Methodological Challenges in the Identification of Drug-Drug Interactions Using Spontaneous Reporting System

Chen Yang Bai

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for Master's Degree in Epidemiology

School of Epidemiology and Public Health

Faculty of Medicine

University of Ottawa

Supervisor: Christopher Gravel

Thesis Advisory Committee Member: Melissa Kampman

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Christopher Gravel, for his unwavering support, guidance, and expertise throughout my Master's journey. His encouragement and critical insights have been invaluable to the successful completion of this thesis. I am also sincerely grateful to my co-supervisor, Dr. Antonios Douros, for his thoughtful feedback and mentorship. I extend my appreciation to my Thesis Advisory Committee member, Dr. Melissa Kampmann, for her time, contributions, and insightful comments.

Special thanks to my colleagues and friends for their encouragement and moral support during challenging times.

Lastly, I thank my family — especially my mother, Ying Pei — for their love, patience, and constant support throughout my academic journey.

Table of Contents

Abstract

Introduction: It has been shown that reporting bias can distort estimates of disproportionate reporting in the single drug setting, however, their influence is unclear when screening spontaneous reporting databases for drug-drug interactions (DDI). Antiviral medications were repurposed to treat COVID-19 during the pandemic period which may have introduced reporting bias. Its impact on DDI signal detection, particularly when using restricted comparator designs for systematic bias mitigation, remains unclear. To investigate this potential we conducted a retrospective bias analysis on a well known DDI.

Methods: We used data from the United States Food and Drug Administration Adverse Event Reporting System (2000Q3–2023Q3) to evaluate changes in disproportionality estimates for lopinavir/ritonavir and atorvastatin with myopathy/rhabdomyolysis. We computed signals using three methods: Concomitant Signal Score (CSS), Omega Shrinkage, and the extended-Bayesian Confidence Propagation Neural Network (BCPNN). Comparisons were conducted across unrestricted and active comparator reference sets (all statins and CYP3A4 statins), pre- and during-pandemic, change in estimates (ACiE) was calculated to quantify differences.

Results: In the unrestricted comparator design, Omega and BCPNN estimates decreased during-pandemic when including Paxlovid (ACiE: –0.37 and –0.24, respectively), potentially by increased background reporting, but increased when excluding Paxlovid (ACiE: 0.84 and 0.16, respectively). In contrast, signal strength increased in all active comparator analyses, particularly with CYP3A4 statins (Omega ACiE: 1.05; BCPNN ACiE: 0.31). CSS results showed a similar trend.

Conclusion: Changes in antiviral indications in response to the COVID-19 pandemic may have altered reporting patterns affecting DDI signal detection.

List of Tables

List of Figures

List of Abbreviations

ADR – Adverse Drug Reaction

ACD – Active Comparator Design

BCPNN – Bayesian Confidence Propagation Neural Network

CK – Creatine Kinase

COVID-19 – Coronavirus Disease 2019

CSS – Concomitant Signal Score

CYP3A – Cytochrome P450 3A Enzyme

CVP – Canada Vigilance Program

DEC – Drug–Event Combination

DDI – Drug–Drug Interaction

DPA – Disproportionality Analysis

EBGM – Empirical Bayes Geometric Mean

EMA – European Medicines Agency

FAERS – FDA Adverse Event Reporting System

FDA – U.S. Food and Drug Administration

HIV – Human Immunodeficiency Virus

IC – Information Component

ICSR – Individual Case Safety Report

IMI PROTECT – Innovative Medicines Initiative Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium

LDL – Low-Density Lipoprotein

MAP – Maximum A Posteriori Estimate

MedDRA – Medical Dictionary for Regulatory Activities

MGPS – Multi-Item Gamma Poisson Shrinker

PASS – Post-Authorization Safety Study

PBRER – Periodic Benefit-Risk Evaluation Report

PD – Pharmacodynamic

PK – Pharmacokinetic

PRR – Proportional Reporting Ratio

PSUR – Periodic Safety Update Report

PT – Preferred Term

RCT – Randomized Controlled Trial

REMS – Risk Evaluation and Mitigation Strategy

RMP – Risk Management Plan

ROR – Reporting Odds Ratio

SDRs – Signals of Disproportionate Reporting

SMQ – Standardized MedDRA Query

SR – Spontaneous Report

URD – Unresctricted Design

WHO – World Health Organization

Chapter I: Background and Introduction

1. Signal Detection in Spontaneous Reporting Systems

1.1 Pharmacovigilance and Drug Safety Evaluation

To demonstrate the safety and efficacy of medications, extensive pre-clinical research and multiple phases of clinical trials, ranging from Phase I to Phase III, are required as part of the evidence submitted to regulatory agencies for market authorization approval[1]. Among the various methods used to evaluate new medical interventions, the randomized controlled trial (RCT) has long been established as the gold standard for establishing drug efficacy[2]. In RCTs, participants are randomly assigned to either a treatment or a control group, which helps ensure the balance of both measured and unmeasured characteristics across study arms. Furthermore, RCTs often incorporate methodological features specifically designed to minimize bias such as blinding of participants and investigators, concealed allocation procedures, and measures to reduce loss to follow-up[3–5]. Together, these design elements enhance the internal validity of RCTs and strengthen the causal inferences that can be drawn regarding a drug's efficacy[6]. However, given that RCT designs are optimized for the estimation of efficacy parameters, evidence on the safety profile of authorized medications requires further post-market study and monitoring.

Despite the rigorous regulatory review of clinical data submitted from pivotal trials conducted under controlled conditions, certain adverse drug reactions (ADR) may only become detectable after a medication is introduced into real-world use[7,8]. In routine clinical practice, medications are prescribed to a much larger and more heterogeneous patient population than those typically enrolled in RCTs. This increased diversity in age, comorbidities, concomitant medications, and other factors often limits the generalizability of RCT findings to real-world populations and settings[2]. Moreover, the relatively short duration and limited sample sizes of many clinical trials may hinder the detection of rare ADRs or those with longer time to onset.

Importantly, RCTs are typically powered to detect efficacy outcomes and not the full spectrum of potential ADRs, which may number in the hundreds. As a result, they often lack the statistical power to detect uncommon or unanticipated ADRs. Consequently, pharmaceutical products are subject to continuous post-market surveillance to monitor their long-term safety and effectiveness in broader populations[9–12].

The "science and activities involved in detecting, assessing, understanding, and ultimately preventing adverse events associated with health products" are collectively referred to as pharmacovigilance by the World Health Organization (WHO)[13]. Following market authorization, pharmaceutical companies are required to engage in ongoing safety monitoring of their product[14–16]. In the European Union and Canada, manufacturers are required to develop and implement Risk Management Plans (RMPs), which outline how identified risks will be monitored and mitigated, often based on safety hypothesis emerging from clinical trial data[17,18]. In the United States, a similar framework exists in the form of Risk Evaluation and Mitigation Strategies (REMS)[19]. To evaluate the updated safety evidence of an approved medication, pharmaceutical manufacturers are required to submit Periodic Safety Update Reports (PSUR), which assess whether emerging safety data warrant further investigation. Lastly, Periodic Benefit-Risk Evaluation Reports (PBRERs) are also required, which provide a comprehensive evaluation of the medication's benefit-risk profile based on updated evidence[15,16,20].

When safety concerns arise, Post-Authorization Safety Studies (PASS), commonly referred to as Phase IV trials, may be voluntarily initiated by a drug manufacture, or mandated by a regulatory. [21–24]. These studies may be requested as part of a PSUR submission or initiated independently in response to emerging safety signals. They frequently take the form of

observational pharmacoepidemiologic investigations aimed at re-evaluating and updating the medication's safety profile in real-world clinical settings[24].

Despite regulatory requirements for the submission of PSURs and PBRERs, newly emerging ADRs may still go undetected. One contributing factor is the scheduled nature of PSURs, which can create temporal gaps in safety surveillance. The frequency of PSUR submission is determined at the time of market authorization and may vary depending on the product's risk profile. For instance, PSURs may be required every six months initially, then annually, and eventually every three years. During these intervals, important safety issues may arise and remain unassessed until the next reporting cycle[16].

A notable example of this limitation is the case of rosiglitazone, an antidiabetic medication that became the subject of major cardiovascular safety concerns. In 2007, Nissen and Wolski published a meta-analysis of clinical trials, suggesting that rosiglitazone was associated with a significant increase in the risk of myocardial infarction and cardiovascular death[25]. These findings, based on publicly available trial data, generated widespread concern and led to regulatory scrutiny. The RECORD trial, a large post-marketing randomized controlled trial sponsored by the manufacturer, was subsequently used to assess this risk[26,27]. This case illustrates how reliance solely on PSURs or sponsor-driven post-market studies can delay the detection and confirmation of serious ADRs. It also highlights the need of real-time safety analyses using publicly accessible data sources, for continuous, proactive monitoring, beyond the scheduled PSUR framework. This is essential to ensure timely identification of emerging safety concerns that may otherwise go unnoticed during the standard reporting intervals.

1.2 Spontaneous Self-Reported Adverse Reactions

Post-market spontaneous ADR reports are routinely collected by pharmaceutical companies and regulatory agencies and are readily available[15,28]. These spontaneous reports (SR), also known as Individual Case Safety Reports (ICSRs), are a foundational data source in pharmacovigilance[29–31]. ICSRs are self-reported cases of undesired and/or unintended effects suspected to be associated with the use of medication(s)[32–34]. They may be submitted by a range of sources, including drug manufacturers, healthcare professionals, and patients or other third parties[32–34].

Each ICSR may contain multiple reported medications and adverse reactions, with varying levels of suspected causality as indicated by the reporter. These reports are collected outside of a hypothesis-driven study design, as they are generated following the experience of an adverse event, either through voluntary submission or in response to mandatory reporting requirements. Unlike administrative or clinical datasets, which are typically generated during routine care, SR data offer a unique perspective by documenting direct user- or observer-reported experiences of medication use, including nuances that might not otherwise be recorded in structured healthcare records. This subjective element of reporter perception can be beneficial, as it may capture early signs of emerging ADRs or rare patient experiences that clinical coding systems might miss. However, it also introduces the potential for reporting errors, misattribution of causality, and inconsistencies in terminology or clinical detail. Reports may reflect the biases, misunderstandings, or assumptions of the reporter, leading to noise or misinformation in the dataset.

Regulatory requirements for submitting spontaneous ADR reports vary across countries and regulatory agencies. However, a common feature among most major regulatory frameworks

is that drug manufacturers are legally obligated to report ADRs as they become aware of a suspected adverse effect associated with one of their products[33,35,36]. In contrast, reporting by healthcare professionals and patients is generally encouraged but remains voluntary in many jurisdictions. A notable exception is Canada's post-market regulatory framework, where the Protecting Canadians from Unsafe Drugs Act, commonly known as Vanessa's Law, grants Health Canada the authority to mandate healthcare institutions to report serious ADRs[37]. Under this legislation, reporting of non-serious ADRs by healthcare professionals and all patient reports remains voluntary.

1.3 Spontaneous Reporting Databases

ICSRs are collected and compiled into large repositories by regulatory agencies. Notable databases includes Health Canada's Canada Vigilance Program (CVP), the U.S. Food and Drug Administration's (FDA) Adverse Event Reporting System (FAERS), and the European Medicines Agency's (EMA) EudraVigilance[35,36,38]. Routine analysis and identification of drug safety signals using these self-reported ADRs is commonly referred to as "passive pharmacovigilance", as the data are collected through voluntary, mostly unsolicited reporting rather than prospective systematic data collection methods[35,36,38].

While their routine collection supports monitoring of previously hypothesized ADRs from clinical trials as part of PSURs, the value of SR data also lies in their potential to enable early detection of emerging, previously undetected, drug safety signals from millions of potential drug-event-combinations (DEC), commonly referred to as signal detection[30]. This safety signal detection process is critical because ICSRs capture real-world, unsolicited experiences of ADRs that may not be fully characterized during pre-market clinical trials. However, the

methodological approaches for quantitative signal detection using SR data are not straightforward, and require the application of specific data mining and statistical approaches.

Key methodological limitations of using SR systems for safety outcome analysis lies in the nature of self-reported data collection[39]. The most prominent issue is underreporting. Studies have estimated that fewer than 10% of serious ADRs and only 2% to 4% of non-serious ADRs are reported to the UK's spontaneous reporting system[30]. Similarly, it has been suggested that the U.S. FDA directly receives reports for fewer than 1% of suspected serious ADRs, although the number of reports submitted to the FDA has steadily increased in recent years[30,35]. This underreporting means that the numerator in any analysis based on SR data, i.e., the number of reported ADR cases, represents only a small subset of the true number of events occurring in the population. Likewise, the denominator, i.e., the number of reported drug exposures, is also an incomplete reflection of total population-level drug use, and is impossible to estimate drug utilization rate from SR data. In addition to underreporting, SR systems suffer from outcome-dependent selection bias since the reporter's perception of the adverse health outcome they are experiencing motivates the reporting behaviour. By definition, a spontaneous report must contain information on this adverse event and suspected medication(s)[40]. Therefore, individuals who experience no adverse events are inherently excluded from the dataset. This creates a fundamental selection bias with outcome-dependent sampling, as reporting behavior is initiated only by the occurrence of an outcome, often influenced by the severity, notoriety, or the reporter's suspicion of causality[41]. As a result, the denominator, drug utilization, is also unknown and cannot be estimated, since reporting is conditional on experiencing and choosing to report the outcome. Yet, signal detection analyses frequently estimate the reporting probability of an adverse event given exposure to a drug, a formulation that mimics a prospective framework[42].

This analytical framing contrasts with the retrospective nature of data collection, where both the exposure and outcome have already occurred prior to reporting.

Taken together, these inherent limitations of self-reported data, particularly underreporting and outcome-dependent selection bias, mean that absolute incidence rates cannot be reliably estimated from SR systems. In addition, the lack of denominator data, missing covariates, and the presence of unmeasured confounding and biasing pathways preclude the use of formal causal inference models to definitively establish drug–event relationships[43,44]. As a result, signal detection in SR systems relies on relative "disproportionality" measures, which assess whether specific DECs are reported at a disproportionate rate than would be expected by under independence between drug and event. Thus, any safety signals identified from SR data should be viewed as generated hypotheses rather than confirmed ADRs as quantitative evidence of heightened reporting cannot be directly used to infer association. These preliminary findings often warrant further investigation using data sources better suited for causal inference and incidence estimation. In practice, regulatory authorities may attempt to verify suspected signals by initiating follow-up studies in large, healthcare databases, such as administrative claims or electronic health records, that enable cohort analysis of the association between drug exposures and subsequent health outcomes[43].

One of the key challenges in interpreting disproportionality-based safety signals is the presence of biases. While some biases are systematic, similar to those encountered in pharmacoepidemiologic studies (e.g., confounding or misclassification), spontaneous reporting systems are particularly prone to reporting biases. Reporting biases arise when external factors, such as media coverage, regulatory warnings, or public health events, artificially inflate or suppress the reporting rates of certain drugs or adverse events. These fluctuations can distort

disproportionality estimates, potentially resulting in false-positive or false-negative signals, and ultimately compromising the reliability of signal detection.

Among the key systematic biases of concern in SR data is confounding by indication, which occurs when the underlying condition for which a drug is prescribed is itself associated with the adverse event of interest[45]. For example, evaluating cardiovascular outcomes in patients using antidiabetic medications against all other drugs can be misleading, as individuals with diabetes inherently carry a higher baseline cardiovascular risk compared to the general population[46]. A similar bias is channeling bias, in which drugs within the same therapeutic class are prescribed to patient subgroups with different prognostic profiles, thereby introducing systematic differences unrelated to the drug's effect[47].

SR data is also subject to reporting biases, which can distort the apparent reporting relationship between drugs and adverse events. One of the well-documented reporting bias is *notoriety bias*, which refers to the surge in reporting of a particular DEC following external stimuli such as media coverage, regulatory actions, or scientific publications[48,49]. This increased attention may artificially elevate reporting frequency, leading to a disproportionality signal that reflects increased awareness rather than a true increase in risk[48,49]. Another important example is the *innocent bystander bias*, which occurs when a drug is frequently co-prescribed with another drug that is the actual cause of the adverse event[50]. In such cases, the innocent drug may be wrongly implicated in spontaneous reports, generating misleading safety signals[50].

The *Weber effect* is a temporal reporting bias characterized by a predictable pattern: the number of spontaneous reports for a new drug typically peaks around the second year after market approval, followed by a gradual decline[51,52]. This phenomenon is hypothesized to affect analyses and may mask or exaggerate early safety signals if not properly accounted for.

However, it has also been shown that the impact of this bias is rare in modern pharmacovigilance databases[51,52]. One of the most studied reporting bias is *masking bias*, which is also known as competition bias, and arises when over-reporting for certain drugs suppresses or obscures the signal of interest for other drugs within the same database[53–56]. In our previous work, we demonstrated that masking bias can alter results in signal detection analyses that implement study designs to mitigate other forms of systematic bias demonstrating potential dependencies between these biasing factors[53]. Overall, the magnitude, direction, and mechanisms of reporting bias can vary, but their presence poses a substantial risk to the validity of pharmacovigilance analyses. These biases may lead to either false-positive signals that prompt unnecessary investigations or false-negative signals that delay detection of emerging safety concerns.

Although SR data are associated with critical limitations as mentioned above, they remain an essential component of the post-market safety surveillance landscape. SR data are a cornerstone of various regulatory pharmacovigilance tools to mandate drug manufacturers to continuously re-evaluate and update the benefit-risk profile of their products. Additionally, the ongoing development and refinement of methodologies to detect safety signals in SR databases is crucial for the timely identification of newly emerging ADR especially those that may not have been evident during pre-approval clinical trials. Importantly, SR databases such as the FAERS, EudraVigilance and CVP are frequently updated and publicly accessible. This makes them relatively transparent and readily available sources of safety data, especially during the early stages following a drug's market authorization, when real-world evidence from other sources such as electronic health records or claims databases may still be limited. In the United States alone, the estimated annual cost of morbidity and mortality related to ADRs exceeds $75 billion, and ADRs consistently rank among the top 10 leading causes of death[43]. These figures

underscore the urgent public health importance of effective pharmacovigilance systems. Consequently, signal detection from SR systems, despite their challenges, continues to play a historically significant and increasingly prominent role in ensuring drug safety and protecting patient health.

1.4 History of Spontaneous Reporting Databases

In our study, we used the U.S FDA FAERS database, one of the first modern SR repository. The origins of FAERS can be traced back to the thalidomide tragedy of the late 1950s and early 1960s, which was one of the most significant drug safety disasters in history. Thalidomide, originally marketed in Europe and other countries as a sedative and treatment for morning sickness during pregnancy, was later found to cause severe congenital malformations, including limb deformities, in thousands of newborns[57]. Although the drug was never approved in the United States, largely due to the efforts of FDA reviewer Dr. Frances Kelsey, this "near-miss" prompted public outrage and catalyzed sweeping regulatory reforms[58]. In response, the U.S. Congress passed the 1962 Kefauver-Harris Amendments to the Federal Food, Drug, and Cosmetic Act, which significantly strengthened the regulatory framework for drug approval[59]. These amendments introduced requirements for demonstrating both safety and efficacy prior to market authorization, established more rigorous standards for clinical trials, mandated informed consent from trial participants, and laid the groundwork for formal systems to monitor adverse drug reactions.

Globally, the thalidomide crisis spurred the formation of national pharmacovigilance programs and the establishment of the World Health Organization's Programme for International Drug Monitoring (PIDM) in 1968[60]. In the United States, the FDA began developing systematic methods to collect and evaluate post-market safety data, leading to the creation of the SR

System, which later evolved into FAERS in 1998. Since then, several FDA regulatory actions have been driven by spontaneous ADR reports[43]. These include strengthened boxed warnings for clozapine (due to myocarditis), nefazodone (due to liver failure), and levomethadyl acetate (due to cardiac arrhythmias). Label warnings were also updated for pioglitazone and rosiglitazone to reflect risks of fluid retention and congestive heart failure[43]. Additionally, the FDA issued a public health advisory regarding itraconazole and terbinafine, citing serious hepatic and potential cardiac adverse events[43]. FAERS now serves as a central repository for voluntary reports of adverse drug events submitted by healthcare professionals, patients, and manufacturers, and is a critical resource for the early detection of drug safety signals.

1.5 The FAERS database

FAERS was utilized in the present study. The submission of ADR reports to the FAERS is mandatory for pharmaceutical manufacturers, but voluntary for healthcare professionals, patients, and other third parties. Reporters to FAERS may include pharmaceutical companies, prescribing clinicians, pharmacists, nurses, patients, caregivers, legal representatives involved in litigation, and others who become aware of a potential adverse drug event. Under U.S. FDA regulations, pharmaceutical manufacturers are required to submit serious unexpected ADRs within 15 calendar days of first learning of the event. Non-serious ADRs must also be submitted in periodic safety reports, typically at intervals determined by the stage of market approval (e.g., quarterly for the first 3 years and annually thereafter)[61]. In contrast, voluntary reporters such as healthcare professionals and patients are encouraged to submit ADRs at any time, with no legally mandated deadline, although timeliness is emphasized in FDA guidance documents[62]. For a ICSR to be considered reportable, the case must include at minimum a patient, a reporter, a medication and an adverse event, while other information may be optional[40]. As of 2024, the

FAERS database contains over 29 million ICSRs, making it one of the largest repositories of post-marketing safety data in the world.

The FAERS database consists of seven anonymized structured files, each capturing a different aspect of ADR reporting. The DEMO file contains patient demographics and administrative details for each case such as age, sex and weight. The DRUG file lists all drugs involved in the report, including their suspected role (i.e. primary suspect, secondary suspect, interacting or concomitant) and dosage information. The REAC file captures reported adverse reactions, which are coded using the Medical Dictionary for Regulatory Activities (MedDRA)[63], a standardized hierarchical vocabulary for ADRs, MedDRA is organized from broad to specific categories: System Organ Class, High-Level Group Terms, High-Level Terms, Preferred Terms (PT), and Lowest Level Terms. In SR systems such as FAERS, ADRs are typically coded at the PT level[63]. The OUTC file records patient outcomes such as death, hospitalization, or disability. The RPSR file indicates the source of the report (e.g., physician, consumer, or lawyer). The THER file includes therapy start and end dates for the drugs involved. Lastly, the INDI file specifies the medical indications for which the drugs were prescribed. The information contained in FAERS data files are listed in **Table 1** below.

Data File	Description
DEMO	Contains patient demographic and administrative information. Each record represents a unique report. Key variables include primaryid, caseid, report date fields (event_dt, mfr_dt, init_fda_dt, fda_dt), patient age (age, age_cod, age_grp), sex (sex), weight (wt, wt_cod), source country (reporter_country, occr_country), and report origin (rept_cod, to_mfr, occp_cod).
DRUG	Contains information on drugs reported in each case. Fields include drugname (reported drug name), prod_ai (active ingredient), role_cod (role of the drug: Primary Suspect [PS], Secondary Suspect [SS], Concomitant [C], or Interacting [I]), route (route of administration), dosage (dose_amt, dose_unit, dose_form, dose_freq), and timing/duration fields (exp_dt, dechal, rechal, etc.). As of Q3 2014,

	prod_ai was added to support analyses using active ingredients instead of brand names.
REAC	Reports adverse drug reactions coded using the Medical Dictionary for Regulatory Activities (MedDRA) at the "Preferred Term" (pt) level[63]. Each record links a reported event to a given primaryid. Also includes drug_rec_act, indicating regulatory action.
OUTC	Captures the outcomes of the ADR report. Outcomes are coded and include: death, hospitalization, disability, life-threatening conditions, required intervention, congenital anomaly, or other (outc_cod).
RPSR	Indicates the source of the report (rpsr_cod). This includes categories such as healthcare professionals (physicians, pharmacists), consumers (patients/caregivers), legal representatives (litigation), or unspecified.
THER	Provides drug therapy timelines. Includes start date (start_dt), end date (end_dt), and therapy duration (dur, dur_cod) for each drug regimen. Can be used to assess treatment windows relative to the onset of adverse events.
INDI	Lists the indication(s) for each reported drug. Like REAC, indications are coded using the MedDRA Preferred Term level (indi_pt), allowing analysis of therapeutic context. Each indication is linked to a specific drug (indi_drug_seq) within a report.

Table 1. Description of FAERS data structure. Abbreviations: FAERS, Federal Adverse Event Reporting System; ADR: Adverse Drug Reaction

To create an analysis-ready dataset from FAERS, the seven individual data files can be linked using the unique case identifier *primaryid*. With the transition from the legacy adverse event reporting system (LAERS) to the current FAERS system in 2012, structural improvements were introduced and another identifier *caseid* was added[35]. It is important to note that *primaryid* reflects a specific version of a case, whereas *caseid* corresponds to the whole individual safety report and remains constant across all versions. As such, multiple *primaryid* entries may exist for a single *caseid*, representing follow-ups or revisions. These duplicates must be identified and removed to avoid inflating counts in downstream analyses. This identifier appears in all files and serves as the common key to merge them into a single, cohesive dataset. By joining on de-duplicated *primaryid*, one can associate each patient's demographic and administrative information (DEMO) with the reported drugs (DRUG), adverse reactions (REAC), outcomes

(OUTC), report sources (RPSR), therapy dates (THER), and indications (INDI). While DEMO, DRUG, and REAC form the core components of every case, the other files may be variably populated depending on the completeness of the report.

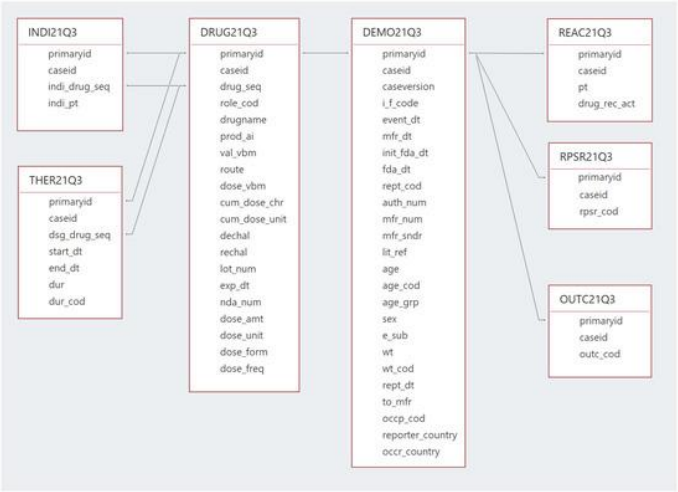

Figure 1. Database structure of the FAERS relational. Source: Khaleel et al[64]

While the FAERS database may appear rich and comprehensive at first glance, containing millions of ICSRs, it is fundamentally limited by the inherent characteristics of SR data. As previously discussed, key limitations such as outcome-dependent selection bias, underreporting, and various systematic and reporting biases significantly restrict the ability of SR databases to support risk estimation or incidence-based analyses. Consequently, traditional epidemiologic methods, such as prospective/retrospective cohort studies, or time-to-event analysis, are not appropriate for FAERS.

Critically, FAERS lacks population representative denominator data indicating how many patients were exposed to each drug and does not contain an unexposed comparison group, both of which are essential for estimating incidence rates or relative risks. Furthermore, since the minimum criteria for an ICSR include only a patient and/or reporter, a suspect drug, and an

adverse event, many key variables are often missing or inconsistently reported. These include therapy start and end dates (THER file), patient age, sex, event onset dates, and other clinical or contextual information necessary for robust analytical modeling. This level of missingness severely limits the feasibility of applying regression-based models, as relevant covariates are often unavailable.

FAERS is also cross-sectional, anonymized, and unlinked, meaning it lacks temporal continuity and cannot be connected to other datasets (e.g., pharmacy claims, clinical outcomes), restricting analyses to case-level using contingency table methods rather than other epidemiologic techniques. While FAERS includes versioned case reports, meaning a single report may appear multiple times over time as it is updated or amended, these updates do not represent longitudinal follow-up in the epidemiologic sense. Each ICSR version reflects a snapshot in time of the case as submitted or modified by the reporter or manufacturer, but there is no systematic, prospective tracking of patient outcomes over time. Instead, case updates are often administrative (e.g., manufacturer follow-up, regulatory queries, or additional information provided post-submission) and are not guaranteed to follow consistent time intervals or contain structured outcome progression data.

In FAERS, ADRs are reported based on the subjective suspicion of an association between a drug and an adverse event, as perceived by the reporter. Reports can be submitted by a wide range of sources, including patients, caregivers, healthcare professionals, pharmaceutical companies, and third parties such as legal representatives. Importantly, there is no requirement for medical confirmation or verification of the suspected drug-event relationship at the time of submission. While some reports, particularly those submitted by physicians, may reflect a clinical assessment to a degree, such evaluation is not systematically required nor documented in

a standardized way across all reports. As a result, the level of suspicion varies widely depending on the reporter's medical background, access to clinical information, and personal beliefs about causality. This means that, at best, investigators can treat ICSRs as a proxy for an ADR, rather than as actual incidences. Ultimately, while FAERS remains an invaluable tool for post-market surveillance, it is best suited for hypothesis generation using disproportionality analysis and data mining techniques. These methods aim to detect patterns of disproportionate reporting of DECs, rather than to provide direct estimates of risk or establish causality.

Although FAERS contains millions of ICSRs, the dataset is inherently high-dimensional and sparse, particularly when viewed through the lens of DECs. FAERS includes thousands of unique drugs and adverse events, and because signal detection are performed at the DEC level, the total number of possible combinations easily reaches millions. For example, with over 20,000 MedDRA PTs and thousands of reported drug names and active ingredients, the number of unique DEC cells, many of which will have counts near zero, is extremely large. While FAERS has captured over 29 million reports cumulatively, only about 2 million reports are submitted annually in recent years, and this number was considerably lower in earlier decades. As a result, even seemingly large sample sizes quickly disperse across the multidimensional space, leaving most DECs with very few observations. This data sparsity poses substantial challenges for signal detection, as the count of reports containing both the drug and the event, is often very small relative to the comparator cell, which contains all other drug-event pairings. Importantly, this means that FAERS is not a "large sample" dataset in the statistical sense when modeling rare DECs. Rather, it should be understood as a highly-sparse, and high-dimensional database. Consequently, widely used signal detection algorithms rely heavily on shrinkage data mining techniques (e.g., Bayesian methods) to stabilize estimates and reduce false positive detection in

low cell-count situations, where traditional epidemiological estimators may struggle due to statistical assumption violations.

1.6 Signal Detection of Single Drug-Event-Combinations

As previously mentioned, traditional epidemiologic methods cannot be reliably applied to SR databases. Instead, data mining techniques to uncover patterns in large datasets, and contingency-table based statistical methods, often called "disproportionality analysis" (DPA) to identify disproportionately high reported DECs, are well suited to suited to identify drug safety signals through signal detection in SR databases[42].

The first step in signal detection involves summarizing DECs into a contingency table, which forms the foundation for subsequent statistical analysis[42]. In this step, each unique pairing of a suspect drug and an associated adverse event reported in an ICSR is treated as a discrete combination. However, before constructing the contingency table, several data cleaning and standardization procedures are essential to ensure validity and reduce misclassification.

First, drug names, which is entered by the reporter as free text in FAERS, may contain typographical errors, inconsistencies in spelling, or refer to various brand names and formulations of the same active substance[35]. To address this, a mapping procedure can be employed to align reported drug names to their corresponding active ingredients, thereby improving consistency and reducing redundancy in the dataset.

Second, adverse events reported in SR systems may be coded by a range of PTs in MedDRA[65]. For instance, a single clinical concept such as myocardial infarction may appear under several PTs, including "acute myocardial infarction", "chronic myocardial infarction", or simply "myocardial infarction". To capture these related terms comprehensively, researchers

often use Standardized MedDRA Queries (SMQs), expert defined groupings of PTs that reflect specific clinical conditions or syndromes[65]. SMQs are widely adopted in pharmacovigilance research and are typically available in both broad and narrow versions. These definitions reflect varying levels of specificity, allowing investigators to select the most appropriate scope for their analysis depending on the clinical context relevant to the ADR being studied[65].

These DECs in the cleaned dataset is then tabulated across the entire database to quantify how frequently each the reporting of each specific drug is linked to the reporting of a particular adverse event. The resulting counts are typically organized into multi-dimensional contingency tables in the form of **Table 2**, which form the basis for various statistical algorithms used to detect disproportionate reporting[42].

	Drugs (D)							
Adverse Event (A)	D_1	D_2	D_3	…	D_b	…	D_B	Total
A_1	n_{11}	n_{21}	n_{31}	…	n_{b1}	…	n_{B1}	$n_{.1}$
A_2	n_{21}	n_{22}	n_{32}	…	n_{b2}	…	n_{B2}	$n_{.2}$
…	…	…	…	…	…	…	…	$n_{.3}$
A_a	n_{1a}	n_{2a}	n_{3a}	…	n_{b3}	…	n_{Ba}	$n_{.a}$
…	…	…	…	…	…	…	…	…
A_A	n_{1A}	n_{2A}	n_{3A}	…	n_{b4}	…	n_{BA}	$n_{.A}$
Total	$n_{1.}$	$n_{2.}$	$n_{3.}$	…	$n_{b.}$	…	$n_{B.}$	$n_{..}$

Table 2. General Contingency Table of All Drugs and Adverse Events in a Spontaneous Database. The contingency table contains “A” numbers of drugs and “B” numbers of adverse events. nAB denotes the frequency of reports of adverse event A and drug B. In the total row and columns, dot “.” represents total counts of report containing that drug or event.

Common DPA algorithms to detect safety signals involve the computation of signal estimators from cell counts derived from a "collapsed" version of the contingency table illustrated in Table 2 into a 2×2 table, as illustrated in Table 3[42]. In this format, the drug of interest is contrasted with the aggregated counts of all other drugs, while the adverse event is compared against the combined frequencies of all other adverse events. This structure simplifies the analysis and enables the application of statistical methods to identify signals of disproportionate reporting for specific DECs.

	Report with Drug	Report of Other Drugs	Total
Report with Event	n_{11}	n_{10}	$n_{1.}$
Report of Other Event	n_{01}	n_{00}	$n_{0.}$
Total	$n_{.1}$	$n_{.0}$	$n_{..}$

Table 3. Collapsed 2x2 Contingency Table Comparing Reference Drug and Adverse events to All Other Drugs and Adverse events. In this table, "1" indicates presence and "0" indicate absence. For example, n11 would denote the frequencies of reports with the drug and adverse events. In the total row and columns, dot "." represents total counts of report containing that drug or event.

Two primary statistical frameworks are commonly employed for the detection of signals of disproportionate reporting (SDRs) in SR used metrics. Frequentist estimates, such as the Reporting Odds Ratio (ROR) and Proportional Reporting Ratio (PRR), which are derived from 2×2 contingency tables that compare the frequency of a specific DEC to all other combinations in the database[66,67]. These estimators assess whether a particular DEC is reported more frequently than would be expected under the assumption of independent reporting between drugs and events (by chance alone). Rather than formal hypothesis testing, these methods use predefined thresholds or rule-based criteria to flag potential safety signals.

The ROR is one of the most widely used disproportionality metrics in pharmacovigilance signal detection[67]. It serves as a reporting-based analogue of the traditional odds ratio used in epidemiologic studies[67,68]. However, rather than estimating an association between exposure and

outcome, the ROR quantifies whether a specific DEC has been reported more frequently than expected by comparing the odds of reporting that DEC versus reporting other events with the same drug or the same event with other drugs[67]. In this context, "more frequently than expected" refers to an elevated odds of observing the target DEC relative to the odds of observing any other combination, based on the background distribution of all reports in the database. Using the collapsed 2x2 table presented in **Table 3.** The ROR is calculated as follows[67]:

$$ROR = \frac{n_{11} \times n_{00}}{n_{10} \times n_{01}} \quad [1]$$

This formula expresses the odds of reporting the event with the drug of interest relative to the odds of reporting the event with all other drugs. A large ROR suggests that the DEC may be reported more frequently in the observed data than would be expected under the assumption of independence between drugs and events[67]. However, a signal is typically considered present only when the lower bound of the 95% confidence interval for the ROR exceeds one. The natural logarithm of the ROR is used to compute the standard error (SE) and construct a 95CI[67]:

$$SE[\ln(ROR)] = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{00}}} \quad [2]$$

$$95CI = \exp(\ln(ROR) \pm 1.96 \times SE[\ln(ROR)]) \quad [3]$$

The PRR is a widely utilized metric in pharmacovigilance for detecting SDRs within SR systems[69]. It assesses whether a specific DEC is reported more frequently relative to all other events associated with the same drug, serving as a reporting-based analogue to the relative risk used in epidemiological studies. The PRR is calculated using the following formula[69]:

$$PRR = \frac{\frac{n11}{n1.}}{\frac{n01}{n0.}} \quad [4]$$

The SE and the 95%CI is computed as follows[69]:

$$SE[\ln(PRR)] = \sqrt{\frac{1}{n_{11}} - \frac{1}{n_{11}+n_{10}} + \frac{1}{n_{10}} - \frac{1}{n_{10}+n_{00}}} \quad [5]$$

$$95CI = \exp(\ln(PRR) \pm 1.96 \times SE[\ln(PRR)]) \quad [6]$$

A lower bound of the 95%CI exceeding 1 is typically considered indicative of a potential safety signal[69]. However, it's crucial to note that the ROR and PRR, are both sensitive to small cell counts, particularly when n_{11} is low[70]. Such scenarios can lead to unstable estimates and inflated disproportionality metrics due to random variation. To mitigate the risk of spurious signals arising from small numbers, researchers often apply minimum threshold criteria for n_{11}. Common adopted thresholds include $n_{11} \geq 3$ or $n_{11} \geq 5$ ensuring sufficient data volume before interpreting a signal as meaningful. Additionally, some protocols incorporate a chi-squared test statistic threshold (e.g., $\chi^2 \geq 4$) to further validate the signal's robustness[71].

In contrast to frequentist methods, the Bayesian approach, exemplified by the Bayesian Confidence Propagation Neural Network (BCPNN), is grounded in Bayes' Theorem and leverages conditional probabilities to evaluate reporting patterns[72]. A central feature of this approach is the incorporation of a prior probability distribution, which represents prior beliefs about the co-occurrence of a drug and an adverse event before any observed data[72]. In BCPNN, a Dirichlet prior is typically applied to the cell probabilities of the 2×2 contingency table, reflecting a neutral expectation of no reporting association, that is, no disproportionate reporting of the DEC[72]. As new reports are introduced, this prior is updated using Bayes' Theorem to form a posterior Dirichlet distribution, which incorporates the observed data into its parameters. The

Information Component (IC), the central disproportionality measure in BCPNN, is computed as a logarithmic transformation of the expected cell probabilities under the posterior distribution[72]. The posterior distribution of the IC is commonly approximated by a normal distribution centered at the maximum a posteriori (MAP) estimate, with credibility intervals derived using empirical or asymptotic methods[72]. This Bayesian updating process not only facilitates sequential learning as more data accumulate but also shrinks extreme estimates toward the prior (null effect), particularly when the number of reports is small. This shrinkage effect helps stabilize signal estimates and reduces the risk of false positives, making Bayesian methods particularly well-suited for signal detection in sparse and high-dimensional SR databases[72].

For a given drug *i* and adverse event *j*, the estimator of the BCPNN, IC, is computed as follows:

$$IC_{ij} = \log_2\left(\frac{p(i,j)}{p(i)\times p(j)}\right) = \log_2\left(\frac{E[\pi_{ij}|n_{ij}]}{E[\pi_{i.}|n_{i.}]\times E[\pi_{.j}|n_{.j}]}\right) \quad [7]$$

Where $E[\pi_{ij}|n_{ij}]$, $E[\pi_{i.}|n_{i.}]$ and $E[\pi_{.j}|n_{.j}]$ are based on pre-defined α_{ij} and β_{ij} hyperparameters.

The expected value of IC_{ij} is given by:

$$E[IC_{ij}|n_{ij}] = \frac{1}{\log 2} \times \ln\left[\frac{(n_{11}+1)(n_{..}+2)^2}{(n_{..}+\beta_{ij}+1)(n_{1.}+1)(n_{.1}+1)}\right] \quad [8]$$

The estimated variance is calculated by:

$$V[IC_{ij}|n_{ij}] = \frac{1}{\log 2} \times \left[\frac{(n_{..}-n_{11}+\beta_{ij})}{(n_{..}+\beta_{ij}+2)} + \frac{(n_{..}-n_{1.}+1)}{(n_{1.}+1)(n_{..}+3)} + \frac{(n_{..}-n_{1.}+1)}{(n_{.1}+1)(n_{..}+3)}\right] \quad [9]$$

An 95% credible interval (95CrI) can be constructed by:

$$95CrI: E[IC_{ij}|n_{ij}] \pm 1.96\sqrt{V[IC_{ij}|n_{ij}]} \quad [10]$$

A lower bound greater than 0 is typically considered an indication of a potential safety signal.

Another Bayesian disproportionality method widely used in pharmacovigilance is the Multi-Item Gamma Poisson Shrinker (MGPS), developed by DuMouchel and colleagues and adopted by the U.S. FDA[73]. The observed number of reports for a drug i and event j, denoted N_{ij}, is modeled as a Poisson random variable with unknown mean μ_{ij}, representing the expected count under the true (but unknown) reporting rate:

$$N_{ij} \sim \text{Poisson}(\mu_{ij}) \qquad \text{for i}^{\text{th}} \text{ drug and j}^{\text{th}} \text{ reaction} \qquad [11]$$

The expected count E_{ij} is calculated assuming independence between drug and event. It's computed as the product of the marginal totals for drug i and event j, divided by the total number of reports[73]:

$$E_{ij} = \frac{N_{i.} \times N_{.j}}{N_{..}} \qquad \text{for i}^{\text{th}} \text{ drug and j}^{\text{th}} \text{ reaction} \qquad [12]$$

The parameter of interest is the true (unknown) reporting ratio λ_{ij} , which compares the true reporting rate μ_{ij} to the expected rate E_{ij}.

$$\lambda_{ij} = \frac{\mu_{ij}}{E_{ij}} \qquad \text{for i}^{\text{th}} \text{ drug and j}^{\text{th}} \text{ reaction} \qquad [13]$$

In MGPS, λ_{ij} is assumed to follow a mixture of two Gamma distributions**.** This empirical Bayes prior is flexible: it allows most DECs to have low rates (shrinking toward 1), while allowing a subset to have higher rates. The mixing weight π is estimated from the entire dataset.

$$\lambda_{ij} \sim \pi \cdot \text{Gamma}(\alpha_1.\beta_1) + (1-\pi) \cdot \text{Gamma}(\alpha_2.\beta_2) \qquad [14]$$

Using Bayes' Theorem, the posterior distribution of the true reporting ratio λ_{ij} is computed by combining the likelihood (based on Poisson assumption) with the prior. This gives a posterior that reflects both the observed data and the distribution of all DECs.

$$p(\lambda_{ij} \mid N_{ij}, E_{ij}) \propto p(N_{ij} \mid \lambda_{ij}, E_{ij}) \cdot p(\lambda_{ij}) \quad [15]$$

The Empirical Bayes Geometric Mean (EBGM) is the posterior mean of the log-transformed reporting ratio, exponentiated back to the original scale. It is a shrinkage-adjusted estimate of RR that stabilizes high or noisy values caused by low counts.

$$\text{EBGM}_{ij} = \exp\left(E\left[\log(\lambda_{ij}) \mid N_{ij}. E_{ij}\right]\right) \quad [16]$$

The 5th (EB05) and 95th (EB95) percentiles of the posterior distribution define a Bayesian credibility interval for $\lambda ij|n$. A lower bound (EB05) > 2 is often used as a threshold for flagging signals. This is more conservative and accounts for posterior uncertainty.

