Concepts of Designing and Implementing Pharmacoepidemiology Studies on the Safety of Systemic Treatments in Dermatology Practice

Sebastian Schneeweiss[1,2] and Maria Schneeweiss[1,2]

The U.S. Food and Drug Administration and clinical guidelines use evidence from pharmacoepidemiology studies to inform prescribing decisions and fill evidence gaps left by randomized controlled trials (RCTs). The long-term safety and infrequent adverse reactions are not well-understood when RCTs are short and involve few patients, as is the case for most systemic immunomodulating drugs in dermatology. A better understanding of the design and implementation of pharmacoepidemiology studies will help practitioners assess the accuracy of etiologic findings and use them with confidence in clinical practice. Conducting pharmacoepidemiology studies follows a structured approach, which we discuss in this article: (i) a design layer connects the research question with the appropriate study design, and considering which hypothetical RCT one ideally would want to conduct reduces inadvertent investigator errors; (ii) a measurement layer transforms longitudinal patient-level data into variables that identify the study population, patient characteristics, treatment, and outcomes; and (iii) the analysis focuses on the causal treatment effect estimation. The review and interpretation of pharmacoepidemiology studies should consider issues beyond a typical review of RCTs, chiefly the lack of baseline randomization and the use of secondary data. Well-designed and well-conducted pharmacoepidemiologic studies complement dermatology practice with critical information on prescribing systemic medications.

JID Innovations (2023);**3:**100226 doi:10.1016/j.xjidi.2023.100226

PHARMACOEPIDEMIOLOGY STUDIES IN DERMATOLOGY COMPLEMENT CRITICAL RANDOMIZED CONTROLLED TRIAL EVIDENCE

Pharmacoepidemiology or real-world evidence (RWE), the understanding of causal treatment effects from electronic data generated by the routine operation of the healthcare system, has gained much attention from United States Food and Drug Administration (FDA), payers, clinical guideline committees, and practicing physicians. Information on the effectiveness of medical products in clinical practice is thought to complement the essential evidence on their efficacy that we gain from randomized controlled trials (RCTs).

Pharmacoepidemiology typically utilizes nonexperimental designs on the basis of secondary data from clinical practice (Eichler et al., 2021). Outside dermatology, much has been written about recent developments in the field, and this text draws from reviews in other medical specialties (Schneeweiss and Patorno, 2021). In dermatology, there are many examples of pharmacoepidemiology studies that helped reassure against or identify new unintended effects that allowed better-informed treatment choices. No matter how evidence is generated, it needs to be internally valid and generalizable to a meaningful target population to be actionable. However, the accuracy of pharmacoepidemiology studies is often questioned, including several prominent studies whose findings contradicted those of RCTs (Grodstein et al., 1996; Hernán et al., 2008). These examples prompted the more rigorous approaches to pharmacoepidemiology that have evolved over the past two decades (Franklin et al., 2021; Wang et al., 2022). A modern approach to pharmacoepidemiology with applications in dermatology is discussed in this article.

[1]*Dermato-Pharmacoepidemiology Work Group, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA*

[2]*These authors contributed equally to this work.*

Correspondence: Sebastian Schneeweiss Dermato-Pharmacoepidemiology Work Group, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1 Brigham Circle (Suite 3030), Boston, Massachusetts 02120, USA. E-mail: sschneeweiss@bwh.harvard.edu

Abbreviations: AD, atopic dermatitis; FDA, United States Food and Drug Administration; PS, propensity score; PSO, psoriasis; RCT, randomized controlled trial; RWE, real-world evidence

Received 3 January 2023; revised 7 August 2023; accepted 8 August 2023

Cite this article as: JID Innovations 2023;3:100226

Large-scale and long-term safety studies in dermatology

Concerns regarding infrequent yet serious unintended consequences of systemic immunomodulating drugs in dermatology, including risks of severe infections, venous thromboembolism, suicidal actions, and cancer occurrence, discourage prescribing of efficacious drugs because a full benefit–risk assessment is difficult (Seyger et al., 2022; van de Kerkhof et al., 2015). Existing RCT evidence is not equally informative on patients with predispositions for infections, immunosuppression, or exposure to seasonally occurring or latent viral infections. Because many immunomodulating drugs are taken for prolonged periods of time, their longer-term effects need to be better understood. Pharmacoepidemiology is well-situated to fill these evidence gaps (Noe and Gelfand, 2018).

Effectiveness studies in dermatology

In some areas, RWE is increasingly used to support the effectiveness claims of drugs. In dermatology, a key challenge is that effects on skin lesions are not well-captured in electronic health records. Specific data-collection mechanisms such as dermatologic disease registries need to be established to include regular measurements of skin manifestations. Although this is feasible in principle (Yiu et al., 2021) and in fact encouraged by the FDA (U.S. Food & Drug Administration, 2018), it is far from efficient because exposure to the drug of interest may be rare, and large numbers of patients need to be followed. Because of the overwhelming superiority of new targeted immunomodulating drugs, small and short-term RCTs will dominate the effectiveness space in dermatology for the foreseeable future. Consequently, less frequent adverse effects may remain undetected by most trials (Noe and Gelfand, 2018).

DATA DERIVED FROM CLINICAL PRACTICE AND THEIR USE FOR RESEARCH ON MEDICATION SAFETY

Modern healthcare systems generate an abundance of electronically stored information on individual patients that results in long-term patient-level digital data streams (Figure 1a). Unlike highly controlled research studies, they reflect clinical practice with its many variations in treatment patterns. Most pharmacoepidemiology studies use such longitudinal data for multiple reasons: (i) they cover populations more representatively than most experimental studies; (ii) they include recording of medical interventions (phototherapy) and drug use (topical, oral, and systemic) in great detail and rely neither on patient consent nor patient recall; (iii) they do not impose experimentation in humans and may be produced faster and at a lower cost than most trials; and (iv) the prospective longitudinal recording of healthcare encounters with well-recorded service dates provides a clear chronology, which is critical for establishing causality in comparative effectiveness studies on the basis of longitudinal data streams. Much has been written about the opportunities and limitations of various data types (Dommasch et al., 2019; Gokhale et al., 2020; Lin and Schneeweiss, 2016; Schneeweiss and Avorn, 2005; Van Beek et al., 2021). Table 1 provides a summary of real-world data types.

Turning real-world data into evidence

It is critical to fully understand a data source and its processing before implementing a study on causal treatment effects (Figure 1b). The process of planning, implementing, and reviewing an RWE study comprises the following three layers that establish a sequential workflow (Figure 2a):

1. A design layer clarifies the basic study design choice, which is best informed by imagining the randomized trial we would ideally perform to answer the research questions and then want to emulate with real-world data—the target trial. When studying the safety of biologics in dermatology, this often guides us to the new-user active-comparator cohort design (Johnson et al., 2013; Ray, 2003), which has been shown multiple times to predict and replicate RCT findings in cases where the design and measurements of the trial can be emulated well (Franklin et al., 2021; Patorno et al., 2019; Schneeweiss et al., 2021a; Wang et al., 2023).
2. A measurement layer transforms the longitudinal patient-level electronic data stream into variables that identify the study population, the pre-exposure health state for confounding control in the absence of baseline randomization, the treatment status, and the treatment-emergent outcomes. Working with secondary data increases the measurement complexity compared with the primary data collection we see in most RCTs.
3. An analysis layer estimates a causal treatment effect, considering the data collection mechanism. Propensity score (PS) analyses to achieve balance in patient characteristics between treatment groups have gained popularity because of their specific suitability to large secondary databases (Webster-Clark et al., 2021). Confounding bias and differential follow-up can be further reduced by additional techniques; well-known inadvertently introduced biases, such as immortal time bias, adjustment for causal intermediates, or reverse causation, should be avoided at all costs (Table 2 provides more details).

