prone data. Although often challenging, validation of linkage quality is critically important as all linkage methods are susceptible to error. The Nordic prescription database networks are examples of highly reliable linkages between encounter data and disease registries with unique identifiers [13] while the Dutch PHARMO system uses probabilistic record linkage methods [14].

Access

Access regulations, costs, and feasibility considerations vary widely between encounter databases and often have a major impact on database choice. Access may, for example, be restricted to certain researchers, such as those working in academia or governmental agencies. Some encounter databases facilitate direct access to either "off-the-shelf" or customized anonymized datasets which may be physically transferred to the researcher's institution or accessed remotely (e.g., select US commercial databases, US governmental databases, or the South Korean HIRA data), while others require in-house data analyses and thus necessitate collaborative agreements with researchers employed by the database custodian or affiliated research institutes (e.g., US health plan databases or Nordic prescription databases). Some databases are directly accessible in anonymized form but

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

require in-house analysis performed by the database custodian when additional "custom" linkages that require personal identifiers have to be implemented (e.g., Truven MarketScan). For studies conducted through the database custodian, it is important to not only consider the attributes of the database itself, but also the data analytic capacity and track record of the in-house research collaborators. While complexity of database structure varies between databases and studies, all work with large encounter databases requires sophisticated programming skills as well as a comprehensive understanding of database-specific details. The latter consideration can be a major advantage of collaborative arrangements that include researchers or programmers from the database custodian.

Costs of data access vary across databases and often within databases, depending on the specific characteristics of the study in question. Fees often vary by size (number of individuals) and complexity (number of files/data sources) of the requested dataset as well as by funding source (e.g., federal versus commercial funding). In-house data analysis often imposes substantial additional costs.

Application processes vary widely as well. While all databases require compliance with data privacy and security restrictions, some may also impose scientific vetting of the research plan or a justification of the benefit of the research to the public. Particularly in projects that require custom linkage with identifiable patient or provider information, close collaboration with the database custodian is needed to obtain necessary approvals and maintain confidentiality. In addition, the time required for the creation of study-specific data-cuts depends on the staffing resources and experience at the database custodian and the complexity of the required dataset. As a result, the duration from the beginning of the application process until the start of the research can vary dramatically between several weeks to multiple years.

In practice, while access considerations and familiarity with a given database are often important drivers of database choice, it is vital never to lose sight of the suitability of the database for the specific research question under study.

Selected Encounter Databases

A selection of widely used encounter databases and database types with their basic characteristics is presented in Table 12.1 and discussed below. Databases will be discussed by region and include US databases, Canadian databases, European databases, and Asian databases.

Encounter Databases in the United States

US encounter databases are arguably both the largest databases available and the most fragmented. Unlike most industrialized nations, the US does not have a uniform health system or universal healthcare coverage, resulting in databases with characteristics that differ markedly from databases in the rest of the world. In 2016, 292 million people, or 91% of the US population, had health insurance coverage, with 28 million uninsured [15]. 216 million people had coverage from private plans (68%), mostly employment-based plans (179 million; 56%). 120 million people (37%) had coverage from a governmental plan; 62 million by Medicaid (19%), 53 million by Medicare (17%), and 15 million had military coverage (5%). Note that these census data-based estimates show some inconsistencies with the reporting from the Centers for Medicare and Medicaid Services (CMS) presented later in the chapter.

Broadly speaking, most employed individuals and their dependents are covered by commercial insurance, adults 65 years and older and qualifying individuals with disabilities are covered by Medicare, and the poor and other disadvantages groups are covered by Medicaid. Furthermore, insurance coverage in the US is not mutually exclusive. In 2016, 22% of the population with

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 12.1 Database characteristics[a]

Type	Government, US	Government, US	Health System Databases, US	Commercial Insurance, US	Government, Canada	Government, Northern Europe	Government, Asia
Examples	Medicare	Medicaid Analytic eXtract (MAX)	Kaiser, Geisinger	HealthCore, MarketScan, Optum, Pharmetrics	Saskatchewan, Quebec	Denmark, Norway, Sweden, Netherlands	South Korea, Taiwan
Networks	Sentinel	Sentinel	HCSRN, Sentinel, PCORnet, VSD	Sentinel, CNODES	CNODES	PROTECT	AsPEN
Population					Province	Country	Country
Relative size	+++	+++	++	+++	++	++	+++
Dwell time	+++	+ to ++	+ to ++	+	+++	++++	++++
Lag in availability	3–4 years	1–2 years	<1/2 year	<1/2 year	Variable	Up to 2 years	Variable
Access	Direct	Direct	In-house	In-house	<1–2 years	In-house	Variable
Retrieval of medical records for validation	Yes	Yes	Yes	Partial	No[b]	Yes	Yes for some databases
Coding, drug	NDC	NDC	NDC	NDC	AHFS	ATC	ATC
Coding, Dx	ICD-9-CM, ICD-10-CM	ICD-9-CM, ICD-10-CM	ICD-9-CM, ICD-10-CM	ICD-9-CM, ICD-10-CM	ICD-9-CM, ICD-10-CM	ICD–8, –9 and -10	ICD-10, ICD-9
Validation	+++	+++	++++	+ to +++	++	++	++
Supplementation	+++	++	++++	+ to +++	++	+++	+++

[a] Drugs and claims only in subset.
[b] Apart from a few rare exceptions, one cannot retrieve medical charts of cases ascertained in a given study. However, can identify patients in medical records in institutions and link back to the database.

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

health insurance had multiple coverage types due to either switches in coverage type or to simultaneous coverage to supplement their primary insurance type.

With the exception of Medicaid programs, which generally provide prescription drug coverage for all beneficiaries, prescription drug insurance is typically provided separately from medical insurance, resulting in subgroups of patients in major databases for whom only pharmacy or medical data are available. Although pharmacy claims are recorded with high accuracy, medication dispensings can be incompletely captured in patients covered by multiple insurance programs or in instances where the co-payment is greater than the cash price of the medication [16,17]. In recent years, several large US retailers have begun to offer low-cost generic medications for as little as $4 for a monthly supply, considerably less than the average tier 1 co-payment ($11 in 2017) [8,18]. Since there is no financial incentive, pharmacies may not submit insurance claims when patients pay cash, resulting in potential underascertainment of low-cost generic medications. To date, empirical studies examining the missingness of dispensings in claims databases have reported a limited impact of such generic drug discount programs [16,17,19]. Payments rates and modalities for medical services vary widely, ranging from fee-for-service to capitated arrangements in which providers receive a fixed payment per patient per unit of time for the delivery of a specified set of services. Detailed claims data are often not available for services or patients covered by such capitated payment models as the payment amount is independent from the specific services provided.

Several large US encounter databases are available and have been widely used for pharmacoepidemiologic research [20–22]. These databases include markedly different groups of the population and often individuals with heterogeneous healthcare coverage are included within the same database. To complicate matters further, significant mobility exists between databases as changing life circumstances (loss of employment, change in employer, disability, reaching age 65/Medicare eligibility) result in changes in insurance coverage. This is often referred to as "churning" and substantially affects the average dwell time of individuals in US encounter databases [23].

US databases generally use the National Drug Code (NDC) for medication data, the Current Procedural Terminology (CPT) coding and Healthcare Common Procedure Coding System (HCPCS) for procedures, and the International Classification of Diseases, Clinical Modification (ICD-CM) system for diagnoses. The US transitioned from ICD-9-CM to ICD-10-CM on October 1, 2015 [24], which has important implications for pharmacoepidemiologic research conducted in US databases. Despite the existence of crosswalks, the performance characteristics of encounter data-based algorithms have to be demonstrated for the new coding system and studies that span the transition date will have to implement multiple coding systems in a single study. Data privacy and security of identifiable healthcare data in the US are governed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) [25].

US Private Insurance Databases

Most healthcare in the US is covered through private insurance, predominantly employer-based insurance. For-profit and not-for-profit insurance companies offer a wide range of plans that vary in characteristics such as premium, co-payment/co-insurance, deductibles, out-of-pocket limits, services covered, drug formularies, and provider choice. Payment systems and business models are complex and undergo continuing change over time. Because most private insurance plans are associated with the employer, many patients frequently change insurance plans due to changes in employment or when employers change their contracted insurance portfolio. Although there are hundreds of health insurance companies in the US, a relatively small number

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of companies provide coverage for a majority of the privately insured population. The great majority of the privately insured population are covered by insurance systems that pay for the care provided by others [20]. Commercial insurance databases derived from these systems are some of the largest databases available for pharmacoepidemiologic research. A smaller group is covered by integrated, often not-for-profit, healthcare delivery systems that assume responsibility for preventive and therapeutic health services to a defined population, often employ group or staff model delivery systems, and frequently operate their own hospitals (e.g., Kaiser Permanente) [22]. Though typically smaller in size, the databases associated with these healthcare systems offer extensive data resources that combine encounter data with detailed clinical data resources, including EHRs and direct access to patients and providers.