A key advantage of MGPS is that it employs a empirical Bayes model, in which the prior distribution hyperparameters are estimated directly from the full dataset[73]. This allows the model to borrow strength across all DECs, thereby stabilizing signal estimates in sparse regions while preserving sensitivity for frequently reported combinations[73]. While both methods apply Bayesian shrinkage to stabilize disproportionality estimates, they differ substantially in their mathematical structure, input requirements, and computational burden. BCPNN works on joint and marginal co-occurrence probabilities and is typically less computationally intensive, making it suitable for rapid screening[72]. MGPS, on the other hand, estimates a mixture of Gamma priors, followed by posterior integration and quantile estimation, which makes it more computationally demanding. However, this complexity yields more flexible modeling capabilities[73].

Various alternative thresholds have been proposed and applied in the literature to define what constitutes a safety signal in disproportionality-based analyses (minimum value of signal strength or lower bound and case counts)[71]. The choice of threshold often depends on the method used (e.g., PRR, ROR, BCPNN, MGPS) and the desired balance between sensitivity and specificity. To this end, Deshpande et al. conducted a systematic literature review to compile and assess the range of thresholds employed in published signal detection studies. **Table 4** summarizes the most commonly applied thresholds across different methodologies[71].

Algorithm	Published Thresholds
ROR	ROR > 1
	ROR_{05} > 2, N > 2
	ROR_{05} > 1, N ≥ 2
	ROR_{05} > 1
PRR	PRR ≥ 3, χ^2 ≥ 4, N ≥ 3
	PRR ≥ 2, χ^2 ≥ 4, N ≥ 3
	PRR ≥ 2, χ^2 > 4, N ≥ 3
	PRR ≥ 2, χ^2 > 4, N > 2
	PRR > 2, χ^2 ≥ 4
	PRR > 1.5, χ^2 ≥ 4, N > 2
	PRR > 1
	PRR_{05} > 1, N ≥ 2
	PRR_{05} > 1
MGPS	EB_{05} ≥ 2, N > 0
	EB_{05} > 2, N > 0
	$EBlog_2$ > 0, N > 0
	EBGM ≥ 2, N > 0
	EBGM > 2, N > 0
	EBAM(L_{95}) > 1
	INTSS > 1
BCPNN	IC > 0
	IC - 2 SD > 0

Table 4. published thresholds for identification of drug safety signal. PRR = Proportional Reporting Ratio; PRR_{05} = Lower 95% confidence interval for PRR; ROR = Reporting Odds Ratio; ROR_{05} = Lower 95% confidence interval for ROR; EBGM = Empirical Bayes Geometric Mean; EB_{05} = 5th percentile of the empirical Bayes posterior gamma mixture (also referred to as $EBGM_{05}$); $EBlog_2$ = Log_2-transformed EBGM; EBAM = Empirical Bayes Arithmetic Mean; L_{95} = Lower bound of 90–95% credibility interval; INTSS = Interaction Signal Score; IC = Information Component (used in BCPNN); IC–2SD = Lower bound of the 95% confidence interval for IC (i.e., IC minus 2 standard deviations); N = Number of case reports. Superscripts a and b denote literature variations in terminology, where a refers to $EBGM_{05}$ and b denotes the base EBGM estimate.

Signal detection in pharmacovigilance does not rely on a single “gold standard” method[42]. Instead, a combination of statistical frameworks is often necessary to ensure that a

diverse range of DEC reporting patterns can be effectively evaluated. This is particularly important in methodological studies, such as the present analysis, which aim to investigate how biases, can impact the detection of safety signals since resulting signals may be different due to varying statistical properties with each method. Bayesian methods, such as BCPNN or MGPS, incorporate prior distributions that regularize estimates when observed data are sparse[42]. This shrinkage effect dampens the influence of highly volatile signal scores that arise from very low cell counts, thereby improving stability and reducing the risk of overestimating associations based purely on random variation. For example, if only a few cases of a DEC were reported but the expected count is near zero, frequentist methods may return extremely high disproportionality scores due to lack of shrinkage. Bayesian models mitigate this by pulling estimates closer to the prior when data are uncertain. Conversely, frequentist methods such as the ROR or PRR offer computational simplicity and transparency. They perform reliably when applied to well-represented drug–event pairs, where large cell counts reduce statistical noise and provide confidence in the observed effect. These methods are particularly effective for quickly identifying common safety signals in large datasets.

By using both frameworks in tandem, researchers can cross-validate findings, compare signal stability across sparse and dense areas of the data, and evaluate how sensitive results are to model assumptions. This dual-approach strategy is especially valuable in settings where methodological rigor is needed to evaluate the influence of bias, assess signal robustness, or explore how different model structures may affect interpretation[74].

1.7 Comparators

In DPA using SR databases, the traditional, default, comparator for is "all other reports in the database." That is, the observed count of a DEC is evaluated against the background of all

other reports, regardless of drug, indication, or patient population[75]. While this approach offers simplicity, it can introduce substantial bias due to differences in baseline risk across drug populations.

As previously mentioned, these systematic biases, such as confounding by indication and channeling bias, arises because patients prescribed different medications often have underlying clinical differences that also affect the risk of certain adverse events. For instance, comparing cardiovascular events associated with an anti-diabetic drug to events associated with unrelated medications ignores the elevated baseline cardiovascular risk in diabetic patients. In such cases, the disproportionality estimate may reflect population risk differences rather than a true drug safety signal.

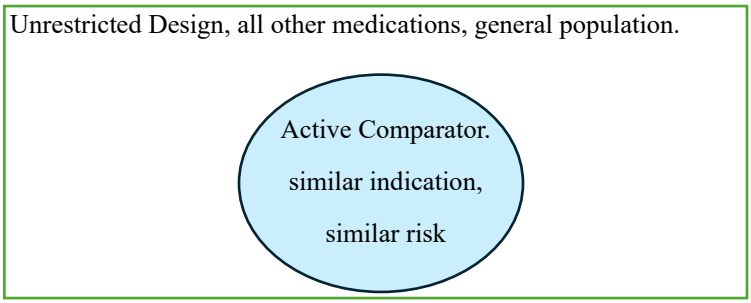

Figure 2: Conceptual diagram comparing unrestricted vs. active comparator design in SR data

To address this, recent studies have advocated for the use of restricted, active comparator designs in SR-based analyses. In this approach, only reports involving drugs within the same therapeutic class or indication are included, and all other reports are excluded from the analysis[46,76]. This narrows the comparator group to a more homogeneous population, improving the validity of the disproportionality estimates by reducing the potential impact of confounding by indication. The concept of using active comparators is well established in pharmacoepidemiology and has recently been extended to pharmacovigilance[46,76]. For example,

Alkabbani et al. demonstrated that restricting analysis to anti-diabetic medications when evaluating cardiovascular outcomes led to more interpretable and potentially valid safety signals than using an unrestricted comparator set[46]. By narrowing the population, the analysis better reflects a common baseline risk, reducing the likelihood that a signal is driven by underlying population differences[46]. **Figure 2** provides a visual illustration of this design.

However, this approach also introduces new challenges. We have previously shown that reporting biases tied to external events, such as regulatory actions or media attention, may exert a greater influence in active comparator designs[53]. This is particularly problematic when the sample size or cardinality of the comparator set is reduced by a large amount, which can magnify the effect of temporal spikes in reporting for certain drugs or events. While active comparator designs enhance internal validity, they may also increase sensitivity to external noise, necessitating careful methodological consideration.

1.8 Bias Evaluation and Mitigation Strategies

An active area of methodological research in passive pharmacovigilance is the evaluation of reporting and systematic biases, and their impact on the validity of signal detection outcomes. For instance, studies by Neha et al. have shown that while some FDA safety alerts and warnings may have a strong and immediate effect on reporting rates, many do not produce any significant change in adverse event reporting or signal strength[49]. Their analysis concluded that notoriety bias, or the inflation of reports following public safety announcements, was not a consistent phenomenon in FAERS, and that overall disproportionality estimates remained largely unaffected by regulatory alerts[49].

In terms of bias mitigation strategies, Bayesian shrinkage methods, such as MGPS and BCPNN, have been shown to be effective in reducing the impact of the innocent bystander effect, where co-prescribed drugs are mistakenly implicated in AE reports[50]. Among various modeling techniques, the LASSO regression has demonstrated particular robustness in minimizing this bias[50]. Additionally, recent studies have reported that the Weber effect, a historical pattern in which reporting peaks in the second year after drug approval, is less commonly observed in modern FAERS data[51,52].

As previously mentioned, to mitigate channeling bias and the innocent bystander effect, the implementation of active comparator designs has been proposed. By restricting analysis to therapeutically similar drugs, researchers aim to reduce bias due to baseline population differences[46]. However, it should be noted that bias mitigation strategies often address only one specific type of bias at a time. For example, masking bias, where disproportionately high reporting of certain drugs suppresses the detectability of other signals, has shown variable effects in the literature. Recently, Bai et al. documented a case of substantial masking when using a restricted reference set, and concluded that reduced sample size and comparator cardinality can exacerbate residual reporting biases, while mitigating systematic biases[53]. **Table 5** summarizes key studies of reporting biases in SR data.

Bias	Definition	Impact	Magnitude of Impact	Supporting literature
Notoriety	Elevated reporting of drug-event combination due to external events	Elevated safety signal of the notoriety drug	Low in FAERS; Potentially large in French database	Neha et al, 2019 Pariente et al, 2007
Innocent by-stander	Drugs are mistaken to be associated with the ADR, since they are prescribed together with the drug that is the ADR's actual cause	Elevated safety signal of the innocent drug	Lowest in Bayesian shrinkage and LASSO methods	Dijkstra et al, 2020
Masking (competition bias)	Drug safety signals are muffled by elevated reporting of other medications in spontaneous reporting databases.	Decreased signal of the study drug	Thought to be low in signal detection with unrestricted comparator; Potentially impactful in restricted comparators.	Bai et al, 2025 Saslvo et al, 2013 Wei et al, 2010 Maignen et al, 2014 Hauben et al, 2017 Pariente et al, 2012
Weber effect	ADR reporting peaks at the end of the second year after a regulatory authority approves a drug	Elevated signal of the drug during the initial two years of approval	Most of the modern adverse event reporting into FAERS does not follow the pattern described by Weber.	Hoffman et al, 2014 Arora et al, 2017
Confounding by indication	Systematic bias. The indication itself is associated with the adverse event of interest	False positive or false negative signal depending on the confounded comparison	Potentially lead to false negatives, could be mitigated through restricted active comparators	Alkabbani et al, 2021

Table 5. List of reporting biases and their characteristics

In summary, the development of methods to evaluate and mitigate reporting biases remains a critical and evolving area of pharmacovigilance research. The IMI PROTECT working group, after five years of summarizing available evidence, concluded that more research is needed to develop and validate algorithmic solutions that quantify and correct for these biases in routine signal detection workflows[44].

2. Drug-Drug Interactions

2.1 Drug-Drug Interactions in Pharmacovigilance

While the previously discussed strategies primarily focus on signal detection for a single DEC, real-world clinical settings often involve the concomitant use of multiple medications, especially in aging populations or those with multiple chronic conditions[77,78]. In these scenarios, each drug may independently contribute to ADRs but more critically, they may also interact in ways that could increase (or sometimes decrease) the risk of ADRs. Such drug–drug interactions (DDIs) pose a substantial and often under-recognized challenge in pharmacovigilance. It has been estimated that up to 30% of unexpected ADRs may be attributed to DDIs, highlighting their considerable burden on healthcare systems and patient safety.[79].

The detection of DDIs introduces several additional layers of complexity compared to single-drug signal detection. Fundamentally, DDIs are classified into two broad mechanistic categories: pharmacokinetic (PK) and pharmacodynamic (PD) interactions. PK interactions occur when one drug alters the PK at the absorption, distribution, metabolism, or excretion level of another, thereby modifying its plasma concentration and exposure profile. For example, a strong inhibitor of a cytochrome P450 (CYP) enzyme may increase the systemic concentration of another drug metabolized by the same enzyme, raising the risk of dose-dependent toxicities[80,81].

In contrast, PD interactions involve functional interplay at the pharmacological target level, where two or more drugs act on the same or on a similar physiological system or pathway. These interactions can be additive (combined effect equals the sum of individual effects), synergistic (combined effect exceeds the sum), or antagonistic (one drug diminishes the effect of another). For instance, combining two serotonergic drugs may elevate the risk of serotonin

syndrome through a synergistic PD mechanism, while using a stimulant and a sedative concurrently may result in an antagonistic PD interaction[82,83].

In the context of DDIs, the interacting medications are typically classified according to their functional roles as the precipitant (or perpetrator) drug and the object (or victim) drug. The precipitant drug is the agent that initiates or modifies the interaction, often by altering the PK or PD profile of the other drug. The object drug, by contrast, is the agent whose pharmacological exposure or effect is altered as a consequence of the interaction[84].

Unlike single drug–adverse event relationships, the detection of DDIs involves the interplay between two or more medications, making the underlying mechanisms more complex and difficult to identify. Moreover, the evaluation of DDIs is often underrepresented in RCTs, as these studies typically exclude patients receiving multiple concomitant medications in order to minimize variability and reduce the risk of confounding. As a result, SR databases have become an increasingly important resource for DDI research[85]. In recent years, the use of SR data for DDI signal detection has gained momentum, with a growing body of literature dedicated to developing and validating analytical methods for identifying DDIs using real-world data[44,85–87]. However, signal detection becomes even more challenging in this context, as all the concerns outlined in **Section 1.8**, such as reporting bias, and confounding, can now affect both drugs simultaneously, potentially with different magnitudes and even opposing directions.

2.2 Study DDI – Atorvastatin and Antivirals

The antiviral medications examined in this study include lopinavir and ritonavir, both of which are protease inhibitors originally developed for the treatment of human immunodeficiency virus (HIV) infection[88]. Ritonavir is often co-administered with lopinavir as a PK enhancer due to its strong inhibition of CYP3A4, which increases the plasma concentration of lopinavir and other co-administered drugs metabolized by CYP3A4[88]. Together, these medications are marketed under brand name Kaletra©[89]. During the early stages of the Coronavirus Disease 2019 (COVID-19) pandemic, the combination of lopinavir/ritonavir gained broader attention as a candidate for drug repurposing, due to its demonstrated in vitro antiviral activity against coronaviruses and its prior use in managing viral infections[90,91]. These agents were administered experimentally in COVID-19 patients under emergency protocols or within clinical trials[90]. However, by October 2020, emerging evidence from the RECOVERY trial demonstrated that lopinavir/ritonavir lacked sufficient efficacy in treating COVID-19 infections, and the combination was never approved for this indication by the US FDA[89,92].

Subsequently, in December, 2021, the U.S. FDA granted Emergency Use Authorization (EUA) for ritonavir in combination with nirmatrelvir for the treatment of mild-to-moderate COVID-19 under the brand name Paxlovid®[93]. This marked a broader introduction of ritonavir into widespread clinical use, now in a pandemic context. Despite its emergency use authorization, access to Paxlovid remained limited during the early rollout due to the requirement for a physician prescription and the need to initiate treatment within five days of symptom onset[94,95]. It was not until July 2022 that pharmacists in the United States were formally authorized to prescribe Paxlovid directly under revised FDA guidelines, improving access to timely treatment[96]. Paxlovid further received full FDA approval for treatment of COVID-19 in

adults on May 25, 2023[95]. Dosage information for Kaletra® and Paxlovid® are presented below in **Table 6.**

Drug	Dosage	Indication	FDA approval Date
Kaletra©	Tablet (lopinavir/ritonavir) • 100mg/25mg x 8 daily • 200mg/50mg x 4 daily Oral solution (lopinavir/ritonavir) • (400mg/100mg)/5mL x 2 daily	HIV infection	15/09/2000
Paxlovid©	Tablet (Nirmatrelvir/ritonavir) • 300mg/100mg x 2 daily Administer for 5 days only	COVID-19 Infection	25/05/2023 (Full approval)

Table 6. Dosage information of Kaletra and Paxlovid

On the other hand, atorvastatin is one of the most widely prescribed statins, indicated for lowering blood cholesterol levels and reducing cardiovascular risk[97]. It exerts its effect by inhibiting HMG-CoA reductase, the rate-limiting enzyme in the mevalonate pathway, which is essential for endogenous cholesterol biosynthesis[97]. While this mechanism is effective in reducing low-density lipoprotein (LDL) cholesterol, it can also disrupt the synthesis of important downstream metabolic products, such as coenzyme Q10 (ubiquinone), which plays a crucial role in mitochondrial energy production within muscle cells.

This disruption is believed to contribute to a class of adverse effects known as statin-associated myopathies, which encompass a spectrum of skeletal muscle toxicities[98]. These range from mild myalgia (muscle pain without elevated creatine kinase (CK) levels), to myopathy (muscle weakness with CK elevation), and in rare but serious cases, rhabdomyolysis[98]. Rhabdomyolysis involves severe muscle breakdown, leading to the release of intracellular components like myoglobin into the bloodstream, which can precipitate acute kidney injury and, if left untreated, be life-threatening[98]. Atorvastatin is primarily metabolized by CYP3A4[98,99]. Following oral administration, atorvastatin undergoes extensive first-pass metabolism, where

CYP3A4 converts it into active and inactive metabolites[98,99]. This metabolic pathway plays a crucial role in determining the systemic exposure and plasma concentration of atorvastatin, directly influencing both its efficacy and toxicity profile[98,99].

Consequently, the co-administration of lopinavir/ritonavir, or nirmatrelvir/ritonavir (the precipitant) with atorvastatin (the object) is expected to result in a PK DDI, driven by ritonavir's potent inhibition of CYP3A4[100]. This interaction can significantly increase plasma concentrations of atorvastatin, which in turn may elevate the risk of statin-induced toxicity such as myopathy and rhabdomyolsis[100101,102]. Due to this well-established PK interaction, clinical guidelines recommend either adjusting the atorvastatin dose or switching to statins with a lower potential for CYP3A4-mediated interactions, such as pravastatin or rosuvastatin, when co-administering with strong CYP3A4 inhibitors[103]. However, such interactions may still be common, as atorvastatin is the most widely prescribed statin and has been estimated to take up 36% of the statin market[104].

2.3 Covid-19 and Antiviral Repurposing

During the COVID-19 pandemic, ritonavir was rapidly repurposed for potential treatment as Paxlovid©, despite originally being developed and approved for the management of HIV infection[105,106]. Using SR data, a previous study investigating COVID-19 as a potential effect modifier found a signal linking the reporting of ritonavir to COVID-19 infections during the early stages of the pandemic[107]. This signal likely reflected not only the increased use of the antivirals but also the broader context in which they were being administered, outside of their original indication for HIV and into a new, rapidly expanding patient population.

As these drugs were repurposed and used more widely, patients who would not normally be exposed to protease inhibitors, such as older adults, individuals with severe comorbidities, or those with no prior antiretroviral therapy history, may began receiving them, often under emergency-use conditions. This shift in the exposed population introduced new dynamics into the SR system. For example, physicians unfamiliar with the typical safety profile of these antivirals might have been more likely to report adverse events out of caution or uncertainty, while patients with COVID-19, who were already severely ill, may have experienced complex clinical trajectories that confounded the attribution of events to specific drugs. Furthermore, the heightened media attention, regulatory interest, and intense clinical scrutiny surrounding COVID-19 treatments likely amplified reporting rates, contributing to elevated reporting of the drug or adverse events, potentially leading to *notoriety bias* for the study DDI of atorvastatin/antiviral for myopathy/rhabdomyolysis. Simultaneously, the novelty of using these antivirals in different populations could have given rise to previously unobserved ADR in the HIV population, or at least the perception of such, which could lead to decreased signal for the study DDI due to *masking effect* by overreporting of other adverse events, previously demonstrated to be possible[108].

3. Objective

3.1 Gap in Literature

As previously mentioned, the evaluation and mitigation of bias in pharmacovigilance has predominantly focused on the single DEC setting. In contrast, the literature addressing bias in the context of DDI detection remains relatively limited. To the best of our knowledge, only a few

studies have attempted to explore methodological challenges specific to DDI analysis in SR systems.

Battini et al. employed a machine learning algorithm to assess the biological plausibility of DDIs, utilizing therapy start and end dates from the THER file in the FAERS database[109]. The algorithm identified whether the suspect drugs were likely co-administered, under the rationale that a reported interaction may be biologically implausible if the two drugs were not taken within overlapping timeframes, for example, if they were administered weeks apart within the same report[109].

Similarly, Kontsioti et al. developed an approach based on a Bayesian framework to assess the biological plausibility of DDIs. The approach incorporated clinically established DDIs extracted from formularies as positive controls, and the concomitant use of medications not known to interact with each other as negative controls. The latter were used to adjust signal detection algorithms[110].

While both studies represent advancements in the validation of reported DDIs through temporal and biological plausibility, it is important to note that neither study addressed the issue of reporting biases, such as *masking*, *notoriety*, or *confounding by co-medication*, in the DDI setting. In addition, while the therapy (THER) file in FAERS can provide potentially useful information, as demonstrated by Battini et al, the large amount of missing data may render complete case analyses necessary[35]. Indeed, the time stamps recorded in the therapy file are often incomplete or inaccurate, limiting their reliability for temporal analyses[35]. Moreover, although positive and negative controls for DDIs may aid in safety evaluation as shown by Kontsioti et al., their applicability to SR data is uncertain, as FAERS is not a clinical dataset but is instead influenced by self-reporting behavior and associated biases. As such, there remains a critical gap

in the literature on how SR behavior and external influences may distort or obscure true DDI safety signals, particularly in the context of polypharmacy and complex treatment regimens.

3.2 Thesis Objective

This study aims to empirically investigate how reporting biases may impact DDI signal detection estimates. We hypothesize that the repurposing of ritonavir for the treatment of COVID-19 during the pandemic period may have altered reporting patterns due to the abrupt change in indication. This may have resulted in a shift in reporting behaviour between two patient subgroups (HIV patients vs COVID-19 patients) with very different clinical characteristics providing a setting to explore this potential influence on a well-established PK DDI – ritonavir and atorvastatin for myopathy/rhabdomyolysis.

First, we will descriptively characterize the annual reporting rates of lopinavir/ritonavir, and nirmetravir/ritonavir prior to and during the COVID-19 pandemic followed by conducting a retrospective bias analysis in the FAERS database to assess whether there were notable changes in the DDI signal detection estimates. We will contrast the DDI signal detection findings between time periods and compare them against their expected results to gain insight into the influence of this potential source of reporting bias.

To enhance the validity of the findings, this study was prospectively registered on the Open Science Framework (https://osf.io/94hnu), outlining the analysis plan a priori.

Chapter II: Statistical Methods and Biases in Signal Detection for DDI

4. Signal Detection Algorithms to Detect DDI

Unlike signal detection in single drug settings, which typically uses a 2×2 contingency table as summarized in **Table 3,** detecting signals for DDIs requires a more complex data structure due to the involvement of at least two drugs. For DDIs, data are often summarized in a 4×2 contingency table as shown in **Table 7,** where the reporting frequencies of both drugs involved in the interaction are evaluated against the reporting of the ADR. In this expanded table format, the rows represent the presence or absence of each drug (e.g., A and B) in the suspected interaction (e.g., Drug A only, Drug B only, both Drug A and Drug B, or neither drug), while the columns represent the presence or absence of the adverse event. This structure allows for a detailed breakdown of the reporting patterns, capturing not only the individual contributions of each drug to the ADR but also their combined reporting rate[111]. In the following sections, we review a subset of commonly used methods for detecting DDIs.

	AE	Other AE	Total
Drug A and B	n_{111}	n_{110}	$n_{11.}$
Drug A only	n_{101}	n_{100}	$n_{10.}$
Drug B only	n_{011}	n_{010}	$n_{01.}$
Other Drugs (neither A or B)	n_{001}	n_{000}	$n_{00.}$
Total	$n_{..1}$	$n_{..0}$	$n_{...}$

Table 7. Collapsed 4x2 Contingency Table Comparing Drug A and Drug B of DDI and Adverse events to All Other Drugs and Adverse events. In this table, "1" indicates presence and "0" indicate absence. For example, n111 would denote the frequencies of reports with both drug A, drug B and the adverse event. In the total row and columns, dot "." represents total counts of report containing that drug or event. AE denotes "Adverse Event".

4.1 Omega Shrinkage

Estimator:

The omega shrinkage method, proposed by Norén et al. (2008), is a disproportionality analysis technique that evaluates potential DDI signals by comparing observed versus expected reporting frequencies assuming a population level additive risk model of the interaction[112]. Specifically, the method calculates the observed reporting rate/frequency of the concomitant

usage f11 —representing the co-reporting frequency of both drugs and the adverse event— against its expected value (E[f11]). The expected value is estimated based on the marginal reporting frequencies of each drug with the adverse event, under the null assumption that the two drugs do not interact[112].

$$\frac{f11}{E[f11]} \quad [1]$$

Although E[f11] is unknown in SR databases, it can be estimated and f11 can be compared with this estimate. To estimate E[f11], we first modeled the occurrence of the adverse event of interest, denoted A, in the underlying population. Let α_0 represent the background risk of A, independent of medication use. This baseline risk may reflect the natural progression of the underlying disease or coincidental events that are temporally associated with treatment but not causally related.

We then consider two drugs, D1 and D2, which may be prescribed individually, together, or not at all. Among individuals who are prescribed neither D1 nor D2, the total risk of experiencing A, denoted p_{00}, is assumed to equal the background risk α_0.

$$p_{00} = \alpha_0 \quad [2]$$

Next, let α_1 represent the excess risk of A associated with exposure to D1, and α_2 represent the excess risk associated with D2. Assuming that the background risk (α_0) and the drug-specific risks (α_1 and α_2) operate independently, the total risk of experiencing the adverse event A when only D1 is used (i.e., in the absence of D2), denoted p_{10}, can be expressed as:

$$p_{10} = 1 - (1 - \alpha_0)(1 - \alpha_1) = \alpha_0 + \alpha_1 - \alpha_0 * \alpha_1 \quad [3]$$

Similarly, for p_{01}:

$$p_{01} = 1 - (1 - \alpha_0)(1 - \alpha_2) \quad [4]$$

Then, the total risk of experiencing the adverse event A when both D1 and D2 are used concurrently, denoted p_{11}, can be calculated under the assumption of independent effects as:

$$p_{11} = 1 - (1 - \alpha_0)(1 - \alpha_1)(1 - \alpha_2) \quad [5]$$

Next, given that both the background risk α_0 and the attributable risk from D1, α_1, can be assumed to be small for a given adverse event, an assumption that is reasonable in the context of large SR systems, their product ($\alpha_0 * \alpha_1$) is negligible compared to either α_0 or α_1 alone. This implies that the following approximation for p_{10}, through an additive risk model, is valid and appropriate for use in SR settings:

$$p_{10} \approx \alpha_0 + \alpha_1 \quad [6]$$

Similarly for p_{01} and p_{11}:

$$p_{01} \approx \alpha_0 + \alpha_2 \quad [7]$$

$$p_{11} \approx \alpha_0 + \alpha_1 + \alpha_2 \quad [8]$$

Due to the absence of reliable information on the total number of drug exposures in the underlying population, along with the inherent under-reporting in SR systems, it is not feasible to directly link population-based probabilities (such as [2], [6], [7], [8]) to the observed relative reporting rates in the database. To derive a comparable reference within the database itself, let A' denote the occurrence of at least one ADR that is not A—such that A and A' are mutually exclusive. Let α'_0 represent the background risk of A'.

If ADRs that could be attributed to D1 or D2 are excluded from A', then the total risk of A' is equal to α'_0 regardless of the drug exposure combination. This allows us to define the probability of observing A'—i.e., the comparator group—consistently across all exposure strata.

$$p'_{00} = \alpha'_0$$

$$p'_{10} = \alpha'_0$$

$$p'_{01} = \alpha'_0 \quad [9]$$

$$p'_{11} = \alpha'_0$$

Next, to derive an estimator for the observed-to-expected ratio of the relative reporting rate in SR systems for the adverse event A, given co-prescription of D1 and D2, we calculate the relevant marginal frequencies. The cell count notation (ie. n_{111}) used here corresponds to the contingency table structure outlined in **Table 7**. Based on these cell counts, the following expression defines the observed relative reporting rate for A associated with the combination of D1 and D2.

$$f_{00} = \frac{n_{001}}{n_{00.}}$$

$$f_{10} = \frac{n_{101}}{n_{10.}}$$

$$f_{01} = \frac{n_{011}}{n_{01.}} \quad [10]$$

$$f_{11} = \frac{n_{111}}{n_{11.}}$$

Next, we derive the estimator for the expected relative reporting rate of A under concomitant usage of D1 and D2, denoted E[f11], based on cases in which at most one of the two drugs was prescribed (i.e., f00, f10, and f01). This expected value serves as the denominator in the observed-to-expected ratio presented in Equation [1].

It is important to note that this estimate may be affected by reporting bias, represented by the unmeasurable factor r. However, since r cannot be quantified using data from SR systems, we proceed by ignoring it. Under this assumption, the expected value of f11 is approximated as follows:

$$E[f_{00}] = E\left[E[f_{00}|n_{00.}]\right]$$

$$= E\left[\frac{\alpha_0 * r}{\alpha_0 * r + \alpha_0{}' * r}\right] \quad [11]$$

$$= \frac{\alpha_0}{\alpha_0 + \alpha'_0}$$

Similarly,

$$E[f_{10}] = \frac{\alpha_0 + \alpha_1}{\alpha_0 + \alpha_1 + \alpha'_0} \quad [12]$$

$$E[f_{01}] = \frac{\alpha_0 + \alpha_2}{\alpha_0 + \alpha_2 + \alpha'_0} \quad [13]$$

$$E[f_{11}] = \frac{\alpha_0 + \alpha_1 + \alpha_2}{\alpha_0 + \alpha_1 + \alpha_2 + \alpha'_0} \quad [14]$$

We can re-express [14] in terms of [11] to [13]:

$$E[f_{11}] = \frac{\alpha_0+\alpha_1+\alpha_2}{\alpha_0+\alpha_1+\alpha_2+\alpha'_0}$$

$$= 1 - \frac{\alpha'_0}{\alpha_0+\alpha_1+\alpha_2+\alpha'_0}$$

$$= 1 - \frac{1}{\frac{\alpha_0+\alpha_1}{\alpha'_0}+\frac{\alpha_2}{\alpha'_0}-\frac{\alpha_0}{\alpha'_0}+1}$$

$$= 1 - \frac{1}{\frac{E[f_{10}]}{1-E[f_{10}]}+\frac{E[f_{01}]}{1-E[f_{01}]}-\frac{E[f_{00}]}{1-E[f_{00}]}+1} \quad [15]$$

Therefore, as an estimator of E[f11], g11, we can use:

$$g_{11} = 1 - \frac{1}{\frac{f_{10}}{1-f_{10}}+\frac{f_{01}}{1-f_{01}}-\frac{f_{00}}{1-f_{00}}+1}$$

However, as α0 or α0 could be misleading negative values, we modify g11 as follows:

$$g_{11} = 1 - \frac{1}{\max\left(\frac{f_{00}}{1-f_{00}},\frac{f_{10}}{1-f_{10}}\right)+\max\left(\frac{f_{00}}{1-f_{00}},\frac{f_{01}}{1-f_{01}}\right)-\frac{f_{00}}{1-f_{00}}+1} \quad [16]$$

Shrinkage:

To construct the omega shrinkage for interaction as seen in SR datasets, we first consider:

$$\Omega_0 = \log_2 \frac{f_{11}}{g_{11}} \quad [17]$$

Given the rarity of adverse events in the context of large SR datasets, the count g11, representing the co-occurrence of both D1, D2, and A', tends to be very small. As a result, the estimate of Ω_0 becomes highly sensitive to spurious associations arising from data sparsity. To mitigate this sensitivity, shrinkage techniques are commonly employed to stabilize the estimates by reducing the influence of random fluctuations.

Following the approach described in [17], we can construct the shrinkage-adjusted estimate by first computing the observed count f11 and its comparator g11 as follows:

$$\frac{f_{11}}{g_{11}} = \frac{n_{111}}{E_{111}} \quad [18]$$

And apply the shrinkage using tuning parameter - α:

$$\Omega = \log_2 \frac{n_{111}+\alpha}{E_{111}+\alpha} \quad [19]$$

Previous studies have shown that setting the shrinkage tuning parameter α to 0.5 provides sufficient regularization to prevent spurious disproportionality signals based on only one or two

reports[112]. This value of alpha was also adopted in our study. Accordingly, the final shrinkage-adjusted estimator is calculated as:

$$\Omega = \log_2 \frac{n_{111}+0.5}{E_{111}+0.5} \quad [20]$$

Uncertainty

Both frequentist and Bayesian implementations are available for omega shrinkage. The frequentist approach calculates the variance as follows based on the Poisson model:

$$\mathrm{Var}(\Omega_0) = \mathrm{Var}(\log_2 \frac{n_{111}}{E_{111}}) \approx \frac{1}{n_{111}\log(2)^2} \quad [21]$$

Applying the central limit theorem, the 95% confidence interval is calculated as follows:

$$\Omega_0 \pm (1.96 * \frac{1}{\sqrt{n_{111}} * \log(2)}) \quad [22]$$

In the Bayesian approach, the credible interval limit can be determined numerically as solution to the following integral function for the posterior quantile μ_q.

$$q = \int_0^{\mu_q} \frac{(E_{111}+\alpha)^{n_{111}+\alpha}}{\Gamma(n_{111}+\alpha)} u^{n_{111}+\alpha-1} e^{-(n_{111}+\alpha)u} du \quad [23]$$

Specifically, the logarithm of the solutions to [23] for q=0.025 and 1=0.975 provides the lower and upper limit, giving credible interval of Ω: $\Omega_{0.025}, \Omega_{0.975}$.

Threshold

In both frequentist and Bayesian approaches, a signal is detected if the lower bound ($\Omega_{0.025}$) of the confidence/credible interval exceeds 0.

4.2 Extended Bayesian Confidence Propagation Neural Network

Also proposed by Norén et al. in 2005, the IC is a Bayesian disproportionality estimator commonly used in signal detection, particularly by the WHO[72]. The IC is estimated through the BCPNN algorithm, which provides both a point estimate and an associated credibility interval. This approach can be extended to detect higher-order interactions, including potential DDIs, by applying the same logic to multi-dimensional contingency tables[113].

The third-order IC applies a Dirichlet prior to the full contingency table in **Table 7**, capturing joint and marginal dependencies. Unlike the omega shrinkage, which models additive risk of DDI, the IC models a multiplicative risk in both single drug and multiple drug context.

$$IC = log_2(\frac{P_{Drug,Event}}{P_{Drug} * P_{Event}})$$

Let x denote drug 1, y denote the adverse event, and z denote the interacting drug 2. The IC for the DDI is defined as the excess disproportionality above what would be expected from the individual DEC:

$$IC_{xyz} = IC_{xy|z} - IC_{xy} = \log_2\left(\frac{P(AE, Drug_1, Drug_2) \cdot P(Drug_1)}{P(Drug_1, Drug_2) \cdot P(AE, Drug_1)}\right)$$

Prior

First, the prior is constructed for the marginal probabilities using Laplace-like smoothing, which involves adding a small constant, 0.5 to the numerator and 0.5×2=1 to the denominator, accounting for the two strata (yes/no) for each binary variable (e.g., drug use or adverse event reporting). This smoothing prevents zero probabilities and is particularly effective in addressing sparsity.

$$q_{1..} = \frac{n_{111} + n_{101} + n_{110} + n_{100} + 0.5}{n_{...} + 1}$$

$$q_{0..} = \frac{n_{011} + n_{001} + n_{010} + n_{000} + 0.5}{n_{...} + 1}$$

$$q_{.1.} = \frac{n_{111} + n_{011} + n_{110} + n_{010} + 0.5}{n_{...} + 1}$$

$$q_{.0.} = \frac{n_{101}+n_{001}+n_{100}+n_{000}+0.5}{n_{...}+1} \quad [23]$$

$$q_{..1} = \frac{n_{111} + n_{101} + n_{011} + n_{001} + 0.5}{n_{...} + 1}$$

$$q_{..0} = \frac{n_{110} + n_{100} + n_{010} + n_{000} + 0.5}{n_{...} + 1}$$

For the pairwise (single drug-event) counts, the smoothed prior is constructed as follows. The smooth constant is 0.25 for the pairwise count numerator and 1 (4 possible strata) for the denominator.

$$q_{11.} = \frac{n_{111}+n_{110}+1/4}{n_{...}+1}, q_{10.} = \frac{n_{101}+n_{100}+1/4}{n_{...}+1}, q_{01.} = \frac{n_{011}+n_{001}+1/4}{n_{...}+1},$$

$$q_{00.} = \frac{n_{010}+n_{000}+1/4}{n_{...}+1}, q_{1.1} = \frac{n_{111}+n_{101}+1/4}{n_{...}+1}, q_{1.0} = \frac{n_{110}+n_{100}+1/4}{n_{...}+1}$$

$$q_{0.1} = \frac{n_{011}+n_{001}+1/4}{n_{...}+1}, q_{0.0} = \frac{n_{010}+n_{000}+1/4}{n_{...}+1}, q_{.11} = \frac{n_{111}+n_{011}+1/4}{n_{...}+1} \quad [24]$$

$$q_{.10} = \frac{n_{110}+n_{010}+1/4}{n_{...}+1}, q_{.01} = \frac{n_{101}+n_{001}+1/4}{n_{...}+1}, q_{.00} = \frac{n_{100}+n_{000}+1/4}{n_{...}+1}$$

Next, the moderating prior for each cell in the contingency table is derived from [23] and [24], under the assumption of no interaction between these variables. Each prior cell count is scaled from a common baseline α.., and reflect the product of the corresponding pairwise terms

divided by the product of the marginal terms. This ensures that the moderating prior is centered at independence, and symmetrically balances the influences across the contingency table.

$$a_{\cdots} = 0.5 \cdot \frac{q_{1\cdot\cdot} \cdot q_{\cdot 1\cdot} \cdot q_{\cdot\cdot 1}}{q_{11\cdot} \cdot q_{1\cdot 1} \cdot q_{\cdot 11}} \quad [26]$$

$$a_{111} = \frac{q_{11\cdot} \cdot q_{1\cdot 1} \cdot q_{\cdot 11}}{q_{1\cdot\cdot} \cdot q_{\cdot 1\cdot} \cdot q_{\cdot\cdot 1}} \cdot a_{\cdots}, a_{110} = \frac{q_{11\cdot} \cdot q_{1\cdot 0} \cdot q_{\cdot 10}}{q_{1\cdot\cdot} \cdot q_{\cdot 1\cdot} \cdot q_{\cdot\cdot 0}} \cdot a_{\cdots}, a_{101} = \frac{q_{10\cdot} \cdot q_{1\cdot 1} \cdot q_{\cdot 01}}{q_{1\cdot\cdot} \cdot q_{\cdot 0\cdot} \cdot q_{\cdot\cdot 1}} \cdot a_{\cdots}, a_{100} = \frac{q_{10\cdot} \cdot q_{1\cdot 0} \cdot q_{\cdot 00}}{q_{1\cdot\cdot} \cdot q_{\cdot 0\cdot} \cdot q_{\cdot\cdot 0}} \cdot a_{\cdots}$$

$$a_{011} = \frac{q_{01\cdot} \cdot q_{0\cdot 1} \cdot q_{\cdot 11}}{q_{0\cdot\cdot} \cdot q_{\cdot 1\cdot} \cdot q_{\cdot\cdot 1}} \cdot a_{\cdots}, a_{010} = \frac{q_{01\cdot} \cdot q_{0\cdot 0} \cdot q_{\cdot 10}}{q_{0\cdot\cdot} \cdot q_{\cdot 1\cdot} \cdot q_{\cdot\cdot 0}} \cdot a_{\cdots}, a_{001} = \frac{q_{00\cdot} \cdot q_{0\cdot 1} \cdot q_{\cdot 01}}{q_{0\cdot\cdot} \cdot q_{\cdot 0\cdot} \cdot q_{\cdot\cdot 1}} \cdot a_{\cdots}, a_{000} = \frac{q_{00\cdot} \cdot q_{0\cdot 0} \cdot q_{\cdot 00}}{q_{0\cdot\cdot} \cdot q_{\cdot 0\cdot} \cdot q_{\cdot\cdot 0}} \cdot a_{\cdots}$$

The posterior parameters, γ, of the Dirichlet distribution after updating the prior calculated in [26] was computed as follows:

$$\gamma_{111} = a_{111} + n_{111}, \gamma_{110} = a_{110} + n_{110}, \gamma_{101} = a_{101} + n_{101}, \gamma_{100} = a_{100} + n_{100}$$

$$\gamma_{011} = a_{011} + n_{011}, \gamma_{010} = a_{010} + n_{010}, \gamma_{001} = a_{001} + n_{001}, \gamma_{000} = a_{000} + n_{000} \quad [27]$$

The posterior expected probabilities derived from the Dirichlet distribution was computed using values from [27] as follow:

$$\mathrm{Ep}_{111} = \frac{\gamma_{111}}{\gamma_{\mathrm{sum}}}$$

$$\mathrm{Ep}_{11\cdot} = \frac{\gamma_{111} + \gamma_{110}}{\gamma_{\mathrm{sum}}}, \mathrm{Ep}_{1\cdot 1} = \frac{\gamma_{111} + \gamma_{101}}{\gamma_{\mathrm{sum}}}, \mathrm{Ep}_{\cdot 11} = \frac{\gamma_{111} + \gamma_{011}}{\gamma_{\mathrm{sum}}}, \quad [28]$$

$$\mathrm{Ep}_{1\cdot\cdot} = \frac{\gamma_{111} + \gamma_{110} + \gamma_{101} + \gamma_{100}}{\gamma_{\mathrm{sum}}}, \mathrm{Ep}_{\cdot 1\cdot} = \frac{\gamma_{111} + \gamma_{110} + \gamma_{011} + \gamma_{010}}{\gamma_{\mathrm{sum}}}, \mathrm{Ep}_{\cdot\cdot 1} = \frac{\gamma_{111} + \gamma_{101} + \gamma_{011} + \gamma_{001}}{\gamma_{\mathrm{sum}}}$$

Estimator

Next, the maximum a posteriori (m.a.p.) estimate is computed, which is better-suited for stratified IC analysis, compared to central estimates.

$$\mathrm{IC}_{\mathrm{MAP}} = \log_2\left(\frac{\mathrm{Ep}_{111}\cdot\mathrm{Ep}_{1..}\cdot\mathrm{Ep}_{.1.}\cdot\mathrm{Ep}_{..1}}{\mathrm{Ep}_{11.}\cdot\mathrm{Ep}_{1.1}\cdot\mathrm{Ep}_{.11}}\right) \quad [29]$$

Uncertainty

To adjust the credibility interval of the IC, the Extended BCPNN framework uses an adaptive correction based on the data concentration within the contingency table. The concentration is quantified by the following ratio, rounded to 1 decimal places:

$$r_{111} = \mathrm{round}\left(\frac{\gamma_{111}}{\min(\gamma_{111}+\gamma_{110}+\gamma_{101},\ \gamma_{111}+\gamma_{011}+\gamma_{101},\ \gamma_{111}+\gamma_{110}+\gamma_{011})},\ 1\right) \quad [30]$$

This ratio r_{111}, reflects how dominant the central triplet cell (n111) is compared to its surrounding cells. A high value (closer to 1) indicates the triplet is highly concentrated, suggesting stronger evidence of a specific DDI–AE combination. A low value indicates more dispersed data and higher uncertainty.

