Clinical Epidemiology

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Framework for Exploration of Statistical Heterogeneity in Multi-Database Studies: A Case Study Using EXACOS-CV Studies

Kirsty Marie Rhodes[1], Edeltraut Garbe[2], Hana Müllerová[3], Paul Ekwaru[4], Nils Kossack[5], Brenda N Baak[6], Muriel Lobier[7], Nathaniel M Hawkins[8], Clementine Nordon[3]

[1]Real-World Science and Analytics, BioPharmaceuticals Medical, AstraZeneca, Cambridge, UK; [2]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany; [3]Respiratory Evidence Strategy, BioPharmaceuticals Medical, AstraZeneca, Cambridge, UK; [4]Medlior Health Outcomes Research Ltd., Calgary, AB, Canada; [5]WIG2 Institute, Leipzig, Germany; [6]PHARMO Institute for Drug Outcomes Research, Utrecht, Netherlands; [7]Biostatistics, Real World Solutions, IQVIA, Espoo, Finland; [8]Division of Cardiology, Centre for Cardiovascular Innovation, University of British Columbia, Vancouver, BC, Canada

Correspondence: Kirsty Marie Rhodes, AstraZeneca, Academy House, Hills Road, Cambridge, CB2 8PA, United Kingdom, Email kirsty.rhodes@astrazeneca.com

Purpose: Multi-database studies may provide heterogeneous results despite using common protocols, leading to challenges in interpretation, but also providing an opportunity to gain insights on populations or healthcare systems. The objectives of these analyses were to develop a framework for exploring sources of statistical heterogeneity and apply it to the multi-database EXACOS-CV (EXAcerbations of COPD and their OutcomeS on CardioVascular diseases) program.

Methods: A conceptual framework to systematically assess sources of statistical heterogeneity in multi-database studies was developed. This framework distinguishes between methodological diversity and true clinical variation. Methodological diversity includes differences in study design and database selection, while true variation considers population and healthcare differences. Possible sources of methodological diversity were identified via a novel checklist and explored. In turn, hypotheses were generated about true variation. The framework and checklist were applied to EXACOS-CV cohort studies in Germany, Canada, the Netherlands, and Spain which deviated least from the common protocol and so were included. Focus was on adjusted hazard ratios (aHR) for post-exacerbation associations with decompensated heart failure (HF) and all-cause death, for which results were most and least heterogeneous, respectively.

Results: Across EXACOS-CV studies, the adjusted hazard ratios (aHR) for HF in the first 1–7 days post-exacerbation, compared to non-exacerbation periods, ranged from 2.6 (95% CI, 2.3, 2.9) in Germany to 72.3 (64.4, 81.2) in Canada, and the association with death, relative to non-exacerbation periods, ranged from 3.5 (2.4, 5.3) in the Netherlands to 22.1 (19.9, 24.4) in Spain. Completed methodological diversity checklists linked differences in aHRs to possible variation in ability to capture pre-existing cardiovascular comorbidities across studies, as well as differences in confounder measurement. Standardizing adjusted models across studies did not fully explain heterogeneity, suggesting other contributing factors. Heterogeneity may result from genuine variation in prevalence of CV disease. It was hypothesized that patients with pre-existing CV disease have more accurate diagnoses and management of post-exacerbation CV events, possibly leading to lower risks of such events.

Conclusion: Multi-database studies can provide directional insights on the study research question while considering healthcare system and population differences. The developed framework aids assessment of heterogeneity sources.

Keywords: heterogeneity, multi-database study, real-world data, chronic obstructive pulmonary disease, exacerbation, cardiovascular outcomes

Introduction

A multi-database study uses two or more secondary databases, usually from different countries, to address common objectives concerning associations between exposures and outcomes of interest. It is usually not possible to pool the patient-level data across databases. The programming and analysis are conducted separately, usually by different teams,

Received: 26 February 2025
Accepted: 28 May 2025
Published: 14 June 2025

as independent studies. Multi-database studies using local data extraction and analysis can show variation in results (ie, statistical heterogeneity) despite using a common study protocol, which leads to challenges in interpretation. Statistical heterogeneity presents itself through estimates of exposure-outcome associations being more different across studies than expected by chance.[1] This may arise due to methodological or database-related differences but could also be attributable to true differences across populations and healthcare systems where the studies are being conducted.

Initiatives, such as the DIVERSE project, will inform the development of guidelines for identifying and reporting heterogeneity in multi-database studies.[2] Example projects have studied methodological and database-related heterogeneity in multi-database studies of drug-outcome associations. The IMI PROTECT project studied the associations between six drug classes and five key adverse events in European and US electronic health records databases,[3] demonstrating the benefits of a common protocol and detailed data specifications in enhancing consistency of findings across databases. Madigan et al systematically studied heterogeneity among multiple databases, holding study methods constant, and explored relative risk estimates for 53 drug-outcome associations.[4] Their findings underscore the sensitivity of observational study results to database selection. From a clinical perspective, it is of interest to explore the differences in results across different healthcare systems and patient populations, by differentiating between true clinical sources of statistical heterogeneity from methodological and database-related variation. While the methods used in the IMI PROTECT project and by Madigan et al are valuable for prospectively exploring sources of methodological and database-related diversity, they would not facilitate post-hoc identification and exploration of methodological and clinical heterogeneity sources in multi-database studies.

Currently, there is no established framework for identifying and exploring sources of statistical heterogeneity in multi-database studies. The aim of the present study was to develop such framework to enable generation of new hypotheses on how true variations across healthcare systems and populations can have an impact on exposure-outcome associations. The framework was developed through application to a case study from the EXACOS-CV (The EXAcerbations of COPD and their OutcomeS on CardioVascular diseases) program, for which methodological details are available elsewhere.[5]

Materials and Methods

Data Source: EXACOS-CV

Exacerbations of COPD are linked to higher risk of cardiovascular (CV) events.[6] The EXACOS-CV program, conducted across multiple countries, is a set of observational studies using a common protocol to examine how the occurrence of CV events changes after a moderate or severe exacerbation of COPD. These studies used secondary data from electronic healthcare and claims databases. The objectives were to determine the association between different time periods (1–7, 8–14, 15–30, 31–180, 181–365 days) following a moderate or severe exacerbation of COPD and severe CV events or all-cause death, compared to time without an exacerbation. Reference data for unexposed periods came from individuals without exacerbations, including all follow-up time of those who never experienced an exacerbation and periods outside 1–365 days after an exacerbation in those who did. This study used already published data; no new human data were collected.