Commercial insurance databases are longitudinal collections of billable healthcare interactions [20]. These databases are maintained by a variety of entities. This includes large insurance companies, often through health data analytics-focused subsidiaries (e.g., Optum Clinformatics/UnitedHealth Group [26]; HealthCore Integrated Research Database/Anthem, Comprehensive Health Insights Outcomes Data/Humana), as well as health information technology companies (e.g., Truven Health MarketScan [27], IQVIA PharmetricsPlus). Commercial insurance databases typically include several millions to tens of millions of individuals cross-sectionally and cumulatively often exceed 100 million unique patients over the life span of the database. Importantly, however, the extremely large sizes of these databases do not necessarily translate directly into the size of pharmacoepidemiologic study cohorts. Given the approximately 30% annual churn rate in commercial insurance coverage and the fact that prescription drug coverage is often separately administered or absent, only approximately 50%, 30%, and 15%, of beneficiaries with medical coverage have continuous medical and pharmacy coverage for 1, 2, and 4+ years, respectively [20].

Another important and often underappreciated feature of commercial insurance databases is the large within-database heterogeneity in data availability, completeness, quality, and ability to link member data to nonencounter data. Within a typical commercial database, members are covered by a variety of insurance products (often from multiple insurance companies), leading to substantial differences in services captured in the database. Drug formularies, which determine coverage and out-of-pocket costs for prescription drugs, for example, vary widely between plans. Similarly, a study that requires data on dental procedures would have to be limited to the subset of beneficiaries with a dental benefit during a specific time period. Completeness and quality of the claims data also depend on the payment model employed by the respective insurance products. As discussed earlier, completeness and accuracy with which services are captured may differ substantially depending on whether services are reimbursed through fee-for-service payments or capitated arrangements. Such capitated arrangements may apply to all medical coverage or be limited to specific services (e.g., specialist visits or mental health services).

The ability to validate or supplement the claims data is also often limited to subgroups of members included in the database. For example, for databases maintained by subsidies of insurance companies, data validation and supplementation may not be permitted for the (sometimes substantial) proportion of individuals in "self-funded" plans, where the employer assumes direct risk for payment and the insurance company only provides administrative services (ASO members). Similarly, the ability to identify patients and validate or supplement patient data depends on the contractual arrangements with the data sources (employers, health plans) and is generally restricted to a limited subset of the full database populations. Given the substantial

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

heterogeneity in multiple data attributes within and between commercial databases, thoughtful consideration of detailed information on members' individual benefit packages is critical to facilitate restriction of the study population to those for whom all necessary data elements and linkages are captured or available in the database.

Several models exist to enable research access to commercial insurance databases. Some databases are directly available in their entirety through licensing arrangements (e.g., MarketScan®), while others are solely accessible on a project-by-project basis via collaborative arrangements involving in-house programmers. Databases available for licensing are deidentified, with all personal identifiers removed, and as such do not support external linkages. Studies that require such linkages for validation or supplementation of the encounter data typically require collaboration with researchers employed by the database custodian. Such collaborations have the added advantage of tapping into the often substantial experience of the custodian research team. Most major commercial insurance providers also participate as data partners for the Sentinel System (see Chapter 25).

Integrated healthcare delivery system databases differ from commercial insurance databases in that they include a defined population whose entire spectrum of care is the responsibility of and provided by the integrated delivery system. Similar to commercial insurance databases, the delivery system databases include pharmacy dispensing data as well as encounter data on diagnoses and procedures from care delivered in both ambulatory and inpatient settings. However, because all care is provided by the delivery system, these databases also have access to full inpatient and outpatient electronic and paper medical records, and have the ability to interact with providers and patients. Although the latter features are also available for subsets of patients in many commercial insurance databases, the uniqueness of integrated delivery systems databases lies in the fact that these linkages cover the entire care received by the patient and are not limited to care received by specific practices or hospitals. Since many EHR systems include information on drugs prescribed, delivery system databases have often access to both prescription and dispensing data, which can be useful for a variety of research questions, such as questions of primary nonadherence [28]. In addition, several integrated healthcare delivery systems include affiliated research centers that maintain a variety of additional data resources such as registries for cancer, diabetes, or cardiovascular disease. Integrated health delivery systems have a long track record of pharmacoepidemiologic research, and many are consortium members in the Health Care System Research Network (HCSRN, formerly known as the HMO Research Network) and data partners for the Sentinel System (see Chapter 25) [22].

US Government

The US government funds healthcare services through several major programs, including Medicare and Medicaid, as well as the Department of Veterans Affairs Healthcare System (VA). In contrast to the VA, which is a large provider of healthcare services operating numerous hospitals, clinics, and nursing homes, Medicaid and Medicare function as payers. Both programs pay directly for services using fee-for-service arrangements, but a large and growing proportion of beneficiaries receives Medicaid (68% in 2016) [29] or Medicare (30% in 2016) [30] coverage administered by private insurance companies through capitated managed care plans. For beneficiaries covered by managed care plans, encounter data for individual services have only recently become available (Medicare) [31] or of mixed completeness and quality (Medicaid) [32] and thus research with Medicaid or Medicare data has historically been restricted to individuals with fee-for-service coverage.

The Centers for Medicare and Medicaid Services (CMS) administer Medicare and Medicaid data and facilitate access to research identifiable files for research purposes. Requests

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for these data files require a research protocol and data use agreement, and are reviewed by CMS's Privacy Board. The application process is managed and supported by the Research Data Assistance Center (ResDAC) at the University of Minnesota, which provides technical assistance to researchers interested in CMS Medicare and Medicaid data. Data access requires payment of fees based on the requested population size as well as the number of data files requested, which can be provided through release of data files to investigators or remotely via the CMS Virtual Research Data Center (VRDC). A mechanism to obtain inpatient hospital and emergency department medical records corresponding to Medicare and Medicaid claims has been described and implemented [33,34]. Medicaid and Medicare data for select states or populations are also available from commercial entities (e.g., IBM Watson Health) [27].

Medicaid is a joint state/federal program intended to provide health coverage for low-income individuals. It is administered separately by each state and state-specific eligibility rules differ within federal regulations. Traditionally, the program has provided coverage limited to certain groups of low-income individuals, including pregnant women, low-income families with children, the chronically disabled, and the elderly. Following the passage of the Affordable Care Act in 2010, about one half of US states have expanded coverage to all individuals under certain income thresholds. In 2016, the average monthly enrollment in Medicaid was 70.9 million (5.7 million aged, 10.6 million blind/disabled, 28 million children, 26.7 million adults including 11.2 million adults eligible through Medicaid expansion) [30]. In 37 states, ≥50% of beneficiaries were covered through private managed care plans [29]. Medicaid coverage for eligible individuals is generally comprehensive although each state, within federally mandated parameters, administers its Medicaid program differently, resulting in variations in Medicaid coverage across the country.

Medicaid Analytic eXtract (MAX) data files include enrollment and claims data for all Medicaid enrollees in the 50 states and the District of Columbia as well as for the approximately 6.5 million (2016) enrollees in the Children's Health Insurance Program (CHIP) which serves uninsured children up to age 19 in families with incomes too high to qualify for Medicaid. MAX files have been produced since 1999 and are available per state per year [35]. MAX data are organized in five files: (1) person summary (demographic characteristics and enrollment information); (2) inpatient (inpatient hospital claims with one record per stay; procedure and diagnosis codes); (3) long-term care (e.g., nursing facility claims); (4) prescription drug (outpatient pharmacy data including national drug code, quantity dispensed, days supply); and (5) other services (e.g., laboratory and other diagnostic claims). MAX data are based on state-level data submitted through the Medicaid Statistical Information System (MSIS) and produced by CMS using extensive editing and quality control. There is a substantial lag of approximately 3–4 years between the end of a calendar year and MAX availability. Because the files are produced by state, some states may have MAX data available sooner than others. Once released, MAX data are final.

Importantly, the state reporting system is currently under transition from MSIS to Transformed-MSIS (T-MSIS). T-MSIS adds new file types (third-party liability, provider, and managed care plan data), new data elements, and modification of existing data elements [36]. One of the intentions of T-MSIS is to improve the capture and quality of encounter data for beneficiaries covered by managed care plans [37]. Data for these beneficiaries have historically been considered not to be up to research standards and have typically been excluded from most pharmacoepidemiologic research [21,32]. Given that a great majority of Medicaid enrollees are now covered under managed care plans, availability of research-quality data for this population (after

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

extensive quality checks and validation studies) would substantially increase the potential of MAX data as a resource for pharmacoepidemiologic research. Gaps in data capture due to periods of ineligibility are common as eligibility is typically determined monthly and changes with income and life circumstances. This issue affects individual eligibility groups differently, with more stable enrollment for those qualifying based on disability and less stable enrollment for low-income adults. Exclusion of beneficiaries without stable enrollment has been implemented based on eligibility files as well as through requirements for Medicaid encounters during specified periods before and after person-time under study.