To account for uncertainty in the MAP estimate of the Information Component, two constants, Ar and Br, are used to calculate the width of the 95% credibility interval, denoted as Δ. Norén et al. derived these constants by simulating the posterior distribution of the IC using 50,000 random monte-carlo draws from the full Dirichlet posterior for each cell configuration. This simulation provided an empirical reference for the 2.5th percentile of the IC distribution (IC 0.025). The values of Ar and Br vary based on the concentration ratio r111, which is calculated as shown in Equation [30]. The following table lists the precomputed constants used to approximate the credibility interval width via interpolation based on the value of r111[113].

r111	Ar	Br
0.0	3.09	2.22
0.1	2.93	2.27
0.2	2.78	2.26
0.3	2.62	2.25
0.4	2.45	2.15
0.5	2.25	2.12
0.6	2.03	2.05
0.7	1.79	1.93
0.8	1.61	1.89
0.9	1.13	1.15
1.0	0.073	-0.081

Table 8. Precomputed constants used to approximate the credibility interval width via interpolation based on the value of r_{111} in the BCPNN algorithm. Ar and Br refers to the constant used to calculate the credible interval.

The Δ value, computed using r111 value and looked up Ar and Br value in **Table 8**, was computed as follow:

$$\Delta = \frac{A_r}{\sqrt{\gamma_{111}}} + B_r \cdot \gamma_{111}^{-3/2} \quad [31]$$

Finally, the credible interval was calculated as:

$$IC_{0.025,\,0.975} = IC_{MAP} \pm \Delta \quad [32]$$

Threshold: A safety signal is identified if the $IC_{0.025}$ exceeds 0.

4.3 Concomitant Signal Score

The Concomitant Signal Score (CSS), proposed by Noguchi et al. in 2020, is a frequentist disproportionality-based metric derived from the PRR and expanded specifically for detecting potential drug-drug interaction signals[114].

The PRR for concomitant usage of D1 and D2 for the adverse event is calculated as follows based on the contingency table in table 4.

$$\mathrm{PRR}_{\mathrm{D}_1\cap\mathrm{D}_2} = \frac{n_{111}/(n_{111}+n_{110})}{(n_{011}+n_{001}+n_{101})/(n_{000}+n_{001}+n_{010}+n_{000}+n_{101}+n_{100})} \quad [33]$$

Similarly:

$$\mathrm{PRR}_{\mathrm{D}_1} = \frac{(n_{111}+n_{101})/(n_{111}+n_{110}+n_{101}+n_{100})}{(n_{001}+n_{011})/(n_{001}+n_{000}+n_{011}+n_{010})} \quad [34]$$

$$\mathrm{PRR}_{\mathrm{D}_2} = \frac{(n_{111}+n_{011})/(n_{111}+n_{110}+n_{011}+n_{010})}{(n_{001}+n_{101})/(n_{001}+n_{000}+n_{101}+n_{100})} \quad [35]$$

Uncertainty

The CSS compares the lower bound of the PRR for the concomitant use of Drug 1 and Drug 2 to the upper bounds of the PRRs for their individual use[114]. Specifically, the lower and upper bounds of the PRRs, as referenced in Equations [33], [34], and [35], were computed as follows:

$$\mathrm{PRR}^{\mathrm{LB}}_{\mathrm{D}_1\cap\mathrm{D}_2} = \exp\left(\log\left(\mathrm{PRR}_{\mathrm{D}_1\cap\mathrm{D}_2}\right) - 1{,}96\cdot\sqrt{\frac{1}{n_{111}} - \frac{1}{n_{111}+n_{110}} + \frac{1}{n_{001}+n_{011}+n_{101}} - \frac{1}{n_{001}+n_{000}+n_{011}+n_{010}}}\right) \quad [36]$$

$$\mathrm{PRR}^{\mathrm{UB}}_{\mathrm{D}_1} = \exp\left(\log\left(\mathrm{PRR}_{\mathrm{D}_1}\right) + 1{,}96\cdot\sqrt{\frac{1}{n_{111}+n_{101}} - \frac{1}{n_{111}+n_{110}+n_{101}+n_{100}} + \frac{1}{n_{001}+n_{011}} - \frac{1}{n_{001}+n_{000}+n_{011}+n_{010}}}\right) \quad [37]$$

$$\mathrm{PRR}^{\mathrm{UB}}_{\mathrm{D}_2} = \exp\left(\log\left(\mathrm{PRR}_{\mathrm{D}_2}\right) + 1{,}96\cdot\sqrt{\frac{1}{n_{111}+n_{011}} - \frac{1}{n_{111}+n_{110}+n_{011}+n_{010}} + \frac{1}{n_{001}+n_{101}} - \frac{1}{n_{001}+n_{000}+n_{101}+n_{100}}}\right) \quad [38]$$

Estimator:

Finally, the CSS was calculated as follows:

$$\mathrm{CSS} = \frac{\mathrm{PRR}^{\mathrm{LB}}_{\mathrm{D_1 \cap D_2}}}{\max\left(\mathrm{PRR}^{\mathrm{UB}}_{\mathrm{D_1}} \ \mathrm{PRR}^{\mathrm{UB}}_{\mathrm{D_2}}\right)}$$

Threshold:

The threshold for identification of signals outlined by Noguchi et al is as follows[114]:

1. The PRR lower bound of concomitant usage of both drugs for the adverse event >1 and,
2. The CSS > 1

5. Comparison of Statistical Methods

As part of a quantitative bias analysis, such as that undertaken in the present thesis, it is essential to include a range of statistical methods with diverse properties and underlying assumptions[46,53,115]. This multipronged approach ensures that the results are not overly dependent on the limitations or strengths of any single model and allows for a more nuanced understanding of how bias may affect signal detection outcomes.

Among the methods evaluated, both the BCPNN and the CSS operate under the assumption of a multiplicative baseline model of risk. In the BCPNN framework, the joint probability (or observed count) of a drug–event or drug–drug–event combination is compared to the product of the marginal probabilities of each component (expected value), consistent with the assumption of independent co-reporting. In CSS the multiplicative relationship is modeled as the PRR of drug-drug-event divided against the DECs. In contrast, the Omega Shrinakge model uses an population level additive model of excess risk.

According to findings from the Innovative Medicines Initiative Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium (IMI PROTECT), which conducted an extensive evaluation of signal detection methodologies in pharmacovigilance, statistical interaction models that assume additive baseline risks may be more appropriate for detecting DDIs in SR systems[44].

In addition, the statistical methods employed in our study differ in their shrinkage properties, which can influence the stability and reliability of signal estimates, particularly in sparse data settings. The BCPNN applies Laplace smoothing by modeling the expected distribution of each cell in the three-way contingency table, effectively shrinking extreme values to mitigate the impact of low counts[113]. In contrast, the Omega Shrinkage method uses a gamma prior, with a tuning parameter of 0.5, to provide shrinkage[112]. Meanwhile, the CSS is a purely frequentist measure and does not incorporate any shrinkage mechanism, which makes it more susceptible to false positive signals, especially in the context of rare events or sparse co-reporting[114].

Overall, incorporating both additive and multiplicative baseline interaction models which are comparable on a ratio scale under Bayesian and frequentist frameworks offers a robust and comprehensive approach for evaluating how different statistical assumptions and shrinkage strategies impact signal detection in a retrospective bias analysis. This multi-method approach not only enhances the interpretability of results but also enables relative comparison across distinct methodologies, strengthening the reliability and generalizability of the findings of a bias analysis. **Table 9** summarizes key characteristics of the statistical methods.

Feature / Criterion	Omega Shrinkage (RRR)	Extended Information Component (IC)	Concomitant Signal Score (CSS)
Statistical Framework	Bayesian	Bayesian	Frequentist
Population Level Interaction Model Assumption	Additive risk	Multiplicative	Multiplicative
Underlying Distribution	Gamma-Poisson	Dirichlet-multinomial	Binomial / Poisson approximation
Smoothing / Prior	Gamma prior via α shrinkage	Laplace-like smoothing/ Dirichlet prior	None (uses raw PRR estimates)
Estimator Type	Log-ratio of observed to expected	MAP of IC from Dirichlet posterior	PRR-based ratio using lower/upper bounds
Uncertainty Measure	Confidence interval (CLT or credible)	Credibility interval via Ar and Br	Confidence intervals of PRR
Signal Detection Threshold	Lower bound of CI > 0	IC0.025 > 0	PRR025 D1∩D2 > 1 and CSS > 1
Handles Sparse Data	Yes (via shrinkage α = 0.5)	Yes (via smoothed priors)	Less robust under low counts
Computational Complexity	Low to moderate	Moderate to high	Low
Proposed By	Norén et al., 2008	Norén et al., 2005	Noguchi et al., 2020

Table 9. Summary of key characteristic of statistical methods used to detect DDI in our study. PRR denotes prioportinoal reporting ratio; CI denotes confidence interval; MAP denotes maximum a posteriori.

6. Conceptual Framework of Reporting Bias in Signal Detection of DDI

As mentioned previously, the mechanisms by which reporting bias may influence signal detection for DDIs in SR systems have not been extensively studied. While the impact of bias has been the subject of substantial methodological research in the context of single drug–event signal detection, its implications for DDIs remain largely underexplored. In this section, we hypothetically examine how various forms of reporting bias and fluctuations in drug utilization may distort the detection of DDI signals, with a particular focus on how such effects may differ based on the pharmacological nature of the interaction (e.g., PK vs. PD) and the comparator design employed in the analysis.

It is important to emphasize that statistical algorithms such as Omega Shrinkage, BCPNN, and CSS are designed to detect disproportionate reporting patterns, but they do not account for the underlying pharmacological mechanisms or biological plausibility of the interaction. As a result, signals may be detected even in cases where no meaningful pharmacological interaction exists, and conversely, true DDIs may be masked by noise or bias in the data. Therefore, the conceptual framework presented below aims to bridge the gap between clinical pharmacology and signal detection methodology.

6.1 Reporting Bias in PK DDI

In PK DDIs, the object drug is typically the direct cause of the ADR, while the precipitant drug modifies the risk by altering the metabolism, elimination or other PK aspects of the object. As a result, the ADR can occur with the object drug alone, independent of any interaction. This means that if the object drug experiences an increase in reporting due to an external event or reporting bias, the DDI signal could increase as the co-reporting of the precipitant medication

would likely increase, too. For instance, notoriety bias, or the innocent bystander effect could elevate the object drug reporting frequency, thereby inflating the DDI signal as well. When the precipitant drug experiences an increase in reporting, the DDI signal may also increase. Since the precipitant itself does not directly cause the ADR, changes in its reporting may or may not affect the DDI signal, however co-reporting of the object medication is likely to increase, which may lead to increased safety signal. In scenarios where both the object and precipitant drugs experience shifts in reporting, whether due to media coverage, regulatory action, or broader clinical use, the potential for distortion in the DDI signal is amplified. If both drugs are over-reported in connection with the same ADR, the signal may be strongly inflated.

However, it should be noted, that the amplified reporting rate of a medication may not be only attributed to the study DDI of interest. Instead, elevated reporting for other ADRs are also likely due to other unrelated external events. In such cases, masking by adverse event, or event-competition bias, may occur, which has demonstrated to be able to decrease safety signals[108].

Pharmacokinetic Interactions			
Object Reporting	**Precipitant Reporting**	**Impact on DDI Signal Detection Estimates**	**Potential Biases**
Increased (for study ADR)	No change	Increase	Notoriety, Innocent-bystander, Webber effect
Increased (for other ADR)	No change	Decrease	Masking
No change	Increased (for study ADR)	Increase	Notoriety, Innocent-bystander, Webber effect
No change	Increased (for other ADR)	Decrease	Masking
Increased (for study ADR)	Increased (for study ADR)	Increase	Notoriety, Innocent-bystander, Webber effect
Increased (for other ADR)	Increased (for study ADR)	Mixed Effect	Notoriety, Innocent-bystander, Webber effect, Masking
Increased (for study ADR)	Increased (for other ADR)	Mixed Effect	Notoriety, Innocent-bystander, Webber effect, Masking

Table 10. Conceptual framework of reporting bias in signal detection for PK DDIs. ADR denotes adverse drug reaction.

In **Table 10** we present all possible ways an increase in reporting that may influence the object, precipitant or both drugs introduced by reporting biases may influence DDI signal detection estimates. This conceptual framework is provided in the setting in which there is an assumed 'true' DDI effect. In the first two columns we specify multiple settings in which a hypothetical increase in the reporting of the drug with the potential reporting biases listed in column 4. In column 3, 'Impact on DDI Signal Detection Estimates', we present our conceptualization of the directionality of the bias on DDI estimates for each setting.

6.2 Reporting Bias in PD DDI

In PD DDIs, both the object and precipitant medications can independently cause the ADR based on their biological mechanism of action. Consequently, the impact of reporting bias may differ substantially from that observed in PK DDIs. Specifically, elevated reporting such as the *innocent bystander effect* or *notoriety bias* affecting only one of the two drugs is less likely to artificially inflate the DDI signal, since the ADR is not biologically plausible in the absence of both agents, and may not be over-reported. In such cases, masking bias by over-reporting of other ADR seems more likely.

Pharmacodynamic Interactions			
Object Reporting	**Precipitant Reporting**	**Interaction Signal**	**Potential Biases**
Increased (for study ADR)	No change	No change	Notoriety, Innocent-bystander, Webber effect
Increased (for other ADR)	No change	Decrease	Masking
No change	Increased (for study ADR)	No change	Notoriety, Innocent-bystander, Webber effect
No change	Increased (for other ADR)	Decrease	Masking
Increased (for study ADR)	Increased (for study ADR)	Increase	Notoriety, Innocent-bystander, Webber effect
Increased (for other ADR)	Increased (for study ADR)	Mixed Effect	Notoriety, Innocent-bystander, Webber effect, Masking
Increased (for study ADR)	Increased (for other ADR)	Mixed Effect	Notoriety, Innocent-bystander, Webber effect, Masking

Table 11. Conceptual framework of reporting bias in signal detection for PD DDIs. ADR denotes adverse drug reaction.

In **Table 11**, we present a conceptual framework illustrating how increases in reporting, introduced by biases, may affect the object, precipitant, or both drugs in the context of PD DDIs. This framework assumes the existence of a 'true' underlying PD interaction. The first two columns outline various hypothetical scenarios involving increased reporting of one or both interacting drugs, while column 3 ('Impact on DDI Signal Detection Estimates') describes the anticipated direction of bias in signal detection. Column 4 specifies the potential reporting biases responsible for these distortions.

6.3 Study Bias Setting

In our study, we aim to investigate whether the COVID-19 pandemic and the repurposing of the HIV antiviral medication, ritonavir, for use in COVID-19 infections had an impact on the DDI signals between the antivirals and atorvastatin for the outcome myopathy/rhabdomyolysis (see **Section 2.2**). This analysis is conducted in the context of a PK DDI, where the antiviral agents act as precipitants, and atorvastatin serves as the object drug. Given the expanded use of ritonavir during the pandemic period, we hypothesize that the precipitant drugs may have

experienced changes in reporting patterns. In contrast, atorvastatin is not hypothesized to be affected by reporting bias during the same timeframe. We refer to our conceptual framework in **Table 10** to characterize the expected behaviour of the bias. This particular DDI would represent the setting of a hypothesized increased reporting of the precipitant (Ritonavir) for the study ADR. In this setting, we expect that detected myopathy/rhabdomyolysis signals for ritonavir and atorvastatin could be influenced by utilization pattern change during the COVID-19 pandemic, and potentially *notoriety or masking bias* related to the precipitant drug due to increased public awareness.

In this study, we adopted an object-restricted reference set approach. Specifically, we focused on statins as the object drug (atorvastatin) class due to their known association with myopathy, the adverse event of interest. Other statins served as comparators, allowing us to evaluate whether the addition of a precipitant (e.g., an antiviral) further elevated the risk. We did not restrict by precipitant class, as the diversity of indications and limited sample size of the object-and-precipitant-restricted-comparator made this infeasible.

CHAPTER III: Reporting Bias in Drug-Drug Interaction Detection Using Spontaneous Reporting Systems: A Bias Analysis of the COVID-19 Impact

Running Title: Reporting Bias in Detection of Drug-Drug Interactions

Informative Title: Reporting Bias in Drug-Drug Interaction Detection Using Spontaneous Reporting Systems: A Bias Analysis of the COVID-19 Impact

Authors: Chen Yang Bai[1], Antonios Douros[1,2,3], Melissa Kampman[5], Christopher A. Gravel[1,2,4]

Affiliations:

1. School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada
2. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada
3. Institute of Clinical Pharmacology and Toxicology, Charité - Universitätsmedizin Berlin, Berlin, Germany
4. Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada
5. Marketed Health Product Directorate, Health Canada

Manuscript type: Original Research Article

Word count manuscript (excluding table/figure captions): 5,023

Word count abstract: 241 (limit 250)

Figures/tables: 10

Keywords: pharmacovigilance; myopathy; rhabdomyolysis; bias; methodology; data analysis; safety

7. Abstract

Introduction: It has been shown that reporting bias can distort estimates of disproportionate reporting in the single drug setting, however, their influence is unclear when screening spontaneous reporting databases for drug-drug interactions (DDI). Antiviral medications were repurposed to treat COVID-19 during the pandemic period which may have introduced reporting bias. Its impact on DDI signal detection, particularly when using restricted comparator designs for systematic bias mitigation, remains unclear. To investigate this potential we conducted a retrospective bias analysis on a well known DDI.

Methods: We used data from the United States Food and Drug Administration Adverse Event Reporting System (2000Q3–2023Q3) to evaluate changes in disproportionality estimates for lopinavir/ritonavir and atorvastatin with myopathy/rhabdomyolysis. We computed signals using three methods: Concomitant Signal Score (CSS), Omega Shrinkage, and the extended-Bayesian Confidence Propagation Neural Network (BCPNN). Comparisons were conducted across unrestricted and active comparator reference sets (all statins and CYP3A4 statins), pre- and during-pandemic, change in estimates (ACiE) was calculated to quantify differences.

Results: In the unrestricted comparator design, Omega and BCPNN estimates decreased during-pandemic when including Paxlovid (ACiE: –0.37 and –0.24, respectively), potentially by increased background reporting, but increased when excluding Paxlovid (ACiE: 0.84 and 0.16, respectively). In contrast, signal strength increased in all active comparator analyses, particularly with CYP3A4 statins (Omega ACiE: 1.05; BCPNN ACiE: 0.31). CSS results showed a similar trend.

Conclusion: Changes in antiviral indications in response to the COVID-19 pandemic may have altered reporting patterns affecting DDI signal detection.

8. Introduction

Spontaneous reporting (SR) databases are powerful resources which can be used to generate hypotheses of potential safety signals for marketed medications. These analyses are subject to intractable systematic reporting biases induced by the self-reported nature of the data. Efforts to characterize and evaluate the impact of systematic reporting biases have primarily focused on algorithms for detecting adverse drug reactions (ADRs) in single drug-event combination (DEC) settings. More recently, the potential to utilize SR databases for drug-drug interaction (DDI) signal detection, arising from the concomitant use of multiple medications has become more common[86,110,111,116–119], however, their potential impact on DDI signal detection remains largely unexplored in the literature[46,53,115]. Indeed, as with all signal detection analyses, such biases can distort DDI safety signals and may lead to the inaccurate identification of signals.

A notable external event that may have introduced temporally dependent changes in reporting biases that could have impacted the mining of SR databases is the COVID-19 pandemic[107,120,121]. In particular, the widespread repurposing of antiviral medications from their original indications to treat COVID-19, may have led to alterations in reporting behaviours[107]. This could have resulted in an expansion of drug indications due to the influence of the pandemic and lack of other available treatments (ie. vaccines), providing an opportunity to study aspects of the reporting behaviour change.

Restricted reference sets, such as the active comparator design, limits the comparator group in signal detection to reports involving drugs with similar therapeutic indications as the drug(s) of interest[46]. By restricting the comparator, active comparator designs (ACD) aim to reduce systematic biases such as confounding by indication or channeling bias[46]. Recent evidence suggests that this restriction may come at the cost of increased residual reporting bias,

particularly masking, due to the reduced size and cardinality of the reference set[53,115]. The impact of repurposing medications to treat COVID-19 during the early period of the pandemic on DDI signal detection has not been previously examined in the literature, and it remains unclear whether such reporting biases exert differential effects on restricted reference sets.

Lopinavir and ritonavir (Kaletra©) are protease inhibitors originally developed for the treatment of human immunodeficiency virus (HIV) infection[89]. Ritonavir is most often used in combination with other antiretrovirals as a pharmacokinetic enhancer due to its potent inhibition of cytochrome P450 3A4 (CYP3A4), which increases the systemic exposure of co-administered drugs metabolized through this pathway[89]. One such medication is atorvastatin, a widely prescribed statin metabolized primarily by CYP3A4[99]. When co-administered with ritonavir, plasma levels of atorvastatin can become significantly elevated, potentially leading to muscle-related toxicities such as myopathy or, in severe cases, rhabdomyolysis[97,100].

During the COVID-19 pandemic, ritonavir, originally approved for the management of HIV, was repurposed for use in treating COVID-19 in combination with nirmatrelvir, under the brand name Paxlovid®[89,93,96]. This may have led to widespread administration in a broader patient population, including older adults and those with comorbidities, under emergency-use conditions[105,106]. As a result, SR patterns likely shifted, with increased adverse event reports influenced by heightened clinical uncertainty, media attention, and regulatory focus.

Given the growing interest in DDI signal detection using SR data and the increasing adoption of restricted reference sets, such as the active comparator designs, it is important to consider the potential impact of reporting bias on drug safety signal detection analyses. To this end, we conducted an evaluation of biases in DDI, hypothesizing that the ritonavir medications would have been influenced by stimulated reporting due to their change in indication, thereby

distorting the DDI safety signals with statins which are a known positive. Additionally, we consider whether reference set restriction to address the potential for confounding by indication may have exacerbated the presence of bias in these analyses.

9. Methods

Data Source

This study utilized data from the United States Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS). In this database, ADR reports are submitted either voluntarily by healthcare professionals and patients or mandatorily by pharmaceutical companies. FAERS comprises seven publicly accessible datasets, which can be linked using unique report identifier[122]. Reported adverse events are encoded using the Medical Dictionary for Regulatory Activities (MedDRA), specifically at the "Preferred Term" (PT) level. For this analysis, MedDRA version 27.0 was employed. Because individual ADRs may be represented by groups of PTs, Standardized MedDRA Queries (SMQs) have been developed to capture clinically meaningful groupings of terms relevant to specific conditions[123]. To ensure data quality, duplicate records were removed, and drug names were cleaned and harmonized to correct for inconsistencies and typographical errors. All drugs listed were included in the analysis, regardless of their suspected role (primary suspect, secondary suspect, concomitant, interacting).

Calendar Time Restriction

We included all FAERS reports submitted quarterly between July 1, 2000 (2000Q3), and October 1, 2023 (2023Q3)[122]. To capture the impact of COVID-19–related drug repurposing and shifts in reporting behavior, we defined several calendar-based analytic intervals. These included

a pre-pandemic period (2000Q3 to 2019Q4), an early pandemic period prior to the U.S. FDA's Emergency Use Authorization (EUA) of Paxlovid® (2020Q1 to 2021Q3)[93]. While Paxlovid received EUA in December 2021, patient access remained limited until July 2022, when U.S. pharmacists were formally authorized to prescribe the medication[96]. As such, we analyzed a period covering the pandemic timeframe before Paxlovid became more accessible (2020Q1 to 2022Q2)[96]. Additional analyses were conducted using full pandemic-era interval (2020Q1 to 2023Q3), corresponding to the official end of the COVID-19 public health emergency.

Disproportionality Analysis

Within each study interval, we conducted three primary disproportionality analyses (DPAs), with each DPA set defined by a unique calendar time window and corresponding comparator strategy. Specifically, DPAs 1–3 represent the pre-COVID period (2000Q3 to 2019Q4), DPAs 4–6 correspond to the pandemic period prior to Paxlovid EUA (2020Q1 to 2021Q3), DPAs 7–9 capture the pandemic period before expanded pharmacist access to Paxlovid (2020Q1 to 2022Q2), DPAs 10–12 represent the full pandemic period excluding Paxlovid (2020Q1 to 2023Q3), and DPAs 13–15 include the full pandemic period with Paxlovid-related reports (2020Q1 to 2023Q3).

Across all intervals, the DPAs varied the comparator group to address potential confounding by indication within the context of a hypothesized pharmacokinetic interaction. Specifically, we defined three comparator sets for the object drug class (statins), which is associated with the adverse drug reaction (myopathy/rhabdomyolysis): (1) all other drugs in the database (unrestricted design), (2) statins only, and (3) statins that are CYP3A4 substrates, as this statin subgroup is most susceptible to the DDI with antivirals. A summary of the calendar intervals, comparator definitions, and corresponding DPA numbers is provided in **Table 12**.

Active comparator definitions were not constructed for the precipitant drug class (antivirals) due to limitations in available sample size.

DPA	Time Frame	Comparator
1	2000Q3 to 2019Q4	Full data
2	2000Q3 to 2019Q4	Statins only
3	2000Q3 to 2019Q4	Statins that are CYP3A4 substrate
4	2020Q1 to 2021Q3	Full data
5	2020Q1 to 2021Q3	Statins only
6	2020Q1 to 2021Q3	Statins that are CYP3A4 substrate
7	2020Q1 to 2022Q2	Full data (excluding Paxlovid)
8	2020Q1 to 2022Q2	Statins only (excluding Paxlovid)
9	2020Q1 to 2022Q2	Statins that are CYP3A4 substrate (excluding Paxlovid)
10	2020Q1 to 2023Q3	Full data (excluding Paxlovid)
11	2020Q1 to 2023Q3	Statins only (excluding Paxlovid)
12	2020Q1 to 2023Q3	Statins that are CYP3A4 substrate (excluding Paxlovid)
13	2020Q1 to 2023Q3	Full data (including Paxlovid)
14	2020Q1 to 2023Q3	Statins only (including Paxlovid)
15	2020Q1 to 2023Q3	Statins that are CYP3A4 substrate (including Paxlovid)

Table 12. Disproportionality analyses conducted in the pre- and post- COVID-19 timeframe. DPA denotes disproportionality analyses. Q denotes quarter.

Adverse Event Definition

To define "myopathy/rhabdomyolysis" we used broad SMQ definitions comparison 59 PTs, respectively. The list of all PTs captured in the SMQ definition can be found in supplementary **Table S1**.

Exposure Definition

To account for alternative spellings and typographical errors in reported drug names, we mapped the "drug name" field in FAERS to their corresponding active ingredients for all drugs included in the study[35]. This standardization was first applied to the drugs of interest, lopinavir/ritonavir (Kaletra©), Nirmatrelvir/ritonavir (Paxlovid©), and atorvastatin (Lipitor©), for the outcome of myopathy and rhabdomyolysis.

The initial comparator group used in the disproportionality analyses consisted of all other drugs reported in FAERS, without restriction, to reflect a broad, population-wide background reporting rate. To emulate ACDs, commonly employed in pharmacoepidemiology research to reduce confounding by indication, where between group differences in patients' underlying risk for adverse events may bias drug safety assessments[46], we constructed a comparator group consisting of all FDA-approved statins. This group included fluvastatin (Lescol©), lovastatin (Mevacor©), lovastatin extended-release (Altoprev©), pitavastatin (Livalo©), pravastatin (Pravachol©), rosuvastatin (Crestor©), and simvastatin (Zocor©), as well as combination products such as lovastatin/niacin extended-release (Advicor©), simvastatin/niacin extended-release (Simcor©), and simvastatin/ezetimibe (Vytorin©)[124].

As an additional subgroup analysis, we further restricted the comparator group to statins that are known CYP3A4 substrates, specifically simvastatin (Zocor©) and lovastatin (Mevacor©), as they are metabolised by the same path way as Atoravastatin [124,125]. These comparator restrictions were designed to approximate a population with comparable baseline risk for myopathy and rhabdomyolysis. However, prior research has noted that the application of restricted comparator sets, such as in ACDs, may may have the undesired effect of amplifying residual reporting biases due to reductions in the sample size and the number of drugs (i.e.,

cardinality) included in the analysis[53,75,115]. The list of drug names and active ingredient can be found in **Table S2.**

Statistical Analysis

We computed annual standardized reporting rates of lopinavir/ritonavir and nirmatrelvir/ritonavir within each year of our study to observe yearly reporting rate changes over time (**Table 13**). To observe the potential for prescribing and utilization pattern changes introduced by the altered indication for COVID-19, we first summarized the top-10 co-reported medications with the antivirals in each year (**Table 14**). We also summarized the demographic characteristics of reporters (e.g., age, sex, reporter type) within each exposure group (antiviral and atorvastatin, antiviral only, atorvastatin only) and time interval to provide context on the composition of the reporting population over time (**Table 15**). We then computed signal detection estimates of disproportionate reporting in each DPA using three estimators: the Concomitant Signal Score (CSS) the Omega Shrinkage the extended Bayesian Confidence Propagation Neural Network (BCPNN) which estimates the Information Component (IC) or the base 2 log of the Relative Reporting Ratio (RRR) under a Bayesian framework.

The CSS evaluates the signal strength of co-prescribed drugs by comparing the lower bound of the PRR for concomitant usage to the upper bounds of the PRRs for each individual drug, highlighting potential signals. In contrast, the Relative Reporting Ratio (RRR) compares the observed co-reporting frequency to the expected frequency under the assumption of independence between drug exposures. While PRR-based approaches are intuitive, they are prone to false positives, especially in sparse data environments. To address this, we implemented Bayesian shrinkage estimators such as Omega Shrinkage and the extended BCPNN. The Omega Shrinkage estimator is based on a baseline additive interaction model, but is interpretable as a

ratio estimator based on a Poisson-gamma model that stabilizes estimates in low-count scenarios. The Extended BCPNN, based on a Dirichlet-multinomial model, quantifies third-order interactions between drug 1, drug 2, and the adverse event under a baseline multiplicative interaction model. It generates a posterior IC with credibility intervals, adjusting for data sparsity. For all Bayesian estimators, a signal was considered detected if the lower bound of the 95% credible interval exceeded zero. For PRR and CSS-based methods, detection required the lower bound of the PRR for concomitant usage (PRRLB) to exceed one, and the CSS to be greater than one, respectively. Results for all estimators, along with their associated intervals, are presented in **Table 17** and **Table 18**.

Bias Analysis

To assess the potential impact of the COVID-19 pandemic on DDI signal estimates, we calculated differences in point estimates across calendar-defined intervals. The pre-pandemic analyses (DPA 1–3) served as benchmark references and were compared against four subsequent pandemic-era sets: the early pandemic period prior to Paxlovid authorization (DPA 4–6), the post-EUA period prior to expanded pharmacist access (DPA 7–9), the full pandemic period excluding Paxlovid (DPA 10–12), and the full pandemic period including Paxlovid (DPA 13–15). For each set of DPAs, we computed the Absolute Change in Estimates (ACiE), defined as the absolute difference in disproportionality metrics between the pre-pandemic and pandemic intervals. These comparisons were made across both the URD and ACD, with results summarized in **Table 19**. Greater absolute differences were interpreted as stronger potential influence of pandemic-related reporting dynamics on the observed DDI signals. All analyses were conducted using R version 4.3.0 (Vienna, Austria).

10. Results

Reporting Patterns of Study Medications

Standardized reporting rates for lopinavir/ritonavir remained consistently low throughout the entire study period, typically registering near 0.00% when expressed as a proportion of total FAERS reports. Although the absolute number of reports modestly increased from 2,687 (0.00%) in 2019 (0.00%) to 5,015 (0.00%) in 2020, likely reflecting early off-label use during the pandemic onset, reporting declined in 2021 (3,612) and further decreased in 2022 (2,737) and 2023 (1,819; through Q3). These counts remained negligible relative to the overall volume of FAERS submissions and did not exceed 0.00% when rounded in percentage terms. In contrast, nirmatrelvir/ritonavir exhibited a markedly different pattern. Following its EUA in late 2021, reporting activity remained minimal that year (18 reports; 0.00%). However, a sharp increase occurred in 2022, with 26,690 reports (13.2%) and in 2023, with 10,900 reports (5.39%; through Q3) through Q3 (5.39%). For brevity, we will suppress the quarters for the rest of the results section. **Figure 3** and **Table 13** visualizes the reporting pattern.

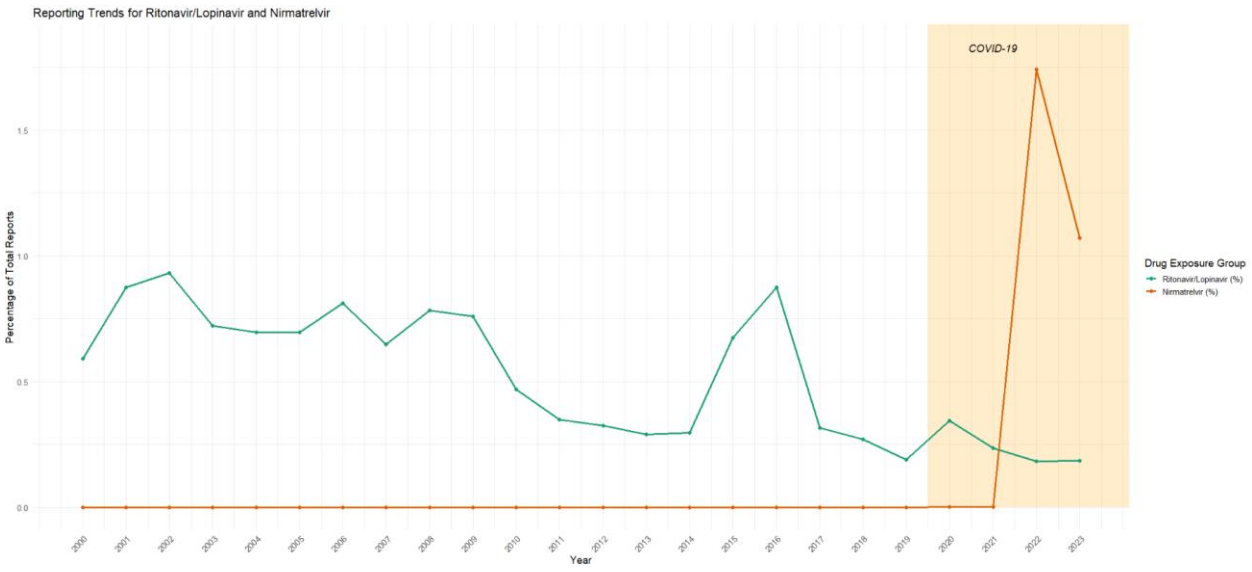

Figure 3. Standardized Reporting Rate of Study Medications. The report counts with antivirals only, statins only, and both antivirals and statins are standardized to the total number of reports received in each year (from quarter 1 to quarter 4). The during-pandemic time frame (after 2020) is highlighted in orange.

year	Ritonavir/Lopinavir	Nirmatrelvir/Ritonavir	Atorvastatin	Total Reports
2000	932 (0.01%)	0 (0%)	5410 (3.43%)	157839
2001	1330 (0.01%)	0 (0%)	6123 (4.03%)	151861
2002	1236 (0.01%)	0 (0%)	4783 (3.61%)	132668
2003	1248 (0.01%)	0 (0%)	5217 (3.02%)	172584
2004	1451 (0.01%)	0 (0%)	7171 (3.44%)	208363
2005	1676 (0.01%)	0 (0%)	8962 (3.72%)	240632
2006	1866 (0.01%)	0 (0%)	8032 (3.50%)	229591
2007	1848 (0.01%)	0 (0%)	8862 (3.11%)	285223
2008	2509 (0.01%)	0 (0%)	8202 (2.56%)	320113
2009	2601 (0.01%)	0 (0%)	8936 (2.61%)	342528
2010	2285 (0%)	0 (0%)	16422 (3.37%)	487006
2011	1901 (0%)	0 (0%)	12754 (2.34%)	544225
2012	2027 (0%)	0 (0%)	15768 (2.53%)	624072
2013	2233 (0%)	0 (0%)	16542 (2.15%)	771034
2014	2571 (0%)	0 (0%)	24356 (2.81%)	865395
2015	8121 (0.01%)	0 (0%)	30039 (2.50%)	1202687
2016	10376 (0.01%)	0 (0%)	33393 (2.81%)	1186656
2017	3883 (0%)	0 (0%)	31911 (2.60%)	1227299
2018	3866 (0%)	0 (0%)	39706 (2.77%)	1434608
2019	2687 (0%)	0 (0%)	41660 (2.93%)	1423222
2020	5015 (0%)	0 (0%)	39921 (2.74%)	1457020
2021	3612 (0%)	18 (0%)	33084 (2.15%)	1535300
2022	2737 (0%)	26690 (13.2%)	32100 (2.09%)	1535693
2023	1819 (0%)	10900 (5.39%)	23848 (2.33%)	1024255

Table 13. Standardized Reporting Rates of Antivirals and Statins. The period following the onset of COVID-19 is highlighted in red.

Next, to explore whether there were observable changes in prescription or utilization patterns introduced by the altered indication, we summarized co-reported medications as a proxy. Prior to the COVID-19 pandemic (2000–2019), the top co-reported medications with ritonavir-containing regimens primarily consisted of HIV antiretrovirals, including emtricitabine, tenofovir, lamivudine, and zidovudine, as well as antibiotic agents such as sulfamethoxazole/trimethoprim used for infection prophylaxis in immunocompromised populations. In 2021, with only limited reports involving Paxlovid, hydroxychloroquine emerged as a top co-reported drug, despite not appearing in the top 10 previously, likely reflecting its early media-driven use during the

pandemic. In 2022–2023, when stratified to exclude Paxlovid, the top co-reported medications continued to reflect traditional HIV therapy patterns. However, when Paxlovid-related reports were included, the top co-reported medications shifted to include general-use drugs, such as atorvastatin, acetaminophen, and aspirin. **Table 14** summarizes the top 10 co-reported medications from 2019 to 2023 in lopinavir/ritonavir versus nirmatrelvir/ritonavir reports. The complete list of co-reported medications for each year is available in supplemental table **S3** and **S4**.