For the present study, we focused on results from Canada,[7] Germany,[8] Spain,[9] and the Netherlands,[10] where deviations from the common protocol[5] were minimal. In each study, people living with COPD were identified from secondary databases between 2014–2018/2019 and included if aged 40 or older. Country-specific databases are described in Supplementary Table 1. The first date of COPD diagnosis defined cohort entry date and marked the start of follow-up. Patients were required to have 12 or 24 months of continuous data before cohort entry. Follow-up lasted until a first outcome of interest, the end of data availability; or end of study period (31 December 2019). EXACOS-CV studied moderate or severe exacerbations of COPD as exposures and non-fatal severe CV events (acute coronary syndrome, arrhythmia, decompensated heart failure (HF), ischemic stroke) and all-cause death as outcomes. Country-specific protocols are described in Supplementary Table 2.

The time until the first occurrence of each outcome was analyzed using time-dependent Cox models. These models estimated the association between each outcome event and the different time periods following a moderate or severe exacerbation (1–7 days, 8–14 days, 15–30 days, 31–180 days, and 181–365 days), compared to periods without exacerbation exposure. Models were fitted with and without adjustment for potential confounders (Supplementary Table 3).

For the case study, we used the results for the outcome with highest statistical heterogeneity. Results for the outcome with lowest statistical heterogeneity served as a reference for comparison.

Framework for Heterogeneity Exploration

We present our framework and the step-by-step approach for heterogeneity exploration.

Step 1: Conceptualizing Statistical Heterogeneity

The conceptual framework distinguishes statistical heterogeneity due to methodological (and database-related) diversity from clinical (true) variation (Figure 1). It was developed by the co-authors of this study, including experts in the different databases, epidemiologists, statisticians, and clinical subject-matter experts.

Each study seeks to estimate the true exposure-outcome association in a target population of interest. A first source of statistical heterogeneity stems from the true differences in population characteristics and healthcare system features,[1] which is termed clinical variation. This means that the exposure-outcome association truly differs across studies due to characteristics specific to a country, such as patient, or healthcare characteristics that modify the strength of the association. For instance, clinical variation occurs if the true exposure-outcome association differs between males and females, and if the distribution of sex truly differs across countries in the population with the disease of interest. Another example of clinical variation is when the true exposure-outcome association differs according to physician's experience; leading to differences in associations observed across countries with different levels of physician experience.[11] The

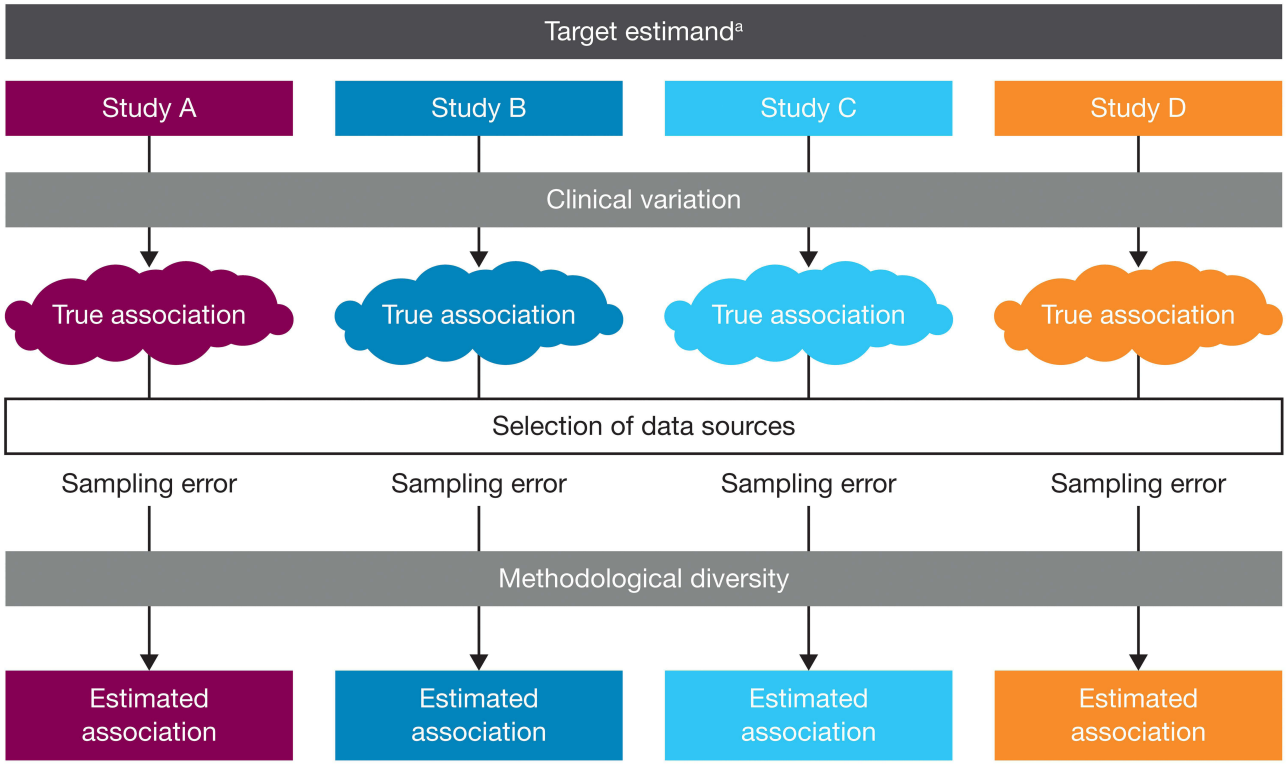

Figure 1 Conceptualization of statistical heterogeneity in a multi-database study. [a]The target estimand is determined by the research question which will consist of the following attributes: population, exposure, endpoint, association measure (eg, unadjusted hazard ratio, adjusted hazard ratio).

https://doi.org/10.2147/CLEP.S520168

importance of clinical sources of heterogeneity can be discussed based on assumptions about effect modification and clinical insight.