Because Medicaid is administered at the state level, state-specific policies (e.g., opioid quantity limits or prior approval requirements) have to be considered in the research design. Medicaid and Medicare data for dually eligible beneficiaries can be linked. Such linkage is important in studies of dual enrollees since Medicaid or Medicare data alone fail to document the full spectrum of care provided to such dual enrollees [38]. Medicaid data for research are also available directly from individual states but access is often limited to researchers with established ties to the specific state Medicaid programs.

Medicare is the federal program that provides healthcare coverage for almost all people 65 years and over as well as for qualifying individuals with permanent disabilities [39]. Medicare coverage consists of four parts: Medicare Part A (Hospital Insurance), Medicare Part B (Medical Insurance), Medicare Part C (Medicare Advantage), and Medicare Part D (Medicare Prescription Drug Coverage). All parts of Medicare coverage require beneficiaries to pay deductibles and some stipulate cost sharing. Part A covers inpatient care in hospitals and skilled nursing facilities, as well as hospice. It is premium free for the great majority of beneficiaries. Part B covers physician and other outpatient services. It is an optional program that requires monthly premiums. Approximately 90% of Medicare beneficiaries enroll in Part B. Part C allows Medicare beneficiaries to enroll in private health plans that administer Part A and B benefits. The large majority of these so-called Medicare Advantage plans also include Part D benefits (i.e., prescription drug coverage). Part C plans are optional and require premiums. In 2016, 30% of Medicare beneficiaries received coverage through Medicare Advantage plans. Importantly, encounter data for Medicare Advantage beneficiaries have only recently become available through CMS (to date solely for service year 2015) [31].

Part D provides outpatient prescription drug coverage. Established in 2006, the program is administered by private companies that provide coverage through hundreds (782 in 2017) of prescription drug plans (PDPs) that differ in formulary coverage and cost sharing. Enrollment in Part D is voluntary and requires a monthly premium that varies between the individual PDPs. Medicare Part D imposes a coverage gap (doughnut hole) that requires beneficiaries to pay a substantial percentage of the cost of their medications (35% and 44% for brand name and generic drugs, respectively, in 2018) until they reach the out-of-pocket spending limit ($5000 in 2018). A large proportion of Medicare beneficiaries have some type of supplemental coverage (employer sponsored, Medicaid, so-called Medigap policies) to reduce out-of-pocket costs from cost-sharing requirements. In 2016, the average monthly Medicare enrollment was 57 million (48 million aged, 9 million disabled) [30]. 17 million beneficiaries were covered through Medicare Advantage and 41 million had a Part D benefit, including 16 million through Medicare Advantage plans [40].

Medicare data are available in several file types that are linkable to each other, as well as to Medicaid data for dually enrolled beneficiaries. File types include Master Beneficiary Summary Files (MBSFs), which include files on demographics and enrollment, chronic conditions, and cost and utilization; Institutional Claims,

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

which include files on inpatient services, skilled nursing facilities, and hospice; Noninstitutional Claims, which include outpatient physician claims (Carrier file) and claims for durable medical equipment; and the Part D event data file, which provides detailed prescription-level outpatient pharmacy claims. Supplementary files provide information on Part D plan characteristics, pharmacies, drugs (crosswalks from First DataBank), prescribers, and formularies.

Since prescription drug data for Medicare have become available after the establishment of Medicare Part D in 2006, Medicare, due to its large and stable population, has become one of the largest and most comprehensive resources for pharmacoepidemiologic research.

Encounter Databases in Canada

Canada, with its population of approximately 36 million, has a universal healthcare program covering all residents regardless of age or income. Program administration is the responsibility of each of its ten provinces. Physician visits, diagnostic tests, procedures (in- or outpatient), and hospitalizations are provided without payment by the patient at the point of care. Encounter data are transactional and consist of billings submitted by healthcare providers on a fee-for-service basis. A small number of physicians may have all or a portion of their activities covered by salary so the services they provide may not be included in the medical services databases. In contrast, public drug coverage programs differ among provinces; programs have been available for varying lengths of time and differ with respect to eligibility criteria as well as characteristics (i.e., copayments and deductibles). Some provinces, such as Saskatchewan and Manitoba, provide coverage for the entire population while in the others, public drug programs restrict coverage to specific segments of the population, such as the elderly, welfare recipients, or those who do not have access to private insurance plans through their employers.

Within each province, three encounter databases are available: (1) beneficiary, (2) medical services, and (3) prescription drugs. These databases are linkable through a unique patient identifier that remains unchanged over time. Additional linkage capacities are available to hospitalization databases, population health surveys [41] or province-specific disease registries [42]. Linkage of hospital charts or outpatient charts for validation of diagnoses or collection of data that are not present in the databases requires approval from the provincial information access commissioner and may not be feasible in all provinces. A number of validation studies of Canadian databases, primarily of diagnoses codes in the medical services databases, can be found in the literature but validation data remain far from comprehensive [43].

Each province maintains its own medical services encounter database, which includes all claims submitted by physicians regardless of setting (inpatient, outpatient, or emergency department) as long as the physician is paid on a fee-for-service basis. The nature of the information in the various provincial medical services databases is similar though differences exist in coding systems, such as the ICD version. For each medical service, the following information is recorded: service (date, description, location, diagnosis, and cost), provider (identifier and specialty). The vast majority of claims are submitted electronically, and the resulting medical services claims databases are populated in real time. In a few provinces, such as Nova Scotia, Manitoba, and British Columbia, mental health services, including psychotherapy, are recorded in a distinct database [44].

Unlike the medical services databases, hospitalization databases are intended for the creation of health statistics rather than for reimbursement purposes. The databases contain clinical data related to hospital discharges from acute or chronic care units, or rehabilitation centers, as well as day surgeries. With the exception of Quebec, which maintains its own hospital discharge database

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(MED-ECHO), all provinces contribute to the Discharge Abstract Database (DAD) maintained by the Canadian Institute for Health Information (CIHI) [45]. The information is therefore homogeneous across provinces. In the hospitalization databases, diagnosis was coded with ICD-9-CM until 31 March 2006 and with ICD-10 thereafter. In the DAD database, information on mental health resources, cancer staging, and reproductive history was added in 2009–2010. Hospitalization databases are typically available six months after the end of the fiscal year (March 31).

Province-specific prescription drug databases record all prescription drugs dispensed in an outpatient setting to individuals covered by the public drug plan. Drugs obtained over the counter, in hospital, in long-term care units, not included in the formulary, or covered only by private insurance programs are not usually included in the database. One exception is PharmaNet in British Columbia that links all pharmacies to a central data system. Every prescription dispensed in the outpatient setting is recorded regardless of coverage; hence, it includes medications covered by the public drug plan and private insurance programs, as well as those acquired out of pocket. Drugs are coded according to the Canadian-specific Drug Information Number (DIN) as well as the American Hospital Formulary Service (AHFS). For each dispensing, the following information is recorded: drug (date of dispensing, drug name, dose per unit, mode of administration, prescribed duration [not recorded in Saskatchewan], cost including dispensing fees), pharmacist (identifier, pharmacy location), and prescriber (identifier, specialty). Indication for a drug prescription is not recorded in any of the dispensing databases. While data and coding systems are similar across provinces, inclusion of individual drugs in the formulary and type of listing (general or restricted) may vary. For each patient, the years of entry and exit from the drug program are available in the beneficiary database. This is important information for studies that include segments of the population whose membership in the drug program may be transitory, such as membership based on income or access to private insurance programs.

Only seven of the 10 Canadian provinces make prescription data available for pharmacoepidemiologic research. Approximately half of these databases are accessible through custodians located in a university setting while the other half are accessible through provincial government agencies. In addition to the drug databases, custodians also act as a repository for other provincial databases and are responsible for their linkage.

Database access varies across provinces. Some provinces (Saskatchewan, Quebec, Nova Scotia) provide raw anonymized datasets to researchers (from academic or industry settings) while others (Ontario, BC) require data to be analyzed in-house by specific research organizations. To maintain confidentiality of the data, no patient, healthcare provider (including pharmacist), or institution identifiers are transmitted to researchers. Additional restrictions are in place in individual provinces. For example, in Quebec only a random sample of approximately 75% of the population eligible for a given study (capped at a maximum of 125 000 eligible patients) may be obtained, and no birthdates are transmitted. Exceptions can be granted through a request to the Provincial Access to Information Commission, which substantially increases the delay in data extraction.