Year	Excluding Paxlovid	Including Paxlovid
2019	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE , 783 (30%)	NA (Not Approved)
	LAMIVUDINE , 723 (27%)	
	DARUNAVIR ETHANOLATE , 613 (23%)	
	TENOFOVIR DISOPROXIL FUMARATE , 489 (19%)	
	ATAZANAVIR SULFATE , 447 (17%)	
	RALTEGRAVIR POTASSIUM , 404 (15%)	
	ZIDOVUDINE , 388 (15%)	
	DARUNAVIR , 368 (14%)	
	SULFAMETHOXAZOLE\TRIMETHOPRIM , 296 (11%)	
	EFAVIRENZ , 275 (10%)	
2020	HYDROXYCHLOROQUINE , 1095 (23%)	NA (Not pproved)
	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE , 915 (19%)	
	LAMIVUDINE , 877 (18%)	
	AZITHROMYCIN ANHYDROUS , 767 (16%)	
	DARUNAVIR ETHANOLATE , 610 (13%)	
	TENOFOVIR DISOPROXIL FUMARATE , 604 (12%)	
	CEFTRIAXONE , 542 (11%)	
	RALTEGRAVIR POTASSIUM , 503 (10%)	
	ZIDOVUDINE , 472 (10%)	
	DARUNAVIR , 443 (9%)	
2021	HYDROXYCHLOROQUINE , 1011 (29%)	HYDROXYCHLOROQUINE (1011 ,29%)
	LAMIVUDINE , 753 (21%)	LAMIVUDINE (753 ,21%)
	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE , 552 (16%)	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (552 ,16%)
	AZITHROMYCIN ANHYDROUS , 505 (14%)	AZITHROMYCIN ANHYDROUS (505 ,14%)
	EFAVIRENZ , 406 (11%)	EFAVIRENZ (406 ,11%)
	DARUNAVIR , 405 (11%)	DARUNAVIR (405 ,11%)

	ATAZANAVIR SULFATE , 396 (11%)	ATAZANAVIR SULFATE (396 ,11%)
	DARUNAVIR ETHANOLATE , 389 (11%)	DARUNAVIR ETHANOLATE (389 ,11%)
	RALTEGRAVIR POTASSIUM , 354 (10%)	RALTEGRAVIR POTASSIUM (354 ,10%)
	ZIDOVUDINE , 347 (10%)	ZIDOVUDINE (347 ,10%)
2022	LAMIVUDINE , 690 (30%)	ATORVASTATIN (1254 ,9%)
	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE , 402 (18%)	ACETAMINOPHEN (958 ,7%)
	DARUNAVIR , 379 (17%)	LEVOTHYROXINE (864 ,6%)
	EFAVIRENZ , 367 (16%)	ASPIRIN (808 ,6%)
	RALTEGRAVIR POTASSIUM , 342 (15%)	LISINOPRIL (776 ,6%)
	TENOFOVIR DISOPROXIL FUMARATE , 341 (15%)	CHOLECALCIFEROL (775 ,6%)
	DARUNAVIR ETHANOLATE , 304 (13%)	AMLODIPINE BESYLATE (763 ,5%)
	ZIDOVUDINE , 290 (13%)	LAMIVUDINE (692 ,5%)
	DOLUTEGRAVIR , 252 (11%)	METFORMIN HYDROCHLORIDE (655 ,5%)
	ATAZANAVIR SULFATE , 233 (10%)	OMEPRAZOLE (627 ,5%)
2023	LAMIVUDINE , 443 (30%)	ATORVASTATIN (548 ,9%)
	DARUNAVIR , 223 (15%)	LAMIVUDINE (444 ,7%)
	TENOFOVIR DISOPROXIL FUMARATE , 218 (15%)	ACETAMINOPHEN (414 ,7%)
	ZIDOVUDINE , 214 (14%)	LEVOTHYROXINE (351 ,6%)
	EFAVIRENZ , 175 (12%)	AMLODIPINE BESYLATE (319 ,5%)
	DOLUTEGRAVIR , 161 (11%)	ASPIRIN (312 ,5%)
	RALTEGRAVIR , 148 (10%)	TACROLIMUS (300 ,5%)
	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE , 147 (10%)	LISINOPRIL (274 ,4%)
	TACROLIMUS , 146 (10%)	LOSARTAN (271 ,4%)
	ABACAVIR , 146 (10%)	METFORMIN HYDROCHLORIDE (258 ,4%)

Table 14. Top 10 co-reported medications with ritonavir-containing regimens from 2019 to 2023, stratified by exclusion vs. inclusion of nirmatrelvir/ritonavir (Paxlovid) reports.

Table 15 presents demographic characteristics across exposure groups and time intervals. As expected, patients exposed to atorvastatin only were consistently older across all periods. Prior to the COVID-19 pandemic, the mean age in the antiviral-only group was 45.80 years (SD: 14.67), notably younger than those in the atorvastatin-only group (64.87 years, SD: 12.46) and the dual-exposure group (52.05 years, SD: 11.50). During the pandemic period prior to expanded Paxlovid access (2020–2022), average ages increased across all groups: 53.52 years (SD: 13.98) for antiviral-only, 66.36 years (SD: 12.59) for atorvastatin-only, and 53.33 years (SD: 10.54) for the dual-exposure group. In the full pandemic period, when stratified by Paxlovid exposure, mean age in the antiviral-only group was 54.89 years (SD: 14.80) when excluding Paxlovid, and 57.40 years (SD: 16.34) when including Paxlovid. For patients with both exposures, the average age was 54.05 years (SD: 10.85) when excluding Paxlovid and 55.78 years (SD: 12.33) when including Paxlovid. The atorvastatin-only group had a mean age of 66.89 years (SD: 12.60). Regarding sex distribution, prior to the pandemic, females accounted for 23.09% (n = 769) of the antiviral-only group, 44.77% (n = 18,307) of the atorvastatin-only group, and 15.00% (n = 27) of the dual-exposure group. During the pandemic period before Paxlovid access expansion, the female proportions were 24.75% (n = 372), 41.58% (n = 4,136), and 27.15% (n = 82), respectively. In the full pandemic period, females represented 26.35% (n = 494) of antiviral-only reports and 26.25% (n = 84) of dual-exposure reports when excluding Paxlovid. Including Paxlovid, females represented 36.72% (n=1,054), of antiviral only reports and 28.49% (n=104) of dual-exposure reports.

	Pre-Covid (2000Q3 to 2019Q4)			During-pandemic (2020Q1 to 2022Q3)			During-pandemic (2020Q1 to 2023Q1) *				
Characteristics	Lopinavir/Ritonavir only (n = 32,364)	Atorvastatin only (n = 291,563)	Anti-virals and Atorvastatin (n = 614)	Lopinavir/Ritonavir only (n =9,716)	Atorvastatin only (n =73,912)	Anti-virals and Atorvastatin (n =466)	Anti-virals (Excluding Paxlovid) (n = 12809)	Anti-virals (Including Paxlovid) (n = 48,913)	Atorvastatin only (n = 108224)	Anti-virals and Atorvastatin (Excluding Paxlovid) (n = 516)	Anti-virals and Atorvastatin (Including Paxlovid) (n =2,020)
Myopathy/Rhabdomyolysis											
Yes, n(%)	3,330 (10.29)	40,888 (14.02)	180 (29.32)	1,503	9,948	302	1875	2870	13810	320	365
No, n(%)	29,034 (89.71)	250,675 (85.98)	434 (70.68)	8,213	63,964	164	10934	46043	94414	196	1655
Age in years Mean (SD)	45.80 (14.67)	64.87 (12.46)	52.05 (11.50)	53.52 (13.98)	66.36 (12.59)	53.33 (10.54)	54.89 (14.8)	57.4 (16.34)	66.89 (12.6)	54.05 (10.85)	55.78 (12.33)
Missing, n(%)	662 (19.88)	10,239 (25.04)	25 (13.89)	403 (26.81)	2006 (20.16)	80 (26.49)	492 (26.24)	660 (23)	2865 (20.75)	80 (25)	84 (23.01)
Female sex, n(%)	769 (23.09)	18,307 (44.77)	27 (15.00)	372 (24.75)	4136 (41.58)	82 (27.15)	494 (26.35)	1054 (36.72)	5855 (42.4)	84 (26.25)	104 (28.49)
Missing, n(%)	224 (6.73)	3,325 (7.89)	8 (4.44)	157 (10.45)	981 (9.86)	3 (0.99)	219 (11.68)	313 (10.91)	6535 (47.32)	233 (72.81)	3 (0.82)
Anti-virals, n(%)	n=3,300	NA	n=180								
Ritonavir	1,996 (60.48)	NA	112 (62.22)	720 (47.9)	NA	168 (55.63)	996 (53.12)	1991 (69.37)	NA	184 (57.5)	229 (62.74)
Lopinavir	1,121 (33.40)	NA	44 (24.44)	618 (41.12)	NA	72 (23.84)	686 (36.59)	686 (23.9)	NA	74 (23.12)	74 (20.27)
Both	213 (6.45)	NA	24 (13.33)	165 (10.98)	NA	62 (20.53)	193 (10.29)	193 (6.72)	NA	62 (19.38)	62 (16.99)
Role of Anti-virals, n(%)		NA	n=185								
Primary suspect drug	764 (21.87)	NA	8 (4.32)	230	NA	3	345	1199	NA	6	34
Secondary suspect drug	1,178 (33.71)	NA	45 (24.43)	437	NA	31	523	596	NA	34	39
Concomitant drug	1,395 (39.93)	NA	125 (67.57)	736	NA	256	810	341	NA	262	267
Interacting drug	157 (4.49)	NA	7 (3.78)	127	NA	13	239	845	NA	19	32
Atorvastatin, n(%)	NA	40,888 (100.00)	180 (100.00)	NA	9,948 (100.00)	302 (100.00)	NA	2870 (100)	13810 (100)	320 (100)	365(100)
Role of Atorvastatin, n(%)	NA	n=42,513	n=180								
Primary suspect drug	NA	15,454 (36.35)	16 (8.89)	NA	1998	16	NA	NA	2901	23	30
Secondary suspect drug	NA	5,059 (11.90)	9 (5.00)	NA	1293	25	NA	NA	1907	27	30
Concomitant drug	NA	21,662 (50.95)	153 (85.00)	NA	6716	260	NA	NA	9139	268	301
Interacting drug	NA	338 (0.80)	2 (1.11)	NA	176	2	NA	NA	231	6	10
N overall medications, mean (SD)	8.84 (8.36)	23.89 (21.48)	10.11 (10.80)	13.07 (14.21)	13.09 (11.85)	30.58 (21.74)	11.82 (13.28)	8.95 (11.67)	12.42 (11.45)	29.78 (21.63)	26.78 (21.57)
Outcome, n(%)	n=4,773										
Hospitalization	1787 (37.44)	15880 (31.94)	78 (29.66)	354 (24.11)	3486 (37.79)	68 (22.74)	519 (28.42)	748 (31.19)	4989 (38.97)	72 (22.78)	84 (24.28)
Life-threatening	329 (6.89)	2223 (4.47)	13 (4.94)	40 (2.72)	359 (3.89)	3 (1)	41 (2.25)	54 (2.25)	503 (3.93)	3 (0.95)	5 (1.45)
Death	384 (8.05)	3410 (6.86)	18 (6.84)	89 (6.06)	544 (5.9)	4 (1.34)	96 (5.26)	108 (4.5)	738 (5.76)	4 (1.27)	6 (1.73)
Disability	163 (3.42)	2999 (6.03)	10 (3.80)	9 (0.61)	303 (3.28)	1 (0.33)	17 (0.93)	37 (1.54)	485 (3.79)	2 (0.63)	4 (1.16)
Required Intervention	114 (2.39)	684 (1.38)	2 (0.76)	0 (0)	15 (0.16)	0 (0)	0 (0)	6 (0.25)	23 (0.18)	0 (0)	1 (0.29)
Congenital Anomaly	44 (0.92)	36 (0.07)	0 (0.00)	1 (0.07)	0 (0)	0 (0)	1 (0.05)	1 (0.04)	1 (0.01)	0 (0)	0 (0)
Other	1952 (40.90)	24485 (49.25)	142 (53.99)	975 (66.42)	4518 (48.98)	223 (74.58)	1152 (63.09)	1444 (60.22)	6064 (47.36)	235 (74.37)	246 (71.1)
Reporting Country, n(%)	n=3,300	n=40,888	n=180								
United States	514 (15.44)	14,130 (34.56)	107 (59.44)	800 (53.23)	3846 (38.66)	248 (82.12)	975 (52)	1561 (54.39)	4734 (34.28)	256 (80)	283 (77.53)
Other	1,045 (31.38)	10,603 (25.93)	20 (11.11)	680 (45.24)	5485 (55.14)	53 (17.55)	857 (45.71)	1117 (38.92)	8048 (58.28)	57 (17.81)	72 (19.73)
Missing	1,771 (53.18)	16,155 (39.51)	53 (29.44)	23 (1.53)	617 (6.2)	1 (0.33)	43 (2.29)	192 (6.69)	1028 (7.44)	7 (2.19)	10 (2.74)

Table 15. Patient Characteristics and Demographic. *Indicates sensitivity analysis

Disproportionality Results of DDI Signal Detection

From 2000–2019, the CSS for lopinavir/ritonavir–atorvastatin–myopathy in the URD was 1.12 (PRRLB: 3.49; DPA 1). In the ACD, CSS values were 1.22 (PRRLB: 1.55; DPA 2) and 1.10 (PRRLB: 1.71; DPA 3) for the statins and CYP3A4 statins subsets, respectively. In the pandemic period prior to Paxlovid© EUA approval, from 2020-2021, CSS values were 4.01 (PRRLB: 15.12; DPA 4), 1.37 (PRRLB: 1.64; DPA 5), and 1.18 (PRRLB: 1.48; DPA 6). In the pandemic period prior to expanded Paxlovid© access, from 2020-2022, CSS values were 3.78 (PRRLB: 14.82; DPA 7), 1.37 (PRRLB: 1.62; DPA 8), and 1.17 (PRRLB: 1.45; DPA 9). In the full pandemic period excluding Paxlovid©, from 2020-2023, CSS values for ritonavir/lopinavir were 3.42 (PRRLB: 13.09; DPA 10), 1.37 (PRRLB: 1.55; DPA 11), and 1.16 (PRRLB: 1.41; DPA 12). In the full pandemic period including Paxlovid©, from 2020-2023, CSS values for ritonavir or nirtrelmavir/lopinavir were 2.65 (PRRLB: 10.14; DPA 13), 1.40 (PRRLB: 1.45; DPA 14), and 1.24 (PRRLB: 1.35; DPA 15).

Omega Shrinkage estimates for the pre-COVID period were 0.66 (95% CrI: 0.44, 0.86; DPA 1), –0.04 (–0.26, 0.17; DPA 2), and –0.56 (–0.78, –0.36; DPA 3). For the 2020–2021 analyses prior to Paxlovid© EUA, the estimates were 1.54 (1.36, 1.71; DPA 4), 0.58 (0.40, 0.75; DPA 5), and 1.01 (0.83, 1.18; DPA6). In the 2020–2022 sensitivity analysis prior to expanded access of paxlovid, the estimates were 1.49 (1.32,1.65; DPA 7), 0.53 (0.36,0.68; DPA 8), and 0.88 (0.71,1.04; DPA 9). In the full COVID period excluding Paxlovid, the estimates were 1.5 (1.34,1.65; DPA 10), 0.48 (0.32,0.63; DPA 11), and 0.87 (0.71,1.02; DPA12). In the full covid period, including Paxlovid (ritonavir/lopinavir or nirmatrelvir), the estimates were 0.29 (0.14,0.44; DPA 13), 0.49 (0.33,0.63; DPA 14), and 0.49 (0.33,0.63 DPA 15).

The BCPNN estimates for the pre-COVID period were -0.02 (95% CrI: -0.25,0.21; DPA 1), -0.03 (-0.24,0.17; DPA 2), and -0.15 (-0.34,0.05; DPA 3). For the 2020–2021 analyses prior to Paxlovid© EUA, the estimates were 0.17 (0,0.35; DPA 4), 0.19 (0.05,0.33; DPA 5), and 0.07 (-0.05,0.2; DPA6). In the 2020–2022 sensitivity analysis prior to expanded access of paxlovid, the estimates were 0.08 (-0.08,0.24; DPA 7), 0.16 (0.03,0.29; DPA 8), and 0.06 (-0.06,0.17; DPA 9). In the full COVID period excluding Paxlovid, the estimates were 0.15 (-0.02,0.31; DPA 10), 0.17 (0.04,0.29; DPA 11), and 0.08 (-0.04,0.19; DPA12). In the full covid period, including Paxlovid, the estimates were -0.26 (-0.41,-0.1; DPA 13), 0.22 (0.08,0.37; DPA 14), and 0.17 (0.02,0.31; DPA 15).

2000Q3 to 2019Q4		
	Myopathy/Rhabdomyolysis	**Other**
Both Drugs	180	434
Antiviral Only	3330	29034
Atorvastatin Only	40888	250675
Neither Drugs	556887	11046786
2020Q1 to 2021Q3		
	Myopathy/Rhabdomyolysis	**Other**
Both Drugs	261	127
Antiviral Only	1183	6449
Atorvastatin Only	7497	47306
Neither Drugs	103765	2486791
2020Q1 to 2022Q2		
	Myopathy/Rhabdomyolysis	**Other**
Both Drugs	302	164
Antiviral Only	1503	8213
Atorvastatin Only	9948	63964
Neither Drugs	138103	3531733
2020Q1 to 2023Q3 (Excluding Paxlovid)		
	Myopathy/Rhabdomyolysis	**Other**
Both Drugs	320	196
Antiviral Only	1875	10934
Atorvastatin Only	13810	94414
Neither Drugs	195952	5234767
2020Q1 to 2023Q3 (Including Paxlovid)		
Other	**Myopathy/Rhabdomyolysis**	**Other**
Both Drugs	365	1655
Antiviral Only	2870	46043
Atorvastatin Only	13810	94414
Neither Drugs	194912	5293595

Table 16. Collapsed 4x2 Contingency Table for Reporting of Antivirals and Atorvastatin.

Timeframe	DPA	Comparator	CSS	PRR Lb
2000Q3-2019Q4	1	Full data	1.12	3.49
	2	Statins	1.22	1.55
	3	CYP3A4 Statins	1.10	1.71
2020Q1-2021Q3	4	Full data	4.01	15.12
	5	Statins	1.37	1.64
	6	CYP3A4 Statins	1.18	1.48
2020Q1-2022Q2	7	Full data	3.78	14.82
	8	Statins	1.37	1.62
	9	CYP3A4 Statins	1.17	1.45
2020Q1-2023 (excluding paxlovid)	10	Full data	3.42	13.09
	11	Statins	1.37	1.55
	12	CYP3A4 Statins	1.16	1.41
2020Q1-2023 (including paxlovid)	13	Full data	2.65	10.14
	14	Statins	1.40	1.45
	15	CYP3A4 Statins	1.24	1.35

Table 17. Signals of DPAs Conducted Using the CSS Method. A signal is identified if the lowerbound of PRR (PRR LB) of the DDI combination exceeds one, and the CSS exceeds 1.

Timeframe	DPA	Comparator	Omega (95% CrI)	BCPNN (95% CrI)
2000Q3-2019Q4	1	Full data	0.66 (0.44,0.86)	-0.02 (-0.25,0.21)
	2	Statins	-0.04 (-0.26,0.17)	-0.03 (-0.24,0.17)
	3	CYP3A4 Statins	-0.56 (-0.78,-0.36)	-0.15 (-0.34,0.05)
2020Q1-2021Q3	4	Full data	1.54 (1.36,1.71)	0.17 (0,0.35)
	5	Statins	0.58 (0.4,0.75)	0.19 (0.05,0.33)
	6	CYP3A4 Statins	1.01 (0.83,1.18)	0.07 (-0.05,0.2)
2020Q1-2022Q2	7	Full data	1.49 (1.32,1.65)	0.08 (-0.08,0.24)
	8	Statins	0.53 (0.36,0.68)	0.16 (0.03,0.29)
	9	CYP3A4 Statins	0.88 (0.71,1.04)	0.06 (-0.06,0.17)
2020Q1-2023 (excluding paxlovid)	10	Full data	1.5 (1.34,1.65)	0.15 (-0.02,0.31)
	11	Statins	0.48 (0.32,0.63)	0.17 (0.04,0.29)
	12	CYP3A4 Statins	0.87 (0.71,1.02)	0.08 (-0.04,0.19)
2020Q1-2023 (including paxlovid)	13	Full data	0.29 (0.14,0.44)	-0.26 (-0.41,-0.1)
	14	Statins	0.49 (0.33,0.63)	0.22 (0.08,0.37)
	15	CYP3A4 Statins	0.49 (0.33,0.63)	0.17 (0.02,0.31)

Table 18. Signals of DPAs Conducted Using the Omega Shrinkage and BCPNN Method. A signal is identified for both methods if the lowerbound of the 95%CrI exceeds 0

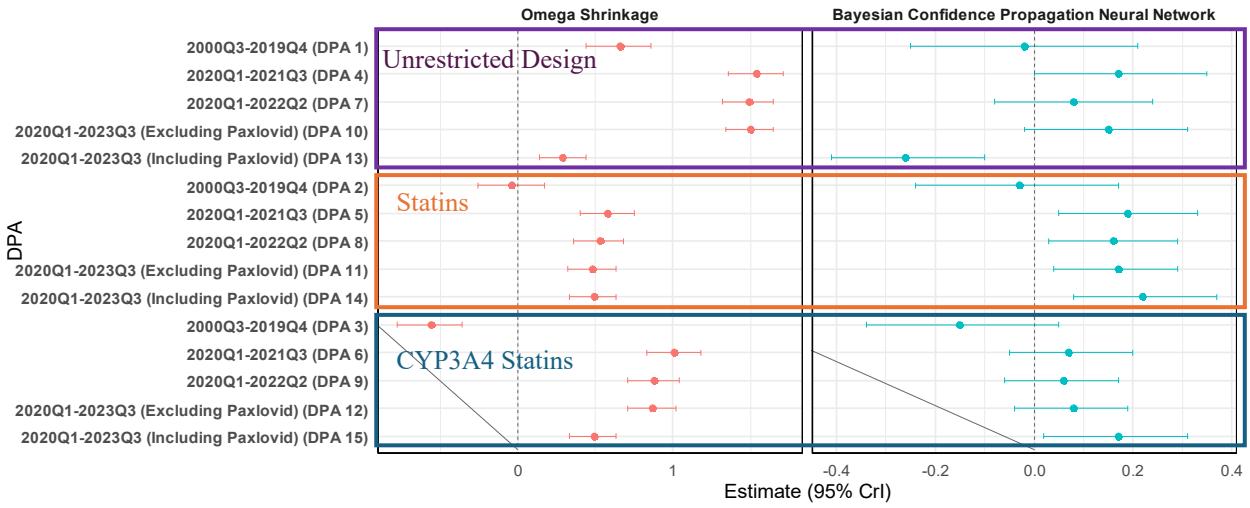

Figure 4. Forest Plot of BCPNN and Omega Shrinkage Signals. Omega Shrinkage and BCPNN signals are computed with unrestricted design, statin comparators and CYP3A4 statin comparators. The resulting signal and 95% CrI are presented.

To quantify changes in signal strength before and after the COVID-19 pandemic, we computed the ACiE between corresponding DPAs for each signal detection method, where positive values indicate an increase in signal strength relative to pre-pandemic levels, and negative values indicate a decrease. For Omega Shrinkage, the ACiEs in the unrestricted comparator group were 0.88 (DPA 4 vs DPA 1) for 2020–2021 vs 2000–2019, 0.83 (DPA 7 vs DPA 1) for 2020–2022 vs 2000–2019, and 0.84 (DPA 10 vs DPA 1) for 2020–2023 vs 2000–2019. However, when Paxlovid-related reports were included, signal strength decreased, with an ACiE of –0.37 (DPA 13 vs DPA 1). In the statin-restricted comparator group, ACiEs were 0.61, 0.61, 0.56, and 0.52 for DPAs 5, 8, 11, and 14 vs DPA 2, respectively. In the CYP3A4 statin-restricted comparator, the ACiEs were 1.57, 1.44, 1.43, and 1.05 for DPAs 6, 9, 12, and 15 vs DPA 3.

For BCPNN, the unrestricted comparator group showed ACiEs of 0.19 (DPA 4 vs DPA 1), 0.10 (DPA 7 vs DPA 1), 0.16 (DPA 10 vs DPA 1), and a decrease of –0.24 (DPA 13 vs DPA 1) with Paxlovid included. For the statin-restricted comparator, ACiEs were 0.22, 0.19, 0.20, and 0.26 for DPAs 5, 8, 11, and 14 vs DPA 2. In the CYP3A4 statin-restricted comparator, ACiEs were 0.22, 0.20, 0.22, and 0.31 for DPAs 6, 9, 12, and 15 vs DPA 3.

For the CSS, ACiEs in the unrestricted comparator group were 2.89 (DPA 4 vs DPA 1), 2.66 (DPA 7 vs DPA 1), 2.30 (DPA 10 vs DPA 1), and 1.53 (DPA 13 vs DPA 1). In the statin-restricted comparator, ACiEs were 0.15, 0.15, 0.14, and 0.18 for DPAs 5, 8, 11, and 14 vs DPA 2, and in the CYP3A4 statin-restricted comparator, ACiEs were 0.08, 0.07, 0.06, and 0.14 for DPAs 6, 9, 12, and 15 vs DPA 3. **Table 19** summarizes the ACiE results.

Comparator	Contrast	Omega	BCPNN	CSS
Unrestricted	4 vs 1	0.88	0.19	2.89
	7 vs 1	0.83	0.10	2.66
	10 vs 1	0.84	0.16	2.30
	13 vs 1	-0.37	-0.24	1.53
Statins	5 vs 2	0.61	0.22	0.15
	8 vs 2	0.61	0.19	0.15
	11 vs 2	0.56	0.20	0.14
	14 vs 2	0.52	0.26	0.18
CYP34A statins	6 vs 3	1.57	0.22	0.08
	9 vs 3	1.44	0.20	0.07
	12 vs 3	1.43	0.22	0.06
	15 vs 3	1.05	0.31	0.14

Table 19. ACiE calculation for contrast of signals between pre and pandemic timeframe.

11. Discussion

We conducted a comprehensive evaluation to assess the potential influence of the COVID-19 pandemic on DDI signal detection using disproportionality measures. Our study applied both frequentist and Bayesian estimators across multiple comparator designs, encompassing unrestricted, statin-restricted, and CYP3A4 statin-restricted reference sets. By examining 15 distinct DPAs, we attempted to characterize changes in signal estimates of lopinavir/ritonavir and atorvastatin with the adverse event of myopathy/rhabdomyolysis. We further adjusted for and studied the impact of the introduction of Paxlovid, where ritonavir reuporposed for treatment of COVID-19 infections in combination with nirmatrelvir, deviating from its original HIV indication.

Our descriptive analysis revealed temporal patterns in the reporting of lopinavir/ritonavir and nirmatrelvir/ritonavir, as well as trends in statin-related reporting. The absolute number of lopinavir/ritonavir reports modestly increased from 2,687 (0.00%) in 2019 to 5,015 (0.00%) in 2020, potentially reflecting early off-label use during the onset of the COVID-19 pandemic—either as part of emergency treatment protocols or due to heightened scientific and media interest. During 2020–2021, in vitro studies and early-phase clinical trials suggested lopinavir/ritonavir as a potential COVID-19 treatment; however, later findings from large-scale trials such as RECOVERY demonstrated limited efficacy, leading to declining clinical use[90,92,105,106]. Correspondingly, FAERS reporting decreased in 2021 (3,612), 2022 (2,737), and 2023 (1,819; through Q3). Despite the early increase, reporting rates for lopinavir/ritonavir remained negligible relative to total FAERS volume, never exceeding 0.00% when expressed as a percentage. In contrast, nirmatrelvir/ritonavir (Paxlovid) exhibited a markedly different pattern. Following its EUA in late 2021, reporting activity remained minimal that year (18 reports;

0.00%). However, a sharp increase occurred in 2022, with 26,690 reports (13.2%), and in 2023, with 10,900 reports (5.39%) through Q3. This spike in reporting coincides with the time period after pharmacist prescribing was authorized[96]. For atorvastatin, reporting rates historically ranged between 2% and 4%, though a gradual downward trend has emerged in recent years. This may reflect shifts in prescribing practices, increasing adoption of alternative statins, or reduced adverse event reporting for long-established therapies.

From the patient demographic data, the average age of the antiviral-only exposure group increased substantially after the onset of COVID-19, rising from 45.80 years pre-pandemic to 57.40 years in the Paxlovid-included group, a shift of nearly 12 years. A similar trend was observed in sex distribution: the proportion of female patients in the antiviral-only group increased from 23.09% to 36.72% during the pandemic period. These shifts further support the hypothesis of a broader post-pandemic utilization pattern, moving beyond the traditionally younger, male-dominated HIV treatment population to an older and more diverse patient group likely treated for COVID-19[126].

A notable shift in the co-reported medications with lopinavir/ritonavir may be suggestive of our hypothesized changes in prescription and utilization patterns. Prior to the pandemic, co-reported drugs were predominantly HIV antiretrovirals and antibacterial agents used to manage opportunistic infections. During the pandemic, hydroxychloroquine emerged as the most frequently co-reported medication for two consecutive years, likely reflecting heightened media attention and widespread off-label prescribing during the early stages of the COVID-19 crisis[127]. In the post-pandemic period, when Paxlovid-related reports were included, co-reported medications shifted toward general-use treatments such as atorvastatin, acetaminophen, and aspirin. However, when Paxlovid was excluded, the co-reported medication profile remained

largely consistent with the pre-pandemic pattern—dominated by antivirals and antibacterial agents.

Notably, a previous study has also identified COVID-19 infection as a potential effect modifier that may introduce statistical reporting interactions with various ADRs[107]. Ritonavir, specifically in the context of off-label use, has been highlighted as one of the agents whose ADR reporting may be influenced by COVID-19 infection status[107]. Taken together, these trends suggest that a broader and more diverse population was exposed to ritonavir during the pandemic, largely for COVID-19 treatment rather than for HIV. The resulting shift in exposure and reporting patterns makes this drug pair a compelling case study for examining how utilization changes and stimulated reporting may impact DDI signal detection in SR systems.

In the URD analyses, an increase in signal strength was observed across most post-COVID intervals—including DPA 4 (prior to Paxlovid EUA), DPA 7 (prior to expanded access), and DPA 10 (full pandemic period excluding Paxlovid), relative to the pre-pandemic baseline. This trend suggests that the increased reporting of lopinavir/ritonavir during the early stages of the pandemic may have provided sufficient volume to reveal a previously undetectable signal for myopathy/rhabdomyolysis, potentially obscured before due to under-reporting, a known limitation of SR systems[128]. However, in the full follow-up period including Paxlovid (DPA 13), a notable decrease in signal strength was observed, indicating that the reporting patterns associated with Paxlovid differ markedly from those of lopinavir/ritonavir. This reversal supports the interpretation that Paxlovid's widespread use may have diluted the observed signal through event competition bias, a variant of masking bias in which an influx of unrelated adverse event reports within the comparator pool suppresses the disproportionality of the target event. This phenomenon has been previously documented and shown to be plausible[108].

In contrast to the URD analyses, all ACD analyses revealed consistent increases in signal strength following the onset of COVID-19. Both the statin-restricted and CYP3A4 statin–restricted comparator groups showed notable post-pandemic rises in signal estimates, regardless of Paxlovid inclusion. This pattern suggests that restricting the comparator population to statin users resulted in a more clinically homogeneous reference group, namely, individuals with hyperlipidemia and baseline risk for myopathy/rhabdomyolysis, thereby enhancing the ability to detect true DDI signals. As lopinavir/ritonavir reporting slightly increased during the pandemic, co-prescription with statins likely became more common, increasing the relative proportion of reports in which the hypothesized interaction was detectable.

Prior to the pandemic, clinicians may have actively avoided co-prescribing atorvastatin with antiviral agents or adjusted dosing regimens to mitigate DDI risk, potentially explaining the absence of a detectable signal in pre-COVID active comparator analyses. In the high-pressure clinical context of COVID-19, however, treatment urgency and limited therapeutic options may have led to less conservative prescribing behavior, resulting in stronger DDI signal emergence in the during-pandemic period for lopinavir/ritonavir.

The largest increase in signal strength during-pandemic-19 was observed in the CYP3A4 statin-restricted comparator group. While the overall trend mirrored that of the broader statin-restricted comparator, the pre-COVID signal estimates were notably lower in this subgroup. This discrepancy may reflect the inherently higher baseline risk for myopathy or rhabdomyolysis among CYP3A4-metabolized statins, such as simvastatin and lovastatin[129], which could reduce the relative contrast needed to detect a signal. Alternatively, as we've previously reported in another study, the smaller sample size and reduced cardinality of the comparator set may have

amplified the influence of reporting biases, making shifts in signal estimates more pronounced[53,115].

Despite the clinical plausibility of the interaction between lopinavir/ritonavir and atorvastatin, differences in signal detection results were observed across methodological approaches. With Omega Shrinkage, the signal was detectable in the unrestricted design pre-COVID but absent in the active comparator setting, likely reflecting a higher baseline risk of myopathy among statin users, which reduces relative contrast when used as the comparator. The broader, more heterogeneous reference set in the unrestricted design may have amplified the risk, making the signal more prominent. Conversely, the BCPNN method failed to detect a signal in either comparator group pre-COVID. This may be attributed to the distinct statistical assumptions underpinning each method: Omega Shrinkage was developed under a baseline additive interaction model, whereas the BCPNN under a multiplicative interaction model. In addition, the Dirichlet of the BCPNN prior may have exerted a higher rate of shrinkage given the full specification of the hyperparameters. These structural differences influence the sensitivity of each method to signal emergence, particularly in the context of changing background risk and reporting behavior. For the CSS signals were detectable across all DPAs, regardless of comparator or timeframe. This is expected, as CSS does not apply shrinkage to the estimate. While the CSS signal increased throughout all analyses, the increase in URD post-covid was lower in the analyses including Paxlovid©. As such, the direction of signal change remained largely consistent across all three statistical methods.

While the ACD aims to reduce confounding by indication and improve signal detection validity through a more homogenous reference group, the loss of signal from URD to ACD is somewhat counterintuitive under the assumption of a true interaction. This attenuation likely

reflects the fact that all statins carry a baseline risk of myopathy; thus, using other statins as comparators narrows the contrast in risk. In some sense, implemented ACD would theoretically decrease the signal, unless PK interaction was much stronger than the baseline myopathy reporting rate in the comparator.

Our study has several strengths. First, we leveraged a large-scale SR database, which provided sufficient power to assess changes in signal strength across defined comparator groups and time periods. Second, by focusing on a clinically established drug-drug interaction (lopinavir/ritonavir and atorvastatin), we were able to evaluate signal behavior in a well-defined scenario of utilization shift during the COVID-19 pandemic. The analytic framework used, contrasting unrestricted and active comparator designs, enabled us to isolate the potential influence of reporting bias stemming from broader exposure and heightened public awareness. To our knowledge, this is the first study to explicitly evaluate the impact of reporting bias on DDI signal detection using pre- and during-pandemic stratification.

However, our findings should be interpreted in light of several limitations. Most notably, while we attributed changes in signal strength to shifts in reporting behavior induced by the COVID-19 pandemic, other unmeasured factors, such as changes in reporting patterns of other medications/events, may also have contributed. We attempted to minimize this impact by implementing a calendar time-based design and making no additional alterations to the analysis framework beyond time period stratification. Nonetheless, residual biases may persist. Additionally, while our selected DDI example is illustrative of how changes in utilization and reporting can distort signal detection, the findings may not generalize to other drug combinations or therapeutic contexts. The extent and nature of reporting bias are highly context-dependent and should be evaluated on a case-by-case basis in future pharmacovigilance research.

12. Conclusion

Our study demonstrates that changes in drug indication over time, such as those brought on by the COVID-19 pandemic, may alter reporting patterns and can meaningfully influence DDI signal detection in SR systems. As shown in our analysis, shifts in exposure and reporting behavior can impact DDI signals, with the direction and magnitude of this effect varying by comparator reference set definition. Researchers should exercise caution when interpreting disproportionality results and carefully select comparator groups that align with the biological, prescribing, and reporting context of the DDI under investigation.

CHAPTER IV: Conclusions and Implications

This thesis provides one of the first extensive evaluations of how pandemic-induced shifts in drug indication leading to alterations in utilization patterns and reporting behaviors can influence DDI signal detection. Using the COVID-19 pandemic, a clinically relevant DDI between Atorvastatin and lopinavir/ritonavir for myopathy/rhabdomyolysis, and leveraging the extensive data available through the FAERS database, we evaluated DPA estimates across both unrestricted and comparator-restricted designs. Our analyses revealed that changes in population exposure and reporting context meaningfully impact the detection of DDIs, and that the direction and magnitude of these impacts vary depending on the comparator group and signal detection method used.

We observed that during the pandemic period, increased use and reporting of ritonavir were accompanied by changes in the concomitant medications, and reduced the DDI signal strength in unrestricted design at least partly due to increased background reporting brought on by repurposing of ritonavir to treat COVID-19 in combination with nirmatrelvir. However, by restricting comparator groups to clinically relevant populations, such as statin or CYP3A4 statin users, signals for the known DDI with atorvastatin and myopathy became more prominent during-pandemic. These findings highlight the dual importance of comparator design and external contextual factors when interpreting pharmacovigilance signals.

Signal detection remains a critical component of post-market safety surveillance, particularly in the period immediately following regulatory approval, in between mandated PSURs, or in the absence of confirmatory real-world evidence. For many newly authorized medications, especially those expedited during public health crises, SR may be the only available source of safety data in the short term. As such, the generation of refined, interpretable, and reporting context-aware signals is essential. These early signals often inform hypotheses for

further epidemiological studies, guide resource allocation, and influence regulatory decision-making. A flawed or biased signal can mislead investigators or delay needed action, whereas a well-calibrated one can accelerate targeted surveillance and ensure patient safety.

Future research should extend this work to other known DDIs, therapeutic areas, and bias contexts, assess the impact of comparator selection across diverse patient populations, and explore algorithmic approaches to detect and adjust for stimulated reporting periods introduced by reporting bias. Validation using structured data sources such as electronic health records or administrative claims databases will also be useful to triangulate findings and reduce residual uncertainty. By better understanding and accounting for the biases inherent in SR data, pharmacovigilance researchers can continue to improve the utility, precision, and actionability of safety signal detection effort.

Reference

1. US FDA. The Drug Development Process. FDA. February 20, 2020. Accessed April 7, 2025. https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process

2. Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research. *BJOG Int J Obstet Gynaecol*. 2018;125(13):1716. doi:10.1111/1471-0528.15199

3. Karanicolas PJ, Farrokhyar F, Bhandari M. Blinding: Who, what, when, why, how? *Can J Surg*. 2010;53(5):345-348.

4. Allocation concealment in randomised controlled trials: are we getting better? | The BMJ. Accessed April 7, 2025. https://www.bmj.com/content/355/bmj.i5663

5. Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. *BMJ*. 2006;332(7547):969-971.

6. Farmer RE, Kounali D, Walker AS, et al. Application of causal inference methods in the analyses of randomised controlled trials: a systematic review. *Trials*. 2018;19:23. doi:10.1186/s13063-017-2381-x

7. Committee on Ethical and Scientific Issues in Studying the Safety ofApproved Drugs, Board on Population Health and Public Health Practice, Institute of Medicine. *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*. National Academies Press (US); 2012. Accessed December 23, 2024. http://www.ncbi.nlm.nih.gov/books/NBK200905/

8. McNaughton R, Huet G, Shakir S. An investigation into drug products withdrawn from the EU market between 2002 and 2011 for safety reasons and the evidence used to support the decision-making. *BMJ Open*. 2014;4(1):e004221. doi:10.1136/bmjopen-2013-004221

9. Canada H. Drug and medical device post market surveillance. May 8, 2018. Accessed December 23, 2024. https://www.canada.ca/en/services/health/drug-health-products/drug-medical-device-highlights-2017/post-market-surveillance.html

10. FDA. Step 5: FDA Post-Market Drug Safety Monitoring. *FDA*. Published online April 18, 2019. Accessed December 23, 2024. https://www.fda.gov/patients/drug-development-process/step-5-fda-post-market-drug-safety-monitoring

11. Pharmacovigilance: post-authorisation | European Medicines Agency (EMA). Accessed December 23, 2024. https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/pharmacovigilance-post-authorisation

12. WHO. Regulation and Prequalification. Accessed December 23, 2024. https://www.who.int/teams/regulation-prequalification/incidents-and-SF/post-market-surveillance

13. Weltgesundheitsorganisation, Collaborating Centre for International Drug Monitoring, eds. *The Importance of Pharmacovigilance: Safety Monitoring of Medicinal Products*. WHO [u.a.]; 2002.