A second source of statistical heterogeneity is the selection and features of data sources such as coding systems and practices, which may lead to the study population not accurately representing the target population, thus studies may vary in degree of bias to external validity, and this could contribute to heterogeneity.

A third source of heterogeneity is methodological diversity, which includes differences across studies in design, variable measurement and analysis methods, indicating that studies are subject to varying degrees of bias to internal validity without necessarily implying true differences in exposure-outcome associations across studies.[1]

Within this conceptual framework, statistical heterogeneity caused by methodological diversity and true clinical variation are considered separately, since clinical variation cannot be easily or directly apprehended. In our framework, the extent to which clinical variation is at play in explaining statistical heterogeneity corresponds to "what cannot be completely explained by methodological diversity". We therefore explored methodological diversity first and evaluated the extent to which it could drive statistical heterogeneity. In a second stage, we discussed how clinical variation may further explain discrepancies in results across countries.

Step 2: Methodological Diversity Exploration

Thompson et al[12] provided checklists for identifying biases in longitudinal observational studies, drawing on the STROBE statement[13] and earlier bias assessment checklists for meta-analysis of randomized trials.[14] We adapted the checklist to focus on differences across studies rather than biases within an individual study and added items that are specific to secondary data studies (Table 1).

Sources of methodological diversity were divided into six categories: selection bias, concerning whether the study population accurately represents the target population; confounding bias, assessing if key confounders have been adjusted for; measurement bias (exposure), evaluating errors in categorizing individuals based on exposure; attrition bias, relating to loss to follow-up; measurement bias (outcome), focused on challenges in accurately measuring outcomes; and other sources of heterogeneity, such as those introduced by the statistical analysis methods used, following a similar approach to the checklist by Thompson et al. Within these categories, specific sources of heterogeneity were further itemized.

In the EXACOS-CV case study, each research team member completed the checklists using individual study statistical analysis plans and clinical study reports, including local methodology details, such as the specific algorithms used to identify people with COPD. Team members indicated whether each checklist item was a concern for methodological diversity and proposed ways to explore each source, using descriptive analysis results, modelling outputs, published literature, and/or national data. Team members met to discuss the completed checklists, reach a consensus on the most relevant methodological diversity concerns, and agree on exploration methods. In turn, the outputs from the explorations were discussed to assess the extent to which methodological diversity could explain statistical heterogeneity across the studies.

Step 3: Clinical Diversity Discussion

In the EXACOS-CV case study, we considered whether true differences in patient or healthcare system characteristics could explain variation in associations between study outcomes and time periods following an exacerbation. This led to new hypotheses about true sources of variation in post-exacerbation outcome incidence (versus no exacerbation) across healthcare systems and populations. Of note, healthcare system characteristics could encompass various features, eg, the way physicians make diagnosis, or usual management of disease across different countries.

Step 4: Defining the Scope of the Exploration

In a multi-database study, there are often numerous sets of results corresponding to different target estimands, such as different analysis populations, exposures, study endpoints, time periods and/or analysis methods. When examining heterogeneity, priority can be given to results where statistical heterogeneity is of particular importance.

In our case study, there were several sets of results, due to the number of exposed periods and outcomes. The importance of statistical heterogeneity across these results was statistically evaluated. For each study, exposed period and outcome, a forest plot of aHR estimates with 95% confidence intervals was used to visually assess the consistency of

Table 1 Checklist for Methodological Diversity Assessment

Item	Checklist Item of Thompson et al[12] (for Identifying Biases within Observational Studies)	Modified item (for Identifying Methodological Diversity across Database Studies)	Is this a Concern? Do you Think there is a Risk of Methodological Heterogeneity? Yes/No	What is the Concern? Please Specify (add as many Concerns as Necessary)
Selection bias				
1.1	Inclusion and exclusion criteria clear?	Are there any differences in eligibility criteria across studies?		
1.2		Does selection of databases lead to not identifying a population that is representative of the target population?		
1.3	Baseline measurements obtained for all participants recruited (ie, no immediate dropouts)?		This item will remain in the general checklist, but is not relevant to EXACOS-CV	
Confounding bias				
2.1	Appropriate choice of confounders (ie, based on importance rather than convenience)?	Are the most important confounders listed? Did the list of confounders differ across studies?		
2.2	Adjustment made for all known important confounders?	Was adjustment made for all pre-specified confounders? Was the adjustment made in a consistent way across studies?		
2.3	Objective method of measuring confounders?	Were there differences across studies in the way confounders were identified and measured? Eg, were there missing data on important confounders in some studies?		
Within item 2.2	Confounders measured accurately?	Delete	Item not adapted since concern is incorporated under item 2.2	
2.4	Appropriate timing for measuring confounders?	Were there differences across studies in time periods over which confounders were measured?		
2.5		Was confounding adjustment successful (eg, did propensity score matching yield covariate balance)?	This item was added to the general checklist, but is not relevant to EXACOS-CV	
Measurement bias (biases in the ways exposures are defined and measured)				
	Was the exposure measure appropriate?	We suggest deleting	Item deleted	
3.1	Objective method of measuring exposure?	Were there differences between studies in the way exposures were identified and measured?		
3.2	Exposure measured accurately?	Is there a risk for misclassification of exposure and non-exposure, and is this risk different across databases?		
3.3	Appropriate timing for measuring exposure?	Were there differences in the timing for measuring exposure across studies?	This item will remain in the general checklist, but is not a concern for EXACOS-CV	
3.4	Was the way that the exposure measure was used in the analysis appropriate?	Were the categorizations of exposure consistent across studies?		
Attrition bias				
4.1	Are the results unlikely to be affected by losses to follow-up?	Does the average follow-up duration differ across studies?		

(*Continued*)

Table 1 (Continued).