LAYER 1: STUDY DESIGN CHOICE

Considerations in selecting etiologic study designs in pharmacoepidemiology

The clinical study question informs study design choices. In most database studies, the design choice is further modified by the limitations of the underlying data sources (Noe and Gelfand, 2018). If we observe fluctuations of treatment status within a patient over time, for example, sporadic use of a topical medication; if that drug has a short hypothesized duration of action; and if we are interested in a rapid-onset outcome, then we may consider a case-crossover design or self-controlled case series (Figure 2b) (Leducq et al., 2020; Maclure, 2007). Most pharmacoepidemiology studies exploit treatment variation between patients due to variations in practice styles and therefore use a cohort study design with concurrent controls, for example, some patients with treatment-recalcitrant atopic dermatitis (AD) start using dupilumab, and others start tralokinumab. Within cohorts, efficient sampling designs such as case-control, case-cohort, or two-stage sampling can be used when information gathering is time consuming or expensive, for example, skin biopsies and serum parameters (Figure 2b) (Schneeweiss, 2010). Treatment variation between groups of patients accessed through higher-level entities, that is, between physicians, hospitals, health plans, etc., can be exploited using instrumental variable analyses, for example, one provider group prefers ustekinumab for the treatment of moderate-to-severe plaque psoriasis (PSO), and another group prefers secukinumab (Brookhart et al., 2006b).

Selecting a comparator group or groups influences the clinical interpretation and may substantially alter the effect size. The comparator needs to be both relevant in the clinical context and a viable alternative to the study drug. Ideally, we want to restrict the comparison population to patients who, in clinical practice, have the same indication as the users of the study agent (Penso et al., 2022; Schneeweiss et al., 2020a).

Cohort studies and target trial thinking to avoid bias

A recommended starting point for planning a pharmacoepidemiology study is to envision the ideal randomized trial to answer the research

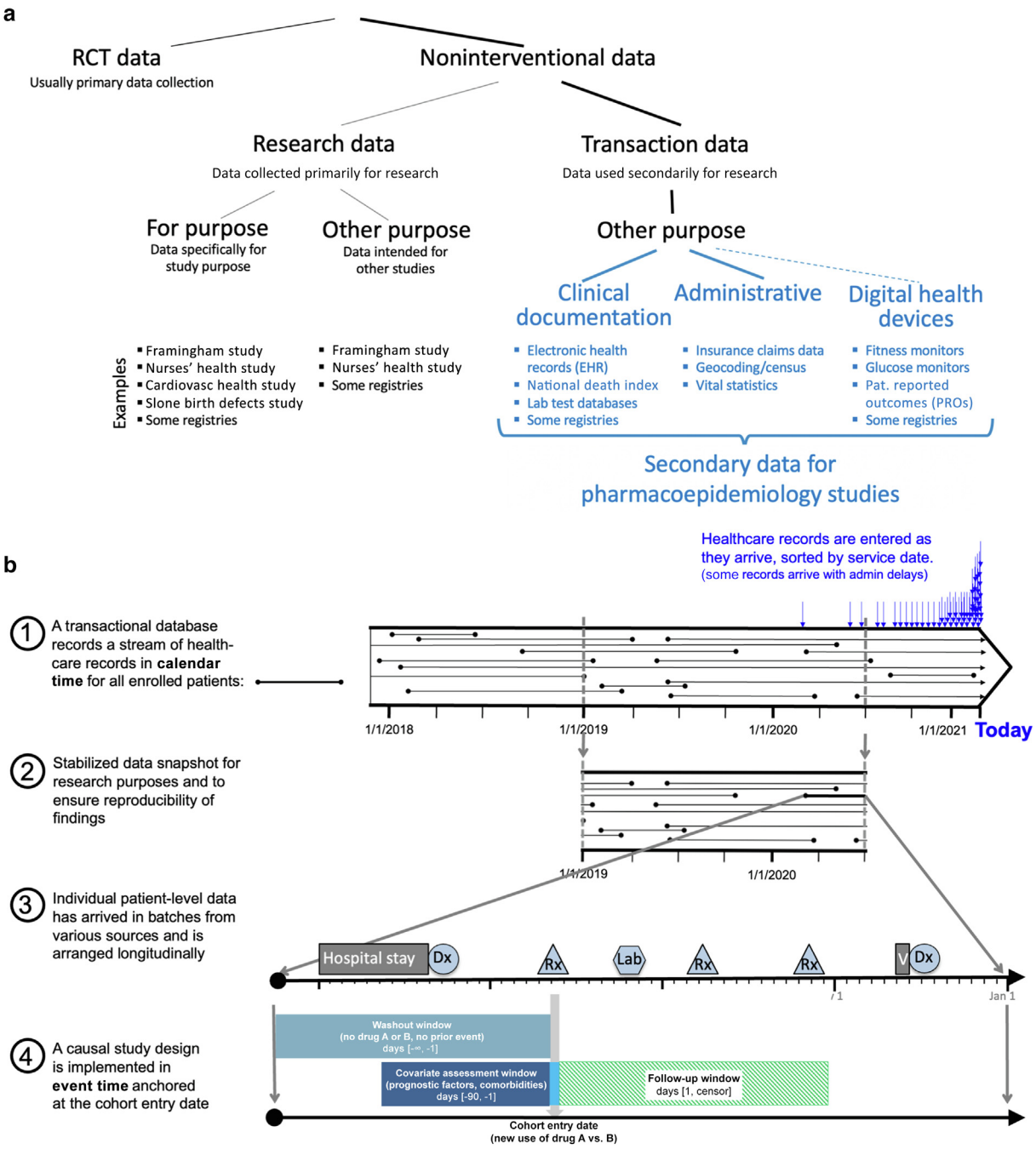


Figure 1. Turning healthcare data sources into pharmacoepidemiology studies. (**a**) A taxonomy of longitudinal data source relevant for pharmacoepidemiology research. (**b**) Typical steps of turning real-time data streams of longitudinal patient data into study data. This figure was adapted from Schneeweiss et al. (2021). Dx denotes disease, and Rx denotes treatment. Lab, laboratory; RCT, randomized controlled trial.

question if it were logistically and ethically possible and then emulate this hypothetical target trial in the design of a pharmacoepidemiology study (Figure 2c). Emulating a target trial encourages clarity in the temporality of measuring patient characteristics before starting treatment and outcomes thereafter, which is critical to enable causal conclusions (Hernán and Robins, 2016). Once a target trial is conceptualized, the design of the trial-emulating pharmacoepidemiology study may reveal weaknesses in data accuracy, completeness, and timing of measurement as well as in causal analysis (Hernán et al., 2016). Trial-emulating pharmacoepidemiology studies often expose tensions between the desire to achieve high generalizability and necessary restrictions to ensure high internal study validity.