14. Health Canada. The Preparing and Submitting Summary Reports for Marketed Drugs and Natural Health Products - Guidance Document for Industry. May 15, 2018. Accessed December 23, 2024. https://www.canada.ca/en/health-canada/services/drugs-health-products/reports-publications/medeffect-canada/preparing-submitting-summary-reports-marketed-drugs-natural-health-products-guidance-industry.html

15. FDA. Providing Postmarket Periodic Safety Reports in the ICH E2C(R2) Format (Periodic Benefit-Risk Evaluation Report). December 19, 2024. Accessed December 23, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-postmarket-periodic-safety-reports-ich-e2cr2-format-periodic-benefit-risk-evaluation

16. EMA. Periodic safety update reports (PSURs) | European Medicines Agency (EMA). Accessed December 23, 2024. https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/pharmacovigilance-post-authorisation/periodic-safety-update-reports-psurs

17. Health Canada. Guidance Document - Submission of Risk Management Plans and Follow-up Commitments. June 26, 2015. Accessed May 7, 2025. https://www.canada.ca/en/health-canada/services/drugs-health-products/reports-publications/medeffect-canada/guidance-document-submission-risk-management-plans-follow-commitments.html

18. EMA. Risk management plans | European Medicines Agency (EMA). April 17, 2013. Accessed May 7, 2025. https://www.ema.europa.eu/en/human-regulatory-overview/marketing-authorisation/pharmacovigilance-marketing-authorisation/risk-management/risk-management-plans

19. FDA. Risk Evaluation and Mitigation Strategies | REMS. FDA. May 16, 2023. Accessed May 7, 2025. https://www.fda.gov/drugs/drug-safety-and-availability/risk-evaluation-and-mitigation-strategies-rems

20. Health Canada. Questions and Answers Regarding the Implementation of a Risk-Prioritized Periodic Safety Update Report Regulatory Review Pilot (PSUR-RRP) at Health Canada. November 22, 2010. Accessed May 7, 2025. https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/product-vigilance/questions-answers-implementation-risk-prioritized-periodic-safety-update-report-regulatory-review-pilot-psur-drug-products.html

21. Health Canada. Post-Authorization Requirements. August 3, 2004. Accessed December 23, 2024. https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/clinical-trials/post-authorization-requirements.html

22. FDA. Post-Approval Studies Program. FDA. August 9, 2024. Accessed December 23, 2024. https://www.fda.gov/medical-devices/postmarket-requirements-devices/post-approval-studies-program

23. EMA. Post-authorisation safety studies (PASS) | European Medicines Agency (EMA). Accessed December 23, 2024. https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/pharmacovigilance-post-authorisation/post-authorisation-safety-studies-pass

24. Roche. Roche | What is a Phase 4 Clinical Trial? And Why is it Important? Accessed December 23, 2024. https://www.roche.com/stories/phase-4-clinical-trials

25. Nissen SE, Wolski K. Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes. *N Engl J Med*. 2007;356(24):2457-2471. doi:10.1056/NEJMoa072761

26. Rosiglitazone Evaluated for Cardiovascular Outcomes — An Interim Analysis | New England Journal of Medicine. Accessed December 23, 2024. https://www.nejm.org/doi/full/10.1056/NEJMoa073394

27. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet Lond Engl*. 2009;373(9681):2125-2135. doi:10.1016/S0140-6736(09)60953-3

28. FDA. Postmarketing Safety Reporting for Human Drug and Biological Products Including Vaccines. April 27, 2020. Accessed May 7, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postmarketing-safety-reporting-human-drug-and-biological-products-including-vaccines

29. Ralph Edwards I. Spontaneous reporting—of what? Clinical concerns about drugs. *Br J Clin Pharmacol*. 1999;48(2):138-141. doi:10.1046/j.1365-2125.1999.00000.x

30. Goldman SA. Limitations and strengths of spontaneous reports data. *Clin Ther*. 1998;20:C40-C44. doi:10.1016/S0149-2918(98)80007-6

31. Sakai T. [Role and Applicability of Spontaneous Reporting Databases in Medical Big Data]. *Yakugaku Zasshi*. 2021;141(2):165-168. doi:10.1248/yakushi.20-00196-1

32. FDA. Individual Case Safety Reports. FDA. February 14, 2025. Accessed May 8, 2025. https://www.fda.gov/industry/fda-data-standards-advisory-board/individual-case-safety-reports

33. Health Canada. Reporting adverse reactions to marketed health products - Guidance document for industry. November 7, 2022. Accessed May 8, 2025. https://www.canada.ca/en/health-canada/services/drugs-health-products/reports-publications/medeffect-canada/reporting-adverse-reactions-marketed-health-products-guidance-industry/guidance-document.html

34. EMA. Mandatory use of ISO/ICH E2B(R3) individual case safety reporting in the EU: hands-on training course using the EudraVigilance system | European Medicines Agency (EMA). June 5, 2024. Accessed May 8, 2025. https://www.ema.europa.eu/en/events/mandatory-use-iso-ich-e2br3-individual-case-safety-reporting-eu-hands-training-course-using-eudravigilance-system-45

35. FDA. FDA Adverse Event Reporting System (FAERS) Public Dashboard. *FDA*. Published online December 7, 2023. Accessed December 23, 2024. https://www.fda.gov/drugs/fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard

36. EMA. EudraVigilance | European Medicines Agency (EMA). April 3, 2024. Accessed December 23, 2024. https://www.ema.europa.eu/en/human-regulatory-overview/research-development/pharmacovigilance-research-development/eudravigilance

37. Health Canada. Overview of Vanessa's Law. October 30, 2014. Accessed May 8, 2025. https://www.canada.ca/en/health-canada/services/drugs-health-products/legislation-guidelines/overview-vannessa-law-protecting-canadians-unsafe-drugs-act-vanessa-law-amendments-food-drugs-act.html

38. Health Canada. Canada Vigilance Program. January 24, 2008. Accessed December 23, 2024. https://www.canada.ca/en/health-canada/services/drugs-health-products/medeffect-canada/canada-vigilance-program.html

39. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions : a systematic review. *Drug Saf*. 2006;29(5):385-396. doi:10.2165/00002018-200629050-00003

40. Gliklich RE, Dreyer NA, Leavy MB. Adverse Event Detection, Processing, and Reporting. In: *Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 3rd Edition*. Agency for Healthcare Research and Quality (US); 2014. Accessed May 8, 2025. https://www.ncbi.nlm.nih.gov/books/NBK208615/

41. Sjölander A. Selection Bias with Outcome-dependent Sampling. *Epidemiol Camb Mass*. 2023;34(2):186-191. doi:10.1097/EDE.0000000000001567

42. Gravel C. *Statistical Methods for Signal Detection in Pharmacovigilance*. Carleton University; 2009. Accessed May 8, 2025. https://hdl.handle.net/20.500.14718/37434

43. Ahmad SR. Adverse Drug Event Monitoring at the Food and Drug Administration. *J Gen Intern Med*. 2003;18(1):57-60. doi:10.1046/j.1525-1497.2003.20130.x

44. Wisniewski AFZ, Bate A, Bousquet C, et al. Good Signal Detection Practices: Evidence from IMI PROTECT. *Drug Saf*. 2016;39(6):469-490. doi:10.1007/s40264-016-0405-1

45. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol*. 1999;149(11):981-983. doi:10.1093/oxfordjournals.aje.a009758

46. Alkabbani W, Gamble JM. Active-comparator restricted disproportionality analysis for pharmacovigilance signal detection studies of chronic disease medications: An example using sodium/glucose cotransporter 2 inhibitors. *Br J Clin Pharmacol*. 2023;89(2):431-439. doi:10.1111/bcp.15178

47. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med*. 1991;10(4):577-581. doi:10.1002/sim.4780100409

48. Pariente A, Gregoire F, Fourrier-Reglat A, Haramburu F, Moore N. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases: the notoriety bias. *Drug Saf*. 2007;30(10):891-898. doi:10.2165/00002018-200730100-00007

49. Neha R, Subeesh V, Beulah E, Gouri N, Maheswari E. Existence of Notoriety Bias in FDA Adverse Event Reporting System Database and Its Impact on Signal Strength. *Hosp Pharm*. 2021;56(3):152-158. doi:10.1177/0018578719882323

50. Dijkstra L. Adverse drug reaction or innocent bystander? A systematic comparison of statistical discovery methods for spontaneous reporting systems - Dijkstra - 2020 - Pharmacoepidemiology and Drug Safety - Wiley Online Library. Accessed May 8, 2025. https://onlinelibrary.wiley.com/doi/10.1002/pds.4970?msockid=371100b59359650e3d6015c292f36401

51. Hoffman KB, Dimbil M, Erdman CB, Tatonetti NP, Overstreet BM. The Weber Effect and the United States Food and Drug Administration's Adverse Event Reporting System (FAERS): Analysis of Sixty-Two Drugs Approved from 2006 to 2010. *Drug Saf*. 2014;37(4):283-294. doi:10.1007/s40264-014-0150-2

52. Arora A, Jalali RK, Vohora D. Relevance of the Weber effect in contemporary pharmacovigilance of oncology drugs. *Ther Clin Risk Manag*. 2017;13:1195-1203. doi:10.2147/TCRM.S137144

53. Bai W, Douros A, Gravel CA. Masking in Active Comparator Designs in Pharmacovigilance: A Retrospective Bias Analysis on the Spontaneous Reporting of Thiazolidinediones and Cardiovascular Events. Accessed December 23, 2024. https://www.authorea.com/users/829042/articles/1223168-masking-in-active-comparator-designs-in-pharmacovigilance-a-retrospective-bias-analysis-on-the-spontaneous-reporting-of-thiazolidinediones-and-cardiovascular-events

54. Hauben M, Maignen F. Does serious consequential masking exist? An update. *Pharmacoepidemiol Drug Saf*. 2017;26(6):727-729. doi:10.1002/pds.4209

55. Maignen F, Hauben M, Dogné JM. A mathematical framework to quantify the masking effect associated with the confidence intervals of measures of disproportionality. *Ther Adv Drug Saf*. 2017;8(7):231-244. doi:10.1177/2042098617704143

56. Pariente A, Avillach P, Salvo F, et al. Effect of competition bias in safety signal generation: analysis of a research database of spontaneous reports in France. *Drug Saf*. 2012;35(10):855-864. doi:10.1007/BF03261981

57. Kim JH, Scialli AR. Thalidomide: the tragedy of birth defects and the effective treatment of disease. *Toxicol Sci Off J Soc Toxicol*. 2011;122(1):1-6. doi:10.1093/toxsci/kfr088

58. Seidman LA, Warren N. Frances Kelsey & Thalidomide in the US: A Case Study Relating to Pharmaceutical Regulations. *Am Biol Teach*. 2002;64(7):495-500. doi:10.2307/4451354

59. Greene JA, Podolsky SH. Reform, Regulation, and Pharmaceuticals — The Kefauver–Harris Amendments at 50. *N Engl J Med*. 2012;367(16):1481-1483. doi:10.1056/NEJMp1210007

60. WHO. Programme for International Drug Monitoring. Accessed April 7, 2025. https://www.who.int/teams/regulation-prequalification/regulation-and-safety/pharmacovigilance/networks/pidm

61. U.S Federal Government. 21 CFR 600.80 -- Postmarketing reporting of adverse experiences. Accessed May 8, 2025. https://www.ecfr.gov/current/title-21/part-600/section-600.80

62. FDA. Postmarketing Adverse Event Reporting Compliance Program. *FDA*. Published online August 9, 2024. Accessed May 8, 2025. https://www.fda.gov/drugs/surveillance/postmarketing-adverse-event-reporting-compliance-program

63. MedDra. Help to Shape the MedDRA Terminology | MedDRA. Accessed March 16, 2025. https://www.meddra.org/

64. Khaleel MA, Khan AH, Ghadzi SMS, Adnan AS, Abdallah QM. A Standardized Dataset of a Spontaneous Adverse Event Reporting System. *Healthcare*. 2022;10(3):420. doi:10.3390/healthcare10030420

65. MedDra. Standardised MedDRA Queries | MedDRA. Accessed May 9, 2025. https://www.meddra.org/standardised-meddra-queries

66. Moore N, Thiessard F, Begaud B. The history of disproportionality measures (reporting odds ratio, proportional reporting rates) in spontaneous reporting of adverse drug reactions. *Pharmacoepidemiol Drug Saf*. 2005;14(4):285-286. doi:10.1002/pds.1058

67. Rothman KJ, Lanes S, Sacks ST. The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiol Drug Saf*. 2004;13(8):519-523. doi:10.1002/pds.1001

68. Szumilas M. Explaining Odds Ratios. *J Can Acad Child Adolesc Psychiatry*. 2010;19(3):227-229.

69. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10(6):483-486. doi:10.1002/pds.677

70. Begaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? *Br J Clin Pharmacol*. 1994;38(5):401-404. doi:10.1111/j.1365-2125.1994.tb04373.x

71. Deshpande G, Gogolak V, Smith SW. Data mining in drug safety: review of published threshold criteria for defining signals of disproportionate reporting. *Pharm Med*. 2010;24(1):37-44.

72. Bate A. Bayesian confidence propagation neural network. *Drug Saf*. 2007;30(7):623-625. doi:10.2165/00002018-200730070-00011

73. Dumouchel W. Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. *Am Stat*. 1999;53(3):177-190. doi:10.1080/00031305.1999.10474456

74. Huang L, Guo T, Zalkikar JN, Tiwari RC. A Review of Statistical Methods for Safety Surveillance. *Ther Innov Regul Sci*. 2014;48(1):98-108. doi:10.1177/2168479013514236

75. Gravel C. Considerations on the use of different comparators in pharmacovigilance: A methodological review - Gravel - 2023 - British Journal of Clinical Pharmacology - Wiley Online Library. Accessed December 23, 2024. https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15802

76. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol*. 2015;11(7):437-441. doi:10.1038/nrrheum.2015.30

77. Polysubstance Use Profiles Among the General Adult Population, United States, 2022 | AJPH | Vol. 115 Issue 5. Accessed May 8, 2025. https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2024.307979?journalCode=ajph

78. Health Canada. Canadian Alcohol and Drugs Survey (CADS): summary of results for 2023. December 27, 2024. Accessed May 8, 2025. https://www.canada.ca/en/health-canada/services/canadian-alcohol-drugs-survey/2023-summary.html

79. Davies's textbook of adverse drug reactions - ProQuest. Accessed December 29, 2024. https://www.proquest.com/docview/195985206?sourcetype=Scholarly%20Journals

80. Levêque D, Lemachatti J, Nivoix Y, et al. [Mechanisms of pharmacokinetic drug-drug interactions]. *Rev Med Interne*. 2010;31(2):170-179. doi:10.1016/j.revmed.2009.07.009

81. Deodhar M, Al Rihani SB, Arwood MJ, et al. Mechanisms of CYP450 Inhibition: Understanding Drug-Drug Interactions Due to Mechanism-Based Inhibition in Clinical Practice. *Pharmaceutics*. 2020;12(9):846. doi:10.3390/pharmaceutics12090846

82. Niu J, Straubinger RM, Mager DE. Pharmacodynamic Drug-Drug Interactions. *Clin Pharmacol Ther*. 2019;105(6):1395-1406. doi:10.1002/cpt.1434

83. MedSafe. Reminder: Interactions Resulting in Serotonin Syndrome. Accessed May 8, 2025. https://www.medsafe.govt.nz/profs/PUArticles/Sep2015/InteractionsSerotoninSyndrome.htm

84. Hennessy S, Leonard C, Gagne J, et al. Pharmacoepidemiologic Methods for Studying the Health Effects of Drug-Drug Interactions (DDIs). *Clin Pharmacol Ther*. 2016;99(1):92-100. doi:10.1002/cpt.277

85. Raschi E. Spotlight commentary: The value of spontaneous reporting systems to detect (the lack of) clinically relevant drug–drug interactions in clinical practice - Raschi - 2023 - British Journal of Clinical Pharmacology - Wiley Online Library. Accessed December 29, 2024. https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/bcp.15780

86. Létinier L, Ferreira A, Marceron A, et al. Spontaneous Reports of Serious Adverse Drug Reactions Resulting From Drug–Drug Interactions: An Analysis From the French Pharmacovigilance Database. *Front Pharmacol*. 2021;11. doi:10.3389/fphar.2020.624562

87. Van Puijenbroek EP, Egberts AC, Meyboom RH, Leufkens HG. Signalling possible drug-drug interactions in a spontaneous reporting system: delay of withdrawal bleeding during concomitant use of oral contraceptives and itraconazole. *Br J Clin Pharmacol*. 1999;47(6):689-693. doi:10.1046/j.1365-2125.1999.00957.x

88. Chandwani A, Shuter J. Lopinavir/ritonavir in the treatment of HIV-1 infection: a review. *Ther Clin Risk Manag*. 2008;4(5):1023-1033.

89. FDA. Drug Approval Package. Accessed March 16, 2025. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2000/21-226_Kaletra.cfm

90. CAO B. A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19 | New England Journal of Medicine. Accessed April 8, 2025. https://www.nejm.org/doi/full/10.1056/NEJMoa2001282

91. Hammond J, Leister-Tebbe H, Gardner A, et al. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *N Engl J Med*. 2022;386(15):1397-1408. doi:10.1056/NEJMoa2118542

92. RECOVERY Collaborative Group. Lopinavir-ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet Lond Engl*. 2020;396(10259):1345-1352. doi:10.1016/S0140-6736(20)32013-4

93. FDA. FDA revises letter of authorization for the emergency use authorization for Paxlovid. *FDA*. Published online August 9, 2024. Accessed June 11, 2025. https://www.fda.gov/drugs/drug-safety-and-availability/fda-revises-letter-authorization-emergency-use-authorization-paxlovid

94. Park A. Why Biden's Test to Treat Program Still Won't Mean You Can Walk Into a Pharmacy and Walk out With a Prescription for a COVID-19 Pill. TIME. March 24, 2022. Accessed June 11, 2025. https://time.com/6159366/covid-19-drug-pharmacies-test-to-treat/

95. FDA. FDA Approves First Oral Antiviral for Treatment of COVID-19 in Adults. FDA. August 9, 2024. Accessed April 13, 2025. https://www.fda.gov/news-events/press-announcements/fda-approves-first-oral-antiviral-treatment-covid-19-adults

96. Adams AJ, Eid DD. Federal pharmacist Paxlovid prescribing authority: A model policy or impediment to optimal care? *Explor Res Clin Soc Pharm*. 2023;9:100244. doi:10.1016/j.rcsop.2023.100244

97. Hukkanen J, Puurunen J, Hyötyläinen T, et al. The effect of atorvastatin treatment on serum oxysterol concentrations and cytochrome P450 3A4 activity. *Br J Clin Pharmacol*. 2015;80(3):473-479. doi:10.1111/bcp.12701

98. Vinci P, Panizon E, Tosoni LM, et al. Statin-Associated Myopathy: Emphasis on Mechanisms and Targeted Therapy. *Int J Mol Sci*. 2021;22(21):11687. doi:10.3390/ijms222111687

99. McIver LA, Siddique MS. Atorvastatin. In: *StatPearls*. StatPearls Publishing; 2025. Accessed May 9, 2025. http://www.ncbi.nlm.nih.gov/books/NBK430779/

100. Chastain DB, Stover KR, Riche DM. Evidence-based review of statin use in patients with HIV on antiretroviral therapy. *J Clin Transl Endocrinol*. 2017;8:6-14. doi:10.1016/j.jcte.2017.01.004

101. Torres PA, Helmstetter JA, Kaye AM, Kaye AD. Rhabdomyolysis: Pathogenesis, Diagnosis, and Treatment. *Ochsner J*. 2015;15(1):58-69.

102. Harati Y. Myopathy - an overview | ScienceDirect Topics. Accessed April 8, 2025. https://www.sciencedirect.com/topics/medicine-and-dentistry/myopathy

103. Sarkar S, Brown TT. Table 5. [Interaction of Antiretroviral Therapy and Statins]. January 21, 2023. Accessed March 16, 2025. https://www.ncbi.nlm.nih.gov/books/NBK567198/table/lipid_hiv.T.interaction_of_antiretroviral/

104. Matyori A, Brown CP, Ali A, Sherbeny F. Statins utilization trends and expenditures in the U.S. before and after the implementation of the 2013 ACC/AHA guidelines. *Saudi Pharm J SPJ*. 2023;31(6):795-800. doi:10.1016/j.jsps.2023.04.002

105. Zequn Z, Yujia W, Dingding Q, Jiangfang L. Off-label use of chloroquine, hydroxychloroquine, azithromycin and lopinavir/ritonavir in COVID-19 risks prolonging the QT interval by targeting the hERG channel. *Eur J Pharmacol*. 2021;893:173813. doi:10.1016/j.ejphar.2020.173813

106. Owa AB, Owa OT. Lopinavir/ritonavir use in Covid-19 infection: is it completely non-beneficial? *J Microbiol Immunol Infect*. 2020;53(5):674-675. doi:10.1016/j.jmii.2020.05.014

107. Hauben M, Hung E, Chen Y. Potential Signals of COVID-19 as an Effect Modifier of Adverse Drug Reactions. *Clin Ther*. 2024;46(1):20-29. doi:10.1016/j.clinthera.2023.10.002

108. Salvo F, Leborgne F, Thiessard F, Moore N, Bégaud B, Pariente A. A potential event-competition bias in safety signal detection: results from a spontaneous reporting research database in France. *Drug Saf*. 2013;36(7):565-572. doi:10.1007/s40264-013-0063-5

109. Battini V, Cocco M, Barbieri MA, et al. Timing Matters: A Machine Learning Method for the Prioritization of Drug-Drug Interactions Through Signal Detection in the FDA Adverse Event Reporting System and Their Relationship with Time of Co-exposure. *Drug Saf*. 2024;47(9):895-907. doi:10.1007/s40264-024-01430-8

110. Magro L, Arzenton E, Leone R, et al. Identifying and Characterizing Serious Adverse Drug Reactions Associated With Drug-Drug Interactions in a Spontaneous Reporting Database. *Front Pharmacol*. 2020;11:622862. doi:10.3389/fphar.2020.622862

111. Noguchi Y, Tachi T, Teramachi H. Review of Statistical Methodologies for Detecting Drug-Drug Interactions Using Spontaneous Reporting Systems. *Front Pharmacol*. 2019;10:1319. doi:10.3389/fphar.2019.01319

112. Norén GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug–drug interaction surveillance. *Stat Med*. 2008;27(16):3057-3070. doi:10.1002/sim.3247

113. Norén GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Stat Med*. 2006;25(21):3740-3757. doi:10.1002/sim.2473

114. Noguchi Y, Aoyama K, Kubo S, Tachi T, Teramachi H. Improved Detection Criteria for Detecting Drug-Drug Interaction Signals Using the Proportional Reporting Ratio. *Pharm Basel Switz*. 2020;14(1):4. doi:10.3390/ph14010004

115. Gravel CA, Bai W, Douros A. Comparators in Pharmacovigilance: A Quasi-Quantification Bias Analysis. *Drug Saf*. 2024;47(8):809-819. doi:10.1007/s40264-024-01433-5

116. Thakrar BT, Grundschober SB, Doessegger L. Detecting signals of drug–drug interactions in a spontaneous reports database. *Br J Clin Pharmacol*. 2007;64(4):489-495. doi:10.1111/j.1365-2125.2007.02900.x

117. Jeong E, Su Y, Li L, Chen Y. Discovering clinical drug-drug interactions with known pharmacokinetics mechanisms using spontaneous reporting systems and electronic health records. *J Biomed Inform*. 2024;153:104639. doi:10.1016/j.jbi.2024.104639

118. Raschi E, Poluzzi E, De Ponti F. Spotlight commentary: The value of spontaneous reporting systems to detect (the lack of) clinically relevant drug–drug interactions in clinical practice. *Br J Clin Pharmacol*. 2023;89(8):2365-2368. doi:10.1111/bcp.15780

119. Gravel CA, Krewski D, Mattison DR, Momoli F, Douros A. Concomitant use of statins and sodium-glucose co-transporter 2 inhibitors and the risk of myotoxicity reporting: A disproportionality analysis. *Br J Clin Pharmacol*. 2023;89(8):2430-2445. doi:10.1111/bcp.15711

120. Montes-Grajales D, Garcia-Serna R, Mestres J. Impact of the COVID-19 pandemic on the spontaneous reporting and signal detection of adverse drug events. *Sci Rep*. 2023;13:18817. doi:10.1038/s41598-023-46275-w

121. Hauben M, Hung E. Effects of the COVID-19 Pandemic on Spontaneous Reporting: Global and National Time-series Analyses. *Clin Ther*. 2021;43(2):360-368.e5. doi:10.1016/j.clinthera.2020.12.008

122. Center for Drug Evaluation and. FDA Adverse Event Reporting System (FAERS) Public Dashboard. *FDA*. Accessed July 20, 2024. https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard

123. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Medical Dictionary for Regulatory Activities. 2024. Accessed July 29, 2024. https://www.meddra.org/

124. FDA. Statins. *FDA*. Published online October 3, 2018. Accessed April 1, 2025. https://www.fda.gov/drugs/information-drug-class/statins

125. Neuvonen P. Drug interactions with lipid-lowering drugs: Mechanisms and clinical relevance. Accessed April 1, 2025. https://oce.ovid.com/article/00003098-200612000-00001

126. U.S. Statistics. HIV.gov. Accessed May 12, 2025. https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics

127. Schwartz IS, Boulware DR, Lee TC. Hydroxychloroquine for COVID19: The curtains close on a comedy of errors. *Lancet Reg Health - Am*. 2022;11:100268. doi:10.1016/j.lana.2022.100268

128. Costa C, Abeijon P, Rodrigues DA, Figueiras A, Herdeiro MT, Torre C. Factors associated with underreporting of adverse drug reactions by patients: a systematic review. *Int J Clin Pharm*. 2023;45(6):1349-1358. doi:10.1007/s11096-023-01592-y

129. Devold HM, Molden E, Skurtveit S, Furu K. Co-medication of statins and CYP3A4 inhibitors before and after introduction of new reimbursement policy. *Br J Clin Pharmacol*. 2009;67(2):234-241. doi:10.1111/j.1365-2125.2008.03345.x

130. Fusaroli M, Salvo F, Khouri C, Raschi E. The reporting of disproportionality analysis in pharmacovigilance: spotlight on the READUS-PV guideline. *Front Pharmacol*. 2024;15. doi:10.3389/fphar.2024.1488725

131. CDC. CDC Museum COVID-19 Timeline. Centers for Disease Control and Prevention. July 8, 2024. Accessed March 31, 2025. https://www.cdc.gov/museum/timeline/covid19.html

132. Drug Approval Package. Accessed March 16, 2025. https://www.accessdata.fda.gov/drugsatfda_docs/nda/99/020702-S018_Lipitor.cfm

133. Gilani B, Cassagnol M. Biochemistry, Cytochrome P450. In: *StatPearls*. StatPearls Publishing; 2025. Accessed April 1, 2025. http://www.ncbi.nlm.nih.gov/books/NBK557698/

134. Empirical bayes screening for multi-item associations | Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. Accessed March 17, 2025. https://dl.acm.org/doi/10.1145/502512.502526

135. Large-scale regression-based pattern discovery: The example of screening the WHO global drug safety database - Caster - 2010 - Statistical Analysis and Data Mining: An ASA Data Science Journal - Wiley Online Library. Accessed March 17, 2025. https://onlinelibrary.wiley.com/doi/10.1002/sam.10078

136. R: The R Project for Statistical Computing. Accessed March 17, 2025. https://www.r-project.org/

Appendix

Supplementary Table S1. Adverse Event Definition (SMQ)

Adverse Event	SMQ pts
Myopathy/Rhabdomyolysis	Diabetic myonecrosis
	Exertional rhabdomyolysis
	Hypothyroid myopathy
	Muscle infarction
	Muscle necrosis
	Myoglobin blood increased
	Myoglobin blood present
	Myoglobin urine present
	Myoglobinaemia
	Myoglobinuria
	Myopathy
	Myopathy toxic
	Necrotising myositis
	Rhabdomyolysis
	Thyrotoxic myopathy
	Acute kidney injury
	Anuria
	Biopsy muscle abnormal
	Blood calcium decreased
	Blood creatine phosphokinase abnormal
	Blood creatine phosphokinase increased
	Blood creatine phosphokinase MM increased
	Blood creatinine abnormal
	Blood creatinine increased
	Chromaturia
	Chronic kidney disease
	Compartment syndrome
	Creatinine renal clearance abnormal
	Creatinine renal clearance decreased
	Diaphragm muscle weakness
	Electromyogram abnormal
	End stage renal disease
	Glomerular filtration rate abnormal
	Glomerular filtration rate decreased
	Haematoma muscle
	Hypercreatininaemia
	Hypocalcaemia
	Muscle discomfort
	Muscle disorder
	Muscle enzyme abnormal
	Muscle enzyme increased
	Muscle fatigue
	Muscle haemorrhage
	Muscle rupture
	Muscle strength abnormal
	Muscular weakness
	Musculoskeletal discomfort
	Musculoskeletal disorder
	Musculoskeletal pain
	Musculoskeletal toxicity
	Myalgia
	Myalgia intercostal
	Myositis
	Oliguria
	Renal failure
	Renal impairment
	Renal tubular necrosis
	Subacute kidney injury
	Tendon discomfort

Supplementary Table S2. Exposure Definition

Active Ingredient	Mapped Drug Names	Drug Set
RITONAVIR	KALETRA	Study Drug
LOPINAVIR	KALETRA	Study Drug
ATORVASTATIN	ATORVALIQ, LIPITOR	Study Drug
FLUVASTATIN	LESCOL	Statin comparator
LOVASTATIN	ALTOPREV, MEVACOR	Statin comparator (CYP3A4)
PITAVASTATIN	LIVALO, ZYPITAMAG	Statin comparator
PRAVASTATIN	PRAVACHOL	Statin comparator
ROSUVASTATIN	CRESTOR, EZALLOR SPRINKLE	Statin comparator
SIMVASTATIN	ZOCOR	Statin comparator (CYP3A4)
LOVASTATIN/NACIN	ADVICOR	Statin comparator
SIMVASTATIN/NACIN	SIMCOR	Statin comparator
SIMVASTATIN/EZTIMIBE	VYTORIN	Statin comparator

Supplementary Table S3. Top-10 co-reported medications yearly with ritonavir in FAERS from 2000 to 2023, including paxlovid.

year	Top Co-Reported Medications	year	Top Co-Reported Medications
2000	STAVUDINE (453 ,50%)	2012	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (626 ,33%)
	LAMIVUDINE (417 ,46%)		ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (412 ,21%)
	DIDANOSINE (284 ,31%)		DARUNAVIR\DARUNAVIR ETHANOLATE (393 ,20%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (202 ,22%)		LAMIVUDINE\ZIDOVUDINE (300 ,16%)
	SAQUINAVIR (198 ,22%)		ISENTRESS\RALTEGRAVIR\RALTEGRAVIR POTASSIUM (278 ,14%)
	EFAVIRENZ (164 ,18%)		LAMIVUDINE (276 ,14%)
	INDINIVIR SULFATE (158 ,17%)		TENOFOVIR DISOPROXIL FUMARATE (223 ,12%)
	ZIDOVUDINE (148 ,16%)		ZIDOVUDINE (192 ,10%)
	INDINAVIR SULFATE (140 ,15%)		SULFAMETHOXAZOLE\TRIMETHOPRIM (148 ,8%)
	AMPRENAVIR (135 ,15%)		ETRAVIRINE (126 ,7%)
2001	STAVUDINE (569 ,45%)	2013	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (751 ,35%)
	LAMIVUDINE (482 ,38%)		DARUNAVIR\DARUNAVIR ETHANOLATE (478 ,22%)
	DIDANOSINE (458 ,36%)		ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (371 ,17%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (302 ,24%)		LAMIVUDINE (265 ,12%)
	EFAVIRENZ (244 ,19%)		ISENTRESS\RALTEGRAVIR\RALTEGRAVIR POTASSIUM (257 ,12%)
	ZIDOVUDINE (212 ,17%)		LAMIVUDINE\ZIDOVUDINE (241 ,11%)
	ABACAVIR\ABACAVIR SULFATE (190 ,15%)		ZIDOVUDINE (224 ,10%)
	NEVIRAPINE (186 ,15%)		TENOFOVIR DISOPROXIL FUMARATE (217 ,10%)
	LAMIVUDINE\ZIDOVUDINE (184 ,15%)		DARUNAVIR (186 ,9%)
	AMPRENAVIR (177 ,14%)		SULFAMETHOXAZOLE\TRIMETHOPRIM (145 ,7%)
2002	STAVUDINE (469 ,39%)	2014	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (493 ,20%)
	LAMIVUDINE (458 ,38%)		LAMIVUDINE (435 ,18%)
	DIDANOSINE (437 ,36%)		EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (369 ,15%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (267 ,22%)		DARUNAVIR ETHANOLATE (349 ,14%)
	TENOFOVIR DISOPROXIL FUMARATE (211 ,18%)		ZIDOVUDINE (314 ,13%)
	EFAVIRENZ (201 ,17%)		ATAZANAVIR SULFATE (296 ,12%)
	LAMIVUDINE\ZIDOVUDINE (163 ,14%)		DARUNAVIR (271 ,11%)
	ZIDOVUDINE (159 ,13%)		LAMIVUDINE\ZIDOVUDINE (266 ,11%)
	AMPRENAVIR (145 ,12%)		ABACAVIR SULFATE\LAMIVUDINE (234 ,9%)
	ABACAVIR\ABACAVIR SULFATE (140 ,12%)		RALTEGRAVIR POTASSIUM (227 ,9%)

2003	LAMIVUDINE (505 ,41%)	**2015**	RIBAVIRIN (3200 ,47%)
	STAVUDINE (382 ,31%)		EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (912 ,13%)
	DIDANOSINE (343 ,28%)		ATAZANAVIR SULFATE (789 ,12%)
	TENOFOVIR DISOPROXIL FUMARATE (298 ,24%)		DASABUVIR (778 ,11%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (264 ,22%)		DARUNAVIR ETHANOLATE (702 ,10%)
	EFAVIRENZ (174 ,14%)		LAMIVUDINE (632 ,9%)
	ZIDOVUDINE (156 ,13%)		ABACAVIR SULFATE\LAMIVUDINE (438 ,6%)
	TENOFOVIR (150 ,12%)		TENOFOVIR DISOPROXIL FUMARATE (392 ,6%)
	LAMIVUDINE\ZIDOVUDINE (149 ,12%)		DARUNAVIR (363 ,5%)
	ABACAVIR SULFATE (126 ,10%)		LAMIVUDINE\ZIDOVUDINE (331 ,5%)
2004	LAMIVUDINE (543 ,38%)	**2016**	RIBAVIRIN (4627 ,51%)
	TENOFOVIR DISOPROXIL FUMARATE (383 ,27%)		DASABUVIR (1370 ,15%)
	DIDANOSINE (355 ,25%)		LAMIVUDINE (760 ,8%)
	STAVUDINE (313 ,22%)		EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (648 ,7%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (266 ,19%)		DARUNAVIR (601 ,7%)
	LAMIVUDINE\ZIDOVUDINE (221 ,16%)		DARUNAVIR ETHANOLATE (568 ,6%)
	EFAVIRENZ (180 ,13%)		ZIDOVUDINE (460 ,5%)
	TENOFOVIR (171 ,12%)		ATAZANAVIR SULFATE (450 ,5%)
	ENFUVIRTIDE (136 ,10%)		ASPIRIN (424 ,5%)
	ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (134 ,9%)		AMLODIPINE BESYLATE (400 ,4%)
2005	LAMIVUDINE (581 ,35%)	**2017**	RIBAVIRIN (929 ,26%)
	TENOFOVIR DISOPROXIL FUMARATE (453 ,28%)		DASABUVIR (720 ,20%)
	DIDANOSINE (389 ,24%)		EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (533 ,15%)
	STAVUDINE (268 ,16%)		LAMIVUDINE (512 ,14%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (267 ,16%)		DARUNAVIR ETHANOLATE (482 ,14%)
	LAMIVUDINE\ZIDOVUDINE (234 ,14%)		DARUNAVIR (286 ,8%)
	ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (229 ,14%)		ATAZANAVIR SULFATE (283 ,8%)
	TENOFOVIR (227 ,14%)		ZIDOVUDINE (259 ,7%)
	ZIDOVUDINE (212 ,13%)		TENOFOVIR (243 ,7%)
	EFAVIRENZ (192 ,12%)		EMTRICITABINE (235 ,7%)
2006	LAMIVUDINE (580 ,32%)	**2018**	DASABUVIR (885 ,24%)
	TENOFOVIR DISOPROXIL FUMARATE (338 ,19%)		EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (669 ,18%)
	DIDANOSINE (337 ,19%)		ATAZANAVIR SULFATE (623 ,17%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (302 ,17%)		RIBAVIRIN (604 ,16%)
	ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (289 ,16%)		LAMIVUDINE (584 ,16%)

	LAMIVUDINE\ZIDOVUDINE (288 ,16%)		DARUNAVIR ETHANOLATE (576 ,15%)
	ZIDOVUDINE (280 ,16%)		DARUNAVIR (513 ,14%)
	STAVUDINE (244 ,14%)		RALTEGRAVIR POTASSIUM (511 ,14%)
	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (242 ,13%)		ZIDOVUDINE (395 ,11%)
	ENFUVIRTIDE (196 ,11%)		TENOFOVIR DISOPROXIL FUMARATE (391 ,10%)
2007	LAMIVUDINE (474 ,27%)	**2019**	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (783 ,30%)
	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (375 ,21%)		LAMIVUDINE (723 ,27%)
	TENOFOVIR DISOPROXIL FUMARATE (326 ,18%)		DARUNAVIR ETHANOLATE (613 ,23%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (316 ,18%)		TENOFOVIR DISOPROXIL FUMARATE (489 ,19%)
	LAMIVUDINE\ZIDOVUDINE (299 ,17%)		ATAZANAVIR SULFATE (447 ,17%)
	DIDANOSINE (262 ,15%)		RALTEGRAVIR POTASSIUM (404 ,15%)
	ZIDOVUDINE (244 ,14%)		ZIDOVUDINE (388 ,15%)
	ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (238 ,13%)		DARUNAVIR (368 ,14%)
	ENFUVIRTIDE (178 ,10%)		SULFAMETHOXAZOLE\TRIMETHOPRIM (296 ,11%)
	DARUNAVIR\DARUNAVIR ETHANOLATE (168 ,9%)		EFAVIRENZ (275 ,10%)
2008	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (583 ,25%)	**2020**	HYDROXYCHLOROQUINE (1095 ,23%)
	LAMIVUDINE (450 ,19%)		EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (915 ,19%)
	LAMIVUDINE\ZIDOVUDINE (365 ,16%)		LAMIVUDINE (877 ,18%)
	TENOFOVIR DISOPROXIL FUMARATE (309 ,13%)		AZITHROMYCIN ANHYDROUS (767 ,16%)
	ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (298 ,13%)		DARUNAVIR ETHANOLATE (610 ,13%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (286 ,12%)		TENOFOVIR DISOPROXIL FUMARATE (604 ,12%)
	ZIDOVUDINE (243 ,10%)		CEFTRIAXONE (542 ,11%)
	DARUNAVIR\DARUNAVIR ETHANOLATE (237 ,10%)		RALTEGRAVIR POTASSIUM (503 ,10%)
	EFAVIRENZ (197 ,8%)		ZIDOVUDINE (472 ,10%)
	DIDANOSINE (190 ,8%)		ACETAMINOPHEN (444 ,9%)
2009	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (734 ,30%)	**2021**	HYDROXYCHLOROQUINE (1011 ,29%)
	LAMIVUDINE (562 ,23%)		LAMIVUDINE (753 ,21%)
	TENOFOVIR DISOPROXIL FUMARATE (430 ,18%)		EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE (552 ,16%)
	ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (377 ,15%)		AZITHROMYCIN ANHYDROUS (505 ,14%)
	ZIDOVUDINE (349 ,14%)		EFAVIRENZ (406 ,11%)
	DARUNAVIR\DARUNAVIR ETHANOLATE (344 ,14%)		DARUNAVIR (405 ,11%)
	LAMIVUDINE\ZIDOVUDINE (339 ,14%)		ATAZANAVIR SULFATE (396 ,11%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (303 ,12%)		DARUNAVIR ETHANOLATE (389 ,11%)

	DIDANOSINE (301 ,12%)		RALTEGRAVIR POTASSIUM (354 ,10%)
	ENFUVIRTIDE (267 ,11%)		ZIDOVUDINE (347 ,10%)
2010	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (612 ,29%)	**2022**	ATORVASTATIN (1254 ,9%)
	ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (403 ,19%)		ACETAMINOPHEN (958 ,7%)
	LAMIVUDINE\ZIDOVUDINE (366 ,17%)		LEVOTHYROXINE (864 ,6%)
	LAMIVUDINE (356 ,17%)		ASPIRIN (808 ,6%)
	TENOFOVIR DISOPROXIL FUMARATE (316 ,15%)		LISINOPRIL (776 ,6%)
	ZIDOVUDINE (305 ,14%)		CHOLECALCIFEROL (775 ,6%)
	DARUNAVIR\DARUNAVIR ETHANOLATE (282 ,13%)		AMLODIPINE BESYLATE (763 ,5%)
	ISENTRESS\RALTEGRAVIR\RALTEGRAVIR POTASSIUM (251 ,12%)		LAMIVUDINE (692 ,5%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (214 ,10%)		METFORMIN HYDROCHLORIDE (655 ,5%)
	EFAVIRENZ (169 ,8%)		OMEPRAZOLE (627 ,5%)
2011	EMTRICITABINE\TENOFOVIR DISOPROXIL FUMARATE\TRUVADA (652 ,36%)	**2023**	ATORVASTATIN (548 ,9%)
	ATAZANAVIR\ATAZANAVIR SULFATE\REYATAZ (421 ,23%)		LAMIVUDINE (444 ,7%)
	LAMIVUDINE (294 ,16%)		ACETAMINOPHEN (414 ,7%)
	DARUNAVIR\DARUNAVIR ETHANOLATE (290 ,16%)		LEVOTHYROXINE (351 ,6%)
	TENOFOVIR DISOPROXIL FUMARATE (266 ,15%)		AMLODIPINE BESYLATE (319 ,5%)
	LAMIVUDINE\ZIDOVUDINE (253 ,14%)		ASPIRIN (312 ,5%)
	ZIDOVUDINE (201 ,11%)		TACROLIMUS (300 ,5%)
	ISENTRESS\RALTEGRAVIR\RALTEGRAVIR POTASSIUM (164 ,9%)		LISINOPRIL (274 ,4%)
	SULFAMETHOXAZOLE\TRIMETHOPRIM (153 ,8%)		LOSARTAN (271 ,4%)
	ATAZANAVIR (147 ,8%)		METFORMIN HYDROCHLORIDE (258 ,4%)

Supplementary Table S4. CSS signal

n111	n110	n11.	n101	n100	n10.	n011	n010	n01.	n001	n000	n00.	n...	year	design	PRR_D1	PRR_D2	PRR_comb	CRR	chi_sq	PRR_lb_combo	CSS
180	3330	3510	434	29034	29468	40888	556887	597775	250675	11046786	11297461	11928214	2000_2019	urd	0.7596	3.080868	4.027193	1.307162	104.3195	3.492892	1.122337
180	184	364	434	433	867	40888	55892	96780	250675	354064	604739	702750	2000_2019	acd statin	1.2001	1.019566	1.719884	1.433118	8.974935	1.550152	1.221256
180	56	236	434	74	508	40888	32407	73295	250675	209897	460572	534611	2000_2019	acd cyp3a4 statin	1.51111	1.025527	1.837277	1.215846	43.65702	1.711002	1.095357
261	1183	1444	127	6449	6576	7497	103765	111262	47306	2486791	2534097	2653379	2020_2021Q3	urd	2.335273	3.686981	16.88438	4.579459	1806.911	15.12376	4.007664
261	127	388	127	156	283	7497	6817	14314	47306	44656	91962	106947	2020_2021Q3	acd statin	1.121347	1.026208	1.759991	1.569532	37.62206	1.641574	1.371962
261	23	284	127	47	174	7497	2996	10493	47306	20593	67899	78850	2020_2021Q3	acd cyp3a4 statin	1.211808	1.033109	1.533564	1.265518	64.08562	1.481073	1.175253
302	1503	1805	164	8213	8377	9948	138103	148051	63964	3531733	3595697	3753930	2020_2022	urd	2.318166	3.844109	16.42934	4.2739	2015.432	14.8189	3.777879
302	154	456	164	191	355	9948	8724	18672	63964	59992	123956	143439	2020_2022	acd statin	1.108804	1.038762	1.725769	1.556423	37.28633	1.615976	1.373314
302	29	331	164	55	219	9948	3723	13671	63964	26812	90776	104997	2020_2022	acd cyp3a4 statin	1.197303	1.038732	1.496873	1.250203	65.90856	1.4474	1.166492
320	1875	2195	196	10934	11130	13810	195952	209762	94414	5234767	5329181	5552268	2020_2023	urd	1.981919	3.762911	14.48316	3.848924	1814.823	13.08608	3.418563
320	180	500	196	226	422	13810	12256	26066	94414	86548	180962	207950	2020_2023	acd statin	1.070593	1.019713	1.653941	1.544884	27.07366	1.548509	1.365711
320	30	350	196	59	255	13810	5176	18986	94414	36607	131021	150612	2020_2023	acd cyp3a4 statin	1.182176	1.013967	1.459374	1.234481	63.67335	1.413088	1.156244
365	2870	3235	1655	46043	47698	13765	194957	208722	92955	5199658	5292613	5552268	2020_2023_2	urd	2.04444	3.762911	11.17605	2.970056	1460.818	10.14482	2.6502
365	242	607	1655	1728	3383	13765	12194	25959	92955	85046	178001	207950	2020_2023_2	acd statin	0.967559	1.019713	1.545943	1.516057	14.68695	1.44879	1.403678
365	49	414	1655	545	2200	13765	5157	18922	92955	36121	129076	150612	2020_2023_2	acd cyp3a4 statin	1.071657	1.013967	1.403746	1.309883	51.92432	1.354883	1.23798
180	3330	3510	434	29034	29468	40888	556887	597775	250675	11046786	11297461	11928214	2000_2019	urd	0.7596	3.080868	4.027193	1.307162	104.3195	3.492892	1.122337
180	184	364	434	433	867	40888	55892	96780	250675	354064	604739	702750	2000_2019	acd statin	1.2001	1.019566	1.719884	1.433118	8.974935	1.550152	1.221256
180	56	236	434	74	508	40888	32407	73295	250675	209897	460572	534611	2000_2019	acd cyp3a4 statin	1.51111	1.025527	1.837277	1.215846	43.65702	1.711002	1.095357
261	1183	1444	127	6449	6576	7497	103765	111262	47306	2486791	2534097	2653379	2020_2021Q3	urd	2.335273	3.686981	16.88438	4.579459	1806.911	15.12376	4.007664
261	127	388	127	156	283	7497	6817	14314	47306	44656	91962	106947	2020_2021Q3	acd statin	1.121347	1.026208	1.759991	1.569532	37.62206	1.641574	1.371962
261	23	284	127	47	174	7497	2996	10493	47306	20593	67899	78850	2020_2021Q3	acd cyp3a4 statin	1.211808	1.033109	1.533564	1.265518	64.08562	1.481073	1.175253