Item	Checklist Item of Thompson et al[12] (for Identifying Biases within Observational Studies)	Modified item (for Identifying Methodological Diversity across Database Studies)	Is this a Concern? Do you Think there is a Risk of Methodological Heterogeneity? Yes/No	What is the Concern? Please Specify (add as many Concerns as Necessary)
4.2	Are the results unlikely to be affected by exclusions from analysis (eg, because of extreme values or missing values of confounders)?	Were there any differences across studies in exclusions from the analysis population due to missing values (eg, due to missing confounder values)?		
Measurement bias (biases in the ways outcomes are defined and measured)				
	Was the outcome measure appropriate?	Item deleted		
5.1	Objective method of measuring outcome?	Were there differences across studies in the way the outcomes were identified and measured?		
5.2	Outcome measured accurately?	Is there a risk for misclassification of outcome (where non-events are classified as events, or events are missing) and is this risk different across databases?		
5.3	Appropriate timing for measuring outcome?	Were there differences in the timing for measuring outcomes across studies?	This item will remain in the general checklist, but is not a concern for EXACOS-CV	
5.4	Was the way that the outcome measure was used in the analysis appropriate?	Is there a possibility of misclassification of the exposure and outcome? Eg, could an exposure be misdiagnosed as an outcome and vice versa? Is this risk particularly important in some studies?		
Other bias suspected		**Other heterogeneity sources suspected**		
6.1	Was the statistical analysis appropriate?	Were there any differences across studies in the analysis methods applied?		
6.2		Were there differences in the implementation of statistical analysis methods across studies (eg, was different computing software used)?		
6.3		Is there a risk that follow-up duration could be over or underestimated (eg, due to unobserved transfers out of GP practice)? Is this risk different across studies?		

results. Non-overlapping 95% confidence intervals suggested that the variation between study results exceeds what can be attributed to sampling error, a situation referred to as statistical heterogeneity.

Instead of the commonly used I^2 statistic,[15] which can be artificially high in large samples and lacks comparability across different sets of results, we measured statistical heterogeneity using between-study variance estimates. For each exposed period and outcome, we estimated between-study variance by a random-effects meta-analysis of log aHRs. We used the restricted maximum likelihood method of estimation, which was deemed to have favorable statistical properties in a simulation study by Langan et al.[16] The analyses were performed using the R software package metafor.[17]

Results

Of all individual outcomes studied in the EXACOS-CV program, results were most heterogeneous for HF and least heterogeneous for all-cause death (Figure 2). In both cases, heterogeneity mainly concerned the early post-exacerbation periods (1–7 days and 8–14 days); aHRs for the association with HF during the 1–7 days post-exacerbation ranged from 2.6 (95% CI, 2.3–2.9) in Germany to 72.3 (64.4–81.2) in Canada; aHRs for associations with death during the 1–7 days post-exacerbation ranged from 3.5 (2.4–5.3) in the Netherlands to 22.1 (19.9–24.4) in Spain. The scope of the case study

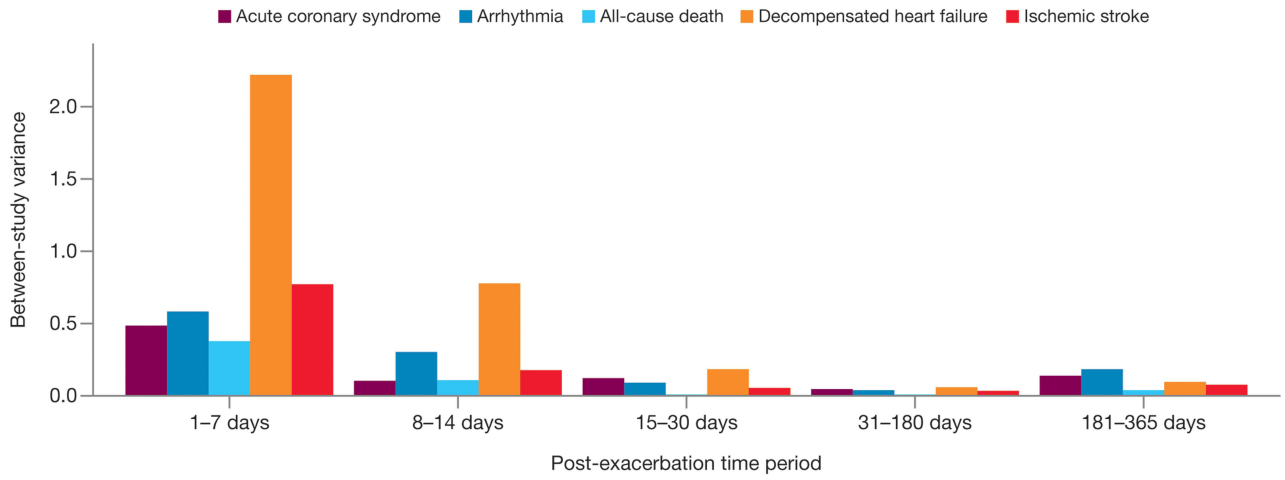

Figure 2 Between-study variance among adjusted hazard ratios for associations between post-exacerbation time periods and the occurrence of cardiovascular events and all-cause death.

was to identify and explore sources of statistical heterogeneity among associations with HF and all-cause death post-exacerbation (Figure 3). Across EXACOS-CV studies, associations between post-exacerbation time periods and outcomes of HF and all-cause death were consistent in the direction of the association but varied in terms of the strength of the association.

Methodological Diversity

The completed checklist of concerns for methodological diversity, reflecting consensus opinion of all co-authors is displayed in Supplementary Table 4. The text that follows focuses on the key findings.