Although Canadian encounter databases are much smaller than US encounter databases, their greatest advantage is that they include a stable population, thereby allowing longer follow-up periods. This is, for example, illustrated through a study on benzodiazepines and Alzheimer's disease, in which a 10-year follow-up was available [46]. The time required for database extraction varies across provinces, ranging from 10–20 weeks to one year, more if a request to the Provincial Access to Information Commission is required.

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Encounter Databases in Europe

Nordic Prescription Databases

The Nordic countries (Denmark, Iceland, Norway, Sweden, and Finland) have tax-supported universal health coverage. All citizens (a combined population of over 25 million people ranging from ~300 000 in Iceland to more than 9 million in Sweden) are provided with unrestricted access to health services including partial or complete reimbursement of medications.

Pharmacies electronically submit information on dispensed prescriptions to national databases without a requirement for informed consent by the patient (available since 1994 in Finland and Denmark, 2004 in Norway, 2005 in Sweden, and 2006 in Iceland) [47]. Unique civil registration codes facilitate unambiguous linkage to various national databases using a central patient router file. Linkable national databases include but are not limited to hospital discharge databases, laboratory data including results, pathology databases, medical birth databases, cancer registries, and cause of death databases, as well as census data, health surveys, biobanks, and patient records. Together, these databases create a federated database network that provides exposure information from the prescription database as well as patient and clinical outcome data from the patient router file and multiple linked autonomous databases.

The prescription databases largely include similar data elements with slight variations between countries. Besides a patient identifier (which also encodes birth year and sex), data include drug data (dispensing date, Nordic article number, a unique identifier similar to the NDC code used in the US, ATC classification, quantity dispensed in defined daily doses), a prescriber identifier (which can be linked to prescriber data such as basic demographics, profession, specialty, practice site), and pharmacy data (name and location). OTC drugs are not included unless they are obtained via prescription. Importantly, some drugs that are also available OTC are used primarily via prescription, to ensure reimbursement [7]. Besides the difference in the age of the databases, the most noticeable difference is the fact that nonreimbursed drugs are not covered by the Finnish database.

Outcome data are primarily based on national hospital discharge databases (registries). While comparable, some differences exist in the age of the patient databases, with the Finnish database dating back to 1969 [48], followed by the Danish (1977) [49], Swedish (1987) [50], and Norwegian registry (2008) [51]. Numerous other databases including cancer, birth, and death, together with pathology and laboratory results, further complement the dataset. Importantly, no large-scale data are available that provide details regarding general practice visits or other nonhospital health services. This is often referred to as a lack of "outpatient" data. However, this term can lead to misunderstandings in the context of the Nordic healthcare model. All hospital databases cover activities within hospital outpatient clinics, and as such all specialized care is covered. However, in all Nordic countries, general practice physicians serve as gatekeepers to specialized care (including both hospital and private practicing specialists). Detailed data, such as diagnoses or laboratory data, are not available. However, data on contacts (without specification for the reason for such contacts) can be obtained.

Rules governing data access vary between the Nordic countries, but generally require collaboration with local researchers. Access to Danish prescription data is particularly restrictive. Consequently, data from the Danish National Prescription Registry [52] cannot leave the data havens provided by Danish authorities. For multinational studies involving Danish individual-level prescription data, pooled analyses require data to be transferred to, for example, Statistics Denmark [53] or metaanalysis techniques to be applied to obtain pooled estimates [54]. Other sources of Danish prescription data are not restricted in the same way, but either only offer local coverage [55,56] or only provide

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

data on reimbursed prescriptions and only cover more recent years [57].

Other European Encounter Databases

Pharmacy-based federated database networks also exist in The Netherlands (PHARMO) [14] and Scotland (Tayside MEMO) [58]. These networks are limited to specific regions of their respective countries and have the ability to link to a number of databases that provide outcome and confounder information similar to those in the Nordic countries. In addition, integrated encounter databases are available in France [59] and some regions of Italy (Lombardy, Tuscany) [60]. The French national claims database, SNIIRAM, captures data for more than 66 million individuals (~98% of the French population) regardless of socioeconomic or employment status. It captures encounter data on outpatient visits, dispensed medication, procedures, chronic conditions, hospital admission diagnoses and procedures, and date of death. Data access, however, is complex.

Encounter Databases in Asia

There are many encounter databases available across the Asia-Pacific region. Many of these are population-wide databases due to the prominence of nationwide healthcare coverage in these countries. For example, South Korea and Taiwan both have single-payer, universal government-run health insurance systems that predominantly operate on a fee-for-service basis and have established national research databases. The National Health Insurance Databases of South Korea and Taiwan are the most well-established and widely used Asian encounter databases. Similar to encounter databases in the US, Canada, and Europe, they capture patient demographic information, medical (in- and outpatient) services and prescription and dispensing data. Encounter databases also exist in Australia and Japan [61]. In Australia, the commonwealth government maintains a dataset of dispensing of subsidized medicines under the Pharmaceutical Benefits Scheme (PBS) and medical services under the Medicare Benefits Schedule (MBS) [62]. A 10% sample of these data, linked longitudinally, is available and has been used for research [63,64]. Additionally, an encounter database of services provided to Australian veterans is maintained by the Australian Department of Veterans Affairs (DVA). These data include all prescriptions dispensed, medical services claimed and hospital visits attended by the veterans, their dependents, and spouses. The DVA data have been used widely for research [65,66].

One of the advantages of databases across the Asia-Pacific region is the consistency of coding systems. For example, encounter databases in South Korea, Taiwan, and Australia all use ATC codes to identify individual medicines and all but Taiwan use ICD-10 codes to identify diagnoses. This allows for comparisons of similar products across different countries without the need to map individual country-specific codes. This has allowed cross-national studies to be conducted using a distributed network approach through the Asian Pharmacoepidemiology Network (AsPEN) [67]. Pharmacoepidemiologic studies using Asian databases have historically been limited due to restrictions in the accessibility of these data. One study found that of 54 encounter databases across the Asia-Pacific region, very few allowed access to raw data [68]. Databases in Australia, Taiwan, and Japan, for example, were considered as having a high level of data accessibility, while South Korea had a medium level and Thailand, China, Malaysia, and Singapore had a low level of accessibility. The level of accessibility can differ for individual databases within the same country; some databases may require a local researcher to access data while others do not provide raw data with only summary-level data available for researchers.

Taiwanese National Health Insurance Research Database

Established in 1995, the National Health Insurance (NHI) program of Taiwan covers approximately

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

23 million individuals, more than 99% of the country's population [69]. The NHI maintains the National Health Insurance Research Database (NHIRD), which is accessible for research. The NHIRD includes but is not limited to patient demographics, prescription and dispensing data, outpatient visits, hospitalizations, and dental care. Data are updated biannually. The NHIRD can be linked to a number of external national databases through a unique and universal personal identification number. Databases available for linkage include numerous registries (birth, death, immunization, cancer, reportable infectious diseases, suicide), population-based screening programs (various cancers, myopia, urine, newborns) as well as regular examinations in school children. Strict procedures for data access and human subject review are in place to assure protection of confidentiality and data security.

South Korean Health Insurance Review and Assessment Data

South Korea has provided universal health coverage since 1989. In 2000, all health insurance systems were integrated into a single national system, creating the National Health Insurance Service (NHIS) and the Health Insurance Review and Assessment Service (HIRA). All healthcare providers are covered under the NHIS and are, with a few exceptions, reimbursed on a fee-for-service basis. Claims are electronically submitted by providers to the HIRA for reimbursement and form the basis for the HIRA database, which contains healthcare utilization and prescribed medications for approximately 50 million individuals [70]. Use of the database was initially limited until it became publicly available for research in 2009.

The HIRA research data include beneficiary ID, basic demographics, procedures, diagnostic tests, all diagnosis received by the beneficiary (coded in KCD6, the Korean Standard Classification of Disease Version 6, which is closely based on the ICD-10 system), in- and outpatient prescriptions (including brand name, generic name, prescription and dispensing date, duration, dose, and route of administration), as well as provider ID and characteristics. Validity of diagnosis data in the HIRA database has been shown to vary according to the severity of the condition (with greater validity for more severe conditions) and the care setting (with higher validity for inpatient than outpatient diagnoses) [71]. HIRA data are available to researchers in academia and government agencies and for those in the private sector such as pharmaceutical companies and medical device companies but access requires in-person consultation at the HIRA and submission of a study proposal. Once approval is given, tailored data extracts with encrypted ID information for protection of privacy are uploaded in a remote access system accessible only by the individual researcher for the study. Importantly, HIRA data are currently available only for a five-year period beginning from the current year although plans exist to expand this period to 10 years.