New-user cohort to identify a clear starting point of treatment

There are several advantages to studying patients at the start of a treatment, which is a key decision-making point for prescribers. Patients in both treatment groups have been evaluated by a physician who concluded that they would benefit from starting or escalating therapy with a newly prescribed drug. This process produces treatment groups that are similar with respect to outcome predictors

Table 1. Frequently Used Data Sources from Clinical Practice and Some of Their Characteristics

Data Source	Merits	Considerations
Health insurance claims data		
Provide a patient-level longitudinal data stream of all encounters with the professional healthcare system from enrollment to disenrollment. Include physician services and hospitalizations, accompanying diagnoses and procedures, and all filled outpatient medication prescriptions, in addition to basic demographic and insurance enrollment information. Contain the billing codes that healthcare providers submit to payers, for example, private health insurance, Medicaid, and Medicare.	Often capture large populations and provide a complete longitudinal record of all encounters. Often linked with other patient-level data, for example, vital statistics, cancer registries, EHRs, and laboratory test results. Several national healthcare systems such as those in Scandinavian countries have universal life-long healthcare coverage.	Some databases from insurers with high membership turnover contain limited longitudinal follow-up, making them less suitable for the study of long-term outcomes. The fact that claims data are transaction data collected for administrative rather than research purposes requires researchers to closely examine whether the measurement of key variables is sufficient for a specific study question, for example, BMI, smoking, alcohol, and family history. Claims data vary throughout the world regarding representativeness, the scope and depth of the included information, data quality, and completeness.
EHR data		
Intended for clinical documentation and contain a wide range of patient health-related information. Obtained from clinicians involved in a patient's care and include both structured and semistructured data (laboratory test results) as well as free text notes.	Contain information-rich clinical data that may be lacking in claims data and may be needed to answer specific study questions, including symptoms, results of physical examinations, laboratory tests and procedures, diagnoses and treatment plans, and medical and social history. Increasingly used in research as a source for detailed clinical information: for example, percentage of body surface area involvement, duration of atopic dermatitis, EASI score, and disease control can be extracted from EHR data, including physician notes.	A limitation of EHR data in the United States is that only patient health information generated within a given provider network is accessible. When patients seek care from physicians or facilities outside that network, that medical information is not retrievable by the investigator. Such data leakage results in an incomplete picture of the patient's care and may lead to bias. Data incompleteness is a common challenge. In addition, there is no universally adopted standard for the types of data that should constitute EHRs. EHR data from highly integrated care systems, for example, Kaiser Permanente, or national healthcare systems, for example, NHS in England, have substantially complete information capture.
Patient registries		
A patient registry uses noninterventional study methods to systematically collect longitudinal information on patients with a particular disease or treatment type. Many registries are based on passive recording mechanisms (EHR); some actively assess specific parameters from patients. They are often centered in academic medical centers, and commercial registries become increasingly available for research. Examples are the American Academy of Dermatology registry (United States), PSOLAR (international), CorEvitas (United States), PsoBest (Germany), and British Association of Dermatologists Biologics and Immunomodulators Register (United Kingdom).	Unlike claims and EHR databases, registries may collect specific and detailed clinical information with high completeness that may be necessary for some RWE studies. This may include results from diagnostic testing, therapies offered and received, family history of disease, behavioral and environmental risk factors, symptoms, and disease progression.	Depending on the registry, it may represent a highly selective patient segment and may not reflect routine practice. The longitudinal record of medication use is often compiled through patient recall or medical records, which are both known to be unreliable. Some registries abstract information from EHRs and are thus subject to the same issues mentioned earlier.
Patient-generated data		
Data from surveys, questionnaires, smartphone apps, and social media that allow continuous data capture. Information is provided mainly by patients rather than by providers.	Questionnaire/survey data sources provide data on QOL measures, which are hard to find in other data sources. May be particularly relevant in pharmacovigilance, particularly rare adverse events associated with treatments, and factors predicting patients' adherence, behaviors, and attitudes. Some data sources include real-time monitoring to allow tracking of selected measures and symptoms.	Use of these sources implies reliance on self-reported variables, leading to recall bias, selective reporting, and missing data on important patient characteristics. Limited generalizability because of selected participants. Limited internal validity because the clinical outcomes reported are often not validated.
Vital statistics		
Vital statistics include records on the date of death and, in some cases, causes of death from physician-issued death certificates, for example, National Death Index compiled by the CDC.	Accurate and complete date of death information is crucial for many studies of clinical outcomes.	Cause of death information is often less reliable and is usually described broad categories, such as cardiovascular death or death due to malignancies and injuries, including self-harm, etc.

(*continued*)

Table 1. Continued

Data Source	Merits	Considerations
Linkage across data types		
Data from two or more sources are linked to bring together complementing information. Appropriate approvals and privacy safeguards need to be in place.	Bringing together data from disparate sources allowing the capture of comprehensive information can be very powerful Typical linkage of insurance claims data with EHRs (see as mentioned earlier) enables a combination of complete longitudinal follow-up that may be lacking in EHRs, with clinical variables that are incomplete in claims.	Validity of results depends on the quality of linkage often through tokenization. Linked data sources are often difficult to maintain and expensive. Challenges in linking data due to different purposes of data collection, discrepancies in data recording, and legal/confidentiality issues.

Abbreviations: BMI, body mass index; CDC, Centers for Disease Control and Prevention; EASI, Eczema Activity Severity Index; HER, electronic health record; NHS, National Health Service; PSOLAR, Psoriasis Longitudinal Assessment and Registry; RWE, real-world evidence.
This table was adapted in parts from Schneeweiss et al. (2021) and from Gokhale et al. (2020).

whether they are observable or not in a given data source (Ray, 2003). The clear temporal sequence of measuring confounders before starting treatment avoids the mistake of adjusting for the consequences of treatment, that is, causal mediators. Because of the well-defined starting point of new-user cohorts, it is possible to assess how hazards vary with duration of treatment. Because the new-user cohort study design closely emulates the standard parallel-group randomized trial (Figure 2d), this familiarity makes it easy to understand (Malone et al., 2018). Examples of such new-user cohort studies are the risk of infection in adults and children with PSO who receive treatment with systemic immunomodulating drugs (Dommasch et al., 2019; Schneeweiss et al., 2020a) or the risk of inflammatory bowel disease in patients using IL-17 inhibitors (Penso et al., 2022).

Active-comparator cohort to inform clinically relevant questions

Physicians can in most circumstances choose between two or more treatment options, and a placebo is not among them. Using active comparators in pharmacoepidemiology studies complements RCT evidence that is often focused on placebo comparisons. There are several examples of successful new-user active-comparator studies in dermatology (Dommasch et al., 2019; Schneeweiss et al., 2021b, 2020a). Non-user comparisons conducted in an attempt to emulate