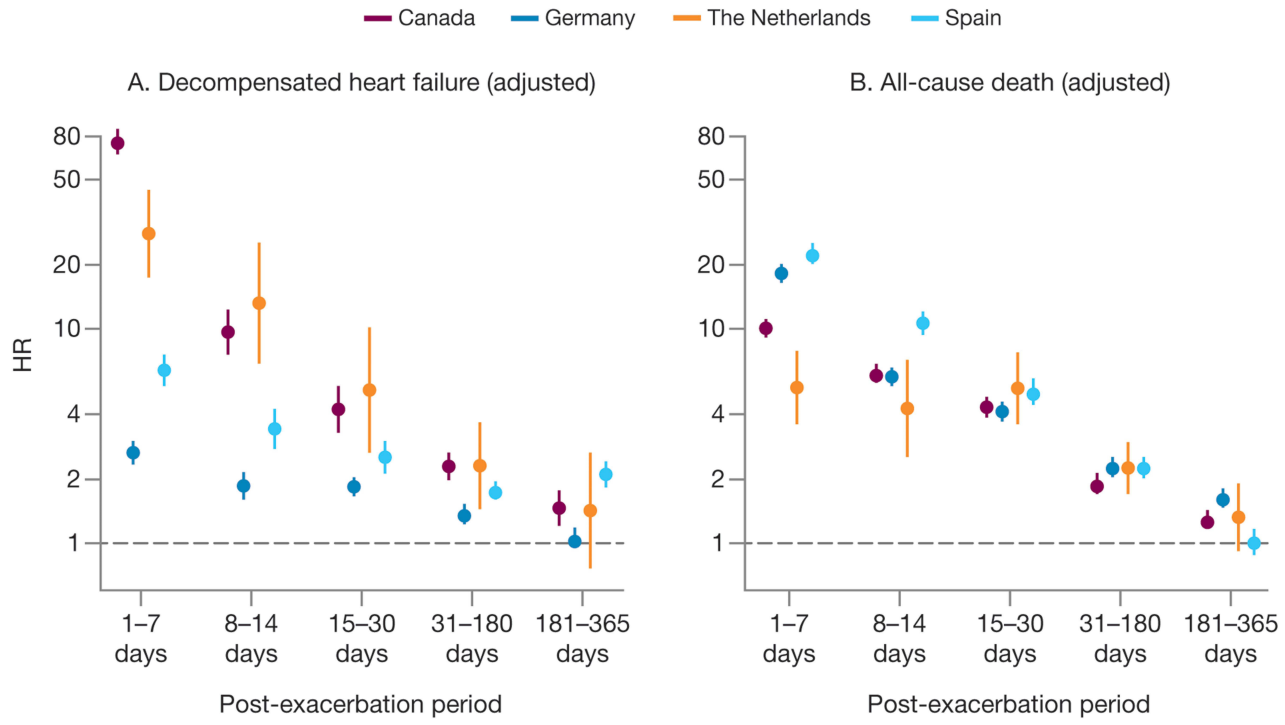

Figure 3 Forest plots of adjusted hazard ratios for association between post-exacerbation time periods and the occurrence of decompensated heart failure (**A**) and all-cause death (**B**).

https://doi.org/10.2147/CLEP.S520168

Selection Bias: Did We Include Non-COPD Patients in the Cohort?

First, we observed that the baseline prevalence of "comorbid asthma" in EXACOS-CV studies was higher in Germany (36.3%), compared to the Netherlands (12%), Spain (10.2%) or Canada (4.3%). The usual prevalence of comorbid asthma in people with COPD is 10–15%[18,19] and thus, inclusion of asthma-only (ie, erroneously diagnosed with COPD) patients in Germany was deemed possible and the exclusion of patients with comorbid asthma was considered as a possibility in Canada. These factors could lead to selection bias, as they compromise the representativeness of the COPD population being studied. If asthma-only (ie, non-COPD) patients were included in the study in Germany, then post-exacerbation associations with both HF and all-cause death might be expected to be underestimated in Germany, because asthma exacerbations are not established as a cardiovascular risk factor and the likelihood of death following an asthma flare-up is not as high as that following a COPD exacerbation. Indeed, estimates for post-exacerbation associations with HF were lowest for Germany (Figure 3). However, estimates of post-exacerbation associations with all-cause death were among the highest in Germany, and so did not support potential over-diagnosis of COPD in Germany as a concern for methodological diversity.

Second, we suspected that the higher baseline prevalence of comorbid asthma could be linked to the different algorithms used to identify people with a diagnosis of COPD in claims and electronic health records databases. In routine clinical practice, misdiagnosis of COPD with asthma is possible. The use of COPD diagnosis codes alone is not specific enough to identify individuals with "true" COPD (and not asthma-only), so algorithms (combination of criteria) are used. The study in Canada used a validated algorithm.[20] Such was not the case in Spain, the Netherlands, and Germany. In Spain and the Netherlands, the algorithm included the conduct of spirometry along with COPD codes (spirometry is the standard tool for diagnosis ascertainment) to enhance the likelihood of including individuals accurately diagnosed with COPD.

Confounding Bias: What Is the Impact of Not Measuring Smoking Accurately?

One possible source of methodological diversity is varying ability to control for confounding (item 2.1, Table 1). Smoking status (current smoker/former smoker/non-smoker) was not fully captured, except in Spain. In the Netherlands, the information was unavailable and in Canada and Germany, the observed low prevalence of smoking reflected missing information. We estimated aHRs for post-exacerbation associations with both HF and all-cause death in Spain, with and without adjustment for smoking status, resulting in similar aHR estimates (Table 2). These data do not support the varying ability to adjust for confounding by smoking status as a concern for statistical heterogeneity.

Table 2 Estimated Hazard Ratios (HR) for Associations Between Post-Exacerbation Periods and Decompensated Heart Failure and All-Cause Death in Spain, with and without Adjustment for Smoking History as a Single Confounder in the Model

	Decompensated Heart Failure		All-Cause Death	
	Unadjusted HR	**Adjusted HR**	**Unadjusted HR**	**Adjusted HR**
Exposure Period				
Prior to first exacerbation	1	1	1	1
1–7 days post-exacerbation	6.57 (5.62–7.69)	6.53 (5.58–7.64)	21.6 (19.5–23.9)	21.4 (19.4–23.7)
8–14 days post-exacerbation	3.62 (2.95–4.43)	3.59 (2.93–4.40)	10.31 (9.09–11.68)	10.2 (9.02–11.6)
15–30 days post-exacerbation	2.61 (2.21–3.08)	2.59 (2.20–3.06)	4.94 (4.38–5.59)	4.91 (4.34–5.54)
31–180 days post-exacerbation	1.83 (1.65–2.02)	1.81 (1.64–2.01)	2.14 (1.95–2.35)	2.13 (1.94–2.34)
181–365 days post-exacerbation	2.14 (1.87–2.44)	2.13 (1.86–2.43)	0.96 (0.83–1.09)	0.95 (0.83–1.09)
>365 days post-exacerbation	1.95 (1.71–2.23)	2.02 (1.79–2.28)	0.96 (0.82–1.12)	0.96 (0.8–1.12)

(*Continued*)

Table 2 (Continued).