Strengths

Encounter databases have a number of strengths in comparison to other data sources for pharmacoepidemiologic research, which explain their broad representation in the literature.

First, automated healthcare databases facilitate the rapid and cost-efficient assembly of extremely large cohorts of patients and provide data on drug exposures, health outcomes, and potential confounding factors. Encounter databases, in particular, are the largest available population-based healthcare databases. Several of the databases discussed in this chapter cumulatively include more than 100 million individuals and provide the ability to rapidly assemble cohorts that are substantially larger than analogous cohorts from EHR databases or *ad hoc* data collection.

Encounter databases thus are uniquely able to address research questions that require the largest possible study sizes. The following example

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

illustrates the differences in cohort sizes for the same study in selected encounter and EHR databases. Filion and colleagues examined proton pump inhibitors and the risk of hospitalization for community-acquired pneumonia among new users of NSAIDs, aged ≥40 years in multiple databases within the Canadian Network for Observational Drug Effect Studies (CNODES) [72]. The respective sizes of study cohorts assembled using a common protocol and allowing multiple cohort entry dates for a single patient were approximately 2.2 million for MarketScan, 1.5 million for the combined Canadian provincial databases, and 0.6 million for the UK GPRD, the largest population-based EHR database. The MarketScan cohort was more than 3.5 times larger than the GRPD cohort, despite not including data on ≥65 year olds who made up around 35% of the total study population.

Second, because encounter databases are population based and provide a comprehensive capture of covered healthcare encounters regardless of the provider, they can support the full range of epidemiologic study designs including cohort, nested case–control, and self-controlled designs. While this strength is shared by a number of other population-based automated databases, it is a critical limitation to nonpopulation-based data sources such as EHR databases of individual institutions or health systems.

Third, many encounter databases facilitate systematic or *ad hoc* linkage to nonencounter data resources, including electronic or paper medical records, disease registries, laboratory results, or patient and provider surveys. Such linkages can support validation of study outcomes and allow supplementation of encounter data with variables such as laboratory results or lifestyle data. In ideal circumstances, such linkages thus provide the ability to take advantage of the size and population-based nature of encounter data, while also accruing the advantages of higher data quality and greater clinical detail available from data sources such as EHRs, disease registries, or patient and provider surveys. Importantly, however, linkage ability and quality vary substantially between individual encounter databases and have to be carefully considered for each study question.

Fourth, many large encounter databases are broadly representative of nations, regions, or particular health systems. As such, they can often serve an important role in facilitating health services and health policy research. Many include very stable populations that facilitate assessment of long-term safety effects and long-term trends in treatment practice and quality. Further, encounter databases from countries or regions with universal health coverage – by definition – are free from selection bias as inclusion in the database is universal.

Fifth, for encounter data generated from fee-for-service payment claims, data elements that directly pertain to the payment amount are subject to auditing and considered highly accurate. This is true for procedure claims (type of procedure performed) [73] as well as for pharmacy claims (date, drug, and quantity dispensed) [74]. Importantly however, the accuracy of procedure data primarily relates to the occurrence of the procedure billed while the accuracy of the clinical indication associated with the procedure may be substantially lower. For example, a validation study that used specific surgical procedure codes in Medicaid data as part of an algorithm to identify cases of hip fracture found in medical record review that while all of the procedures billed for were actually performed, some of the procedures were used to correct orthopedic conditions other than hip fracture [75]. A further advantage of pharmacy data compared to prescription data recorded in EHR databases (see Chapter 13) is the fact that prescription dispensings are one step closer to ingestion than what was prescribed and thus are subject to a lesser degree of exposure misclassification [76]. The accuracy of encounter data generated by administrative processes not

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

related to payment is less well established and likely to vary depending on the existence and rigor of quality assurance processes.

Sixth and last, data capture processes in encounter data are automated and independent of the study question and hypothesis, greatly diminishing the likelihood of recall or assessment biases.

Limitations

Encounter databases are primarily intended and maintained for payment or other administrative purposes, and therefore are subject to important limitations when used for research.

First, one of the greatest concerns when using encounter databases for pharmacoepidemiologic research is the uncertain validity of diagnostic information (see Chapters 11 and 37) [49]. While these concerns apply to all diagnostic encounter data, they are amplified for diagnoses recorded in the outpatient setting where diagnosis is typically not directly linked to a particular level of payment. It is thus critically important for all encounter-based research to validate diagnostic data (for both outcomes and important confounders) against external gold standards such as the medical record or disease registries. These gold standards, of course, may not be correct either when compared to research-grade diagnoses as employed by randomized controlled trials.

Second, encounter data lack clinical detail such as markers of disease severity (e.g., blood pressure, ejection fraction) and lifestyle factors (tobacco and alcohol use, body mass index, physical activity). Oftentimes, data elements are available (e.g., diagnostic codes for obesity or smoking status) but of extremely low sensitivity. For example, a study using data from the National Health and Nutrition Examination Survey to validate diagnosis of obesity in Medicare claims found that claims-based diagnostics codes fail to identify a great majority of patients with obesity (sensitivity of 18%) [77]. Though still far from perfect, clinical details such as disease severity and lifestyle factors are generally better captured by paper or electronic medical records. Because such clinical detail is often critical for confounding adjustment, methods that minimize unmeasured or residual confounding (self-controlled designs, active comparator new-user designs, instrumental variable analyses, propensity score calibration) are of great importance to encounter-based pharmacoepidemiologic research (see Chapter 43).

Third, while limitations of encounter databases can often be overcome by facilitating linkage to nonencounter data such as EHRs, disease registries, or laboratory results, such linkages are typically time-consuming and costly and, in many cases, only available to subsets of the database population. Further, when compared to population-based EHR databases, the resulting linked/enriched encounter data typically remain less comprehensive, and validation is often restricted to small samples often with poor response/retrieval rates.

Fourth, in certain situations, medication dispensing information may not capture data for specific drugs or drug classes. This may include drugs excluded from reimbursement, drugs that are primarily obtained over the counter, as well as low-cost generic drugs that are paid for out of pocket because the cash price is lower than the required co-payment. This may result in misclassification of exposure, such that some patients will appear not to be exposed to a medicine when in fact they were. Nonreimbursable drugs as well as low-cost generics are often better captured in EHR databases, which contain information on all prescriptions written. However, the disadvantage of prescription information is that not all prescriptions will be dispensed and will result in misclassification of exposure, such that some patients will appear to be exposed to a medicine when in fact they were not.

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fifth and last, due to the fragmentation of the US healthcare system, many large US encounter databases lack representativeness of the general population and feature significant turnover and short dwell times (e.g., US private insurance databases, MAX) [20,21,78].

Particular Applications

Encounter databases have been used in thousands of pharmacoepidemiologic publications, many of which have shaped clinical medicine or regulatory decision making. These databases have supported work across a wide spectrum of areas including drug safety, comparative effectiveness, drug utilization and health services research, methods and validation, as well as pharmacoeconomics. Descriptions of numerous specific applications of individual databases can be found in the 5th edition of *Pharmacoepidemiology* [14,20–22,45].This section outlines some typical activities involved in encounter database studies and presents some of the considerations in choosing the optimal encounter database when multiple options are available or assessing the suitability of a specific database for a given research question.

Typical Activities Involved in Studies Using Encounter Databases

Although encounter databases vary in data structure, coding schemes, and numerous other specifics, a number of activities are typical across all such databases [20]. Virtually all pharmacoepidemiologic studies of encounter databases require *linkage of records between data files and over time*. Records from different data domains, such as membership, outpatient services, inpatient services, and pharmacy, are linked so that an individual's entire set of encounters over the study period can be available for analysis. Another ubiquitous step in the conduct of pharmacoepidemiologic studies involves the *aggregation of drug, diagnosis, and procedure codes into meaningful study variables*. Exposures, outcomes, potential confounders, and inclusion/exclusion criteria for study are defined via code lists using drug, diagnosis, and procedure codes, or combinations thereof. These code lists are typically study and database specific using the coding schemes utilized by the respective database and drugs approved and available for the study population during the study period. It is often desirable to use previously validated algorithms for the definition of study outcomes and important confounding variables. Such algorithms often combine diagnostic codes, drug codes, and procedure codes for more accurate measures of disease (see Chapter 37).