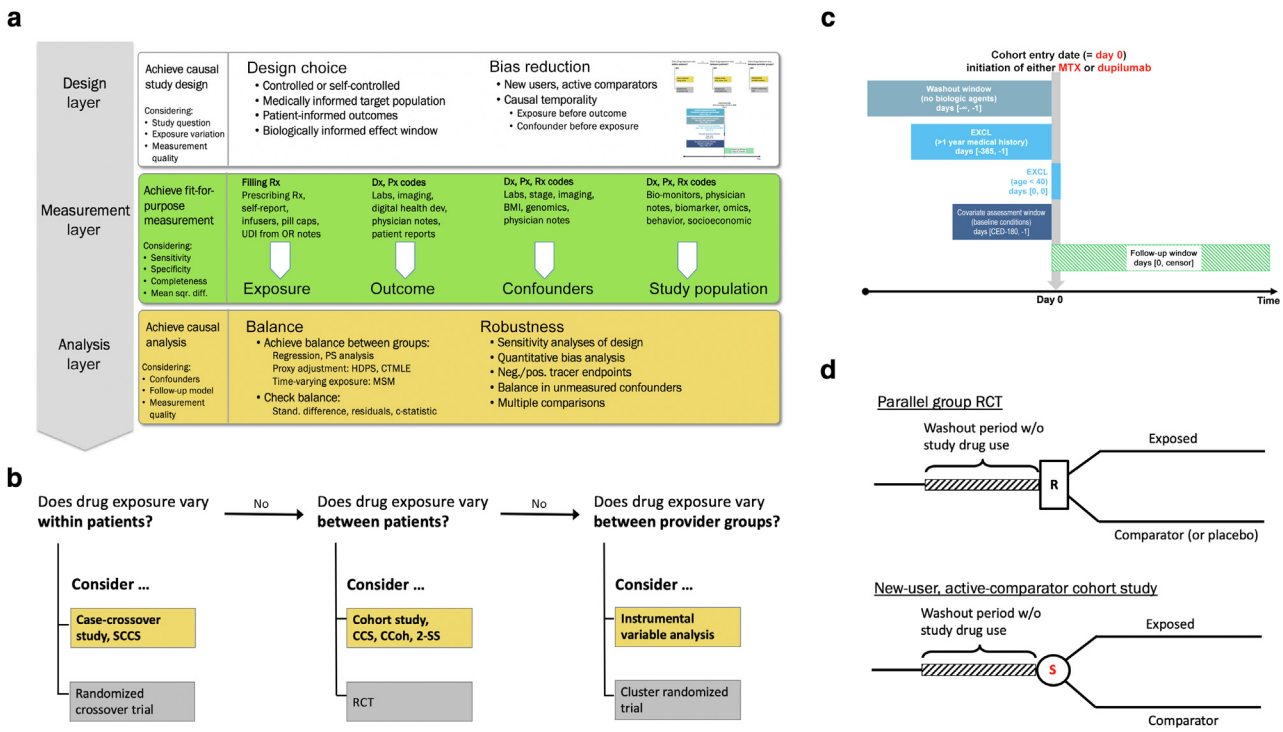
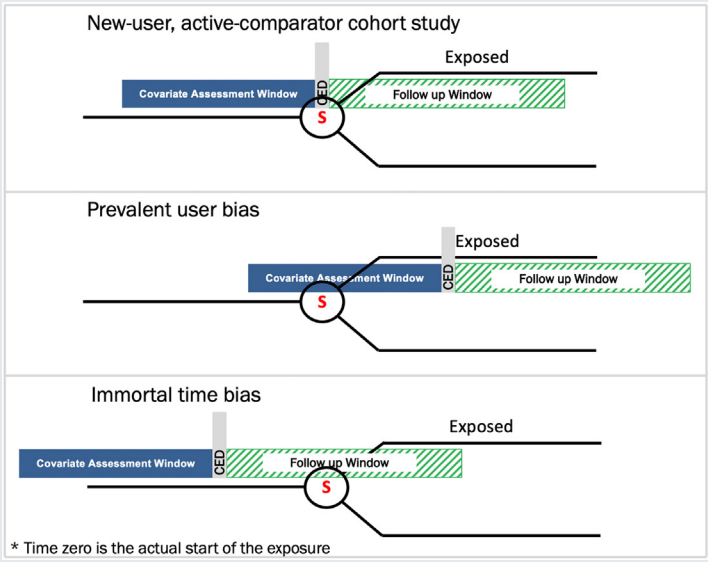


Figure 2. Conducting pharmacoepidemiology studies on treatment effects: study design and implementation. (**a**) A three-layered approach to turning longitudinal electronic healthcare data into a causal cohort study design. (**b**) The study question and sources of treatment variation guide study design choices. (**c**) Graphical illustration of longitudinal study design choices in pharmacoepidemiology. (**d**) Schematic of a parallel-group RCT and the corresponding cohort design emulating a (hypothetical) target RCT. This figure was adapted from Schneeweiss et al. (2021). For more information, see Schneeweiss (2010). Dx denotes disease, and Rx denotes treatment. 2-SS, two-stage sampling; CCoh, case-cohort sampling; CCS, case-control sampling; CTMLE, collaborative targeted maximum likelihood estimation; HDPS, high-dimensional propensity score; MSM, marginal structural model; Neg., negative; pos., positive; RCT, randomized controlled trial; SCCS, self-controlled case series.

Table 2. Inadvertently Introduced Biases

Term Used in This Article	Description
Biases caused by misaligning time zero: immortal time bias and prevalent user bias	 **Cohort entry date (CED) is aligned** with the start of follow-up at time zero,* the actual treatment start. Pre-treatment covariate assessment is in line with a causal study design (see **Figure 2D**) **CED is after time zero**, i.e. patients are assigned treatment status by considering ongoing (*prevalent*) treatment that started in the past. Selective dropout before CED will cause bias, particularly if a treatment effect changes with time, gets weaker or stronger with longer treatment. **CED is before time zero**, i.e. patients are assigned their treatment status by looking into the future. To determine future treatment status patients will need to be alive in the future, they are *immortal* for that period. If this is differential it causes bias. Remedy: Always align study follow-up with the beginning of treatment. Failing to do so does not automatically cause bias but increases its likelihood.
Adjustment for causal intermediates	Standard approaches to confounding adjustment involve measuring pretreatment predictors of the study outcome and adjusting for those variables in a regression model or other statistical models. However, adjusting for variables after treatment has started to run the risk of adjusting for consequences of the treatment, which may be precursors of the outcome of interest. Such adjustment for intermediates usually diminishes the true effect size on the outcome of interest. In a hypothetical example study on the risk of conjunctivitis among patients starting dupilumab, adjustment for the use of eye drops after the start of dupilumab, which may be indicative of soon-to-be-diagnosed conjunctivitis, would reduce the estimate of the true risk of conjunctivitis. Remedy: Adjust only for pre-exposure covariates. More complex methods are available for time-varying exposures.
Reverse causation	A bias occurring when a causal chronology, that is, first treatment, then effect, is not properly considered, and the treatment is falsely identified as a consequence of the condition and not the reverse. A hypothetical example study that would record the occurrence of psoriasis and the use of ustekinumab at any point during a calendar year may find a correlation between the two. However, it is not that ustekinumab is causing psoriasis (= reverse causation), but instead that ustekinumab is used in response to psoriasis. Remedy: Ensure proper consideration of the treatment–outcome chronology.

placebo-controlled trials often suffer from strong treatment selection bias. Persons receiving treatment differ from those not receiving it in ways that are difficult to completely measure and control analytically. Such strong confounding also occurs when comparing two different treatment modalities, for example, systemic treatment versus topical treatment or phototherapy (Schneeweiss et al., 2021b).