	Decompensated Heart Failure		All-Cause Death	
	Unadjusted HR	**Adjusted HR**	**Unadjusted HR**	**Adjusted HR**
Exposure Period				
Ex-smoker		1		1
Smoker		0.753 (0.68–0.83)		0.81 (0.74–0.88)
Never		0.97 (0.89–1.07)		1.03 (0.95–1.12)

Confounding Bias: Could Differences in Adjusted Models Explain Heterogeneity?

There were differences in the handling of covariates in the models across countries (item 2.2, Table 1). For instance, COPD-related medications were categorized differently and time-varying confounders were updated annually in some studies and on occurrence in others (see Supplementary Table 4). Applying the same minimally adjusted model to each study's database reduced between-study variance among aHRs (Figure 4), but statistical heterogeneity remained substantial for the association with HF during the 1–7 days and 8–14 days post-exacerbation and for the association with all-cause death in the 1–7 days post-exacerbation. These findings suggest that differences in confounder-adjusted models are a source of statistical heterogeneity, but do not fully explain the variation in study results.

Confounding Bias: Was Pre-Existing Cardiovascular Disease Over-Reported in Claims Data?

A concern was the varying ability to capture pre-existing cardiovascular comorbidities, as key confounders (item 2.3, Table 1). In Germany, with highest observed prevalence of pre-existing cardiac conditions (Table 3), the proportion of patients with hypertension (75%) and heart failure (28%) was observed to be higher than in COSYCONET (hypertension: 51%, heart failure: 5%), a multi-center cohort study based on primary collected data.[21,22] In Canada, the observed prevalence of baseline heart failure (9.4%) and other cardiac conditions and risk factors (such as dyslipidaemia and hypertension) at baseline were lower than epidemiologically expected. Findings indicate between-study variation in the degree of measurement error in key confounders, which may have led to differences in the estimates of associations between post-exacerbation periods and outcomes in confounder-adjusted models.

Measurement Bias: Was There Misdiagnosis of COPD Exacerbation and Decompensated HF?

Within individual studies, the association with HF during the 1–7 days post-exacerbation was interpreted with caution due to these conditions having similar signs and symptoms and therefore being at risk of misdiagnosis (item 5.4,

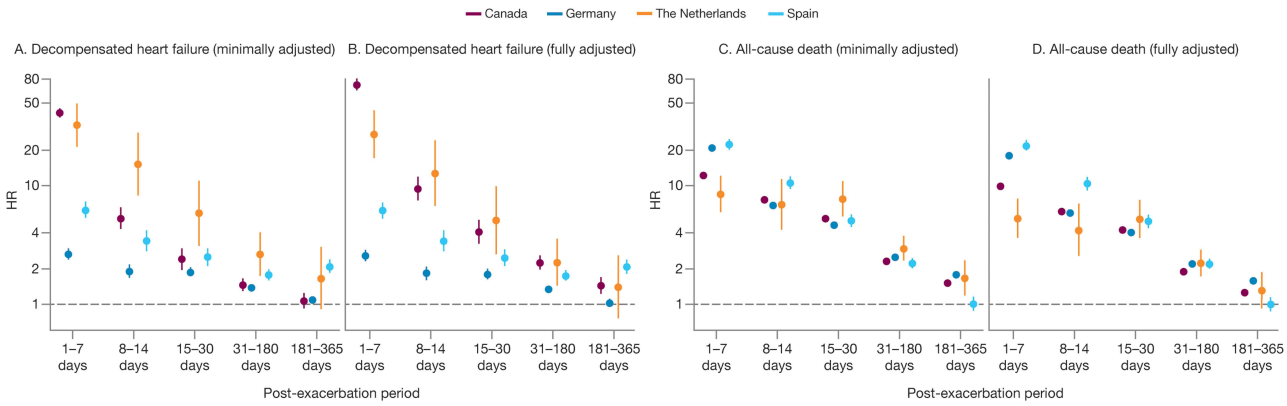

Figure 4 Forest plots of minimally adjusted and fully adjusted hazard ratios for association between post-exacerbation time periods and the occurrence of decompensated heart failure (**A** and **B**) and all-cause death (**C** and **D**).

https://doi.org/10.2147/CLEP.S520168

Table 3 Baseline Characteristics of EXACOS-CV Study Cohorts

	Canada	Germany	The Netherlands	Spain
Sample size	**N=142,787**	**N=126,795**	**N=8020**	**N=24,393**
Cohort entry year[a]				
2012	n/a	n/a	n/a	n/a
2013	n/a	n/a	n/a	n/a
2014	68.3%	36.5%	24.0%	57.4%
2015	7.6%	22.7%	22.0%	9.7%
2016	7.7%	17.2%	21.0%	10.2%
2017	7.5%	13.0%	18.0%	11.5%
2018	7.1%	10.7%	16.0%	11.2%
2019	1.8%	0%	0%	0%
Age: Mean (SD)	68.1 (12.3)	66.5 (12.0)	65 (11)	67.9 (11.6)
Male	51.7%	60.0%	53.0%	78.2%
Urban area (vs rural)	61.7%	79.2%		
Newly diagnosed	37.1%	36.6%	100%	52.1%
Number of exacerbations in the previous 12 months				
None	81.8%	72.9%	100.0%	82.6%
1	14.5%	15.7%	0%	8.8%
2+	3.6%	11.4%	0%	8.6%
Number of GP visits in the previous 12 months Mean (SD)	12.4 (15.5)	13.5 (14.9)	6 (7)	7 (3.5)
Comorbidities				
Obesity		34.9%	11%	
Smoking history	NE	37.7%	36.0%	82.4%
Diabetes mellitus type-2	19.3%	30.9%	13.0%	17.2%
Disorders of lipoprotein metabolism and other lipidemias	22.0%	56.6%	19.0%	50.3%
Ischemic heart diseases	25.5%	36.5%	17.0%	17.8%
Hypertensive disease	46.5%	75.2%	37.0%	59.8%
Heart failure	13.1%	27.9%	7.0%	19.1%
Pulmonary edema	2.9%	0.9%	<0.5%	3.4%
Pulmonary hypertension	1.2%	4.8%	<0.5%	3.4%
Venous thromboembolism	2.5%	10.4%	4.0%	7.5%
Cerebrovascular disease	9.3%	23.5%	10.0%	9.9%
Arrhythmia	15.5%	28.1%	11.0%	11.3%
Asthma	4.3%	36.3%	12.0%	10.2%