Together with demographic information, these study-specific variables (e.g., drug classes, disease states) facilitate the *creation of the study population*. Study populations often consist of (new) users of specific drugs or drug classes within individuals who meet specific inclusion and exclusion criteria based on their encounter-derived medical history. Once the study population is identified in the dataset, analytic plans often specify the construction of longitudinal histories. Exposure, occurrence of outcome events, and presence of confounding factors are measured over time, typically in temporal relation to the study's index date. This facilitates the assessment of exposure periods and person-time at risk, and allows calculation of incidence rates and measures of association. If additional data not available in the encounter database are required, complementary information may be gathered through linkage to electronic medical records, data obtained directly from patients or their physicians from surveys, retrieval of paper medical records, or data routinely collected in disease, immunization, or national vital registries.

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Deciding Between Individual Encounter Databases

Database choice or evaluation of suitability of a single database should involve consideration of all database attributes relevant to the research question under study [3]. Some of the key attributes that differentiate individual encounter databases are shown in Table 12.1 and discussed below.

Target Population

The database should capture a large and representative sample of the target population (e.g., patients exposed to a particular drug) to adequately address the study question. For example, Stroup and colleagues aimed to examine the effectiveness of initiating treatment with either clozapine or a standard antipsychotic among adults with evidence of treatment-resistant schizophrenia using national US Medicaid data [79]. On first glance, this might not be an obvious choice as the adult Medicaid population is highly selective and often transient. However, Medicaid covers approximately two-thirds of all US adults with schizophrenia because most patients with severe schizophrenia qualify for disability [80]. In addition, because these individuals qualify for Medicaid because of disability rather than because of their economic condition, they are typically stably enrolled without breaks in coverage. While non-US encounter databases might have provided similarly large numbers of stably enrolled patients with schizophrenia, the authors sought a US database because of the pronounced differences in psychiatric treatment practice between US and most other countries. The study was conducted as a 1:1 propensity score matched cohort study and found that clozapine-treated patients compared to patients treated with a standard antipsychotic had a decreased risk of psychiatric hospital admission (hazard ratio 0.78, 95% confidence interval (CI) 0.69–0.88) but an increased risk of diabetes mellitus (hazard ratio 1.63, 95% CI 0.98–2.70).

Database Size

The database should be large enough to provide sufficient power to answer the research question, that is, to detect a meaningful difference between treatment groups (should a difference truly exist). This assessment should be based not on the size of the overall database but rather the size of the actual study cohort, that is, the cohort after exclusion of individuals for whom required data elements are unavailable (e.g., after exclusion of individuals under capitated payment plans), and after application of inclusion and exclusion criteria (e.g., sufficient uninterrupted baseline period).

A study by Shin et al. aimed to determine the risk of cardiovascular conditions in children and adolescents with ADHD associated with use of methylphenidate [81]. As the outcome was rare, the South Korean HIRA database of over 50 million participants was used. From this large population database, 144 258 patients aged less than 18 with a diagnosis of ADHD were retrieved. Of these, 114 657 were new users of methylphenidate and 1224 had an incident cardiovascular event. Due to the rare outcome, a self-controlled case series design was used which, compared to other designs, has the advantage of requiring fewer patients for similar power (see Chapter 43).

Ability to Validate Outcomes

Because encounter data are primarily collected for administrative purposes, the ability to validate or adjudicate outcome definitions derived from these data is essential for pharmacoepidemiologic studies. Outcome validation should generally be performed as part of any encounter-based study unless the outcome measures have previously been validated for the database. However, the ability to validate outcomes, through reliable linkage to external gold standards such as the medical record or disease registries, varies markedly between databases and is often a major consideration for database selection.

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lo Re and colleagues, for example, conducted a series of postauthorization safety studies to examine the safety (hospitalization for major adverse cardiovascular events, acute kidney injury, acute liver failure, infections, and severe hypersensitivity events) of saxagliptin compared to other oral antidiabetic drugs in patients with type 2 diabetes [82,83]. The studies were conducted separately in two EHR databases (Clinical Practice Research Datalink, The Health Improvement Network) and two encounter databases (Medicare, HealthCore Integrated Research Database). One of the requirements for the choice of encounter databases in this study was the ability to obtain inpatient medical records for outcome adjudication. Using a new-user active comparator cohort design, the study found no evidence of increased risk of any of the outcome events within any of the four databases.

Other outcomes are notoriously undercoded in encounter data and require development of custom algorithms. For example, using data from Quebec, Moride et al. developed and validated a case detection algorithm for suicide attempts in youth through a review of medical charts [84]. The following algorithm was used: diagnostic code of injury or intoxication with a location of service in the ED, followed by a psychiatric consult or a psychiatric diagnosis (psychiatric diagnoses consisting of depression, eating disorder, schizophrenia, ADHD, substance abuse, others) within two days of the ED visit. This algorithm had a sensitivity of 70% and a specificity of 97.6%.

Availability of Nonstandard Encounter Data

While all encounter databases provide information on medical services and prescription drugs, studies often require encounter data on services that are not universally available in all databases. For example, Gupta and colleagues examined opioid prescribing practices among US dentists from 2010 to 2015 using the MarketScan database [85]. Because dental services are not captured for all individuals in the database, the study population was appropriately restricted to those with simultaneous enrollment in a medical and a dental plan.

Ability to supplement with non-encounter data: Studies using encounter data may require clinical detail not available from encounter data often for the purpose of confounding adjustment or to supplement outcome identification. The ability to perform linkages that allow enrichment of the dataset with non-encounter data is thus vital and often a decisive consideration in choosing a study database. For example, Huybrechts and colleagues examined the comparative mortality risk of individual antipsychotics in elderly nursing home residents using data for US nursing home residents dually eligible for Medicaid and Medicare [86]. Clinical variables such as cognitive function or behavioral symptoms of dementia are important potential confounders but poorly measured in encounter databases. Linkage to the Minimum Data Set (MDS, available from CMS), a federally mandated health assessment tool used in nursing homes that captures information on physical, psychological, and psychosocial functioning, active clinical diagnoses, health conditions, treatments, and services, allowed the inclusion of these important covariates into the study. Using a propensity score-adjusted new-user cohort design, the authors showed that compared to initiators of risperidone, initiators of haloperidol had an increased mortality risk and initiators of quetiapine had a decreased mortality risk.

As another example, a Swedish-Danish study investigated the risks associated with being admitted to an emergency department with suspected poisoning, most often psychotropics or analgesics. Leveraging the ability to link data on admissions and prescription fills to a dataset including detailed ECGs on those admitted to the hospital, they could estimate not only the occurrence of QTc prolongation within the population but also to what extent QTc prolongation as a marker was associated with 30-day mortality [87].

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The Future

Pharmacoepidemiologic research with encounter databases has become more and more widely used and involves an increasing number of databases in a growing number of regions of the world. This trend is expected to continue, particularly as encounter databases become available in regions for which currently no data are available. In addition, the following three major factors are likely to shape the future of encounter databases: (1) advances in information technology (IT), (2) privacy regulations, and (3) changing healthcare systems. Advances in IT will continue to expand the boundaries of data storage and processing, and increasingly facilitate linkages with new and more complex sources of data, including biomarkers, social media, web searches, and around-the-clock biometric information from wearables. In addition, automated tools for data visualization and analysis of health data are becoming more accessible.

The potential for rapid development of progressively complex, detailed, and complete data resources is likely to be counteracted by increasingly strict regulations governing data privacy. These regulations will vary substantially between countries and are likely subject to rapid change.

Last, and maybe most importantly, encounter-based data are a secondary byproduct of administrative systems, created to support the local healthcare system; research applications are secondary uses. As such, encounter-based healthcare data will continue to be subject to changes in the healthcare systems that generate the data. Again, these changes are likely to vary drastically between countries and over time.

For example, the US healthcare environment is undergoing enormous transformation. Historically, healthcare providers in the US have been paid using a fee-for-service approach, where providers bill health insurance companies for the cost of the services they provide, generally justifying those bills with diagnoses. These paid claims represent the core of these encounter databases. However, the net result of this approach is that the more providers do, the more they are paid, which may result in overservicing and wasted resources. The result has been a large incentive to increase utilization, and rapidly increasing costs in the US for providing healthcare, made worse by an aging population. Under this model, the levels of expenditure are unsustainable. This has led to a shift from a fee-for-service model to a "per patient per month" payment system, so-called "population health", which of course switches the incentive to providing less care. In order to attempt to address that, incentives are being put in place to ensure that people are not receiving *too little* care, referred to as "value health". The US is in the middle of this transition now, varying greatly in different parts of the country. However, in response, there has been a remarkable consolidation of physician practices, hospitals, etc., in order to achieve sufficient scale to create the needed extensive and costly data infrastructure, and to assume the large risk associated with population health. Many other initiatives are under way as well, to limit the increasing costs of medical care. The results will likely be large changes over the next few years in the data as part of US encounter databases.