Dealing with treatment sequencing for chronic skin conditions

A complication to the new-user cohort design occurs when guideline-recommended treatment escalation correlates with severity of the condition. For example, in chronic inflammatory skin diseases, a patient may start out with a topical agent, escalate to a first-line systemic agent, and finally use a second-line targeted biologic agent as the condition progresses. In a study of systemic biologic agents, this means that very few or no patients will be treatment naïve when starting biologic treatment. There are two practical solutions. First, the population of interest are subjects with a common treatment and, upon consultation with a physician, switch to a biologic or an alternative treatment. This design compares switchers with other switchers, a clinically relevant question when deciding how to best escalate treatment (Schneeweiss, 2010). Second, one may seek to compare the outcomes of escalating treatment with those of staying on the current treatment. Although the prevalent new-user design provides a framework for that (Suissa et al., 2017), it should be kept in mind that in clinical practice, there are often good reasons why some patients escalate and others do not, which raises questions of comparability.

LAYER 2: MEASUREMENT CONSIDERATIONS WHEN WORKING WITH SECONDARY HEALTHCARE DATA

Patient data need to be fit for purpose in a given study to support causal conclusions on treatment effects. This comes down to the quality of measuring treatment, outcomes, and confounding factors, as listed in Figure 2a (Lash et al., 2014). It is fundamental to understand that the advantage of primary data collection, costly and time consuming as it is, is that the investigator controls what to measure, how to measure, and when to measure. In working with secondary data, we trade that level of control for much larger and more representative

data and therefore must concern ourselves with data completeness, accuracy, and timing of measurement.

Identifying the study population

Clear identification of the study population is important in assessing the generalizability of findings. For example, in RWE on the treatment of AD, a suitable population may be identified by diagnostic information (Abuabara et al., 2017; Lee et al., 2021) or in combination with starting with a diagnosis-specific treatment, for example, dupilumab (Iyer et al., 2023). Exclusions follow on the basis of the desired age range, the absence or presence of certain diagnoses, and the presence of markers of severity of the AD. Typical markers of interest are age at onset, body surface area involved, systemic corticosteroid use, number of topical agent prescriptions, and prior treatments with nonbiologic immunomodulators, to name a few recurrent themes. If those markers are vital to the interpretation of findings, one would identify a data source that captures them.

Treatment

For pharmacoepidemiologic studies, it is fundamental to record the start and end of the treatment of interest. Electronic pharmacy dispensing records avoid the limitations of patient recall (Table 1) and are considered largely accurate in recording the start of a drug exposure. Pharmacists fill prescriptions with little room for interpretation and are reimbursed by insurers on the basis of detailed, complete, and accurate electronically submitted claims (West et al., 1995). Specific use patterns may still vary substantially depending on the information better recorded in physician notes, particularly where topical as-needed medications and over-the-counter medications are concerned.

Outcomes

Clinically meaningful adverse events, such as serious infections, rapid-onset alopecia, or cancer, are more accurately recorded in secondary data than improvements in skin condition, an important reason why pharmacoepidemiology is more reliable when studying unintended effects. It should be noted that not all claims databases allow multiyear follow-up because enrollees may lose or switch insurance plans and can be lost to further follow-up (Table 1).

Because claims data often lack detailed clinical information, researchers must consider the incomplete recording of study outcomes and the resulting misclassification bias. Generally, a lack of specificity of the outcome measurement is worse than a lack of sensitivity. A relative risk estimate is unbiased by outcome misclassification if the specificity of the outcome assessment is 100%, even if the sensitivity is substantially lower, as long as the misclassification is nondifferential (Rothman and Poole, 1988). Although this approach increases study accuracy, it may also result in fewer events and thus lower statistical power. Confronted with this validity—precision trade off, it is recommended to choose the more accurate approach even if it results in less precise findings.

As an example, hospital discharge diagnoses, such as serious infections, have a high positive predictive value for capturing serious infections or venous thromboembolism, two endpoints of concern in treatments with immunomodulating drugs (Ammann et al., 2018; Sahli et al., 2016; Schneeweiss et al., 2007b).

Confounding factors

Confounding factors are risk factors for an outcome that are imbalanced between the treatment groups. Baseline risk factors are measured before treatment initiation to avoid adjusting for causal mediators (VanderWeele, 2019). In longer-term follow-up studies, some relevant outcome predictors change with time, such as the intermittent use of steroid tapers, which can be addressed with structural regression methods (Robins et al., 2000). Some challenges in healthcare databases include the complete and accurate measurement of important outcome predictors. Misclassified or unobserved confounder information leads to bias by residual confounding, which is addressed in the following section on data analysis.

Missing data

Missing data are issues that cut across all aspects of measurement discussed in the previous sections. If critical data items were not recorded at all or recorded with substantial missingness or misclassification, the data will not be fit for purpose. Studies similar to the validation studies quoted earlier may quantify the amount of missingness or misclassification, on the basis of which one can make an assessment of whether a study can still be performed and whether data imputation strategies should be applied (Lash et al., 2014; Schneeweiss, 2006).

LAYER 3: DATA ANALYSIS

Before starting an analysis, investigators should be precise about the effect they are trying to estimate. There is a range of causal parameters that can be estimated in longitudinal studies. In this paper, we focus on those most relevant for pharmacoepidemiology studies in dermatology.

Causal effect of interest

The as-treated effect. This is the effect of initiating the study treatment and continuing to receive it. Patients' follow-up time is censored at discontinuation of the initial treatment. The numerical value of the as-treated effect from a given study takes into account the duration of treatment persistence. In most situations, the as-treated effect is of great interest to both patients and physicians because it informs on the expected treatment effect while the patient is actually being treated.

The effect of complex treatment strategies. In many chronic conditions, such as AD, PSO, and hidradenitis suppurativa, it is recommended to start and switch therapy, change dose, or even pause treatment depending on treatment results. One may therefore be interested in estimating the effect of a longitudinal treatment strategy instead of analyzing the effect of treatment with a single drug.

The as-started effect. This is the effect of the initial treatment choice, regardless of whether that treatment continued over a given period of time, similar to the intention-to-treat effect in RCTs. The magnitude of the as-started effect from a given pharmacoepidemiology study depends on the specific patterns of deviation from the initial treatment choice

Table 3. Some Technical Terms Used in Etiologic Pharmacoepidemiology Studies

Term Used in This Article	Description
95% confidence interval	The range of possible effect estimates due to chance. The 95% CI of a risk ratio means that if the study were repeated 100 times, 95% of the resulting risk ratio estimates would be within that range.
Bias	Distortion of the observed finding from the true causal treatment effect in the study population. Three sources of bias are generally considered: (i) chance, typically quantified in 95% CIs; (ii) information bias caused by misclassification of exposure or outcome; and (iii) confounding bias (see below).
Confounding	A systematic bias that occurs in comparative effectiveness research if risk factors for the outcome of interest are imbalanced between the treatment groups being compared.
Generalizability	The extent to which the results from the study sample apply to a target population that may be composed differently.
Noninterventional study	An approach that does not interfere with the choice of treatment or measurement of study variables. This is in contrast to experimental studies, in which the treatment is assigned, and measurements are actively performed on patients. Synonyms include nonexperimental study and observational study.
Rate	The fraction of people who experience an event over an observed person time. The rate of newly occurring melanoma is X per 1,000 person years.
Rate ratio (RR)	The ratio of two rates. By convention, the rate in the exposed group is divided by the rate in the comparison or referent group. The rate ratio is unitless. A synonym is incidence rate ratio. A rate ratio can be estimated as a hazard ratio (HR), a reasonable approximation in most practical situations. In case-control sampling designs, odds ratios (OR) are estimates of the rate ratio.
Risk	The proportion (or percentage) of subjects who develop the outcome of interest over a defined period of time. The 10-year risk of developing melanoma is X%. A synonym is cumulative incidence.
Risk ratio (RR)	The ratio of two risks. By convention, the risk in the exposed group is divided by the risk in the comparison or referent group. The risk ratio is unitless. Synonyms are relative risk or cumulative incidence ratio.
Target trial	The ideal yet hypothetical randomized trial that would best answer the research question at hand. By trying to emulate this hypothetical target trial as best as possible, using nonrandomized data from clinical practice, investigators (i) avoid frequent inadvertently introduced biases such as immortal time bias and mediator adjustment (Table 2) and (ii) recognize study limitations, including data-related biases such as measurement error of the exposure and misclassification of the outcome, thus becoming able to discuss them using the target trial as a benchmark.
Time-related biases	Inadvertently introduced biases (other than the biases mentioned earlier) created by misaligning the start of a cohort experience (time 0): (i) prevalent user bias happens when time 0 is set sometime after the treatment started; (ii) immortal-time bias happens when time 0 is set sometime before the study treatment is determined.
Validity	The extent to which a study estimates what it set out to estimate. Two types are distinguished: (i) internal validity of a study, which means the absence of bias in the analysis of the actual study population, and (ii) external validity of a study, which means the study finding is generalizable to the population it was intended to apply to.