(*Continued*)

Table 3 (Continued).

	Canada	Germany	The Netherlands	Spain
Chronic kidney disease	10.6%	20.8%	5.0%	10.7%
Severe mental illness	6.2%	15.9%	NE	14.9%
Anxiety disorder	16.2%	45.8%	NE	37.9%
Severe mental illness or anxiety disorder	NE	NE	20.0%	NE
Medication use in the past 12 months				
COPD-related:				
Long-acting inhaler (any)	NE	NE	27.0%	92.0%
Long-acting inhaled COPD drug use as single therapy, n (%)	31.40%	NE	NE	NE
Long-acting inhaled COPD drug use as combination therapy, n (%)	33.40%	NE	NE	NE
LABA	NE	49.4%	NE	66.1%
LAMA	NE	29.4%	NE	43.5%
ICS	NE	43.4%	NE	59.2%
Short-Acting inhaler (any)	37.1%		20.0%	96.1%
SABA	NE	40.5%	NE	95.5%
SAMA	NE	2.7%	NE	10.4%
Roflumilast or Theophylline	0.9%	6.5%	<0.5%	1.8%
Cardiac medication (any)	61.0%	NE	47.0%	91.4%
Antithrombotic agents	NE	NE	NE	15.2%
Antiarrhythmics, class I and II	NE	2.2%	NE	10.6%
Digitalis glycosides	NE	3.9%	NE	inc above
Antihypertensives (others)	NE	5.2%	NE	5.4%
Diuretics	NE	35.1%	NE	44.2%
Beta blocking agents	NE	41.2%	NE	48.0%
Calcium channel blockers	NE	25.1%	NE	11.4%
Agents acting on the renin-angiotensin system	NE	57.0%	NE	52.6%
Metabolic medication (any)	43.6%	41.5%	34.0%	62.3%
Lipid modifying agents	NE	33.3%	NE	53.3%
Antidiabetic agents	NE	19.3%	NE	19.0%

Notes: [a]The structure of the data in Germany differed from that in other studies, with dates of GP visits being quarters of years rather than the specific date, month and year.
Abbreviation: NE, not evaluated.

Table 1). The aHRs for associations with HF during the 1–7 days post-exacerbation period were substantially higher for Canada and the Netherlands than for Germany and Spain. Further, the percentage reduction in aHR from the 1–7 day post-exacerbation period to the 8–14 days period varied across studies and was highest for Canada at 87% followed by the Netherlands at 53%. For Germany and Spain, respective reductions of 29% and 45% were observed, which were

https://doi.org/10.2147/CLEP.S520168

lower than the reductions in aHR for all-cause death, for which there is no risk of misdiagnosis. These results point to differences across studies in the extent of misdiagnosis of COPD exacerbations and HF as a heterogeneity source, and additionally suggest variation in immediate risk of death following an exacerbation.

Clinical Diversity Discussion

Methodological diversity across studies could not explain entirely the statistical heterogeneity observed. The research team discussed which real differences in population and healthcare system characteristics could be at play. For instance, it was hypothesized that local differences in identification or management of COPD exacerbations (exposure) or concurrent cardiovascular disease could influence post-exacerbation risk of cardiovascular outcomes.

Significant differences in baseline characteristics were observed, which may modify associations between post-exacerbation time periods and HF (Table 3): the proportion of male patients ranged from 52% in Canada to 78% in Spain; patients in the Netherlands were younger, on average (only newly diagnosed patients with COPD were included); prevalence of cardiac conditions was higher in Germany and Spain (for example, hypertensive disease prevalence ranged from 37% in the Netherlands to 75% in Germany); and cardiac medication prevalence ranged from 47% in the Netherlands to 91% in Spain.

We discussed which factors could truly differ across studies and lead to varying levels of misdiagnosis between exacerbations and HF. In EXACOS-CV estimates of aHRs for HF during these periods were lowest for Germany, followed by Spain. Germany showed highest prevalence of cardiac comorbidity that may not be fully explained by coding practices or measurement issues. In Spain, higher cardiac medication use was observed. The hypothesis is that there is an association between the prevalence of diseases in a country, and its pattern of management; for instance, in countries where CV diseases are more prevalent, diagnoses of CV disease may be more accurate, or CV events may be more efficiently managed. Similarly, in countries where CV disease are better managed, the disease is more diagnosed and appears to be more prevalent.

Another hypothesis is that patients with higher severity of cardiac disease are at higher risk of death immediately after an exacerbation; thus, potentially explaining the higher aHRs for all-cause death during 1–7 days post-exacerbation in Germany and Spain, where prevalence of cardiac comorbidity and cardiac medication usage is higher.

Discussion

In this case study using the example of the EXACOS-CV multi-database study program, we present a novel framework for identifying and systematically exploring sources of statistical heterogeneity and generate potential hypotheses on sources of clinical variation across studies. This separated methodological (including database-related) sources of heterogeneity from true sources of variation so that new hypotheses about the roles of differences in patient demographics, underlying severity of disease and healthcare system characteristics could be generated. The framework was applied to the EXACOS-CV multi-database study but has general applicability to studying heterogeneity among secondary data studies.