Supplementary Table S5. Omega Shrinkage signal

n111	n110	n11.	n101	n100	n10.	n011	n010	n01.	n001	n000	n00.	n...	year	design	omega	omega_lb_fr	omega_ub_fr	omega_lb_by	omega_ub_by
180	434	614	184	433	617	40888	250675	291563	55892	354064	409956	702750	2000_2019	acd statin	-0.03699	-0.24776	0.173771	-0.25554	0.16597
180	434	614	3330	29034	32364	40888	250675	291563	556887	11046786	11603673	11928214	2000_2019	urd	0.65977	0.449007	0.870533	0.441225	0.862732
180	434	614	56	74	130	40888	250675	291563	32407	209897	242304	534611	2000_2019	acd cyp3a4 statin	-0.56332	-0.77408	-0.35255	-0.78186	-0.36035
261	127	388	1183	6449	7632	7497	47306	54803	103765	2486791	2590556	2653379	2020_2021Q3	urd	1.537493	1.362464	1.712522	1.357094	1.70714
261	127	388	127	156	283	7497	47306	54803	6817	44656	51473	106947	2020_2021Q3	acd statin	0.576948	0.401919	0.751977	0.396549	0.746595
261	127	388	23	47	70	7497	47306	54803	2996	20593	23589	78850	2020_2021Q3	acd cyp3a4 statin	1.005653	0.830624	1.180682	0.825254	1.1753
302	164	466	1503	8213	9716	9948	63964	73912	138103	3531733	3669836	3753930	2020_2022	urd	1.487513	1.324798	1.650228	1.320157	1.645576
302	164	466	154	191	345	9948	63964	73912	8724	59992	68716	143439	2020_2022	acd statin	0.52691	0.364196	0.689625	0.359554	0.684973
302	164	466	29	55	84	9948	63964	73912	3723	26812	30535	104997	2020_2022	acd cyp3a4 statin	0.877312	0.714597	1.040027	0.709956	1.035374
320	196	516	1875	10934	12809	13810	94414	108224	195952	5234767	5430719	5552268	2020_2023	urd	1.497935	1.339863	1.656007	1.335482	1.651616
320	196	516	180	226	406	13810	94414	108224	12256	86548	98804	207950	2020_2023	acd statin	0.478616	0.320544	0.636688	0.316163	0.632298
320	196	516	30	59	89	13810	94414	108224	5176	36607	41783	150612	2020_2023	acd cyp3a4 statin	0.868566	0.710494	1.026638	0.706113	1.022248
365	1655	2020	2870	46043	48913	13765	92955	106720	194957	5199658	5394615	5552268	2020_2023_2	urd	0.293087	0.145079	0.441095	0.141238	0.437245
365	1655	2020	242	1728	1970	13765	92955	106720	12194	85046	97240	207950	2020_2023_2	acd statin	0.485576	0.337568	0.633584	0.333728	0.629734
365	1655	2020	49	545	594	13765	92955	106720	5157	36121	41278	150612	2020_2023_2	acd cyp3a4 statin	0.485576	0.337568	0.633584	0.333728	0.629734
180	434	614	184	433	617	40888	250675	291563	55892	354064	409956	702750	2000_2019	acd statin	-0.03699	-0.24776	0.173771	-0.25554	0.16597
180	434	614	3330	29034	32364	40888	250675	291563	556887	11046786	11603673	11928214	2000_2019	urd	0.65977	0.449007	0.870533	0.441225	0.862732
180	434	614	56	74	130	40888	250675	291563	32407	209897	242304	534611	2000_2019	acd cyp3a4 statin	-0.56332	-0.77408	-0.35255	-0.78186	-0.36035
261	127	388	1183	6449	7632	7497	47306	54803	103765	2486791	2590556	2653379	2020_2021Q3	urd	1.537493	1.362464	1.712522	1.357094	1.70714
261	127	388	127	156	283	7497	47306	54803	6817	44656	51473	106947	2020_2021Q3	acd statin	0.576948	0.401919	0.751977	0.396549	0.746595
261	127	388	23	47	70	7497	47306	54803	2996	20593	23589	78850	2020_2021Q3	acd cyp3a4 statin	1.005653	0.830624	1.180682	0.825254	1.1753
302	164	466	1503	8213	9716	9948	63964	73912	138103	3531733	3669836	3753930	2020_2022	urd	1.487513	1.324798	1.650228	1.320157	1.645576

Bai 2025

Supplementary Table S6. BCPNN signal

n111	n110	n11.	n101	n100	n10.	n011	n010	n01.	n001	n000	n00.	n...	year	design	IC	IC_LB	IC_UB
180	3330	3510	434	29034	29468	40888	556887	597775	250675	11046786	11297461	11928214	2000_2019	urd	-0.01674	-0.24765	0.21417
180	184	364	434	433	867	40888	55892	96780	250675	354064	604739	702750	2000_2019	acd statin	-0.03471	-0.24257	0.173139
180	56	236	434	74	508	40888	32407	73295	250675	209897	460572	534611	2000_2019	acd cyp3a4 statin	-0.14576	-0.3417	0.050178
261	1183	1444	127	6449	6576	7497	103765	111262	47306	2486791	2534097	2653379	2020_2021Q3	urd	0.174424	0.001977	0.346872
261	127	388	127	156	283	7497	6817	14314	47306	44656	91962	106947	2020_2021Q3	acd statin	0.185405	0.045765	0.325044
261	23	284	127	47	174	7497	2996	10493	47306	20593	67899	78850	2020_2021Q3	acd cyp3a4 statin	0.072256	-0.05376	0.198274
302	1503	1805	164	8213	8377	9948	138103	148051	63964	3531733	3595697	3753930	2020_2022	urd	0.082319	-0.07795	0.242587
302	154	456	164	191	355	9948	8724	18672	63964	59992	123956	143439	2020_2022	acd statin	0.156854	0.027086	0.286623
302	29	331	164	55	219	9948	3723	13671	63964	26812	90776	104997	2020_2022	acd cyp3a4 statin	0.055168	-0.06194	0.172274
320	1875	2195	196	10934	11130	13810	195952	209762	94414	5234767	5329181	5552268	2020_2023	urd	0.145106	-0.01895	0.309165
320	180	500	196	226	422	13810	12256	26066	94414	86548	180962	207950	2020_2023	acd statin	0.168322	0.042272	0.294372
320	30	350	196	59	255	13810	5176	18986	94414	36607	131021	150612	2020_2023	acd cyp3a4 statin	0.077904	-0.03584	0.191653
365	2870	3235	1655	46043	47698	13765	194957	208722	92955	5199658	5292613	5552268	2020_2023_2	urd	-0.25811	-0.41169	-0.10453
365	242	607	1655	1728	3383	13765	12194	25959	92955	85046	178001	207950	2020_2023_2	acd statin	0.222982	0.077246	0.368718
365	49	414	1655	545	2200	13765	5157	18922	92955	36121	129076	150612	2020_2023_2	acd cyp3a4 statin	0.168123	0.022387	0.313858
180	3330	3510	434	29034	29468	40888	556887	597775	250675	11046786	11297461	11928214	2000_2019	urd	-0.01674	-0.24765	0.21417
180	184	364	434	433	867	40888	55892	96780	250675	354064	604739	702750	2000_2019	acd statin	-0.03471	-0.24257	0.173139
180	56	236	434	74	508	40888	32407	73295	250675	209897	460572	534611	2000_2019	acd cyp3a4 statin	-0.14576	-0.3417	0.050178
261	1183	1444	127	6449	6576	7497	103765	111262	47306	2486791	2534097	2653379	2020_2021Q3	urd	0.174424	0.001977	0.346872
261	127	388	127	156	283	7497	6817	14314	47306	44656	91962	106947	2020_2021Q3	acd statin	0.185405	0.045765	0.325044
261	23	284	127	47	174	7497	2996	10493	47306	20593	67899	78850	2020_2021Q3	acd cyp3a4 statin	0.072256	-0.05376	0.198274

Registered OSF Protocol

Registration DOI: https://doi.org/10.17605/OSF.IO/94HNU

Reporting Bias in Drug-Drug Interaction Detection Using Spontaneous Reporting Systems: A Bias Analysis of the COVID-19 Impact

Authors: Chen Yang Bai[1], Antonios Douros[1,2,3], Christopher A. Gravel[1,2,4]

Affiliations:

School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada

Institute of Clinical Pharmacology and Toxicology, Charité - Universitätsmedizin Berlin, Berlin, Germany

Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada

Background

Reporting bias has been shown to impact signal detection studies in unexpected ways[53,75,115]. Previous efforts to evaluate and mitigate systematic and reporting biases in signal detection using SR databases have primarily focused on adverse drug reaction (ADR) detection in single Drug-Event-Combinations [44,53,75,130]. However, there has been minimal research on the impact of these biases in drug-drug interaction (DDI) detection to date.

The nature of a DDI - whether pharmacokinetic (PK) or pharmacodynamic (PD)—may influence the extent to which reporting bias introduces spurious findings in signal detection studies. One notable event that may have altered spontaneous reporting patterns was the widespread repurposing of antiviral medications for the treatment of COVID-19 during the early stages of the pandemic[107]. However, the impact of this event on the detection of DDI signals involving antiviral agents remains largely unknown. It is therefore plausible that reporting biases introduced during this period may have distorted DDI safety signals, potentially influencing evidence-based decision-making by regulators and the broader public.

Objective

This study aims to investigate the impact of reporting biases on DDI detection by conducting a retrospective bias analysis centered around the potential changes in reporting patterns before and after the COVID-19 pandemic using the U.S. FDA Adverse Event Reporting System (FAERS). We will investigate the following research questions:

We will investigate changes to the reporting patterns pre/post the pandemic for repurposed antivirals.

We will assess whether there were changes in the findings of DDI signal detection studies for these antivirals pre and post the COVID pandemic.

We will contrast DDI signal detection findings pre/post pandemic and compare against their expected results.

Methods

Data Source

We propose to use the FAERS database, which is a collection of suspected ADR reports maintained by the FDA which is updated quarterly and is mandatory for pharmaceutical companies and voluntary for healthcare professionals, patients, and other third parties[35]. FAERS data is anonymized, publicly accessible, and consists of seven linked data files containing information listed below:

- DEMO: Including patient demographics and administration information
- Drug: Reported drug(s) information

- Each drug is flagged with different suspicion levels: Primary Suspect (PS), Secondary Suspect (SS) or Concomitant medication (C).
- REAC: Reported adverse drug reaction(s)
- Adverse events in FAERS are coded using the Medical Dictionary for Regulatory Activities at the "preferred term" (pt) level.
- OUTC: Reported patient outcomes
- RPSR: Information about source(s) of reports
- THER: Therapy start and end date
- INDI: Information about the indication

The FAERS data structure is briefly summarized in **Supplementary Table S1**.

Study DDI

The DDI to be analyzed in this study involves antiviral medications (lopinavir/ritonavir) and statins (atorvastatin, rosuvastatin). Lopinavir and ritonavir are protease inhibitors primarily used for the treatment of HIV. Ritonavir is often co-administered with lopinavir as a pharmacokinetic enhancer due to its strong inhibition of cytochrome P450 3A4 (CYP3A4), which increases the plasma concentration of lopinavir and other co-administered drugs metabolized by CYP3A4[88].

Atorvastatin is widely prescribed statins for lowering cholesterol by inhibiting HMG-CoA reductase, an enzyme critical for cholesterol biosynthesis. Atorvastatin is primarily metabolized by CYP3A4[97], as such the concomitant use of lopinavir/ritonavir with atorvastatin is expected to lead to a pharmacokinetic DDI due to CYP3A4 inhibition by ritonavir[100]. This DDI should

increase atorvastatin plasma concentration and possibly increase the risk of statin-induced toxicity such as myopathy and, in severe cases, rhabdomyolysis[100].

Given this pharmacokinetic interaction, current clinical guidelines recommend either dose adjustment or the use of alternative statins with lower interaction potential when co-administering these medications[103]. **Table 1.** describes dosing recommendations published in endotext[103].

Statin	Antiretroviral Medication	Recommendation
Atorvastatin	Lopinavir/Ritonavir	Titrate atorvastatin dose carefully and administer the lowest effective dose while monitoring for toxicities. Do not exceed 20 mg atorvastatin daily.

Table 1. Dosing Recommendation of Atorvastatin in Co-prescription with Protease Inhibitors

Study Time Frame

Our study timeframe will span from July 1st, 2000 (2000Q3) to June 30th, 2023 (end of 2023Q2), aligning with the market authorization dates of the antiviral-statin DDI under investigation. The follow-up period for our statistical analysis will begin on the latest FDA market authorization date of the study medications and will end with the CDC's declaration of the end of the COVID-19 pandemic[131]. A detailed description of the medication and regulatory timeframe is provided in **Table 2.**

Drug	Market Authorization Date
Lopinavir	September 15th, 2000[89]
Ritonavir	June 19th, 1999[89]

Atorvastatin	December 2nd, 1999[132]

Table 2. Regulatory Approval of Study Drugs. The FDA regulatory approval date of study drugs have been added and referenced to the publicly available drug approval package.

Exposure Definitions

To account for alternative spellings and misspellings of reported product names, we will map the "drug name" field in FAERS to their respective active ingredients for all drugs included in the study[35]. This process will first be conducted for the study drugs – lopinavir/ritonavir (Kaletra©) and atorvastatin (Lipitor©) for the outcome of myopathy/rhabdomyolysis.

The first comparator in our signal detection algorithm will include all other drugs listed in FAERS reports without restriction. Second, To construct active comparator designs (ACD)—commonly used in pharmacovigilance to mitigate confounding by indication, where patients' baseline risk for an adverse event influences drug safety estimates[46]—we propose to first construct a comparator set of all FDA approved statins, which comprises of fluvastatin (Lescol©), lovastatin (Mevacor©) lovastatin extended-release (Altoprev©), pitavastatin (Livalo©), pravastatin (Pravachol©), rosuvastatin (Crestor©), and simvastatin (Zcor©). We will also include combination statin products, including lovastatin/niacin extended-release (Advicor©), simvastatin/nacin extended release (Simcor©), and simvastatin/ezetimibe (Vytorin©)[124]. As a further subgroup analysis, we will further define a ACD set comprising of only statins that are CYP3A4 inhibitors, including simvastatin (Zcor©) and lovastatin (Mevacor©)[124,125]. These approaches aimed to create a patient population with a similar risk profile for myopathy and rhabdomyolysis. However, previous studies have demonstrated the impact of residual reporting biases may be

magnified under restricted comparator sets such as the ACD due to decrease in sample size and the number of drugs (cardinality) in the analysis[53,75].

Finally, as part of sensitivity analyses, we will exclude reports with CYP3A4 inhibitors from the analyses, filtering out all reports with reported usage of amiodarone (Pacerone©), amitriptyline (Elavil©, Endep©, Vanatrip©), aprepitant (Aponvie, Cinvanti, Emend), carvedilol (Coreg©), chloramphenicol (Ocu-Chlor©), cimetidine (Tagamet©), ciprofloxacin (Cipro©, Proquin XR©), clarithromycin (Biaxin©), codeine, donepezil (Aricept©), fluvoxamine (Luvox©), haloperidol (Haldol©), imatinib (Gleevec©, Imkeldi©), ketoconazole (Nizoral©), metoprolol (Kapspargo Sprinkle©, Lopressor©, Toprol-XL©), paroxetine (Brisdelle©, Paxil©, Pexeva©), risperidone (Risperdal©), tramadol (ConZip©, Qdolo©, Ultram©), and verapamil (Calan©, Isoptin SR©, Cerelan©)[133]. The list of drug names to be mapped to active ingredients in our study can be found in **Supplementary Table S2.**

Adverse Event Definitions

Since adverse events in FAERS can be reported using multiple PTs, we will implement the broad and narrow Standard MedDRA Queries (SMQ)[63] to capture relevant cases of myopathy and rhabdomyolysis, comprising 59 PTs. The full list of PTs in the SMQ definition is provided in **Supplementary Table S3**.

Bias Setting

During the early stages of the COVID-19 pandemic, ritonavir and lopinavir were repurposed for COVID-19 treatment and frequently used off-label, in contrast to their indicated use for HIV[105]. A previous study looking at COVID-19 as a potential effect modifier of repurposed COVID-19 medications identified a signal between the reporting of these medications and COVID-19

infections during this period[107]. As a result, a much broader and more diverse population may have been exposed to these antivirals, potentially leading to different reporting behaviors compared to the originally indicated patient population, leading to distorted safety signals due to reporting biases.

To isolate relevant calendar time periods for the purpose of evaluating the impact of reporting bias potentially introduced by the COVID-19 pandemic, we will conduct 12 disproportionality analyses (denoted DPA1-12). First, we will create two non-overlapping timeframes within the follow-up period as follows:

2000Q3–2019Q4, aligning with the drug approval dates mentioned in **Section 3.3** and representing the pre-pandemic period (DPA 1).

2020Q1–2023Q3, corresponding to beginning and the end of COVID-19 pandemic per the Centre for Disease Control (DPA 4)[131].

To examine the effect of the active comparator set, we will implement the ACD as described in **Section 3.4** (DPA 2-3). Finally, we will combine both timeframe and comparator restrictions to evaluate the impact of COVID-19 within the ACD (DPA 5-6). As a sensitivity analysis, reports containing other commonly used CYP3A4 inhibitors will be excluded and each of the above analyses will be repeated (DPA 7-12). **Table 3** summarizes all proposed DPAs and the associated parameters.

DPA	Time Frame	Comparator
1	2000Q3 to 2019Q4	Full data
2	2000Q3 to 2019Q4	Statins only
3	2000Q3 to 2019Q4	Statins that are CYP3A4 substrate
4	2020Q1 to 2023Q3	Full data
5	2020Q1 to 2023Q3	Statins only
6	2020Q1 to 2023Q3	Statins that are CYP3A4 substrate
7*	2000Q3 to 2019Q4	Full data, excluding CYP3A4 inhibitors
8*	2000Q3 to 2019Q4	Statins only, excluding CYP3A4 inhibitors
9*	2000Q3 to 2019Q4	Statins that are CYP3A4 substrate; excluding CYP3A4 inhibitors
10*	2020Q1 to 2023Q3	Full data, excluding CYP3A4 inhibitors
11*	2020Q1 to 2023Q3	Statins only, excluding CYP3A4 inhibitors
12*	2020Q1 to 2023Q3	Statins that are CYP3A4 substrate; excluding CYP3A4 inhibitors

Table 3. Disproportionality analyses conducted in the pre- and post- COVID-19 timeframe. DPA denotes disproportionality analyses. Q denotes quarter. * Indicates this DPA is conducted as a sensitivity analysis.

Statistical Analysis

The following signal detection algorithms will be used to conduct DPAs within these 12 study settings. A broad set of signal detection methods are proposed to mitigate the influence of the statistical properties of a given algorithm on the findings.

- Omega Shrinkage (Ω)[112]
- Multi-item Gamma Poisson Shrinkage (MGPS)[134]
- Extended Bayesian Confidence Propagation Neural Network (E-BCPNN)[113]
- Bayesian Logistical Regression (LR-B)[135]
- Logistical Regression – Additive interaction (LR-A)[111]

- Log-Binomial Regression – Multiplicative interaction (LR-M)[111]
- Concomitant Signal Score (CSS)[111]

The description of the procedures can be found in **Table 4.** All statistical analysis will be completed in R, version 4.3[136].

Signal Detection Algorithm	Procedure
Omega Shrinkage	Construct 4x2 contingency table Compute Ω estimator Compute frequentist 95% CI Compute bayesian 95% CrI
MGPS	Using the openEBGM package: Compute counts and expected value Squash data Compute hyperparameters through maximization procedure Compute EBGM and associated 95% CrI
E-BCPNN	Construct 4x2 contingency table Compute information component Compute associated 95% CrI
LR-B	Construct analytic ready dataset (1/0 indicator, per id) Construct prior distribution (L1 penalty, Laplace prior) Construct logistical regression model with interaction Derive estimate and 95% CrI through MCMC
LR-A	Construct analytic ready dataset (1/0 indicator, per id) Construct logistical regression model with intereaction Derive estimate and 95% CI through MLE

LR-M	Construct analytic ready dataset (1/0 indicator, per id) Construct log-binomial regression model with interaction Derive estimate and 95% CI through MLE
CSS	Construct 4x2 contingency table Compute PRR of single drug-AE reaction and 95% CI Compute CSS

Table 4. Signal Detection Algorithms Applied in the Study. CI denotes confidence interval, CrI denotes credible interval, EBGM denotes empirical bayesian geometric mean, MCMC denotes monte-carlo Markov chain, MLE denotes maximum likelihood estimation. The available codes are uploaded to RPubs.

Bias Analysis

To evaluate the impact of biases, we will relatively compare the findings form the following analyses:

Before (DPA 1) vs after (DPA 4) COVID.

Impact of ACD implementation pre-COVID timeframe (DPA 1 vs 2 vs 3).

Impact of ACD implementation during-pandemic timeframe (DPA 4 vs 5 vs 6).

Impact of ACD implementation before (DPA 2,3) vs after (DPA 5,6) covid.

Impact of implementing the sensitivity analysis (DPA 1-3 vs DPA 7-8; DPA 4-6 vs DPA 9-12).

Additional Information

Anticipated Start Date

April 4th, 2025

Anticipated Completion Date

December 1st, 2025

Funding sources/sponsor

None

Conflict of Interest

None

Ethics Disclosure

None (usage of publicly available and anonymized data)

Analytic R Codes:

Script 1: Automated Extraction and Unzipping of Server Data

```
library(tools)

# Define input and output paths
zip_dir <- "H:/Chrome Download"
output_base <- "H:/DDI Analysis/bias analysis/raw/yearly"

# Identify and sort ZIP files by numeric order
zip_files <- list.files(zip_dir, pattern = "\\.zip$", full.names = TRUE)
zip_numbers <- as.integer(gsub("\\D", "", basename(zip_files)))
sorted_zip_files <- zip_files[order(zip_numbers)]

# Initialize progress bar
total_files <- length(sorted_zip_files)
pb <- txtProgressBar(min = 0, max = total_files, style = 3)

# Extract ZIP files sequentially into indexed folders
for (i in seq_along(sorted_zip_files)) {
  zip_path <- sorted_zip_files[i]
  output_folder <- file.path(output_base, as.character(i - 1))
  dir.create(output_folder, recursive = TRUE, showWarnings = FALSE)
  cat(sprintf("Extracting %s to %s\n", basename(zip_path), output_folder))
  unzip(zip_path, exdir = output_folder)
  setTxtProgressBar(pb, i)
}

close(pb)
cat("All files extracted.\n")

# Standardize CSV filenames within each extracted folder
```

```
for (i in 0:23) {
  folder <- file.path(output_base, as.character(i))
  files <- list.files(folder, pattern = "\\.csv$", full.names = TRUE)

  drug_file <- files[grepl("drug.*\\.csv$", basename(files), ignore.case = TRUE)]
  if (length(drug_file) == 1) {
    file.rename(drug_file, file.path(folder, "drug.csv"))
  } else if (length(drug_file) > 1) {
    warning(sprintf("Multiple drug CSV files found in folder %s", i))
  }

  reac_file <- files[grepl("reac.*\\.csv$", basename(files), ignore.case = TRUE)]
  if (length(reac_file) == 1) {
    file.rename(reac_file, file.path(folder, "reac.csv"))
  } else if (length(reac_file) > 1) {
    warning(sprintf("Multiple reac CSV files found in folder %s", i))
  }
}
```

Script 2: Mapping Drug Names

```
library(data.table)

# ====================== #
# Define Study Drug Sets
# ====================== #

# Study drugs of interest
study <- c("RITONAVIR", "LOPINAVIR", "ATORVASTATIN")

# All statins
statins <- c(
  "FLUVASTATIN", "LOVASTATIN", "PITAVASTATIN", "PRAVASTATIN",
  "ROSUVASTATIN", "SIMVASTATIN", "LOVASTATIN_NACIN",
  "SIMVASTATIN_NACIN", "SIMVASTATIN_EZETIMIBE"
)

# Known CYP3A4 inhibitors
cyp3a4_inhibitors <- c(
  "AMIODARONE", "AMITRIPTYLINE", "APREPITANT", "CARVEDILOL", "CHLORAMPHENICOL",
  "CIMETIDINE", "CIPROFLOXACIN", "CLARITHROMYCIN", "CODEINE", "DONEPEZIL",
  "FLUVOXAMINE", "HALOPERIDOL", "IMATINIB", "KETOCONAZOLE", "METOPROLOL",
  "PAROXETINE", "RISPERIDONE", "TRAMADOL", "VERAPAMIL"
)

# ============================ #
# Define Brand Name Mappings
# ============================ #

# Drug name aliases
ritonavir <- c("KALETRA")
lopinavir <- c("KALETRA")
```

```
atorvastatin <- c("ATORVALIQ", "LIPITOR")

# Statin brands
fluvastatin <- c("LESCOL")
lovastatin <- c("ALTOPREV", "MEVACOR")
pitavastatin <- c("LIVALO", "ZYPITAMAG")
pravastatin <- c("PRAVACHOL")
rosuvastatin <- c("CRESTOR", "EZALLOR SPRINKLE")
simvastatin <- c("ZOCOR")
lovastatin_nacin <- c("ADVIDCOR")
simvastatin_nacin <- c("SIMCOR")
simvastatin_ezetimibe <- c("VYTORIN")

# CYP3A4 inhibitor brands
amiodarone <- c("PACERONE")
amitriptyline <- c("ELAVIL", "ENDEP", "VANATRIP")
aprepitant <- c("APONVIE", "CINVANTI", "EMEND")
carvedilol <- c("COREG")
chloramphenicol <- c("OCU-CHLOR")
cimetidine <- c("TAGAMET")
ciprofloxacin <- c("CIPRO", "PROQUIN XR")
clarithromycin <- c("BIAXIN")
codeine <- c("CODEINE")
donepezil <- c("ARICEPT")
fluvoxamine <- c("LUVOX")
haloperidol <- c("HALDOL")
imatinib <- c("GLEEVEC", "IMKELD")
ketoconazole <- c("NIZORAL")
metoprolol <- c("KAPSPARGO SPRINKLE", "LOPRESSOR", "TOPROL-XL")
paroxetine <- c("BRISDELLE", "PAXIL", "PEXEVA")
risperidone <- c("RISPERDAL")
tramadol <- c("CONZIP", "QDOLO", "ULTRAM")
verapamil <- c("CALAN", "ISOPTIN SR", "CERELAN")
```

```r
# ================================== #
# Drug Mapping and Harmonization
# ================================== #

# Process data for each timeframe
for (time_frame in c("2000_2019", "2020_2023")) {
  cat(paste0("processing ", time_frame, "\n"))

  # Read drug file
  drug_file_path <- paste0("H://DDI Analysis//bias analysis//raw//", time_frame, "//drug.csv")
  cat("reading drug file \n")
  drug <- fread(drug_file_path)

  # Subset relevant columns
  cat("subsetting drug file \n")
  key_col <- if ("caseid" %in% colnames(drug)) "caseid" else "primaryid"
  drug <- drug[, .SD, .SDcols = c(key_col, "drugname", "trade_name", "active_ingredient", "role_cod")]
  gc()

  # Function to tag drugs by name or trade name
  tag_drug <- function(targets, label) {
    drug[grepl(paste(targets, collapse = "|"), active_ingredient, ignore.case = TRUE) |
         grepl(paste(targets, collapse = "|"), drugname, ignore.case = TRUE) |
         grepl(paste(targets, collapse = "|"), trade_name, ignore.case = TRUE),
         active_ingredient := label]
  }

  cat("Tagging study DDI drugs\n")
  tag_drug(ritonavir, "RITONAVIR")
  tag_drug(lopinavir, "LOPINAVIR")
  tag_drug(atorvastatin, "ATORVASTATIN")
```

```
cat("Tagging statins\n")
tag_drug(fluvastatin, "FLUVASTATIN")
tag_drug(lovastatin, "LOVASTATIN")
tag_drug(pitavastatin, "PITAVASTATIN")
tag_drug(pravastatin, "PRAVASTATIN")
tag_drug(rosuvastatin, "ROSUVASTATIN")
tag_drug(simvastatin, "SIMVASTATIN")
tag_drug(lovastatin_nacin, "LOVASTATIN_NACIN")
tag_drug(simvastatin_nacin, "SIMVASTATIN_NACIN")
tag_drug(simvastatin_ezetimibe, "SIMVASTATIN_EZETIMIBE")

cat("Tagging CYP3A4 inhibitors\n")
tag_drug(amiodarone, "AMIODARONE")
tag_drug(amitriptyline, "AMITRIPTYLINE")
tag_drug(aprepitant, "APREPITANT")
tag_drug(carvedilol, "CARVEDILOL")
tag_drug(chloramphenicol, "CHLORAMPHENICOL")
tag_drug(cimetidine, "CIMETIDINE")
tag_drug(ciprofloxacin, "CIPROFLOXACIN")
tag_drug(clarithromycin, "CLARITHROMYCIN")
tag_drug(codeine, "CODEINE")
tag_drug(donepezil, "DONEPEZIL")
tag_drug(fluvoxamine, "FLUVOXAMINE")
tag_drug(haloperidol, "HALOPERIDOL")
tag_drug(imatinib, "IMATINIB")
tag_drug(ketoconazole, "KETOCONAZOLE")
tag_drug(metoprolol, "METOPROLOL")
tag_drug(paroxetine, "PAROXETINE")
tag_drug(risperidone, "RISPERIDONE")
tag_drug(tramadol, "TRAMADOL")
tag_drug(verapamil, "VERAPAMIL")

# Drop original name fields
```

```
  drug[, c("drugname", "trade_name") := NULL]

  # Save processed drug file
  fwrite(drug, paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//drug.csv"))
  rm(drug)
  gc()
}

# ============================ #
# Merge Drug and Reaction Files
# ============================ #

for (time_frame in c("2000_2019", "2020_2023")) {
  drug <- fread(paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//drug.csv"))
  reac <- fread(paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//reac.csv"))

  # Merge on appropriate ID field
  id_field <- if ("caseid" %in% colnames(drug)) "caseid" else "primaryid"
  merged <- drug[reac, on = id_field, allow.cartesian = TRUE]
  merged <- unique(merged)

  fwrite(merged, paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "//merged.csv"))

  rm(drug, reac, merged)
}
```

Script 3: Merging Files

```
# ====================== #
# Merge Drug + Reaction #
# ====================== #

library(data.table)

# ---------- Pre-COVID Period (2000–2019) ---------- #

time_frame <- "2000_2019"
cat("Processing time frame:", time_frame, "\n")

# Read preprocessed drug and reaction files
drug <- fread(paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//drug.csv"))
reac <- fread(paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//reac.csv"))

# Merge on primaryid (many-to-many allowed)
merged <- drug[reac, on = "primaryid", allow.cartesian = TRUE]

# Remove duplicate rows after merge
merged <- unique(merged)

# Save merged file
fwrite(merged, paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "//merged.csv"))

# Clean up
rm(drug, reac, merged)
gc()



# ---------- During-pandemic Period (2020–2023) ---------- #
```

```
time_frame <- "2020_2023"
cat("Processing time frame:", time_frame, "\n")

# Read during-pandemic drug and reaction files
drug <- fread(paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//drug.csv"))
reac <- fread(paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//reac.csv"))

# Merge on primaryid (many-to-many allowed)
merged <- drug[reac, on = "primaryid", allow.cartesian = TRUE]

# Remove duplicate rows
merged <- unique(merged)

# Save merged output
fwrite(merged, paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "//merged.csv"))

# Clean up
rm(drug, reac, merged)
gc()
```

Script 4: Assigning SMQ

```
# ================================ #
```

```
# Define PTs for SMQ Aggregation #
# ================================ #

{
  # Standardize all terms to uppercase to ensure consistent matching
  myopathy_rhabdomyolysis_smq <- toupper(c(
    "Diabetic myonecrosis",
    "Exertional rhabdomyolysis",
    "Hypothyroid myopathy",
    "Muscle infarction",
    "Muscle necrosis",
    "Myoglobin blood increased",
    "Myoglobin blood present",
    "Myoglobin urine present",
    "Myoglobinaemia",
    "Myoglobinuria",
    "Myopathy",
    "Myopathy toxic",
    "Necrotising myositis",
    "Rhabdomyolysis",
    "Thyrotoxic myopathy",
    "Acute kidney injury",
    "Anuria",
    "Biopsy muscle abnormal",
    "Blood calcium decreased",
    "Blood creatine phosphokinase abnormal",
    "Blood creatine phosphokinase increased",
    "Blood creatine phosphokinase MM increased",
    "Blood creatinine abnormal",
    "Blood creatinine increased",
    "Chromaturia",
    "Chronic kidney disease",
    "Compartment syndrome",
```

"Creatinine renal clearance abnormal",
"Creatinine renal clearance decreased",
"Diaphragm muscle weakness",
"Electromyogram abnormal",
"End stage renal disease",
"Glomerular filtration rate abnormal",
"Glomerular filtration rate decreased",
"Haematoma muscle",
"Hypercreatininaemia",
"Hypocalcaemia",
"Muscle discomfort",
"Muscle disorder",
"Muscle enzyme abnormal",
"Muscle enzyme increased",
"Muscle fatigue",
"Muscle haemorrhage",
"Muscle rupture",
"Muscle strength abnormal",
"Muscular weakness",
"Musculoskeletal discomfort",
"Musculoskeletal disorder",
"Musculoskeletal pain",
"Musculoskeletal toxicity",
"Myalgia",
"Myalgia intercostal",
"Myositis",
"Oliguria",
"Renal failure",
"Renal impairment",
"Renal tubular necrosis",
"Subacute kidney injury",
"Tendon discomfort",
"Rhabdomyolysis"

```
  ))
}

# ================================ #
# Pre-COVID Period (2000–2019)    #
# ================================ #

# Set time frame
time_frame <- "2000_2019"
cat(paste0("Processing ", time_frame, "\n"))

library(data.table)

# Read raw reaction file
reac <- fread(paste0("H://DDI Analysis//bias analysis//raw//", time_frame, "//reac.csv"))

# Keep only primaryid and preferred term (PT)
cat("Subsetting reac file\n")
reac <- reac[, .(primaryid, pt)]

# Convert PTs to uppercase for standardized matching
reac$pt <- toupper(reac$pt)

# Reassign any PT matching SMQ definition to "MYOPATHY/RHABDOMYOLYSIS"
cat("Assigning SMQ label\n")
reac[pt %in% myopathy_rhabdomyolysis_smq, pt := "MYOPATHY/RHABDOMYOLYSIS"]

# Save the cleaned and standardized reaction file
fwrite(reac, paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//reac.csv"))



# ================================ #
# During-pandemic Period (2020–2023)   #
```

```
# ================================ #

# Set time frame
time_frame <- "2020_2023"
cat(paste0("Processing ", time_frame, "\n"))

# Read raw reaction file
reac <- fread(paste0("H://DDI Analysis//bias analysis//raw//", time_frame, "//reac.csv"))

# Keep only primaryid and preferred term (PT)
cat("Subsetting reac file\n")
reac <- reac[, .(primaryid, pt)]

# Convert PTs to uppercase for standardized matching
reac$pt <- toupper(reac$pt)

# Reassign any PT matching SMQ definition to "MYOPATHY/RHABDOMYOLYSIS"
cat("Assigning SMQ label\n")
reac[pt %in% myopathy_rhabdomyolysis_smq, pt := "MYOPATHY/RHABDOMYOLYSIS"]

# Save the cleaned and standardized reaction file
fwrite(reac, paste0("H://DDI Analysis//bias analysis//processed//", time_frame, "//reac.csv"))
```