Encounter-based data are an important resource for pharmacoepidemiologic research. These data are comprehensive and often have a high level of quality as they are collected for payment purposes. As these data are generated for purposes other than research, consideration of their applicability, completeness, and generalizability needs to be carefully weighed against their convenience. As with any data source, careful consideration should be given to the issues of bias and confounding (see Chapter 3) which are not problems diminished by the increased size of the database.

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

References

1 Szklo M. Population-based cohort studies. *Epidemiol Rev* 1998; **20**(1): 81–90.

2 Strom BL. How should one perform pharmacoepidemiologic studies? Choosing among the available alternatives. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 364–76.

3 Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, LoCasale R. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012; **21**(1): 1–10.

4 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**(9): 915–20.

5 Schneeweiss S, Patrick AR, Sturmer T, *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 2007; **45**(10 Suppl 2): S131–42.

6 Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 537–54.

7 Schmidt M, Hallas J, Friis S. Potential of prescription registries to capture individual-level use of aspirin and other nonsteroidal anti-inflammatory drugs in Denmark: trends in utilization 1999–2012. *Clin Epidemiol* 2014; **6**: 155–68.

8 Choudhry NK, Shrank WH. Four-dollar generics – increased accessibility, impaired quality assurance. *N Engl J Med* 2010; **363**(20): 1885–7.

9 Centers for Medicare & Medicaid Servicers. Transition to Part D Coverage of Benzodiazepines and Barbiturates Beginning in 2013. www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/BenzoandBarbituratesin2013.pdf (accessed April 8, 2019).

10 Mikkelsen KH, Knop FK, Frost M, Hallas J, Pottegard A. Use of antibiotics and risk of type 2 diabetes: a population-based case-control study. *J Clin Endocrinol Metab* 2015; **100**(10): 3633–40.

11 Mason CA, Tu S. Data linkage using probabilistic decision rules: a primer. *Birth Defects Res A Clin Mol Teratol* 2008; **82**(11): 812–21.

12 Zhu Y, Matsuyama Y, Ohashi Y, Setoguchi S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *J Biomed Inform* 2015; **56**: 80–6.

13 Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014; **29**(8): 541–9.

14 Herings RMC, Pedersen L. Pharmacy-based medical record linkage systems. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 271–86.

15 Barnett JC, Berchick ER. Current Population Reports, P60-260, Health Insurance Coverage in the United States: 2016. Washington, DC: US Government Printing Office, 2017. www.census.gov/content/dam/Census/library/publications/2017/demo/p60-260.pdf (accessed April 8, 2019).

16 Cepeda MS, Fife D, Denarie M, Bradford D, Roy S, Yuan Y. Quantification of missing prescriptions in commercial claims databases: results of a cohort study. *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 386–92.

17 Zhou L, Stearns SC, Thudium EM, Alburikan KA, Rodgers JE. Assessing Medicare Part D claim completeness using medication self-reports: the role of veteran status and Generic Drug Discount Programs. *Med Care* 2015; **53**(5): 463–70.

18 Claxton G, Rae M, Long M, Damico A. Kaiser Family Foundation, Employer Health Benefits, 2018 Annual Survey. http://files.kff.org/

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

attachment/Report-Employer-Health-Benefits-Annual-Survey-2018 (accessed April 9, 2019).
19 Roberto PN, Stuart B. Out-of-plan medication in Medicare Part D. *Am J Manag Care* 2014; **20**(9): 743–8.
20 Seeger J, Daniel GW. Commercial insurance databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 189–208.
21 Hennessy S, Palumbo Freeman C, Cunningham F. US Government claims databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 209–23.
22 Andrade SE, Raebel MA, Boudreau D, *et al*. Health maintenance organizations/health plans. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 163–88.
23 Austic EA, Lawton E, Riba M, Udow-Phillips, M. Insurance churning. Cover Michigan Survey 2015. Ann Arbor, MI: Center for Healthcare Research and Transformation, 2016.
24 International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). Centers for Disease Control and Prevention; 2016. www.cdc.gov/nchs/icd/icd10cm.htm (accessed April 9, 2019).
25 Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191 (21 August 1996).
26 www.optum.com/content/dam/optum/resources/productSheets/Clinformatics_for_Data_Mart.pdf (accessed April 9, 2019).
27 Hansen LG, Chang S. Health Research Data for the Real World: The MarketScan Databases. July 2011. Truven Health Analytics. http://truvenhealth.com/portals/0/assets/PH_11238_0612_TEMP_MarketScan_WP_FINAL.pdf (accessed April 9, 2019).
28 Reynolds K, Muntner P, Cheetham TC, *et al*. Primary non-adherence to bisphosphonates in an integrated healthcare setting. *Osteoporos Int.* 2013; **24**(9): 2509–17.
29 Medicaid Managed Care Enrollment and Program Characteristics, 2016. www.medicaid.gov/medicaid/managed-care/downloads/enrollment/2016-medicaid-managed-care-enrollment-report.pdf (accessed April 9, 2019).
30 CMS Fast Facts. January 2018. www.cms.gov/fastfacts/ (accessed April 9, 2019).
31 Research Data Assistance Center (ResDAC). Medicare Advantage (Part C) Encounter Data RIFs Final Version. https://www.resdac.org/cms-news/medicare-advantage-part-c-encounter-data-rifs-final-version (accessed July 29, 2019).
32 Byrd VL, Dodd AH. Assessing the Usability of Encounter Data for Enrollees in Comprehensive Managed Care 2010–2011. Washington, DC: Centers for Medicare & Medicaid Services, 2015. www.mathematica-mpr.com/our-publications-and-findings/publications/assessing-the-usability-of-encounter-data-for-enrollees-in-comprehensive-managed-care-2010-2011 (accessed April 9, 2019).
33 Hennessy S, Leonard CE, Bilker WB. Researchers and HIPAA. *Epidemiology* 2007; **18**(4): 518.
34 Hennessy S, Leonard CE, Freeman CP, *et al*. Validation of diagnostic codes for outpatient-originating sudden cardiac death and ventricular arrhythmia in Medicaid and Medicare claims data. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 555–62.
35 Ruttner L, Borck R, Nysenbaum J, Williams S. Guide to MAX Data. Medicaid Policy Brief #21. Mathematica Policy Research, 2015. www.mathematica-mpr.com/our-publications-and-findings/publications/guide-to-max-data (accessed April 9, 2019).
36 Transformed Medicaid Statistical Information System (T-MSIS). www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/index.html (accessed April 9, 2019).

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

37 Submitting Accurate and Complete Encounter Data (Managed Care). www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/?entry=43416 (accessed April 9, 2019).

38 Hennessy S, Bilker WB, Weber A, Strom BL. Descriptive analyses of the integrity of a US Medicaid claims database. *Pharmacoepidemiol Drug Saf* 2003; **12**(2): 103–11.

39 Kaiser Family Foundation. An Overview of Medicare. Issue Brief. November 2017. http://files.kff.org/attachment/issue-brief-an-overview-of-medicare (accessed April 9, 2019).

40 Kaiser Family Foundation. Medicare Part D in 2016 and Trends over Time. Publication #8915. http://files.kff.org/attachment/Report-Medicare-Part-D-in-2016-and-Trends-over-Time (accessed April 9, 2019).

41 Smith PM, Stock SR, McLeod CB, Koehoorn M, Marchand A, Mustard CA. Research opportunities using administrative databases and existing surveys for new knowledge in occupational health and safety in Canada, Quebec, Ontario and British Columbia. *Can J Public Health* 2010;**101**(Suppl 1):S46–52.

42 Downey W, Beck P, McNutt M, Stang M, Osei W, Nichol. Health databases in Saskatchewan. In: Strom BL, ed. *Pharmacoepidemiology*, 3rd edn. Chichester: John Wiley & Sons Ltd, 2000, pp. 325–45.

43 Abou Chakra CN, Moride Y, Greenfield B, *et al*. Validation of claims databases for the ascertainment of adverse events: from descriptive to predictive methodologies. *Pharmacoepidemiol. Drug Saf* 2009; **18**: S1–S273.

44 Population Data BC. www.popdata.bc.ca/data/ (accessed April 9, 2019).

45 Moride Y, Metge CJ. Canadian provincial databases. In: Strom BL, Kimmel SE, Hennessy S, eds. *Pharmacoepidemiology*, 5th edn. Chichester: Wiley-Blackwell, 2012, pp. 259–69.

46 Billioti de Gage S, Moride Y, Ducruet T, *et al*. Benzodiazepine use and risk of Alzheimer's disease: case-control study. *BMJ* 2014; **349**: g5205.

47 Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sorensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol* 2010; **106**(2): 86–94.