The explanations are meant as a primer for practitioners. Further reading is recommended.

during follow-up. As patients discontinue treatment, their exposure status will still be categorized according to the initial treatment. This handling avoids issues arising from informative censoring but leads to exposure-misclassified person time.

Several key terms relevant to the analysis of causal treatment effects are explained in Table 3.

Channeling of treatment and the resulting confounding of causal treatment effects

Physicians choose treatments in light of disease severity and information prognostic for the outcome that is available at the time of prescribing. The factors influencing the treatment decision vary by physician and with time and frequently involve clinical, functional, or behavioral characteristics of patients that may not be completely recorded in healthcare databases. Failure to balance those prognostic factors results in confounding bias. Because treatment selection according to disease severity and prognosis is an integral part of practicing medicine, the resulting bias can be strong and needs to be thoroughly addressed.

Such patient-level factors are not to be confused with physician-prescribing preference, which occurs if the same patient sees two different physicians who prescribe different treatments. Such variation is independent of patient predictors of the outcome and does typically not lead to confounding bias. In fact, it can be utilized in instrumental variable analyses, a technique that reduces the confounding caused by observed and unobserved factors (Brookhart et al., 2006b).

Analyzing comparable patients

Restriction to similar patients. In the absence of random treatment assignment, a key challenge is to achieve a balance of risk factors for the outcome between treatment groups. Restriction is a common and effective analytic tool to make treatment groups more comparable and thereby reduce residual confounding. Some restrictions are obvious because they are made by explicit criteria, such as limiting the study population to adults aged 18–30 years to study the safety of systemic biologic medications in young adults and excluding those with concomitant rheumatic conditions to not confuse the medications indication (Schneeweiss et al., 2020b). Other restrictions, such as matching on a confounder summary score, either a PS, or a disease risk score, are frequently used in pharmacoepidemiology. Restrictions reduce bias and

have implications for the generalizability of findings such as many trials (Schneeweiss et al., 2007a).

Propensity score (PS) analyses. PSs are multivariable balancing tools that can efficiently balance large numbers of covariates, even if low study event counts would impede outcome regression models. PS analyses have emerged as an effective tool for adjusting large numbers of potential confounders when using data from clinical practice to best reduce confounding. They fit the target trial paradigm because the PS emulates the randomization process using the observed data. In a new-user cohort study, the PS is the estimated probability of starting treatment A versus starting treatment B, conditional on all observed pretreatment patient characteristics. The PS is usually estimated with logistic regression, and strategies for variable selection are well-described (Brookhart et al., 2006a). Once a PS is estimated on the basis of all measured covariates, the actual adjustment is done in a second step. Typical strategies include adjustment for quintiles of the PS, matching, or regression weighting by PS (Webster-Clark et al., 2021).

Matching on PS in a cohort study has practical advantages that may outweigh the theoretical drawback of not using the full dataset when not all eligible patients match. Matching excludes patients with extreme PS values (close to 100% or close to 0% likelihood to receive treatment), where there is little clinical ambivalence in treatment choice. Trimming such patients from the analysis reduces residual confounding and informs investigators on the treatment effects among those patients where treatment seems interchangeable (Walker et al., 2013). In contrast to regression outcome models, fixed-ratio PS-matched analyses such as the frequently used 1:1 matching allow the investigator to demonstrate the balance achieved in each of the potential confounders. Standardized differences of covariates have gained popularity in PS-matching analyses (Franklin et al., 2014). Fixed-ratio matching in cohort studies does not require matched analyses to obtain unbiased results (Rothman, 2008). In settings with very few events, fine stratification by PS may be preferred (Desai et al., 2017).

Any patient information measured before the start of treatment can be considered a potential confounding factor. Because the measurement of these factors is not in the investigator's control when using secondary data, confounding from unobservable factors (e.g., severity of AD) can be reduced by observing and adjusting for measurable proxies of the underlying confounders (e.g., topical treatment vs. biologic treatment). The unobserved confounders are adjusted to the amount such proxy measurements are correlated with the confounding factor of interest (Gelman et al., 2013; Wooldridge, 2002). Examples of well-measured proxies are the use of oxygen canisters (correlated with frail health), regular use of preventative services (correlated with health-seeking behavior), systemic glucocorticoid use (correlated with disease exacerbation), etc. This produces high-dimensional covariate spaces with several thousand pretreatment covariates, which can be reduced to those that are likely confounders by established variable-reduction techniques and finally used for confounding control through a PS model (Karim et al., 2018; Schneeweiss et al., 2017, 2009b). The performance of the resulting high-dimensional PS adjustment is equal to and often superior to investigator-specified covariates in terms of bias reduction across a range of research questions and versatile in a variety of data sources and coding systems (Schneeweiss, 2018; Zhou et al., 2017).

Subgroup analyses and treatment effect modification

Large healthcare databases make it possible to stratify an analysis by multiple clinical factors that are relevant to both prescribers and patients, for example, a study examining the risk of myocardial infarction in patients with PSO stratified by mild PSO versus severe PSO (Gelfand et al., 2006). Another example would be including an asthma subgroup or recent oral steroid use subgroup when studying a drug used by patients with AD (Schneeweiss et al., 2021a). General recommendations for studying heterogeneous treatment effects apply to RWE as well as to RCTs (Lesko et al., 2018; Segal et al., 2023). A particular concern remains posthoc screening for effect modification, which may produce false-positive findings despite fairly conservative statistical tests for interaction. Signals of effect modification should be confirmed in subsequent studies using other data sources and require close coordination between clinical science and statistical modeling (Zhang et al., 2015).

Sensitivity analyses

Sensitivity analyses help investigators to better understand the robustness of study findings. Important sensitivity analyses constitute variations of key study design parameters that should be preplanned in the study protocol, including their justification, and other tests, whether the study results may change if misclassification or confounding would have remained uncontrolled (Schneeweiss and Avorn, 2005). An important tool for detecting the impact of unobserved confounding on the validity of findings is quantitative bias analysis. Bias analyses of residual confounding seek to determine how strong and how imbalanced an unobserved confounding factor would have to be between exposure groups to explain the observed effect (Patorno et al., 2018; Schneeweiss, 2006). Lash et al. (2014) proposed a comprehensive approach that considers several systematic errors simultaneously, combining sensitivity analyses for confounding, misclassification, and selection bias in one process.