Script 5: Counting Annual Data

```
library(data.table)
```

```
ritonavir <- c("PAXLOVID","RITONAVIR","KALETRA")
lopinavir <- c("PAXLOVID","LOPINAVIR","KALETRA")
atorvastatin <- c("ATORVASTATIN","ATORVALIQ", "LIPITOR")

myopathy_rhabdomyolysis_smq <- toupper(c(
  "Diabetic myonecrosis",
  "Exertional rhabdomyolysis",
  "Hypothyroid myopathy",
  "Muscle infarction",
  "Muscle necrosis",
  "Myoglobin blood increased",
  "Myoglobin blood present",
  "Myoglobin urine present",
  "Myoglobinaemia",
  "Myoglobinuria",
  "Myopathy",
  "Myopathy toxic",
  "Necrotising myositis",
  "Rhabdomyolysis",
  "Thyrotoxic myopathy",
  "Acute kidney injury",
  "Anuria",
  "Biopsy muscle abnormal",
  "Blood calcium decreased",
  "Blood creatine phosphokinase abnormal",
  "Blood creatine phosphokinase increased",
  "Blood creatine phosphokinase MM increased",
  "Blood creatinine abnormal",
  "Blood creatinine increased",
  "Chromaturia",
  "Chronic kidney disease",
  "Compartment syndrome",
  "Creatinine renal clearance abnormal",
```

```
  "Creatinine renal clearance decreased",
  "Diaphragm muscle weakness",
  "Electromyogram abnormal",
  "End stage renal disease",
  "Glomerular filtration rate abnormal",
  "Glomerular filtration rate decreased",
  "Haematoma muscle",
  "Hypercreatininaemia",
  "Hypocalcaemia",
  "Muscle discomfort",
  "Muscle disorder",
  "Muscle enzyme abnormal",
  "Muscle enzyme increased",
  "Muscle fatigue",
  "Muscle haemorrhage",
  "Muscle rupture",
  "Muscle strength abnormal",
  "Muscular weakness",
  "Musculoskeletal discomfort",
  "Musculoskeletal disorder",
  "Musculoskeletal pain",
  "Musculoskeletal toxicity",
  "Myalgia",
  "Myalgia intercostal",
  "Myositis",
  "Oliguria",
  "Renal failure",
  "Renal impairment",
  "Renal tubular necrosis",
  "Subacute kidney injury",
  "Tendon discomfort",
  "Rhabdomyolysis"
))
```

```
#Define the folder path
raw_folder <- ("H://DDI Analysis//bias analysis//raw//yearly//")
processed_folder <- ("H://DDI Analysis//bias analysis//processed//yearly//")


merge_and_clean <- function(drug,reac){ #change caseid, primaryid as necessary

 drug <- drug_table[,.(primaryid,role_cod,drugname,trade_name,active_ingredient)]

 reac <- reac_table[,.(primaryid,pt)] #only need these columns

 reac$pt <- toupper(reac$pt)

 cat("Assigning SMQ", "\n")

 reac[pt %in% myopathy_rhabdomyolysis_smq, pt := "MYOPATHY/RHABDOMYOLYSIS"]

 cat("Tagging Study DDI drugs", "\n")

 # Study DDI
 cat(" - RITONAVIR\n")
 drug[grepl(paste(ritonavir, collapse = "|"), active_ingredient, ignore.case = TRUE) |
       grepl(paste(ritonavir, collapse = "|"), drugname, ignore.case = TRUE) |
       grepl(paste(ritonavir, collapse = "|"), trade_name, ignore.case = TRUE),
     active_ingredient := "RITONAVIR"]
```

```
  cat(" - LOPINAVIR\n")
  drug[grepl(paste(lopinavir, collapse = "|"), active_ingredient, ignore.case = TRUE) |
         grepl(paste(lopinavir, collapse = "|"), drugname, ignore.case = TRUE) |
         grepl(paste(lopinavir, collapse = "|"), trade_name, ignore.case = TRUE),
       active_ingredient := "LOPINAVIR"]

  cat(" - ATORVASTATIN\n")
  drug[grepl(paste(atorvastatin, collapse = "|"), active_ingredient, ignore.case = TRUE) |
         grepl(paste(atorvastatin, collapse = "|"), drugname, ignore.case = TRUE) |
         grepl(paste(atorvastatin, collapse = "|"), trade_name, ignore.case = TRUE),
       active_ingredient := "ATORVASTATIN"]


  drug[,drugname:=NULL]
  drug[,trade_name:=NULL]


  merged <- drug[reac, on = "primaryid", allow.cartesian = TRUE] #merging demo_reac and demo_drug by caseid#

  merged$pt <- toupper(merged$pt)

  return(merged)
}

for (i in 0:23){
  cat("Processing Folder:", i, "\n")

  drug_table <- fread(paste0(raw_folder,i,"//drug.csv"))
  reac_table <- fread(paste0(raw_folder,i,"//reac.csv"))

  merged <- merge_and_clean(drug_table,reac_table)

  fwrite(merged, paste0(processed_folder, i, ".csv")) #Stored as 1.csv
```

```
  cat(sprintf("Saved merged file: %s/%s.csv\n", processed_folder, i))

}



################## Counting ####################

get_table <- function(d1, d2, ae, merged){
  merged$d1i <- ifelse(merged$active_ingredient %in% d1, 1, 0)
  merged$d2i <- ifelse(merged$active_ingredient == d2, 1, 0)
  merged$aei <- ifelse(merged$pt == ae, 1, 0)

  merged <- merged[, .(id, d1i, d2i, aei)]
  flat <- unique(merged[, .(d1i = max(d1i), d2i = max(d2i), aei = max(aei)), by = id])

  # Count subgroups with AE present
  n111 <- nrow(flat[d1i == 1 & d2i == 1 & aei == 1])
  n101 <- nrow(flat[d1i == 1 & d2i == 0 & aei == 1])
  n011 <- nrow(flat[d1i == 0 & d2i == 1 & aei == 1])
  n001 <- nrow(flat[d1i == 0 & d2i == 0 & aei == 1])

  # Totals for combinations (regardless of AE)
  n11. <- nrow(flat[d1i == 1 & d2i == 1])
  n10. <- nrow(flat[d1i == 1 & d2i == 0])
  n01. <- nrow(flat[d1i == 0 & d2i == 1])
  n00. <- nrow(flat[d1i == 0 & d2i == 0])

  # Derived margins
  n1.. <- n11. + n10.
  n.1. <- n11. + n01.
  n... <- nrow(flat)
```

```
  table <- data.table(
    n111 = n111,
    n11. = n11.,
    n101 = n101,
    n10. = n10.,
    n011 = n011,
    n01. = n01.,
    n1.. = n1..,
    n.1. = n.1.,
    n... = n...
  )

  return(table)
}



processed_folder <- ("H://DDI Analysis//bias analysis//processed//yearly//")

results_together <- data.table()

for (i in 0:23) {

  cat(paste0("iteration", i, "\n"))

  merged <- fread(paste0(processed_folder, paste0(i, ".csv")))

  merged[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINA
VIR"]
  merged[,pt:=toupper(pt)]
  merged[pt %in% myopathy_rhabdomyolysis_smq, pt := "MYOPATHY/RHABDOMYOLYSIS"]



  setnames(merged, "primaryid", "id")
```

```
  counts <- get_table("RITONAVIR/LOPINAVIR", "ATORVASTATIN", "MYOPATHY/RHABDOMYOLYSIS",
merged)


  counts$year <- 2000 + i


  results_together <- rbind(results_together, counts)
}


fwrite(results_together, "H://DDI Analysis//bias analysis//yearly counts//count.csv")




### plotting ###


library(data.table)
library(ggplot2)


# Load the data
result_together <- fread("H://DDI Analysis//bias analysis//yearly counts//count.csv")


# Calculate numeric percentages
result_together[, `:=`(
  n1_percent = 100 * `n1..` / `n...`,
  n_1_percent = 100 * `n.1.` / `n...`
)]


# Melt into long format
plot_data_long <- melt(
  result_together[, .(year, n1_percent, n_1_percent)],
  id.vars = "year",
  variable.name = "group",
  value.name = "percentage"
)
```

```
# Rename groups for legend clarity
plot_data_long[, group := factor(group,
                                 levels = c("n1_percent", "n_1_percent"),
                                 labels = c("Ritonavir/Lopinavir (%)", "Atorvastatin (%)")
)]

# Plot
ggplot(plot_data_long, aes(x = year, y = percentage, color = group)) +
  # COVID-19 period shading (2020+)
  annotate("rect", xmin = 2019.5, xmax = Inf, ymin = 0, ymax = Inf,
           fill = "orange", alpha = 0.2) +
  annotate("text", x = 2021, y = max(plot_data_long$percentage) * 1.05,
           label = "COVID-19", color = "black", size = 5, fontface = "italic") +

  geom_line(size = 1.2) +
  geom_point(size = 2) +
  labs(
    x = "Year",
    y = "Percentage of Total Reports",
    color = "Drug Exposure Group",
    title = "Exposure Trends for Ritonavir/Lopinavir and Atorvastatin"
  ) +
  scale_color_manual(values = c("#1b9e77", "#d95f02")) +
  theme_minimal(base_size = 14) +
  scale_x_continuous(breaks = result_together$year) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```r
library(data.table)

result_together <- fread("H://DDI Analysis//bias analysis//yearly counts//count.csv")


# Create formatted table with 2 decimal places for percentages
results_summary <- results_together[, .(
  year,
  `n111 (%)` = sprintf("%d (%.2f%%)", n111, 100 * n111 / n11.),
  `n11. (%)` = sprintf("%d (%.2f%%)", n11., 100 * n11. / n...),
  `n10. (%)` = sprintf("%d (%.2f%%)", n10., 100 * n10. / n...),
  `n01. (%)` = sprintf("%d (%.2f%%)", n01., 100 * n01. / n...),
  `n1.. (%)` = sprintf("%d (%.2f%%)", n1.., 100 * n1.. / n...),
  `n.1. (%)` = sprintf("%d (%.2f%%)", n.1., 100 * n.1. / n...),
  `n...` = n...
)]

# View the result
results_summary


library(ggplot2)

# Make long-form data for stacking
plot_data <- results_together[, .(
  year,
  AE = n111,
  No_AE = n11. - n111
)]

# Melt into long format
```

```
plot_data_long <- melt(plot_data, id.vars = "year", variable.name = "group", value.name = "count")



ggplot(plot_data_long, aes(x = year, y = count, fill = group)) +
  # COVID-19 shading (2020+)
  annotate("rect", xmin = 2019.5, xmax = Inf, ymin = 0, ymax = Inf,
           fill = "orange", alpha = 0.3) +

  # Add COVID-19 text annotation
  annotate("text", x = 2021, y = max(plot_data_long$count) * 1.1,
           label = "COVID-19", color = "black", size = 5, fontface = "italic") +

  geom_bar(stat = "identity", position = "stack", color = NA) +
  labs(
    x = "Year",
    y = "Number of Reports",
    fill = "Report Type"
  ) +
  scale_fill_manual(values = c("AE" = "#d73027", "No_AE" = "#d9d9d9"),
                    labels = c("Myopathy/Rhabdomyolysis", "Other AE")) +
  scale_x_continuous(breaks = unique(plot_data_long$year)) +
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1)
  )



fwrite(results_summary, "H://DDI Analysis//bias analysis//yearly counts//summary.csv")
```

```
library(data.table)
library(ggplot2)

# Load data
results_summary <- fread("H://DDI Analysis//bias analysis//yearly counts//count.csv")

# Calculate percentages
results_summary[, `:=`(
  `n11_percent` = (`n11.` / `n...`) * 100,
  `n10_percent` = (`n10.` / `n...`) * 100,
  `n01_percent` = (`n01.` / `n...`) * 100
)]

# Melt for plotting
plot_long <- melt(results_summary[, .(year, n11_percent, n10_percent, n01_percent)],
                  id.vars = "year",
                  variable.name = "group",
                  value.name = "percentage")

# Relabel for legend
plot_long[, group := factor(group,
                            levels = c("n11_percent", "n10_percent", "n01_percent"),
                            labels = c("Antivirals and Statins (%)", "Antivirals only (%)", "Statins only (%)"))]

# Plot
ggplot(plot_long, aes(x = year, y = percentage, color = group)) +
  # COVID-19 shading (2020+)
  annotate("rect", xmin = 2019.5, xmax = Inf, ymin = 0, ymax = Inf,
           fill = "orange", alpha = 0.3) +
  annotate("text", x = 2021, y = max(plot_long$percentage) * 1.1,
           label = "COVID-19", color = "black", size = 5, fontface = "italic") +
  geom_line(size = 1.2) +
```

```
  geom_point(size = 2) +
  scale_color_manual(
    values = c(
      "Antivirals and Statins (%)" = "#1b9e77",
      "Antivirals only (%)" = "#d95f02",
      "Statins only (%)" = "#7570b3"
    )
  ) +
  labs(
    x = "Year",
    y = "Percentage of Total Reports",
    color = "Group",
    title = "Percentage of Report Subgroups Over Time"
  ) +
  theme_minimal(base_size = 14) +
  scale_x_continuous(breaks = results_summary$year) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
library(data.table)
library(ggplot2)

# Load data
results_summary <- fread("H://DDI Analysis//bias analysis//yearly counts//count.csv")
```

```
# Calculate subgroup-specific percentages
results_summary[, `:=`(
  n111_percent = (n111 / `n11.`) * 100,
  n101_percent = (n101 / `n10.`) * 100,
  n011_percent = (n011 / `n01.`) * 100
)]

# Reshape for plotting
plot_long <- melt(results_summary[, .(year, n111_percent, n101_percent, n011_percent)],
                  id.vars = "year",
                  variable.name = "group",
                  value.name = "percentage")

# Update group names for legend
plot_long[, group := factor(group,
                            levels = c("n111_percent", "n101_percent", "n011_percent"),
                            labels = c("Antivirals + Statins", "Antivirals Only", "Statins Only"))]

# Plot
ggplot(plot_long, aes(x = year, y = percentage, color = group)) +
  # COVID-19 period shading
  annotate("rect", xmin = 2019.5, xmax = Inf, ymin = 0, ymax = Inf,
           fill = "orange", alpha = 0.3) +
  annotate("text", x = 2021, y = max(plot_long$percentage) * 1.1,
           label = "COVID-19", color = "black", size = 5, fontface = "italic") +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  scale_color_manual(
    values = c(
      "Antivirals + Statins" = "#1b9e77",
      "Antivirals Only" = "#d95f02",
      "Statins Only" = "#7570b3"
    )
```

```
  ) +
  labs(
    x = "Year",
    y = "Percentage with Myopathy/Rhabdomyolysis",
    color = "Subgroup",
    title = "Myopathy/Rhabdomyolysis Rates by Drug Subgroup Over Time"
  ) +
  theme_minimal(base_size = 14) +
  scale_x_continuous(breaks = results_summary$year) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
library(data.table)
```

```
get_co_drugs <- function(merged){

  merged <- merged[,.(id,active_ingredient)]

  merged <- unique(merged)

  merged[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINA
VIR"]
```

```
  antiviral_id <- unique(merged[active_ingredient == "RITONAVIR/LOPINAVIR"] $id)

  antiviral_reports <- unique(merged[id%in%antiviral_id])

  other_drugs <- antiviral_reports[!active_ingredient == "RITONAVIR/LOPINAVIR"]

  top_active_ingredients <- other_drugs[, .N, by = active_ingredient][
    order(-N)][
      1:10]

  top_active_ingredients$percent <- top_active_ingredients$N/length(unique(other_drugs$id))


  return(top_active_ingredients)

}



processed_folder <- ("H://DDI Analysis//bias analysis//processed//yearly//")

results_together <- data.table()

for (i in 0:23) {

  cat(paste0("iteration", i, "\n"))

  merged <- fread(paste0(processed_folder, paste0(i, ".csv")))

  setnames(merged, "primaryid", "id")

  counts <- get_co_drugs(merged)
```

```
  counts$year <- 2000 + i

  results_together <- rbind(results_together, counts)
}



fwrite(results_together, "H://DDI Analysis//bias analysis//yearly counts//coprescribed drugs.csv")



library(data.table)

# Load the table you generated
coprescribed <- fread("H://DDI Analysis//bias analysis//yearly counts//coprescribed drugs.csv")

# Get the unique drugs across all years
unique_drugs <- unique(coprescribed$active_ingredient)

# You can now manually classify them. Here's an example based on known drugs:
hiv <- c(
  "STAVUDINE", "LAMIVUDINE", "DIDANOSINE", "SAQUINAVIR", "EFAVIRENZ",
  "INDINIVIR SULFATE", "ZIDOVUDINE", "INDINAVIR SULFATE", "AMPRENAVIR",
  "ABACAVIR\\ABACAVIR SULFATE", "NEVIRAPINE", "LAMIVUDINE\\ZIDOVUDINE",
  "TENOFOVIR DISOPROXIL FUMARATE", "TENOFOVIR", "ABACAVIR SULFATE",
  "ENFUVIRTIDE", "ATAZANAVIR\\ATAZANAVIR SULFATE\\REYATAZ",
  "EMTRICITABINE\\TENOFOVIR DISOPROXIL FUMARATE\\TRUVADA",
  "DARUNAVIR\\DARUNAVIR ETHANOLATE", "ISENTRESS\\RALTEGRAVIR\\RALTEGRAVIR POTASSI
UM",
  "ATAZANAVIR", "ETRAVIRINE", "DARUNAVIR", "EMTRICITABINE\\TENOFOVIR DISOPROXIL FUMA
RATE",
  "DARUNAVIR ETHANOLATE", "ATAZANAVIR SULFATE", "ABACAVIR SULFATE\\LAMIVUDINE",
  "RALTEGRAVIR POTASSIUM", "EMTRICITABINE"
)
```

```
antibiotics <- c(
  "SULFAMETHOXAZOLE\\TRIMETHOPRIM", "AZITHROMYCIN ANHYDROUS", "CEFTRIAXONE"
)

antivirals_nonhiv <- c(
  "RIBAVIRIN",      # HCV + RSV
  "DASABUVIR",      # HCV
  "HYDROXYCHLOROQUINE"  # Off-label for COVID
)

general_meds <- c(
  "ASPIRIN", "AMLODIPINE BESYLATE", "ACETAMINOPHEN", "ATORVASTATIN",
  "LEVOTHYROXINE", "LISINOPRIL", "CHOLECALCIFEROL", "METFORMIN HYDROCHLORIDE",
  "OMEPRAZOLE", "TACROLIMUS", "LOSARTAN"
)



coprescribed[, indication := fifelse(
  active_ingredient %in% hiv, "Antiviral (HIV)",
  fifelse(active_ingredient %in% antibiotics, "Antibiotic",
          fifelse(active_ingredient %in% antivirals_nonhiv, "Antiviral (Non-HIV)",
                  "General Use")))
]



library(data.table)
library(ggplot2)

# Load your file
coprescribed <- fread("H://DDI Analysis//bias analysis//yearly counts//coprescribed drugs.csv")

# Drug classifications
```

```
hiv <- c(
  "STAVUDINE", "LAMIVUDINE", "DIDANOSINE", "SAQUINAVIR", "EFAVIRENZ",
  "INDINIVIR SULFATE", "ZIDOVUDINE", "INDINAVIR SULFATE", "AMPRENAVIR",
  "ABACAVIR\\ABACAVIR SULFATE", "NEVIRAPINE", "LAMIVUDINE\\ZIDOVUDINE",
  "TENOFOVIR DISOPROXIL FUMARATE", "TENOFOVIR", "ABACAVIR SULFATE",
  "ENFUVIRTIDE", "ATAZANAVIR\\ATAZANAVIR SULFATE\\REYATAZ",
  "EMTRICITABINE\\TENOFOVIR DISOPROXIL FUMARATE\\TRUVADA",
  "DARUNAVIR\\DARUNAVIR ETHANOLATE", "ISENTRESS\\RALTEGRAVIR\\RALTEGRAVIR POTASSI
UM",
  "ATAZANAVIR", "ETRAVIRINE", "DARUNAVIR", "EMTRICITABINE\\TENOFOVIR DISOPROXIL FUMA
RATE",
  "DARUNAVIR ETHANOLATE", "ATAZANAVIR SULFATE", "ABACAVIR SULFATE\\LAMIVUDINE",
  "RALTEGRAVIR POTASSIUM", "EMTRICITABINE"
)

antibiotics <- c(
  "SULFAMETHOXAZOLE\\TRIMETHOPRIM", "AZITHROMYCIN ANHYDROUS", "CEFTRIAXONE"
)

antivirals_nonhiv <- c(
  "RIBAVIRIN", "DASABUVIR", "HYDROXYCHLOROQUINE"
)

general_meds <- c(
  "ASPIRIN", "AMLODIPINE BESYLATE", "ACETAMINOPHEN", "ATORVASTATIN",
  "LEVOTHYROXINE", "LISINOPRIL", "CHOLECALCIFEROL", "METFORMIN HYDROCHLORIDE",
  "OMEPRAZOLE", "TACROLIMUS", "LOSARTAN"
)

# Assign indications
coprescribed[, indication := fifelse(
  active_ingredient %in% hiv, "Antiviral (HIV)",
  fifelse(active_ingredient %in% antibiotics, "Antibiotic",
          fifelse(active_ingredient %in% antivirals_nonhiv, "Antiviral (Non-HIV)",
```

```
                "General Use")))
]

# Ensure indication is a factor with consistent levels
coprescribed[, indication := factor(indication, levels = c(
  "Antiviral (HIV)", "Antibiotic", "Antiviral (Non-HIV)", "General Use"
))]

# Re-summarize by year and indication
summary_class <- coprescribed[, .(
  total_percent = sum(percent, na.rm = TRUE)
), by = .(year, indication)]

# Fix COVID shading
covid_start <- which(levels(factor(summary_class$year)) == "2020")

# Plot
ggplot(summary_class, aes(x = factor(year), y = total_percent, fill = indication)) +
  geom_bar(stat = "identity", position = position_dodge()) +

  annotate("rect", xmin = covid_start - 0.5, xmax = covid_start + 3.5,
           ymin = 0, ymax = Inf, fill = "orange", alpha = 0.2) +

  annotate("text", x = covid_start + 1.5, y = max(summary_class$total_percent) * 1.05,
           label = "COVID-19", color = "black", size = 5, fontface = "italic") +

  labs(
    x = "Year",
    y = "Percentage of Co-Reported Drugs (top-10)",
    fill = "Indication"
  ) +
  scale_fill_manual(values = c(
    "Antiviral (HIV)" = "#1b9e77",
```

```
  "Antibiotic" = "#d95f02",
  "Antiviral (Non-HIV)" = "#7570b3",
  "General Use" = "#e7298a"
 )) +
 theme_minimal(base_size = 13) +
 theme(axis.text.x = element_text(angle = 45, hjust = 1))



library(data.table)
library(ggplot2)

# Load the data
plot_data <- fread("H://DDI Analysis//bias analysis//signals//plot.csv")

# Desired DPA order
desired_order <- c(9,6,3,8,5,2,7,4,1)



# Assign group labels
plot_data[, DPA_Group := fifelse(DPA %in% c(1, 2, 3), "Pre-Covid",
                                 fifelse(DPA %in% c(4, 5, 6), "During-pandemic", "During-pandemic*"))]

# Create labeled factor for y-axis
plot_data[, DPA_Label := paste0(DPA_Group, " (DPA ", DPA, ")")]
plot_data[, DPA_Label := factor(DPA_Label, levels = plot_data[match(desired_order, DPA), DPA_Label])]

# Reshape to long format
bcpnn <- plot_data[, .(DPA, DPA_Label, Estimate = BCPNN, LB = BCPNN_LB, UB = BCPNN_UB, Method = "Bayesian Confidence Propagation Neural Network")]
omega <- plot_data[, .(DPA, DPA_Label, Estimate = Omega, LB = Omega_LB, UB = Omega_ub, Method = "Omega Shrinkage")]
long_data <- rbind(omega, bcpnn)
long_data[, Method := factor(Method, levels = c("Omega Shrinkage", "Bayesian Confidence Propagation Neural Network"))]
```

```
# Plot with bold y-axis labels
ggplot(long_data, aes(x = Estimate, y = DPA_Label, color = Method)) +


 geom_point(position = position_dodge(width = 0.6), size = 3) +
 geom_errorbarh(aes(xmin = LB, xmax = UB), height = 0.2,
                position = position_dodge(width = 0.6)) +
 geom_vline(xintercept = 0, linetype = "dashed", color = "gray40") +
 facet_wrap(~Method, scales = "free_x") +
 labs(
  x = "Estimate (95% CrI)",
  y = "DPA"
 ) +
 theme_minimal(base_size = 14) +
 theme(
  legend.position = "none",
  panel.border = element_rect(color = "black", fill = NA),
  axis.line.x = element_line(color = "black"),
  axis.line.y = element_line(color = "black"),
  strip.text = element_text(face = "bold"),
  axis.text.y = element_text(face = "bold")  # <-- This line bolds y-axis labels
 )
```

Script 6: Sub-setting to Active Comaprator

```
library(data.table)

# === PARAMETERS ===
time_frame <- "2000_2019"
input_path <- paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "//merged.csv")
output_urd <- paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "/")
output_acd <- paste0("H://DDI Analysis//bias analysis//acd//", time_frame, "/")



# === DRUG GROUPS ===
statins <- c(
  "ATORVASTATIN",
  "FLUVASTATIN",
  "LOVASTATIN",
  "PITAVASTATIN",
  "PRAVASTATIN",
  "ROSUVASTATIN",
  "SIMVASTATIN",
  "LOVASTATIN_NACIN",
  "SIMVASTATIN_NACIN",
  "SIMVASTATIN_EZETIMIBE"
)

non_cyp3A4_statins <- c(
```

```
  "FLUVASTATIN",
  "PITAVASTATIN",
  "PRAVASTATIN",
  "ROSUVASTATIN"
)

statins_1 <- c(
  "ATORVASTATIN",
  "SIMVASTATIN",
  "SIMVASTATIN_NACIN",
  "LOVASTATIN",
  "LOVASTATIN_NACIN"
)

cyp3a4_inhibitors <- c(
  "AMIODARONE", "AMITRIPTYLINE", "APREPITANT", "CARVEDILOL", "CHLORAMPHENICOL",
  "CIMETIDINE", "CIPROFLOXACIN", "CLARITHROMYCIN", "CODEINE", "DONEPEZIL",
  "FLUVOXAMINE", "HALOPERIDOL", "IMATINIB", "KETOCONAZOLE", "METOPROLOL",
  "PAROXETINE", "RISPERIDONE", "TRAMADOL", "VERAPAMIL"
)


# === READ DATA ===
data <- fread(input_path)
setnames(data, "primaryid", "id", skip_absent = TRUE)


# === URD_2: URD WITHOUT CYP3A4 INHIBITORS ===
cat("Creating: URD without CYP3A4 inhibitors\n")
data_urd_2 <- data[!id %in% data[active_ingredient %in% cyp3a4_inhibitors, unique(id)]]
fwrite(data_urd_2, paste0(output_urd, "merged_1.csv"))
rm(data_urd_2)
```

```
# === URD_3: URD WITH ONLY CYP3A4 STATINS
cat("Creating: URD WITH ONLY CYP3A4 STATINS\n")
data_urd_3 <- data[!id %in% data[active_ingredient %in% non_cyp3A4_statins, unique(id)]]
fwrite(data_urd_3, paste0(output_urd, "merged_2.csv"))
rm(data_urd_3)


# === ACD_1: All Statins ===
cat("Creating: ACD statins (all statin users)\n")
data_acd_1 <- data[id %in% data[active_ingredient %in% statins, unique(id)]]
fwrite(data_acd_1, paste0(output_acd, "merged_1.csv"))
rm(data_acd_1)


# === ACD_2: CYP3A4-Metabolized Statins (simva/lova) ===
cat("Creating: ACD CYP3A4 statins\n")
data_acd_2 <- data[id %in% data[active_ingredient %in% statins_1, unique(id)]]
fwrite(data_acd_2, paste0(output_acd, "merged_2.csv"))
rm(data_acd_2)


# === ACD_3: Statins WITHOUT CYP3A4 inhibitors ===
cat("Creating: ACD statins without CYP3A4 inhibitors\n")
statin_ids <- data[active_ingredient %in% statins, unique(id)]
inhibitor_ids <- data[active_ingredient %in% cyp3a4_inhibitors, unique(id)]
data_acd_3 <- data[id %in% statin_ids & !id %in% inhibitor_ids]
fwrite(data_acd_3, paste0(output_acd, "merged_3.csv"))
rm(data_acd_3)


# === ACD_4: CYP3A4 Statins WITHOUT CYP3A4 inhibitors ===
cat("Creating: ACD CYP3A4 statins without CYP3A4 inhibitors\n")
statin1_ids <- data[active_ingredient %in% statins_1, unique(id)]
```

```
data_acd_4 <- data[id %in% statin1_ids & !id %in% inhibitor_ids]
fwrite(data_acd_4, paste0(output_acd, "merged_4.csv"))
rm(data_acd_4)

gc()
cat("All subsets generated.\n")




library(data.table)

# === PARAMETERS ===
time_frame <- "2020_2023"
input_path <- paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "//merged.csv")
output_urd <- paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "/")
output_acd <- paste0("H://DDI Analysis//bias analysis//acd//", time_frame, "/")



# === DRUG GROUPS ===
statins <- c(
  "ATORVASTATIN",
  "FLUVASTATIN",
  "LOVASTATIN",
  "PITAVASTATIN",
  "PRAVASTATIN",
  "ROSUVASTATIN",
  "SIMVASTATIN",
  "LOVASTATIN_NACIN",
  "SIMVASTATIN_NACIN",
  "SIMVASTATIN_EZETIMIBE"
)
```

```
non_cyp3A4_statins <- c(
  "FLUVASTATIN",
  "PITAVASTATIN",
  "PRAVASTATIN",
  "ROSUVASTATIN"
)

statins_1 <- c(
  "ATORVASTATIN",
  "SIMVASTATIN",
  "SIMVASTATIN_NACIN",
  "LOVASTATIN",
  "LOVASTATIN_NACIN"
)

cyp3a4_inhibitors <- c(
  "AMIODARONE", "AMITRIPTYLINE", "APREPITANT", "CARVEDILOL", "CHLORAMPHENICOL",
  "CIMETIDINE", "CIPROFLOXACIN", "CLARITHROMYCIN", "CODEINE", "DONEPEZIL",
  "FLUVOXAMINE", "HALOPERIDOL", "IMATINIB", "KETOCONAZOLE", "METOPROLOL",
  "PAROXETINE", "RISPERIDONE", "TRAMADOL", "VERAPAMIL"
)



# === READ DATA ===
data <- fread(input_path)
setnames(data, "caseid", "id", skip_absent = TRUE)



# === URD_2: URD WITHOUT CYP3A4 INHIBITORS ===
cat("Creating: URD without CYP3A4 inhibitors\n")
data_urd_2 <- data[!id %in% data[active_ingredient %in% cyp3a4_inhibitors, unique(id)]]
fwrite(data_urd_2, paste0(output_urd, "merged_1.csv"))
```

```
rm(data_urd_2)

# === URD_3: URD WITH ONLY CYP3A4 STATINS
cat("Creating: URD WITH ONLY CYP3A4 STATINS\n")
data_urd_3 <- data[!id %in% data[active_ingredient %in% non_cyp3A4_statins, unique(id)]]
fwrite(data_urd_3, paste0(output_urd, "merged_2.csv"))
rm(data_urd_3)


# === ACD_1: All Statins ===
cat("Creating: ACD statins (all statin users)\n")
data_acd_1 <- data[id %in% data[active_ingredient %in% statins, unique(id)]]
fwrite(data_acd_1, paste0(output_acd, "merged_1.csv"))
rm(data_acd_1)


# === ACD_2: CYP3A4-Metabolized Statins (simva/lova) ===
cat("Creating: ACD CYP3A4 statins\n")
data_acd_2 <- data[id %in% data[active_ingredient %in% statins_1, unique(id)]]
fwrite(data_acd_2, paste0(output_acd, "merged_2.csv"))
rm(data_acd_2)


# === ACD_3: Statins WITHOUT CYP3A4 inhibitors ===
cat("Creating: ACD statins without CYP3A4 inhibitors\n")
statin_ids <- data[active_ingredient %in% statins, unique(id)]
inhibitor_ids <- data[active_ingredient %in% cyp3a4_inhibitors, unique(id)]
data_acd_3 <- data[id %in% statin_ids & !id %in% inhibitor_ids]
fwrite(data_acd_3, paste0(output_acd, "merged_3.csv"))
rm(data_acd_3)


# === ACD_4: CYP3A4 Statins WITHOUT CYP3A4 inhibitors ===
```

```
cat("Creating: ACD CYP3A4 statins without CYP3A4 inhibitors\n")
statin1_ids <- data[active_ingredient %in% statins_1, unique(id)]
data_acd_4 <- data[id %in% statin1_ids & !id %in% inhibitor_ids]
fwrite(data_acd_4, paste0(output_acd, "merged_4.csv"))
rm(data_acd_4)

gc()
cat("All subsets generated.\n")




library(data.table)

# === PARAMETERS ===
time_frame <- "2020_2023Q1"
input_path <- paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "//merged.csv")
output_urd <- paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "/")
output_acd <- paste0("H://DDI Analysis//bias analysis//acd//", time_frame, "/")



# === DRUG GROUPS ===
statins <- c(
  "ATORVASTATIN",
  "FLUVASTATIN",
  "LOVASTATIN",
  "PITAVASTATIN",
  "PRAVASTATIN",
  "ROSUVASTATIN",
  "SIMVASTATIN",
  "LOVASTATIN_NACIN",
  "SIMVASTATIN_NACIN",
```

```
  "SIMVASTATIN_EZETIMIBE"
)

non_cyp3A4_statins <- c(
  "FLUVASTATIN",
  "PITAVASTATIN",
  "PRAVASTATIN",
  "ROSUVASTATIN"
)

statins_1 <- c(
  "ATORVASTATIN",
  "SIMVASTATIN",
  "SIMVASTATIN_NACIN",
  "LOVASTATIN",
  "LOVASTATIN_NACIN"
)

cyp3a4_inhibitors <- c(
  "AMIODARONE", "AMITRIPTYLINE", "APREPITANT", "CARVEDILOL", "CHLORAMPHENICOL",
  "CIMETIDINE", "CIPROFLOXACIN", "CLARITHROMYCIN", "CODEINE", "DONEPEZIL",
  "FLUVOXAMINE", "HALOPERIDOL", "IMATINIB", "KETOCONAZOLE", "METOPROLOL",
  "PAROXETINE", "RISPERIDONE", "TRAMADOL", "VERAPAMIL"
)



# === READ DATA ===
data <- fread(input_path)
setnames(data, "caseid", "id", skip_absent = TRUE)



# === URD_2: URD WITHOUT CYP3A4 INHIBITORS ===
cat("Creating: URD without CYP3A4 inhibitors\n")
```

```
data_urd_2 <- data[!id %in% data[active_ingredient %in% cyp3a4_inhibitors, unique(id)]]
fwrite(data_urd_2, paste0(output_urd, "merged_1.csv"))
rm(data_urd_2)

# === URD_3: URD WITH ONLY CYP3A4 STATINS
cat("Creating: URD WITH ONLY CYP3A4 STATINS\n")
data_urd_3 <- data[!id %in% data[active_ingredient %in% non_cyp3A4_statins, unique(id)]]
fwrite(data_urd_3, paste0(output_urd, "merged_2.csv"))
rm(data_urd_3)



# === ACD_1: All Statins ===
cat("Creating: ACD statins (all statin users)\n")
data_acd_1 <- data[id %in% data[active_ingredient %in% statins, unique(id)]]
fwrite(data_acd_1, paste0(output_acd, "merged_1.csv"))
rm(data_acd_1)



# === ACD_2: CYP3A4-Metabolized Statins (simva/lova) ===
cat("Creating: ACD CYP3A4 statins\n")
data_acd_2 <- data[id %in% data[active_ingredient %in% statins_1, unique(id)]]
fwrite(data_acd_2, paste0(output_acd, "merged_2.csv"))
rm(data_acd_2)



# === ACD_3: Statins WITHOUT CYP3A4 inhibitors ===
cat("Creating: ACD statins without CYP3A4 inhibitors\n")
statin_ids <- data[active_ingredient %in% statins, unique(id)]
inhibitor_ids <- data[active_ingredient %in% cyp3a4_inhibitors, unique(id)]
data_acd_3 <- data[id %in% statin_ids & !id %in% inhibitor_ids]
fwrite(data_acd_3, paste0(output_acd, "merged_3.csv"))
rm(data_acd_3)
```

```
# === ACD_4: CYP3A4 Statins WITHOUT CYP3A4 inhibitors ===
cat("Creating: ACD CYP3A4 statins without CYP3A4 inhibitors\n")
statin1_ids <- data[active_ingredient %in% statins_1, unique(id)]
data_acd_4 <- data[id %in% statin1_ids & !id %in% inhibitor_ids]
fwrite(data_acd_4, paste0(output_acd, "merged_4.csv"))
rm(data_acd_4)

gc()
cat("All subsets generated.\n")
```

Script 7: Generate Patient Demographic Table

```
library(data.table)

# === Load and Prepare Data ===
time_frame <- "2020_2023"
data <- fread(paste0("H://DDI Analysis//bias analysis//urd//", time_frame, "//merged.csv"))
setnames(data, "caseid", "id")

d1_study <- c("RITONAVIR", "LOPINAVIR")
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# === Signal Detection Function (Unchanged) ===
get_table <- function(d1, d2, ae, merged){
  merged$d1i <- ifelse(merged$active_ingredient %in% d1, 1, 0)
  merged$d2i <- ifelse(merged$active_ingredient == d2, 1, 0)
  merged$aei <- ifelse(merged$pt == ae, 1, 0)
  flat <- merged[, .(d1i = max(d1i), d2i = max(d2i), aei = max(aei)), by = id]

  ids_n111  <- flat[d1i == 1 & d2i == 1 & aei == 1, id]
```

```
  ids_n101 <- flat[d1i == 1 & d2i == 0 & aei == 1, id]
  ids_n011 <- flat[d1i == 0 & d2i == 1 & aei == 1, id]

  n111 <- nrow(flat[d1i == 1 & d2i == 1 & aei == 1])
  n110 <- nrow(flat[d1i == 1 & d2i == 1 & aei == 0])
  n101 <- nrow(flat[d1i == 1 & d2i == 0 & aei == 1])
  n100 <- nrow(flat[d1i == 1 & d2i == 0 & aei == 0])
  n011 <- nrow(flat[d1i == 0 & d2i == 1 & aei == 1])
  n010 <- nrow(flat[d1i == 0 & d2i == 1 & aei == 0])
  n001 <- nrow(flat[d1i == 0 & d2i == 0 & aei == 1])
  n000 <- nrow(flat[d1i == 0 & d2i == 0 & aei == 0])

  n..1 <- nrow(flat[aei == 1])
  n..0 <- nrow(flat[aei == 0])
  n11. <- n111 + n110
  n10. <- n101 + n100
  n01. <- n011 + n010
  n00. <- nrow(flat[d1i == 0 & d2i == 0])
  n... <- nrow(flat)

  contingency <- rbind(
    c(n111, n110, n11.),
    c(n101, n100, n10.),
    c(n011, n010, n01.),
    c(n001, n000, n00.),
    c(n..1, n..0, n...)
  )
  rownames(contingency) <- c("Drug D1 and Drug D2", "Only Drug D1", "Only Drug D2", "Neither", "Total")
  colnames(contingency) <- c("Target AE", "All Other AEs", "Total")

  return(list(
    contingency_table = contingency,
    cell_counts = list(n111=n111,n110=n110,n101=n101,n100=n100,n011=n011,n010=n010,n001=n001,n000=n000,
n..1=n..1,n..0=n..0,n11.=n11.,n10.=n10.,n01.=n01.,n00.=n00.,n...=n...),
```

```
  id_lists = list(ids_n111 = ids_n111, ids_n101 = ids_n101, ids_n011 = ids_n011)
 ))
}



# === Demographics + Outcomes Function (Unchanged) ===
demo <- fread(paste0("H://DDI Analysis//bias analysis//raw//", time_frame, "//demo.csv"))
demo <- demo[, .(caseid, age, age_cod, sex, occr_country)]
setnames(demo, "caseid", "id")
demo$age <- as.numeric(demo$age)

outc <- fread(paste0("H://DDI Analysis//bias analysis//raw//", time_frame, "//outc.csv"))
outc <- outc[, .(caseid, outc_cod)]
setnames(outc, "caseid", "id")

# Age to years conversion
age_conversion <- c(YR=1, MON=1/12, DY=1/365, HR=1/(365*24), WK=1/52, DEC=10, SEC=1/(365*24*60*60), MIN=1/(365*24*60))
demo[, age_years := ifelse(age_cod %in% names(age_conversion), age * age_conversion[age_cod], NA)]

get_demographics_outcomes <- function(ids, group_name) {
  subset_demo <- demo[id %in% ids][order(id, -(!sex %in% c("F", "M")))][!duplicated(id)]
  mean_age <- round(mean(subset_demo$age_years, na.rm = TRUE), 2)
  sd_age <- round(sd(subset_demo$age_years, na.rm = TRUE), 2)
  missing_age <- sum(is.na(subset_demo$age_years))
  total_age <- nrow(subset_demo)
  percent_missing_age <- round(100 * missing_age / total_age, 2)

  sex_table <- table(factor(subset_demo$sex, levels = c("F", "M")))
  sex_missing <- sum(is.na(subset_demo$sex) | !(subset_demo$sex %in% c("F", "M")))
  total_sex <- sum(sex_table) + sex_missing

  country_us <- sum(subset_demo$occr_country == "US", na.rm = TRUE)
  country_missing <- sum(is.na(subset_demo$occr_country) | subset_demo$occr_country == "")
```

```
  country_other <- total_age - country_us - country_missing

  country_summary <- data.table(group = group_name, US_count = country_us, US_percent = round(100 * country_
us / total_age, 2), Other_count = country_other, Other_percent = round(100 * country_other / total_age, 2), Missing
_count = country_missing, Missing_percent = round(100 * country_missing / total_age, 2))


  subset_outc <- outc[id %in% ids][!duplicated(id)]

  outc_map <- c(HO="Hospitalization", LT="Life Threatening", DE="Death", OT="Other", DS="Disability", RI="R
equired Intervention", CA="Congenital Anomaly")

  subset_outc[, outc_cat := ifelse(outc_cod %in% names(outc_map), outc_cod, NA)]

  outcome_table <- table(factor(subset_outc$outc_cat, levels = names(outc_map)))

  outcome_missing <- sum(is.na(subset_outc$outc_cat))

  total_outcomes <- sum(outcome_table) + outcome_missing


  outc_summary <- as.data.table(outcome_table)

  setnames(outc_summary, c("outc_cod", "count"))

  outc_summary[, `:=`(group = group_name, description = outc_map[outc_cod], percent = round(100 * count / total_
outcomes, 2))]

  outc_summary <- rbind(outc_summary, data.table(outc_cod = NA, count = outcome_missing, group = group_nam
e, description = "Missing", percent = round(100 * outcome_missing / total_outcomes, 2)))


  demo_row <- data.table(group = group_name, mean_age = mean_age, sd_age = sd_age, missing_age_count = miss
ing_age, missing_age_percent = percent_missing_age, female_count = sex_table["F"], female_percent = round(100
* sex_table["F"] / total_sex, 2), male_count = sex_table["M"], male_percent = round(100 * sex_table["M"] / total_s
ex, 2), missing_sex_count = sex_missing, missing_sex_percent = round(100 * sex_missing / total_sex, 2), total = tot
al_sex)


  return(list(demo = demo_row, outc = outc_summary, country = country_summary))

}


# === Run Combined Workflow ===

table <- get_table(d1_study, d2_study, ae_study, data)

list2env(table$cell_counts, envir = .GlobalEnv)


# Constructing the contingency table using values from environment

four_by_two <- data.table(

  ` ` = c("D1D2", "D1", "D2", "D1D2'", "Total"),

  `AE` = c(n111, n101, n011, n001, n..1),
```

```
  `AE'` = c(n110, n100, n010, n000, n..0),
  `Total` = c(n11., n10., n01., n00., n...)
)

# Display the table
print(four_by_two)

n101 <- table$id_lists$ids_n101
n111 <- table$id_lists$ids_n111
n011 <- table$id_lists$ids_n011

res_1 <- get_demographics_outcomes(n101, "Antiviral only")
res_2 <- get_demographics_outcomes(n111, "Antiviral + Atorvastatin")
res_3 <- get_demographics_outcomes(n011, "Atorvastatin only")

# Final summaries
demo_summary_all <- rbindlist(list(res_1$demo, res_2$demo, res_3$demo))
outcome_summary_all <- rbindlist(list(res_1$outc, res_2$outc, res_3$outc))
country_summary_all <- rbindlist(list(res_1$country, res_2$country, res_3$country))

# === Print Results ===
print(demo_summary_all)
print(outcome_summary_all)
print(country_summary_all)



table(demo[id%in%n101]$sex)

test <- demo[id%in%n101]



length(unique(demo$id))
```

```
length(unique(data$id))



demo_summary_all[, female_n_perc := paste0(female_count, " (", female_percent, ")")]
demo_summary_all[, male_n_perc := paste0(male_count, " (", male_percent, ")")]
demo_summary_all[, missing_sex_n_perc := paste0(missing_sex_count, " (", missing_sex_percent, ")")]

demo_summary_all[, age_summary := paste0(mean_age, " (", sd_age, ")")]
demo_summary_all[, missing_age_n_perc := paste0(missing_age_count, " (", missing_age_percent, ")")]

demo_summary_table <- demo_summary_all[, .(
  group,
  age_summary,
  missing_age_n_perc,
  female_n_perc,
  male_n_perc,
  missing_sex_n_perc
)]



outcome_summary_all[, n_perc := paste0(count, " (", percent, ")")]
outcome_summary_table <- dcast(
  outcome_summary_all,
  group ~ description,
  value.var = "n_perc"
)

get_role_distribution <- function(dt, ids, drug) {
  dt <- unique(dt[id %in% ids & active_ingredient %in% drug, .(id, role_cod)])
  res <- dt[, .N, by = role_cod][order(role_cod)]
  return(res)
```

```
}



drug_distribution_n101 <- get_role_distribution(data, n101, drug = c("LOPINAVIR","RITONAVIR"))
drug_distribution_n111 <- get_role_distribution(data, n111, drug = c("LOPINAVIR","RITONAVIR"))
drug_distribution_n011 <- get_role_distribution(data, n011, drug = "ATORVASTATIN")
drug_distribution_n111 <- get_role_distribution(data, n111, drug = "ATORVASTATIN")



get_antiviral_combo <- function(dt, ids) {
  dt_sub <- unique(dt[id %in% ids & active_ingredient %in% c("LOPINAVIR", "RITONAVIR"), .(id, active_ingre
dient)])

  wide <- dcast(dt_sub, id ~ active_ingredient, fun.aggregate = length, value.var = "active_ingredient")

  # Fill missing columns with 0
  if (!"LOPINAVIR" %in% names(wide)) wide[, LOPINAVIR := 0]
  if (!"RITONAVIR" %in% names(wide)) wide[, RITONAVIR := 0]

  # Categorize exposures
  wide[, group := fifelse(LOPINAVIR > 0 & RITONAVIR > 0, "BOTH",
                          fifelse(LOPINAVIR > 0, "LOPINAVIR only",
                                  fifelse(RITONAVIR > 0, "RITONAVIR only", "None")))]

  # Count and format
  summary <- wide[, .N, by = group][order(group)]
  summary[, percent := round(100 * N / sum(N), 2)]
  summary[, formatted := paste0(N, " (", percent, ")")]

  return(summary[, .(group, formatted)])
}



anti_viral_n111 <- get_antiviral_combo(data, n111)
```

```
anti_viral_n101 <- get_antiviral_combo(data, n101)



get_country_distribution <- function(dt, ids) {
  dt <- unique(dt[id %in% ids, .(id, occr_country)])

  dt[, country_group := fifelse(occr_country == "US", "US",
                                fifelse(is.na(occr_country) | occr_country == "", "Missing", "Other"))]

  res <- dt[, .N, by = country_group][order(country_group)]
  res[, percent := round(100 * N / sum(N), 2)]
  res[, formatted := paste0(N, " (", percent, ")")]

  return(res[, .(country_group, formatted)])
}

country_n111 <- get_country_distribution(demo,n111)
country_n101 <- get_country_distribution(demo,n101)
country_n011 <- get_country_distribution(demo,n011)



get_avg_ingredient_per_id <- function(dt, ids) {
  # Subset and deduplicate
  dt <- unique(dt[id %in% ids, .(id, active_ingredient)])

  # Count number of ingredients per id
  count_per_id <- dt[, .N, by = id]

  # Compute average
  avg_count <- mean(count_per_id$N)
  sd_count <- sd(count_per_id$N)
```

```
  return(data.table(avg = avg_count, sd = sd_count))
}

drug_count_n101 <- get_avg_ingredient_per_id(data, n101)
drug_count_n011 <- get_avg_ingredient_per_id(data, n011)
drug_count_n111 <- get_avg_ingredient_per_id(data, n111)









fwrite(demo_summary_table, "H://DDI Analysis//bias analysis//tableone//2020_2023//demo.csv")
fwrite(outcome_summary_table, "H://DDI Analysis//bias analysis//tableone//2020_2023//outc.csv")

fwrite(drug_distribution_n101, "H://DDI Analysis//bias analysis//tableone//2020_2023//drug_counts_n101.csv")
fwrite(role_summary_all, "H://DDI Analysis//bias analysis//tableone//2020_2023//role_summary.csv")
```