48 National Institute for Health and Welfare. Care Register for Health Care. https://thl.fi/en/web/thlfi-en/statistics/information-on-statistics/register-descriptions/care-register-for-health-care#data (accessed April 9, 2019).

49 Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015; **7**: 449–90.

50 Swedish National Patient Register. www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish (accessed April 9, 2019).

51 Guidelines for the Distribution of Data from the Norwegian Patient Register. https://helsedirektoratet.no/Documents/English/guidelines.pdf (accessed April 9, 2019).

52 Pottegard A, Schmidt SAJ, Wallach-Kildemoes H, Sorensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol* 2017; **46**(3): 798.

53 Karlstad O, Zoega H, Furu K, *et al*. Use of drugs for ADHD among adults – a multinational study among 15.8 million adults in the Nordic countries. *Eur J Clin Pharmacol* 2016; **72**(12): 1507–14.

54 Yoshida K, Gruber S, Fireman BH, Toh S. Comparison of privacy-protecting analytic and data-sharing methods: a simulation study. *Pharmacoepidemiol Drug Saf* 2018; **27**(9): 1034–41.

55 Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. *Clin Epidemiol* 2010; **2**: 273–9.

56 Hallas J, Hellfritzsch M, Rix M, Olesen M, Reilev M, Pottegard A. Odense

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Pharmacoepidemiological Database: a review of use and content. *Basic Clin Pharmacol Toxicol* 2017; **120**(5): 419–25.

57 Johannesdottir SA, Horvath-Puho E, Ehrenstein V, Schmidt M, Pedersen L, Sorensen HT. Existing data sources for clinical epidemiology: the Danish National Database of Reimbursed Prescriptions. *Clin Epidemiol* 2012; **4**: 303–13.

58 Wei L, Parkinson J, MacDonald TM. *The Tayside Medicines Monitoring unit (MEMO)*. In: Strom BL, ed. Pharmacoepidemiology, 4th edn. Chichester: Wiley-Blackwell, 2005, pp. 323–36.

59 Bezin J, Duong M, Lassalle R, *et al*. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2017; **26**(8): 954–62.

60 Coloma PM, Schuemie MJ, Trifiro G, *et al*. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf* 2011; **20**(1): 1–11.

61 Lai EC, Man KK, Chaiyakunapruk N, *et al*. Brief Report: databases in the Asia-Pacific Region: the potential for a distributed network approach. *Epidemiology* 2015; **26**(6): 815–20.

62 Page E, Kemp-Casey A, Korda R, Banks E. Using Australian Pharmaceutical Benefits Scheme data for pharmacoepidemiological research: challenges and approaches. *Public Health Res Pract* 2015; **25**(4): e2541546.

63 Department of Health. Sources of Epidemiological Data for Use in Generating Utilisation Estimates 2015. Canberra: Commonwealth of Australia. www.pbs.gov.au/info/industry/useful-resources/sources (accessed April 9, 2019).

64 Lalic S, Gisev N, Bell JS, Korhonen MJ, Ilomaki J. Predictors of persistent prescription opioid analgesic use among people without cancer in Australia. *Br J Clin Pharmacol* 2018; **84**(6): 1267–78.

65 Pratt N, Roughead EE, Ramsay E, Salter A, Ryan P. Risk of hospitalization for hip fracture and pneumonia associated with antipsychotic prescribing in the elderly: a self-controlled case-series analysis in an Australian health care claims database. *Drug Saf* 2011; **34**(7): 567–75.

66 Pratt NL, Ramsay EN, Kalisch Ellett LM, Nguyen TA, Barratt JD, Roughead EE. Association between use of multiple psychoactive medicines and hospitalization for falls: retrospective analysis of a large healthcare claim database. *Drug Saf* 2014; **37**(7): 529–35.

67 Roughead EE, Chan EW, Choi NK, *et al*. Variation in association between thiazolidinediones and heart failure across ethnic groups: retrospective analysis of large healthcare claims databases in six countries. *Drug Saf* 2015; **38**(9): 823–31.

68 Milea D, Azmi S, Reginald P, Verpillat P, Francois C. A review of accessibility of administrative healthcare databases in the Asia-Pacific region. *J Mark Access Health Policy* 2015; **3**.

69 Hsing AW, Ioannidis JP. Nationwide Population Science: lessons from the Taiwan National Health Insurance Research Database. *JAMA Intern Med* 2015; **175**(9): 1527–9.

70 Kim JA, Yoon S, Kim LY, Kim DS. Towards actualizing the value potential of Korea Health Insurance Review and Assessment (HIRA) data as a resource for health research: strengths, limitations, applications, and strategies for optimal use of HIRA data. *J Korean Med Sci* 2017; **32**(5): 718–28.

71 Park BJ, Sung JH, Park KD, Seo SW, Kim SW. *Report of the Evaluation for Validity of Discharged Diagnoses in Korean Health Insurance Database*. Seoul: Seoul National University, 2003, pp. 19–52.

72 Filion KB, Chateau D, Targownik LE, *et al*. Proton pump inhibitors and the risk of hospitalisation for community-acquired pneumonia: replicated cohort studies with meta-analysis. *Gut* 2014; **63**(4): 552–8.

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

73 Ko CW, Dominitz JA, Green P, Kreuter W, Baldwin LM. Accuracy of Medicare claims for identifying findings and procedures performed during colonoscopy. *Gastrointest Endosc* 2011; **73**(3): 447–53 e441.

74 Levy AR, O'Brien BJ, Sellors C, Grootendorst P, Willison D. Coding accuracy of administrative drug claims in the Ontario Drug Benefit database. *Can J Clin Pharmacol* 2003; **10**(2): 67–71.

75 Wysowski DK, Baum C. The validity of Medicaid diagnoses of hip fracture. *Am J Public Health* 1993; **83**(5): 770.

76 Pottegard A, Christensen R, Houji A, *et al*. Primary non-adherence in general practice: a Danish register study. *Eur J Clin Pharmacol* 2014; **70**(6): 757–63.

77 Lloyd JT, Blackwell SA, Wei, II, Howell BL, Shrank WH. Validity of a claims-based diagnosis of obesity among Medicare beneficiaries. *Eval Health Prof* 2015; **38**(4): 508–17.

78 Crystal S, Akincigil A, Bilder S, Walkup JT. Studying prescription drug use and outcomes with medicaid claims data: strengths, limitations, and strategies. *Med Care* 2007; **45**(10 Supl 2): S58–65.

79 Stroup TS, Gerhard T, Crystal S, Huang C, Olfson M. Comparative effectiveness of clozapine and standard antipsychotic treatment in adults with schizophrenia. *Am J Psychiatry* 2016; **173**(2): 166–73.

80 Khaykin E, Eaton WW, Ford DE, Anthony CB, Daumit GL. Health insurance coverage among persons with schizophrenia in the United States. *Psychiatr Serv* 2010; **61**(8): 830–4.

81 Shin JY, Roughead EE, Park BJ, Pratt NL. Cardiovascular safety of methylphenidate among children and young people with attention-deficit/hyperactivity disorder (ADHD): nationwide self controlled case series study. *BMJ* 2016; **353**: i2550.

82 Lo Re V, Carbonari DM, Saine ME, *et al*. Postauthorization safety study of the DPP-4 inhibitor saxagliptin: a large-scale multinational family of cohort studies of five outcomes. *BMJ Open Diabetes Res Care* 2017; **5**(1): e000400.

83 Lo Re V 3rd, Haynes K, Ming EE, *et al*. Safety of saxagliptin: rationale for and design of a series of postmarketing observational studies. *Pharmacoepidemiol Drug Saf* 2012; **21**(11): 1202–15.

84 Moride Y, Lynd LD, Ducruet H, Li H, Tournier M, Greenfield B. Antidepressants and risk of suicide or self-harm in Canadian youth: a study involving common data models in Quebec and British Columbia. *Pharmacoepidemiol Drug Saf* 2014; **23**(S1): S10–S11.

85 Gupta N, Vujicic M, Blatz A. Opioid prescribing practices from 2010 through 2015 among dentists in the United States: what do claims data tell us? *J Am Dent Assoc* 2018; **149**(4): 237–45 e236.

86 Huybrechts KF, Gerhard T, Crystal S, *et al*. Differential risk of death in older residents in nursing homes prescribed specific antipsychotic drugs: population based cohort study. *BMJ* 2012; **344**: e977.

87 Schade Hansen C, Pottegard A, Ekelund U, *et al*. Association between QTc prolongation and mortality in patients with suspected poisoning in the emergency department: a transnational propensity score matched cohort study. *BMJ Open* 2018; **8**(7): e020036.

10.1002/9781119413431.ch12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781119413431.ch12 by University Of California, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License