Sometimes there is uncertainty about the correct length of the exposure risk window on the basis of the pharmacology of the study agent and the underlying biology. For example, when the potential effect of tacrolimus on the risk of cancer was under study, there was a debate about whether tacrolimus was a tumor inducer, which would require a very long exposure risk window until the cancer may become clinically apparent, versus a tumor promoter, which could have a shorter risk window (Schneeweiss et al., 2009a). Varying the length of the exposure risk window is recommended to understand the robustness of the findings and is easy to accomplish in cohort studies (Solomon et al., 2008).

Another sensitivity analysis concerns the potential for informative censoring. Patients switch or end treatments because they did not have the intended effect or they experienced signs of a side effect. The stronger such treatment discontinuation is associated with the occurrence of the

outcome, the more strongly biased an as-treated analysis is. Inverse probability of censoring weighting is an analytic tool to correct for bias (Toh et al., 2012). However, it requires that patient characteristics predict the treatment discontinuation reasonably well, which is not often the case.

CONCLUSION

Conducting etiologic pharmacoepidemiology studies follows a structured approach, which we explained in this paper: (i) a design layer connects the research question with the appropriate study design, and contemplating the ideal RCT reduces investigator errors; (ii) a measurement layer transforms longitudinal patient-level data into variables that identify the study population, pre-exposure patient characteristics, treatment status, and outcomes; and (iii) the analysis focuses on estimation of the causal treatment effect.

All studies on the effectiveness and safety of treatments are based on an understanding of the biological nature and medical practice represented in an abstract model, that is, a study design with the necessary measurements, and a statistical model for data analysis. As in a randomized trial, this requires many assumptions and simplifications of a complex world. Producers of evidence in medicine, experimental or otherwise, and decision makers who wish to improve the practice of medicine should be aware of the resulting uncertainty as they plan actions.

ORCIDs

Sebastian Schneeweiss: http://orcid.org/0000-0003-2575-467X
Maria Schneeweiss: http://orcid.org/0000-0002-4304-5271

CONFLICT OF INTEREST

SS is participating in investigator-initiated grants to the Brigham and Women's Hospital from Boehringer Ingelheim and UCB, unrelated to the topic of this study. He is a consultant to Aetion, a software manufacturer, in which he owns equity. He is an advisor to Temedica GmbH, a patient-oriented data generation company. He is a principal investigator of the United States Food and Drug Administration Sentinel Innovation Center. His interests were declared, reviewed, and approved by the Brigham and Women's Hospital in accordance with their institutional compliance policies. MS is participating in investigator-initiated grants to the Brigham and Women's Hospital from UCB, unrelated to the topic of this study. Her interests were declared, reviewed, and approved by the Brigham and Women's Hospital in accordance with their institutional compliance policies.

ACKNOWLEDGMENTS

This work was funded by grant R01-AR080194 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases and the Division of Pharmacoepidemiology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School (Boston, MA).

REFERENCES

Abuabara K, Magyari AM, Hoffstad O, Jabbar-Lopez ZK, Smeeth L, Williams HC, et al. Development and validation of an algorithm to accurately identify atopic eczema patients in primary care electronic health records from the UK. J Invest Dermatol 2017;137:1655–62.

Ammann EM, Cuker A, Carnahan RM, Perepu US, Winiecki SK, Schweizer ML, et al. Chart validation of inpatient International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) administrative diagnosis codes for venous thromboembolism (VTE) among intravenous immune globulin (IGIV) users in the Sentinel Distributed Database. Medicine (Baltimore) 2018;97:e9960.

Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol 2006a;163:1149–56.

Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology 2006b;17:268–75.

Desai RJ, Rothman KJ, Bateman BT, Hernandez-Diaz S, Huybrechts KF. A propensity-score-based fine stratification approach for confounding adjustment when exposure is infrequent. Epidemiology 2017;28:249–57.

Dommasch ED, Kim SC, Lee MP, Gagne JJ. Risk of serious infection in patients receiving systemic medications for the treatment of psoriasis [published correction appears in JAMA Dermatol 2019;155:865. JAMA Dermatol 2019;155:1142–52.

Eichler HG, Pignatti F, Schwarzer-Daum B, Hidalgo-Simon A, Eichler I, Arlett P, et al. Randomized controlled trials versus real world evidence: neither magic nor myth. Clin Pharmacol Ther 2021;109:1212–8.

Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K, et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative. Circulation 2021;143:1002–13.

Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. Stat Med 2014;33:1685–99.

Gelfand JM, Neimann AL, Shin DB, Wang X, Margolis DJ, Troxel AB. Risk of myocardial infarction in patients with psoriasis. JAMA 2006;296:1735–41.

Gelman A, Carlin J, Stern H, Dunson DB, Vehtari A, Rubin D. Bayesian data analysis. New York, NY: Chapman and Hall/CRC; 2013.

Gokhale M, Stürmer T, Buse JB. Real-world evidence: the devil is in the detail. Diabetologia 2020;63:1694–705.

Grodstein F, Stampfer MJ, Manson JE, Colditz GA, Willett WC, Rosner B, et al. Postmenopausal estrogen and progestin use and the risk of cardiovascular disease [published correction appears in N Engl J Med 1996;335:1406]. N Engl J Med 1996;335:453–61.

Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology 2008;19:766–79.

Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016;183:758–64.

Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol 2016;79:70–5.

Iyer S, Patel N, Sanfilippo E, Dellavalle RP, Silverberg JI. Assessment of the validity of international classification of disease tenth revision codes for atopic dermatitis. Arch Dermatol Res 2023;315:879–84.

Johnson ES, Bartman BA, Briesacher BA, Fleming NS, Gerhard T, Kornegay CJ, et al. The incident user design in comparative effectiveness research. Pharmacoepidemiol Drug Saf 2013;22:1–6.

Karim ME, Pang M, Platt RW. Can we train machine learning methods to outperform the high-dimensional propensity score algorithm? Epidemiology 2018;29:191–8.

Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. Int J Epidemiol 2014;43:1969–85.

Leducq S, Caille A, Le Cleach L, Samimi M, Tavernier E, Maruani A, et al. Research techniques made simple: randomized controlled trials for topical drugs in dermatology: when and how should we use a within-person design? J Invest Dermatol 2020;140:931–8.e1.

Lee H, He M, Cho SK, Bessette L, Tong AY, Merola JF, et al. Validation of claims-based algorithms to identify patients with psoriasis. Pharmacoepidemiol Drug Saf 2021;30:868–74.

Lesko CR, Henderson NC, Varadhan R. Considerations when assessing heterogeneity of treatment effect in patient-centered outcomes research. J Clin Epidemiol 2018;100:22–31.

Lin KJ, Schneeweiss S. Considerations on the analysis of longitudinal electronic heath records linked to claims data to study the effectiveness and safety of drugs. Clin Pharmacol Ther 2016;100:147–59.

Maclure M. 'Why me?' versus 'why now?'–differences between operational hypotheses in case-control versus case-crossover studies. Pharmacoepidemiol Drug Saf 2007;16:850–3.

Malone DC, Brown M, Hurwitz JT, Peters L, Graff JS. Real-world evidence: useful in the real world of US payer decision making? How? When? And What Studies? Value Health 2018;21:326–33.