Script 8: BCPNN

```
library(data.table)

# === Function to generate BCPNN-compatible count table ===
get_ic_table <- function(d1, d2, ae, merged){
  merged$d1i <- ifelse(merged$active_ingredient == d1, 1, 0)
  merged$d2i <- ifelse(merged$active_ingredient == d2, 1, 0)
  merged$aei <- ifelse(merged$pt == ae, 1, 0)
  merged <- merged[, .(id, d1i, d2i, aei)]
  flat <- merged[, .(d1i = max(d1i), d2i = max(d2i), aei = max(aei)), by = id]

  # Original counts
    n111 = nrow(flat[d1i == 1 & d2i == 1 & aei == 1])
    n110 = nrow(flat[d1i == 1 & d2i == 1 & aei == 0])
    n101 = nrow(flat[d1i == 1 & d2i == 0 & aei == 1])
    n100 = nrow(flat[d1i == 1 & d2i == 0 & aei == 0])
    n011 = nrow(flat[d1i == 0 & d2i == 1 & aei == 1])
    n010 = nrow(flat[d1i == 0 & d2i == 1 & aei == 0])
    n001 = nrow(flat[d1i == 0 & d2i == 0 & aei == 1])
    n000 = nrow(flat[d1i == 0 & d2i == 0 & aei == 0])
    n... = nrow(flat)

  return(list(
    cell_counts = list(
      n111 = n111, n110 = n101, n101 = n110, n011 = n011,
      n100 = n100, n010 = n001, n001 = n010, n000 = n000,
      n... = n...
    )
  ))

}
get_ic_result <- function(count_row) {
```

```
with(count_row, {
  # Smoothed marginal probabilities
  q1.. <- (n111 + n110 + n101 + n100 + 0.5) / (n... + 1)
  q0.. <- (n011 + n010 + n001 + n000 + 0.5) / (n... + 1)
  q.1. <- (n111 + n011 + n101 + n001 + 0.5) / (n... + 1)
  q.0. <- (n110 + n010 + n100 + n000 + 0.5) / (n... + 1)
  q..1 <- (n111 + n110 + n011 + n001 + 0.5) / (n... + 1)
  q..0 <- (n101 + n100 + n010 + n000 + 0.5) / (n... + 1)

  # Smoothed pairwise probabilities
  q11. <- (n111 + n110 + 0.25) / (n... + 1)
  q10. <- (n101 + n100 + 0.25) / (n... + 1)
  q01. <- (n011 + n001 + 0.25) / (n... + 1)
  q00. <- (n010 + n000 + 0.25) / (n... + 1)
  q1.1 <- (n111 + n101 + 0.25) / (n... + 1)
  q1.0 <- (n110 + n100 + 0.25) / (n... + 1)
  q0.1 <- (n011 + n001 + 0.25) / (n... + 1)
  q0.0 <- (n010 + n000 + 0.25) / (n... + 1)
  q.11 <- (n111 + n011 + 0.25) / (n... + 1)
  q.10 <- (n110 + n010 + 0.25) / (n... + 1)
  q.01 <- (n101 + n001 + 0.25) / (n... + 1)
  q.00 <- (n100 + n000 + 0.25) / (n... + 1)

  # Step 2: Moderating prior
  a... <- 0.5 * ((q1.. * q.1. * q..1) / (q11. * q1.1 * q.11))

  a111 <- ((q11. * q1.1 * q.11) / (q1.. * q.1. * q..1)) * a...
  a110 <- ((q11. * q1.0 * q.10) / (q1.. * q.1. * q..0)) * a...
  a101 <- ((q10. * q1.1 * q.01) / (q1.. * q.0. * q..1)) * a...
  a100 <- ((q10. * q1.0 * q.00) / (q1.. * q.0. * q..0)) * a...
  a011 <- ((q01. * q0.1 * q.11) / (q0.. * q.1. * q..0)) * a...
  a010 <- ((q01. * q0.0 * q.10) / (q0.. * q.1. * q..0)) * a...
  a001 <- ((q00. * q0.1 * q.01) / (q0.. * q.0. * q..1)) * a...
```

```
a000 <- ((q00. * q0.0 * q.00) / (q0.. * q.0. * q..0)) * a...

# Posterior gammas
gam111 <- a111 + n111; gam110 <- a110 + n110
gam101 <- a101 + n101; gam100 <- a100 + n100
gam011 <- a011 + n011; gam010 <- a010 + n010
gam001 <- a001 + n001; gam000 <- a000 + n000

gamma_sum <- sum(gam111, gam110, gam101, gam011, gam100, gam010, gam001, gam000)

# Joint and marginal posteriors
Ep111 <- gam111 / gamma_sum
Ep11. <- (gam111 + gam110) / gamma_sum
Ep1.1 <- (gam111 + gam101) / gamma_sum
Ep.11 <- (gam111 + gam011) / gamma_sum
Ep1.. <- (gam111 + gam110 + gam101 + gam100) / gamma_sum
Ep.1. <- (gam111 + gam110 + gam011 + gam010) / gamma_sum
Ep..1 <- (gam111 + gam101 + gam011 + gam001) / gamma_sum

ic_map <- log2((Ep111 * Ep1.. * Ep.1. * Ep..1) / (Ep11. * Ep1.1 * Ep.11))

r111 <- round(gam111 / pmin(gam111 + gam110 + gam101,
                            gam111 + gam011 + gam101,
                            gam111 + gam110 + gam011), 1)

# Lookup tables
id <- c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)
Alist <- c(3.09, 2.93, 2.78, 2.62, 2.45, 2.25, 2.03, 1.79, 1.61, 1.13, 0.073)
Blist <- c(2.22, 2.27, 2.26, 2.25, 2.15, 2.12, 2.05, 1.93, 1.89, 1.15, -0.081)

Ar <- Alist[which(id == r111)]
Br <- Blist[which(id == r111)]
```

```
    delta <- Ar / sqrt(gam111) + Br * gam111^(-1.5)

    data.table(
      IC = ic_map,
      IC_LB = ic_map - delta,
      IC_UB = ic_map + delta
    )
  })
}



time_frame <- ("2000_2019")

design <- ("urd")

# --- First time window; urd
time_frame <- "2000_2019"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "primaryid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0 <- get_ic_table(d1_study, d2_study, ae_study, data_0)
list2env(table_0$cell_counts, envir = .GlobalEnv)

counts_0 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
```

```
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = design
)

# Apply get_ic_result to count row
result_0 <- cbind(counts_0, get_ic_result(counts_0))




### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2000_2019"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_1 <- get_ic_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
```

```
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)
result_0_1 <- cbind(counts_0_1, get_ic_result(counts_0_1))
results_together <- rbind(result_0, result_0_1)



rm(data_0_1)
rm(result_0_1)
rm(counts_0_1)
rm(table_0_1)
gc()



### --- First time window; urd without non-cyp3a4 statins
time_frame <- "2000_2019"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_0_2 <- get_ic_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)
```

```
counts_0_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd cyp3a4 statins"
)

# Apply get_ic_result to count row
result_0_2 <- cbind(counts_0_2, get_ic_result(counts_0_2))

results_together <- rbind(results_together,result_0_2)

rm(data_0_2)
rm(result_0_2)
rm(counts_0_2)
rm(table_0_2)
gc()



### --- First time window; acd statins
design <- ("acd")
time_frame <- "2000_2019"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
```

```
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_1 <- get_ic_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)

counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_ic_result to count row
result_1_1 <- cbind(counts_1_1, get_ic_result(counts_1_1))

results_together <- rbind(results_together,result_1_1)

rm(data_1_1)
gc()




### --- First time window; acd cyp3a4 statins
```

```
design <- ("acd")
time_frame <- "2000_2019"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_2 <- get_ic_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)

counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)

# Apply get_ic_result to count row
result_1_2 <- cbind(counts_1_2, get_ic_result(counts_1_2))

results_together <- rbind(results_together,result_1_2)

rm(data_1_2)
```

```
gc()




### --- First time window; acd statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2000_2019"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_3 <- get_ic_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)


counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin no cyp3a4 inhibitor"
)


# Apply get_ic_result to count row
```

```
result_1_3 <- cbind(counts_1_3, get_ic_result(counts_1_3))


results_together <- rbind(results_together,result_1_3)


rm(data_1_3)
gc()




### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2000_2019"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_4 <- get_ic_table(d1_study, d2_study, ae_study, data_1_4)
list2env(table_1_4$cell_counts, envir = .GlobalEnv)


counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
```

```
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)

# Apply get_ic_result to count row
result_1_4 <- cbind(counts_1_4, get_ic_result(counts_1_4))

results_together <- rbind(results_together,result_1_4)

rm(data_1_4)
gc()



################################################################################################################

time_frame <- ("2020_2023")

design <- ("urd")

# --- First time window; urd
time_frame <- "2020_2023"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "caseid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"
```

```
# Count table
table_0 <- get_ic_table(d1_study, d2_study, ae_study, data_0)
list2env(table_0$cell_counts, envir = .GlobalEnv)

counts_0 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = design
)

# Apply get_ic_result to count row
result_0 <- cbind(counts_0, get_ic_result(counts_0))

results_together <- rbind(results_together,result_0)

rm(data_0)
gc()




### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2020_2023"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]
```

```
# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_1 <- get_ic_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)

# Apply get_ic_result to count row
result_0_1 <- cbind(counts_0_1, get_ic_result(counts_0_1))

results_together <- rbind(results_together,result_0_1)

rm(data_0_1)
gc()

### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2020_2023"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
```

```
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_0_2 <- get_ic_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)


counts_0_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd cyp3a4 statins"
)


# Apply get_ic_result to count row
result_0_2 <- cbind(counts_0_2, get_ic_result(counts_0_2))


results_together <- rbind(results_together,result_0_2)


rm(data_0_2)
gc()


### --- First time window; acd statins
```

```
design <- ("acd")
time_frame <- "2020_2023"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_1 <- get_ic_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)

counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_ic_result to count row
result_1_1 <- cbind(counts_1_1, get_ic_result(counts_1_1))

results_together <- rbind(results_together,result_1_1)

rm(data_1_1)
```

```
gc()




### --- First time window; acd cyp3a4 statins
design <- ("acd")
time_frame <- "2020_2023"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_2 <- get_ic_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)


counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)
```

```
# Apply get_ic_result to count row
result_1_2 <- cbind(counts_1_2, get_ic_result(counts_1_2))


results_together <- rbind(results_together,result_1_2)


rm(data_1_2)
gc()




### --- First time window; acd statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_3 <- get_ic_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)


counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
```

```
  n... = n...,
  year = time_frame,
  design = "acd statin no cyp3a4 inhibitor"
)

# Apply get_ic_result to count row
result_1_3 <- cbind(counts_1_3, get_ic_result(counts_1_3))

results_together <- rbind(results_together,result_1_3)

rm(data_1_3)
gc()



### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_4 <- get_ic_table(d1_study, d2_study, ae_study, data_1_4)
list2env(table_1_4$cell_counts, envir = .GlobalEnv)

counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
```

```
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)

# Apply get_ic_result to count row
result_1_4 <- cbind(counts_1_4, get_ic_result(counts_1_4))

results_together <- rbind(results_together,result_1_4)

rm(data_1_4)
gc()



################################################################################



time_frame <- ("2020_2023Q1")

design <- ("urd")

# --- First time window; urd
time_frame <- "2020_2023Q1"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "caseid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
```

```
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0 <- get_ic_table(d1_study, d2_study, ae_study, data_0)
list2env(table_0$cell_counts, envir = .GlobalEnv)

counts_0 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = design
)

# Apply get_ic_result to count row
result_0 <- cbind(counts_0, get_ic_result(counts_0))

results_together <- rbind(results_together,result_0)

rm(data_0)
gc()




### --- First time window; urd without cyp3a4 inhibitors
```

```
time_frame <- "2020_2023Q1"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_1 <- get_ic_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)

# Apply get_ic_result to count row
result_0_1 <- cbind(counts_0_1, get_ic_result(counts_0_1))

results_together <- rbind(results_together,result_0_1)

rm(data_0_1)
gc()
```

```
### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2020_2023Q1"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_2 <- get_ic_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)

counts_0_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd cyp3a4 statins"
)

# Apply get_ic_result to count row
result_0_2 <- cbind(counts_0_2, get_ic_result(counts_0_2))

results_together <- rbind(results_together,result_0_2)
```

```
rm(data_0_2)
gc()

### --- First time window; acd statins
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_1 <- get_ic_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)

counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_ic_result to count row
```

```
result_1_1 <- cbind(counts_1_1, get_ic_result(counts_1_1))


results_together <- rbind(results_together,result_1_1)


rm(data_1_1)
gc()




### --- First time window; acd cyp3a4 statins
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_2 <- get_ic_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)


counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
```

```
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)

# Apply get_ic_result to count row
result_1_2 <- cbind(counts_1_2, get_ic_result(counts_1_2))

results_together <- rbind(results_together,result_1_2)

rm(data_1_2)
gc()




### --- First time window; acd statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_3 <- get_ic_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)

counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
```

```
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin no cyp3a4 inhibitor"
)

# Apply get_ic_result to count row
result_1_3 <- cbind(counts_1_3, get_ic_result(counts_1_3))

results_together <- rbind(results_together,result_1_3)

rm(data_1_3)
gc()



### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_4 <- get_ic_table(d1_study, d2_study, ae_study, data_1_4)
```

```
list2env(table_1_4$cell_counts, envir = .GlobalEnv)


counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)

# Apply get_ic_result to count row
result_1_4 <- cbind(counts_1_4, get_ic_result(counts_1_4))

results_together <- rbind(results_together,result_1_4)

rm(data_1_4)
gc()



fwrite(results_together,paste0("H://DDI Analysis//bias analysis//signals//bcpnn.csv"))
```

Script 9: Omega Shrinkage

```
library(data.table)

get_table <- function(d1, d2, ae, merged){

  merged$d1i <- ifelse(merged$active_ingredient %in% d1, 1, 0)
  merged$d2i <- ifelse(merged$active_ingredient == d2, 1, 0)
  merged$aei <- ifelse(merged$pt == ae, 1, 0)

  merged <- merged[, .(id, d1i, d2i, aei)]

  print(length(unique(merged$id)))

  flat <- unique(merged[, .(d1i = max(d1i), d2i = max(d2i), aei = max(aei)), by = id])

  print(length(unique(flat$id)))



  # Get ids for groups of interest
  ids_n11 <- flat[d1i == 1 & d2i == 1, id]
  ids_n101 <- flat[d1i == 1 & d2i == 0, id]
  ids_n011 <- flat[d1i == 0 & d2i == 1, id]

  # Individual cell counts
  n111 <- nrow(flat[d1i == 1 & d2i == 1 & aei == 1])
  n110 <- nrow(flat[d1i == 1 & d2i == 1 & aei == 0])

  n101 <- nrow(flat[d1i == 1 & d2i == 0 & aei == 1])
  n100 <- nrow(flat[d1i == 1 & d2i == 0 & aei == 0])

  n011 <- nrow(flat[d1i == 0 & d2i == 1 & aei == 1])
  n010 <- nrow(flat[d1i == 0 & d2i == 1 & aei == 0])
```

```
  n001 <- nrow(flat[d1i == 0 & d2i == 0 & aei == 1])
  n000 <- nrow(flat[d1i == 0 & d2i == 0 & aei == 0])

  # Marginal cell counts
  n..1 <- nrow(flat[aei == 1])
  n..0 <- nrow(flat[aei == 0])

  n11. <- length(ids_n11)
  n10. <- length(ids_n101)
  n01. <- length(ids_n011)
  n00. <- nrow(flat[d1i == 0 & d2i == 0])

  # Total
  n... <- nrow(flat)

  # Construct contingency table
  contingency <- rbind(
    c(n111, n110, n11.),
    c(n101, n100, n10.),
    c(n011, n010, n01.),
    c(n001, n000, n00.),
    c(n..1, n..0, n...)
  )
  rownames(contingency) <- c("Drug D1 and Drug D2", "Only Drug D1", "Only Drug D2", "Neither Drug D1 nor D
rug D2", "Total")
  colnames(contingency) <- c("Target AE", "All Other AEs", "Total")

  return(list(
    contingency_table = contingency,
    cell_counts = list(
      n111 = n111, n110 = n110, n101 = n101, n100 = n100,
      n011 = n011, n010 = n010, n001 = n001, n000 = n000,
      n..1 = n..1, n..0 = n..0, n11. = n11., n10. = n10.,
```

```
      n01. = n01., n00. = n00., n... = n...
    ),
    id_lists = list(
      ids_n11 = ids_n11,
      ids_n101 = ids_n101,
      ids_n011 = ids_n011
    )
  ))
}

get_omega_result <- function(count_row) {
  with(count_row, {
    # Observed rates
    f00 <- n001 / n00.
    f10 <- n101 / n10.
    f01 <- n011 / n01.
    f11 <- n111 / n11.

    # Expected
    max1 <- max(f00 / (1 - f00), f10 / (1 - f10))
    max2 <- max(f00 / (1 - f00), f01 / (1 - f01))
    g11 <- 1 - 1 / (max1 + max2 - f00 / (1 - f00) + 1)
    E111 <- g11 * n11.

    # Omega
    omega <- log2((n111 + 0.5) / (E111 + 0.5))

    # Frequentist CI
    omega_se <- 1.96 / (log(2) * sqrt(n111))
    omega_lb_fr <- omega - omega_se
    omega_ub_fr <- omega + omega_se

    # Bayesian CI
```

```
    alpha <- n111 + 0.5
    beta <- E111 + 0.5
    omega_lb_by <- log2(qgamma(0.025, shape = alpha, rate = beta))
    omega_ub_by <- log2(qgamma(0.975, shape = alpha, rate = beta))

    return(data.table(
      omega = omega,
      omega_lb_fr = omega_lb_fr,
      omega_ub_fr = omega_ub_fr,
      omega_lb_by = omega_lb_by,
      omega_ub_by = omega_ub_by
    ))
  })
}



time_frame <- ("2000_2019")


design <- ("urd")


# --- First time window; urd
time_frame <- "2000_2019"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "primaryid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_0 <- get_table(d1_study, d2_study, ae_study, data_0)
```

```
list2env(table_0$cell_counts, envir = .GlobalEnv)


counts_0 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = design
)


# Apply get_omega_result to count row
result_0 <- cbind(counts_0, get_omega_result(counts_0))


rm(data_0)
gc()




### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2000_2019"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"
```

```
# Count table
table_0_1 <- get_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)

# Apply get_omega_result to count row
result_0_1 <- cbind(counts_0_1, get_omega_result(counts_0_1))

results_together <- rbind(result_0,result_0_1)

rm(data_0_1)
gc()



### --- First time window; urd without non-cyp3a4 statins
time_frame <- "2000_2019"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
```

```
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_2 <- get_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)

counts_0_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "urd only cyp3a4 statins"
)

# Apply get_omega_result to count row
result_0_2 <- cbind(counts_0_2, get_omega_result(counts_0_2))

results_together <- rbind(results_together,result_0_2)

rm(data_0_2)
gc()



### --- First time window; acd statins
design <- ("acd")
```

```
time_frame <- "2000_2019"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_1 <- get_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)

counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_omega_result to count row
result_1_1 <- cbind(counts_1_1, get_omega_result(counts_1_1))

results_together <- rbind(results_together,result_1_1)

rm(data_1_1)
gc()
```

```
### --- First time window; acd statins
design <- ("acd")
time_frame <- "2000_2019"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_1 <- get_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)

counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_omega_result to count row
result_1_1 <- cbind(counts_1_1, get_omega_result(counts_1_1))

results_together <- rbind(results_together,result_1_1)
```

```
rm(data_1_1)
gc()



### --- First time window; acd cyp3a4 statins
design <- ("acd")
time_frame <- "2000_2019"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_2 <- get_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)


counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)
```

```
# Apply get_omega_result to count row
result_1_2 <- cbind(counts_1_2, get_omega_result(counts_1_2))


results_together <- rbind(results_together,result_1_2)


rm(data_1_2)
gc()




### --- First time window; acd statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2000_2019"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_3 <- get_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)


counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
```

```
  n... = n...,
  year = time_frame,
  design = "acd statin no cyp3a4 inhibitor"
)

# Apply get_omega_result to count row
result_1_3 <- cbind(counts_1_3, get_omega_result(counts_1_3))

results_together <- rbind(results_together,result_1_3)

rm(data_1_3)
gc()



### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2000_2019"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_4 <- get_table(d1_study, d2_study, ae_study, data_1_4)
list2env(table_1_4$cell_counts, envir = .GlobalEnv)

counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
```

```
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)

# Apply get_omega_result to count row
result_1_4 <- cbind(counts_1_4, get_omega_result(counts_1_4))

results_together <- rbind(results_together,result_1_4)

rm(data_1_4)
gc()



################################################################################################################

time_frame <- ("2020_2023")

design <- ("urd")

# --- First time window; urd
time_frame <- "2020_2023"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "caseid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]
```

```
# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0 <- get_table(d1_study, d2_study, ae_study, data_0)
list2env(table_0$cell_counts, envir = .GlobalEnv)

counts_0 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = design
)

# Apply get_omega_result to count row
result_0 <- cbind(counts_0, get_omega_result(counts_0))

results_together <- rbind(results_together,result_0)

rm(data_0)
gc()
```

```
### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2020_2023"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_1 <- get_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)

# Apply get_omega_result to count row
result_0_1 <- cbind(counts_0_1, get_omega_result(counts_0_1))

results_together <- rbind(results_together,result_0_1)
```

```
rm(data_0_1)
gc()



### --- First time window; urd without non-cyp3a4 statins
time_frame <- "2020_2023"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_2 <- get_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)

counts_0_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "urd only cyp3a4 statins"
)

# Apply get_omega_result to count row
result_0_2 <- cbind(counts_0_2, get_omega_result(counts_0_2))
```

```
results_together <- rbind(results_together,result_0_2)


rm(data_0_2)
gc()









### --- First time window; acd statins
design <- ("acd")
time_frame <- "2020_2023"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_1 <- get_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)


counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
```

```
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_omega_result to count row
result_1_1 <- cbind(counts_1_1, get_omega_result(counts_1_1))

results_together <- rbind(results_together,result_1_1)

rm(data_1_1)
gc()




### --- First time window; acd cyp3a4 statins
design <- ("acd")
time_frame <- "2020_2023"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_2 <- get_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)
```

```
counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)

# Apply get_omega_result to count row
result_1_2 <- cbind(counts_1_2, get_omega_result(counts_1_2))

results_together <- rbind(results_together,result_1_2)

rm(data_1_2)
gc()




### --- First time window; acd statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"
```

```
# Count table
table_1_3 <- get_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)

counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd statin no cyp3a4 inhibitor"
)

# Apply get_omega_result to count row
result_1_3 <- cbind(counts_1_3, get_omega_result(counts_1_3))

results_together <- rbind(results_together,result_1_3)

rm(data_1_3)
gc()



### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]
```

```
# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_4 <- get_table(d1_study, d2_study, ae_study, data_1_4)
list2env(table_1_4$cell_counts, envir = .GlobalEnv)

counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)

# Apply get_omega_result to count row
result_1_4 <- cbind(counts_1_4, get_omega_result(counts_1_4))

results_together <- rbind(results_together,result_1_4)

rm(data_1_4)
gc()


################################################################################
```

```
time_frame <- ("2020_2023Q1")


design <- ("urd")


# --- First time window; urd
time_frame <- "2020_2023Q1"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "caseid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_0 <- get_table(d1_study, d2_study, ae_study, data_0)
list2env(table_0$cell_counts, envir = .GlobalEnv)


counts_0 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
```

```
  design = design
)

# Apply get_omega_result to count row
result_0 <- cbind(counts_0, get_omega_result(counts_0))

results_together <- rbind(results_together,result_0)

rm(data_0)
gc()




### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2020_2023Q1"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_1 <- get_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
```

```
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)

# Apply get_omega_result to count row
result_0_1 <- cbind(counts_0_1, get_omega_result(counts_0_1))

results_together <- rbind(results_together,result_0_1)

rm(data_0_1)
gc()




### --- First time window; urd without non-cyp3a4 statins
time_frame <- "2020_2023Q1"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_2 <- get_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)

counts_0_2 <- data.table(
```

```
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "urd only cyp3a4 statins"
)

# Apply get_omega_result to count row
result_0_2 <- cbind(counts_0_2, get_omega_result(counts_0_2))

results_together <- rbind(results_together,result_0_2)

rm(data_0_2)
gc()




### --- First time window; acd statins
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"
```

```
# Count table
table_1_1 <- get_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)

counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_omega_result to count row
result_1_1 <- cbind(counts_1_1, get_omega_result(counts_1_1))

results_together <- rbind(results_together,result_1_1)

rm(data_1_1)
gc()




### --- First time window; acd cyp3a4 statins
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
```

```
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_2 <- get_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)


counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)


# Apply get_omega_result to count row
result_1_2 <- cbind(counts_1_2, get_omega_result(counts_1_2))


results_together <- rbind(results_together,result_1_2)


rm(data_1_2)
gc()
```

```
### --- First time window; acd statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_3 <- get_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)

counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd statin no cyp3a4 inhibitor"
)

# Apply get_omega_result to count row
result_1_3 <- cbind(counts_1_3, get_omega_result(counts_1_3))

results_together <- rbind(results_together,result_1_3)
```

```
rm(data_1_3)
gc()



### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_4 <- get_table(d1_study, d2_study, ae_study, data_1_4)
list2env(table_1_4$cell_counts, envir = .GlobalEnv)

counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n11.,
  n101 = n101, n100 = n100, n10. = n10.,
  n011 = n011, n010 = n010, n01. = n01.,
  n001 = n001, n000 = n000, n00. = n00.,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)
```

```
# Apply get_omega_result to count row
result_1_4 <- cbind(counts_1_4, get_omega_result(counts_1_4))

results_together <- rbind(results_together,result_1_4)

rm(data_1_4)
gc()



fwrite(results_together,paste0("H://DDI Analysis//bias analysis//signals//omega.csv"))
```

Script 10: CSS

```
library(data.table)

# === Function to generate BCPNN-compatible count table ===
get_css_table <- function(d1, d2, ae, merged){
  merged$d1i <- ifelse(merged$active_ingredient == d1, 1, 0)
  merged$d2i <- ifelse(merged$active_ingredient == d2, 1, 0)
  merged$aei <- ifelse(merged$pt == ae, 1, 0)
  merged <- merged[, .(id, d1i, d2i, aei)]
  flat <- merged[, .(d1i = max(d1i), d2i = max(d2i), aei = max(aei)), by = id]

  # Original counts
  n111 = nrow(flat[d1i == 1 & d2i == 1 & aei == 1])
  n110 = nrow(flat[d1i == 1 & d2i == 1 & aei == 0])
  n101 = nrow(flat[d1i == 1 & d2i == 0 & aei == 1])
  n100 = nrow(flat[d1i == 1 & d2i == 0 & aei == 0])
  n011 = nrow(flat[d1i == 0 & d2i == 1 & aei == 1])
```

```
  n010 = nrow(flat[d1i == 0 & d2i == 1 & aei == 0])
  n001 = nrow(flat[d1i == 0 & d2i == 0 & aei == 1])
  n000 = nrow(flat[d1i == 0 & d2i == 0 & aei == 0])
  n... = nrow(flat)
  n11. = n111 + n110
  n10. = n101 + n100
  n01. = n011 + n010
  n00. = n001 + n000

  return(list(
    cell_counts = list(
      n111 = n111, n110 = n101, n101 = n110, n011 = n011,
      n100 = n100, n010 = n001, n001 = n010, n000 = n000,
      n11. = n11., n10. = n10., n01. = n01., n00. = n00.,
      n... = n...
    )
  ))

}
get_css_result <- function(count_row) {
  with(count_row, {

    # PRR for D1 ∩ D2
    PRR_comb <- (n111 / (n111 + n110)) / ((n011 + n001 + n101) / (n000 + n001 + n010 + n000 + n101 + n100))

    # PRR for D1 alone
    PRR_D1 <- ((n111+n101)/(n111 + n110 + n101 + n100))/ ((n001 + n011) / (n001 + n000 + n011 + n010))

    # PRR for D2 alone
    PRR_D2 <- ((n111 + n011)/(n111+n110 + n011+n010))/((n001 + n101)/(n001+n000 + n101+n100))

    # CRR
    CRR <- PRR_comb / max(PRR_D1, PRR_D2)
```

```
  # χ²
  numerator_inner <- abs(n111 * (n000 + n010 + n100) - n110 * (n001 + n011 + n101)) - (n... / 2)
  numerator <- (n... * numerator_inner^2)

  denom_1 <- (n111 + n110)
  denom_2 <- (n111 + n101 + n011 + n001)
  denom_3 <- (n000 + n001 + n010 + n011 + n100 + n101)
  denom_4 <- (n000 + n010 + n100 + n110)

  # === Final chi-squared value ===
  chi_sq_0 <- numerator / (denom_1)
  chi_sq_1 <- chi_sq_0 / denom_2
  chi_sq_2 <- chi_sq_1 / denom_3
  chi_sq <- chi_sq_2 / denom_4

  # CSS

  PRR_lb_combo <- exp(log(PRR_comb)-1.96*sqrt((1/n111) - (1/(n111+n110)) + (1/(n001+n011+n101)) - (1/(n00
1+n000 + n011+n010))))
  PRR_ub_d1 <- exp(log(PRR_D1)+1.96*sqrt((1/(n111+n101)) - (1/(n111 + n110 + n101 + n100)) + (1/(n001 + n0
11)) - (1/(n001 + n000 + n011 + n010))))
  PRR_ub_d2 <- exp(log(PRR_D2)+1.96*sqrt((1/(n111 + n011)) - (1/(n111+n110 + n011+n010)) + (1/(n001 + n10
1)) - (1/(n001+n000 + n101+n100))))

  CSS <- PRR_lb_combo/max(PRR_ub_d1,PRR_ub_d2)

  # Return
  data.table(
    PRR_D1 = PRR_D1,
    PRR_D2 = PRR_D2,
    PRR_comb = PRR_comb,
    CRR = CRR,
    chi_sq = chi_sq,
```

```
    PRR_lb_combo = PRR_lb_combo,
    CSS = CSS
  )
 })
}



time_frame <- ("2000_2019")

design <- ("urd")

# --- First time window; urd
time_frame <- "2000_2019"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "primaryid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0 <- get_css_table(d1_study, d2_study, ae_study, data_0)
list2env(table_0$cell_counts, envir = .GlobalEnv)

counts_0 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
```

```
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = design
)

# Apply get_css_result to count row
result_0 <- cbind(counts_0, get_css_result(counts_0))

rm(data_0)
gc()




### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2000_2019"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_1 <- get_css_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
```

```
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)

# Apply get_css_result to count row
result_0_1 <- cbind(counts_0_1, get_css_result(counts_0_1))

results_together <- rbind(result_0,result_0_1)

rm(data_0_1)
gc()



### --- First time window; urd cyp3a4 statins
time_frame <- "2000_2019"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_2 <- get_css_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)

counts_0_2 <- data.table(
```

```
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd cyp3a4 statins"
)

# Apply get_css_result to count row
result_0_2 <- cbind(counts_0_2, get_css_result(counts_0_2))

results_together <- rbind(results_together,result_0_2)

rm(data_0_2)
gc()



### --- First time window; acd statins
design <- ("acd")
time_frame <- "2000_2019"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
```

```
table_1_1 <- get_css_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)

counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_css_result to count row
result_1_1 <- cbind(counts_1_1, get_css_result(counts_1_1))

results_together <- rbind(results_together,result_1_1)

rm(data_1_1)
gc()




### --- First time window; acd cyp3a4 statins
design <- ("acd")
time_frame <- "2000_2019"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]
```

```
# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_2 <- get_css_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)

counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)

# Apply get_css_result to count row
result_1_2 <- cbind(counts_1_2, get_css_result(counts_1_2))

results_together <- rbind(results_together,result_1_2)

rm(data_1_2)
gc()



### --- First time window; acd statins no cyp3a4 inhibitors
```

```
design <- ("acd")
time_frame <- "2000_2019"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_3 <- get_css_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)

counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin no cyp3a4 inhibitor"
)

# Apply get_css_result to count row
result_1_3 <- cbind(counts_1_3, get_css_result(counts_1_3))

results_together <- rbind(results_together,result_1_3)

rm(data_1_3)
```

```
gc()




### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2000_2019"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_4 <- get_css_table(d1_study, d2_study, ae_study, data_1_4)
list2env(table_1_4$cell_counts, envir = .GlobalEnv)


counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)


# Apply get_css_result to count row
```

```
result_1_4 <- cbind(counts_1_4, get_css_result(counts_1_4))


results_together <- rbind(results_together,result_1_4)


rm(data_1_4)
gc()




################################################################################################
##############


time_frame <- ("2020_2023")


design <- ("urd")


# --- First time window; urd
time_frame <- "2020_2023"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "caseid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVI
R"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_0 <- get_css_table(d1_study, d2_study, ae_study, data_0)
list2env(table_0$cell_counts, envir = .GlobalEnv)


counts_0 <- data.table(
```

```
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = design
)

# Apply get_css_result to count row
result_0 <- cbind(counts_0, get_css_result(counts_0))

results_together <- rbind(results_together,result_0)

rm(data_0)
gc()




### --- First time window; urd without cyp3a4 inhibitors
time_frame <- "2020_2023"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"
```

```
# Count table
table_0_1 <- get_css_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)

# Apply get_css_result to count row
result_0_1 <- cbind(counts_0_1, get_css_result(counts_0_1))

results_together <- rbind(results_together,result_0_1)

rm(data_0_1)
gc()



### --- First time window; urd cyp3a4 statins
design <- ("urd")
time_frame <- "2020_2023"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
```

```
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_2 <- get_css_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)

counts_0_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd with cyp3a4 statins"
)

# Apply get_css_result to count row
result_0_2 <- cbind(counts_0_2, get_css_result(counts_0_2))

results_together <- rbind(results_together,result_0_2)

rm(data_0_2)
gc()


### --- First time window; acd statins
design <- ("acd")
time_frame <- "2020_2023"
```

```
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_1 <- get_css_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)


counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)


# Apply get_css_result to count row
result_1_1 <- cbind(counts_1_1, get_css_result(counts_1_1))


results_together <- rbind(results_together,result_1_1)


rm(data_1_1)
gc()
```

```
### --- First time window; acd cyp3a4 statins
design <- ("acd")
time_frame <- "2020_2023"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_2 <- get_css_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)

counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)

# Apply get_css_result to count row
result_1_2 <- cbind(counts_1_2, get_css_result(counts_1_2))
```

```
results_together <- rbind(results_together,result_1_2)


rm(data_1_2)
gc()




### --- First time window; acd statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_3 <- get_css_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)


counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
```

```
  design = "acd statin no cyp3a4 inhibitor"
)

# Apply get_css_result to count row
result_1_3 <- cbind(counts_1_3, get_css_result(counts_1_3))

results_together <- rbind(results_together,result_1_3)

rm(data_1_3)
gc()



### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_4 <- get_css_table(d1_study, d2_study, ae_study, data_1_4)
list2env(table_1_4$cell_counts, envir = .GlobalEnv)

counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
```

```
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)

# Apply get_css_result to count row
result_1_4 <- cbind(counts_1_4, get_css_result(counts_1_4))

results_together <- rbind(results_together,result_1_4)

rm(data_1_4)
gc()

################################################################################




time_frame <- ("2020_2023Q1")

design <- ("urd")

# --- First time window; urd
time_frame <- "2020_2023Q1"
data_0 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged.csv"))
setnames(data_0, "caseid", "id")
data_0[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
```

```
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0 <- get_css_table(d1_study, d2_study, ae_study, data_0)
list2env(table_0$cell_counts, envir = .GlobalEnv)

counts_0 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = design
)

# Apply get_css_result to count row
result_0 <- cbind(counts_0, get_css_result(counts_0))

results_together <- rbind(results_together,result_0)

rm(data_0)
gc()




### --- First time window; urd without cyp3a4 inhibitors
```

```
time_frame <- "2020_2023Q1"
data_0_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_0_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_1 <- get_css_table(d1_study, d2_study, ae_study, data_0_1)
list2env(table_0_1$cell_counts, envir = .GlobalEnv)

counts_0_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd no cyp3a4 inhibitors"
)

# Apply get_css_result to count row
result_0_1 <- cbind(counts_0_1, get_css_result(counts_0_1))

results_together <- rbind(results_together,result_0_1)

rm(data_0_1)
gc()
```

```
### --- First time window; urd cyp3a4 statins
time_frame <- "2020_2023Q1"
data_0_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_0_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_0_2 <- get_css_table(d1_study, d2_study, ae_study, data_0_2)
list2env(table_0_2$cell_counts, envir = .GlobalEnv)

counts_0_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "urd cyp3a4 statins"
)

# Apply get_css_result to count row
result_0_2 <- cbind(counts_0_2, get_css_result(counts_0_2))

results_together <- rbind(results_together,result_0_2)
```

```
rm(data_0_2)
gc()

### --- First time window; acd statins
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_1 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_1.csv"))
data_1_1[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_1 <- get_css_table(d1_study, d2_study, ae_study, data_1_1)
list2env(table_1_1$cell_counts, envir = .GlobalEnv)

counts_1_1 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin"
)

# Apply get_css_result to count row
```

```
result_1_1 <- cbind(counts_1_1, get_css_result(counts_1_1))


results_together <- rbind(results_together,result_1_1)


rm(data_1_1)
gc()




### --- First time window; acd cyp3a4 statins
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_2 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_2.csv"))
data_1_2[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]


# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"


# Count table
table_1_2 <- get_css_table(d1_study, d2_study, ae_study, data_1_2)
list2env(table_1_2$cell_counts, envir = .GlobalEnv)


counts_1_2 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
```

```
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin"
)

# Apply get_css_result to count row
result_1_2 <- cbind(counts_1_2, get_css_result(counts_1_2))

results_together <- rbind(results_together,result_1_2)

rm(data_1_2)
gc()



### --- First time window; acd statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_3 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_3.csv"))
data_1_3[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_3 <- get_css_table(d1_study, d2_study, ae_study, data_1_3)
list2env(table_1_3$cell_counts, envir = .GlobalEnv)

counts_1_3 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
```

```
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd statin no cyp3a4 inhibitor"
)

# Apply get_css_result to count row
result_1_3 <- cbind(counts_1_3, get_css_result(counts_1_3))

results_together <- rbind(results_together,result_1_3)

rm(data_1_3)
gc()



### --- First time window; acd cyp3a4 statins no cyp3a4 inhibitors
design <- ("acd")
time_frame <- "2020_2023Q1"
data_1_4 <- fread(paste0("H://DDI Analysis//bias analysis//", design, "//", time_frame, "//merged_4.csv"))
data_1_4[active_ingredient %in% c("RITONAVIR", "LOPINAVIR"), active_ingredient := "RITONAVIR/LOPINAVIR"]

# Study setup
d1_study <- "RITONAVIR/LOPINAVIR"
d2_study <- "ATORVASTATIN"
ae_study <- "MYOPATHY/RHABDOMYOLYSIS"

# Count table
table_1_4 <- get_css_table(d1_study, d2_study, ae_study, data_1_4)
```

```
list2env(table_1_4$cell_counts, envir = .GlobalEnv)


counts_1_4 <- data.table(
  d1 = d1_study,
  d2 = d2_study,
  ae = ae_study,
  n111 = n111, n110 = n110, n11. = n111 + n110,
  n101 = n101, n100 = n100, n10. = n101 + n100,
  n011 = n011, n010 = n010, n01. = n011 + n010,
  n001 = n001, n000 = n000, n00. = n001 + n000,
  n... = n...,
  year = time_frame,
  design = "acd cyp3a4 statin no cyp3a4 inhibitor"
)

# Apply get_css_result to count row
result_1_4 <- cbind(counts_1_4, get_css_result(counts_1_4))

results_together <- rbind(results_together,result_1_4)

rm(data_1_4)
gc()



fwrite(results_together,paste0("H://DDI Analysis//bias analysis//signals//css.csv"))
```