Noe MH, Gelfand JM. Research techniques made simple: pharmacoepidemiology research methods in dermatology. J Invest Dermatol 2018;138:e13–8.

Patorno E, Gopalakrishnan C, Franklin JM, Brodovicz KG, Masso-Gonzalez E, Bartels DB, et al. Claims-based studies of oral glucose-lowering medications can achieve balance in critical clinical variables only observed in electronic health records. Diabetes Obes Metab 2018;20:974–84.

Patorno E, Schneeweiss S, Gopalakrishnan C, Martin D, Franklin JM. Using real-world data to predict findings of an ongoing phase IV cardiovascular outcome trial: cardiovascular safety of linagliptin versus glimepiride. Diabetes Care 2019;42:2204–10.

Penso L, Bergqvist C, Meyer A, Herlemont P, Weill A, Zureik M, et al. Risk of inflammatory bowel disease in patients with psoriasis and psoriatic arthritis/ankylosing spondylitis initiating interleukin-17 inhibitors: a nationwide population-based study using the French national health data system. Arthritis Rheumatol 2022;74:244–52.

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

Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550–60.

Rothman KJ. Modern epidemiology. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.

Rothman KJ, Poole C. A strengthening programme for weak associations. Int J Epidemiol 1988;17:955–9.

Sahli L, Lapeyre-Mestre M, Derumeaux H, Moulis G. Positive predictive values of selected hospital discharge diagnoses to identify infections responsible for hospitalization in the French national hospital database. Pharmacoepidemiol Drug Saf 2016;25:785–9.

Schneeweiss MC, Huang JT, Wyss R, Schneeweiss S, Merola JF. Serious infection risk in children with psoriasis on systemic treatment: a propensity score-matched population-based study. J Am Acad Dermatol 2020a;82:1337–45.

Schneeweiss MC, Kim SC, Wyss R, Schneeweiss S, Merola JF. Dupilumab and the risk of conjunctivitis and serious infection in patients with atopic dermatitis: a propensity score-matched cohort study. J Am Acad Dermatol 2021a;84:300–11.

Schneeweiss MC, Perez Chada L, Gottlieb AB, Merola JF. Older adults on systemic treatment for psoriasis and risk of infection: a propensity score-matched population-based study. Br J Dermatol 2020b;183:564–6.

Schneeweiss MC, Perez-Chada L, Merola JF. Comparative safety of systemic immunomodulatory medications in adults with atopic dermatitis. J Am Acad Dermatol 2021b;85:321–9.

Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf 2006;15:291–303.

Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol Drug Saf 2010;19:858–68.

Schneeweiss S. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. Clin Epidemiol 2018;10:771–88.

Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;58:323–37.

Schneeweiss S, Doherty M, Zhu S, Funch D, Schlienger RG, Fernandez-Vidaurre C, et al. Topical treatments with pimecrolimus, tacrolimus and medium- to high-potency corticosteroids, and risk of lymphoma. Dermatology 2009a;219:7–21.

Schneeweiss S, Eddings W, Glynn RJ, Patorno E, Rassen J, Franklin JM. Variable selection for confounding adjustment in high-dimensional covariate spaces when analyzing healthcare databases. Epidemiology 2017;28:237–48.

Schneeweiss S, Patorno E. Conducting real-world evidence studies on the clinical outcomes of diabetes treatments [published correction appears in Endocr Rev 2021;42:873. Endocr Rev 2021;42:658–90.

Schneeweiss S, Patrick AR, Stürmer T, Brookhart MA, Avorn J, Maclure M, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. Med Care 2007a;45:S131–42.

Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data [published correction appears in Epidemiology 2018;29:e63–4]. Epidemiology 2009b;20:512–22.

Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. J Clin Epidemiol 2007b;60:397–409.

Segal JB, Varadhan R, Groenwold RHH, Li X, Nomura K, Kaplan S, et al. Assessing heterogeneity of treatment effect in real-world data [published correction appears in Ann Intern Med 2023;176:882]. Ann Intern Med 2023;176:536–44.

Seyger MMB, Augustin M, Sticherling M, Bachhuber T, Fang J, Hetherington J, et al. Physician-reported clinical unmet needs, burden and treatment patterns of paediatric psoriasis patients: a US and EU real-world evidence study. Acta Derm Venereol 2022;102:adv00660.

Solomon DH, Lunt M, Schneeweiss S. The risk of infection associated with tumor necrosis factor alpha antagonists: making sense of epidemiologic evidence. Arthritis Rheum 2008;58:919–28.

Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. Pharmacoepidemiol Drug Saf 2017;26:459–68.

Toh S, Reichman ME, Houstoun M, Ross Southworth M, Ding X, Hernandez AF, et al. Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system [published correction appears in Arch Intern Med 2013;173:14. Arch Intern Med 2012;172:1582–9.

U.S. Food & Drug Administration. Framework for FDA's Real-World Evidence Program. 2018. https://www.fda.gov/media/120060/download. accessed September 12, 2023.

Van Beek MJ, Swerlick RA, Mathes B, Hruza GJ, Resneck J Jr, Pak HS, et al. The 2020 annual report of DataDerm: the database of the American Academy of Dermatology. J Am Acad Dermatol 2021;84:1037–41.

van de Kerkhof PC, Reich K, Kavanaugh A, Bachelez H, Barker J, Girolomoni G, et al. Physician perspectives in the management of psoriasis and psoriatic arthritis: results from the population-based multinational assessment of Psoriasis and Psoriatic Arthritis survey. J Eur Acad Dermatol Venereol 2015;29:2002–10.

VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol 2019;34:211–9.

Walker AM, Patrick A, Lauer M, Hornbrook M, Marin M, Platt R, et al. A Tool for assessing the feasibility of comparative effectiveness research. Comp Eff Res 2013;3:11–20.

Wang SV, Schneeweiss S, , RCT-DUPLICATE Initiative, Franklin JM, Desai RJ, Feldman W, et al. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. JAMA 2023;329:1376–85.

Wang SV, Sreedhara SK, Schneeweiss S, REPEAT Initiative. Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. Nat Commun 2022;13:5126.

Webster-Clark M, Stürmer T, Wang T, Man K, Marinac-Dabic D, Rothman KJ, et al. Using propensity scores to estimate effects of treatment initiation decisions: state of the science. Stat Med 2021;40:1718–35.

West SL, Savitz DA, Koch G, Strom BL, Guess HA, Hartzema A. Recall accuracy for prescription medications: self-report compared with database information. Am J Epidemiol 1995;142:1103–12.

Wooldridge JM. Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press; 2002. p. 752.

Yiu ZZN, Mason KJ, Hampton PJ, Reynolds NJ, Smith CH, Lunt M, et al. Randomized trial replication using observational data for comparative effectiveness of secukinumab and ustekinumab in psoriasis: a study from the British Association of Dermatologists biologics and immunomodulators register. JAMA Dermatol 2021;157:66–73.

Zhang Y, Laber EB, Tsiatis A, Davidian M. Using decision lists to construct interpretable and parsimonious treatment regimes. Biometrics 2015;71:895–904.

Zhou M, Wang SV, Leonard CE, Gagne JJ, Fuller C, Hampp C, et al. Sentinel modular program for propensity score-matched cohort analyses: application to glyburide, glipizide, and serious hypoglycemia. Epidemiology 2017;28:838–46.