Sources of concern about heterogeneity in multi-database studies have been discussed in previous literature, with others highlighting the need for data source tailored choices in such studies.[23] Within our heterogeneity framework, a novel contribution is a formal checklist of considerations about methodology diversity across secondary data studies, which will aid identification of concerns for heterogeneity in multi-database studies, prompting exploration thereof and potentially leads to generation of new hypotheses about factors that truly differ across studies and their impact on the exposure-outcome associations.

In the case study presented, results of the four included EXACOS-CV cohort studies were consistent in terms of the direction of post-exacerbation associations with HF and all-cause death. This increases confidence in conclusions of the individual studies, demonstrating a sharp increase in the likelihood of these outcomes following a COPD exacerbation that is sustained up to 1 year following the exacerbation. The mechanisms underpinning the association between an exacerbation and the increased risk of HF decompensation are multiple and well understood.[24] What is unclear, is whether the observed differences in strength of association can be explained by true variation in populations and

healthcare systems characteristics across countries and/or databases. The present study provides a framework allowing for this type of exploration which, in turn, the generation of hypothesis on drivers of statistical heterogeneity.

Statistical heterogeneity among EXACOS-CV concerns the strength of associations. Concerns about methodological diversity included the potential over-diagnosis of COPD (inclusion of patients with asthma and without COPD) in Germany and over-reporting of cardiovascular comorbidity at baseline in German claims data due to reimbursement-related reasons. The former concern was not supported by the data; the inclusion of asthma-only (non-COPD) patients would be expected to lead to underestimation of aHRs of all-cause death, as the likelihood of death following an asthma flare-up is not as high as this following a COPD exacerbation.[25] This was not the case as statistical heterogeneity for the outcome of all-cause death was very low. The latter concern could not be ruled out; the prevalence of cardiovascular comorbidities was higher than observed in an observational study using primarily collected data. However, the higher medication use observed in Germany and Spain would rule out at least some concerns about over-recording for reimbursement issues. It is possible that variation in prevalence of comorbidities may be in part due to under-coding in the database used in the Canadian study and over-coding in German claims data, as well as the inclusion of only newly diagnosed COPD patients (with less cardiovascular comorbidity) in the Netherlands, and true differences across countries in the characteristics of individuals.

Behavioral factors tend to be poorly captured in secondary databases. There were differences in the distributions of age and sex across EXACOS-CV study populations, and it could be expected that these populations differ in lifestyle factors. We cannot rule out the inability to capture smoking status as a concern for heterogeneity, despite the lack of influence of adjusting for smoking on the results of the study in Spain, since smoking information even in Spain was rather limited, with no information on quantity or duration.

Findings of our explorations showed that difference in adjusted models is a concern for statistical heterogeneity among EXACOS-CV studies. Heterogeneity was reduced by applying the same adjusted model to each study's database. While differences across studies in the list of confounders adjusted for were not enough to fully explain the observed heterogeneity among EXACOS-CV studies, we have identified several factors which may collectively contribute to the observed heterogeneity. EXACOS-CV studies consistently showed an increased risk of decompensation of HF following exacerbations of COPD.

We observed that the association with decompensation HF during the 1–7 day period following an exacerbation was lowest in magnitude for Germany and Spain, which showed the highest levels of cardiovascular comorbidity and cardiovascular medication use. In contrast, the association with all-cause death during the same period following an exacerbation was highest in magnitude for Germany and Spain. True higher prevalence of cardiovascular diseases could suggest that post-exacerbation diagnoses of decompensated HF were more accurate (lower risk of misclassification of exposure and outcome) in those two countries. However, we were unable to identify sources of bias which would additionally explain the higher association with death in Germany and Spain.

There are limitations to our explorations of heterogeneity among EXACOS-CV studies. It was not possible to fully quantity the extent to which between-study heterogeneity can be explained by each possible source. Analyses of patient-level data would have helped in the assessment of heterogeneity sources; for example, in quantifying heterogeneity explained by stratifying on history of cardiac comorbidity. However, this research was limited in terms of the exploratory analyses that could be done, due to the requirements for additional resourcing. In addition, we studied each individual concern for heterogeneity in turn; it is possible that multiple factors act together to cause between-study heterogeneity, but this would be difficult to investigate. A further limitation is that the impact of healthcare system-related characteristics on between-study heterogeneity could not be quantitively explored, since we had no information on healthcare system-related characteristics such as hospital type (tertiary/general hospitals), access to healthcare and physician characteristics.

Conclusion

Our framework for exploration of statistical heterogeneity provides a way to conceptualize variation among results of multi-database studies and disentangle true sources of variation from methodological (including database-related)

variation. Methodological diversity checklists have been generated for use in multi-database programs, as well as considerations about true sources of variation that may lead to generation of new hypotheses.

Heterogeneity among multi-database study results is a source of information, providing directional insights on the research question while considering healthcare system and population differences.

Ethics Statement

This article used reported data from published studies.[7–10] The analysis focuses solely on statistical examination of aggregate-level data from these publications. No patient-level data, new primary data collection, patient interaction, or interventions were involved. Consequently, Institutional Review Board (IRB) review was not sought, consistent with standard procedures for studies using aggregate-level data from existing publications.

Acknowledgments

The authors thank Ignacio Hernández from Atrys Health for aiding our investigations by performing additional analyses of EXACOS-CV study data in Spain. We acknowledge support from CMC Connect, a division of IPG Health Medical Communications and funded by AstraZeneca, for producing the figures.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was funded by AstraZeneca.

Disclosure

KR and CN are employees of AstraZeneca and hold shares/stock options of AstraZeneca. EG is a scientific consultant to AstraZeneca. BNB is an employee of the PHARMO Institute for Drug Outcomes Research. This independent research institute performs financially supported studies for government and related healthcare authorities and several pharmaceutical companies. HM is an employee of AstraZeneca and holds shares/stock options of AstraZeneca, GSK and Haleon; PE is an employee of Medlior Health Outcomes Research Ltd. which received funding for the EXACOS-CV study from AstraZeneca. ML is an employee of IQVIA which received funding for the EXACOS-CV study from AstraZeneca. NK is an employee of WIG2 Institute, which received funding for the EXACOS-CV study from AstraZeneca. NH has received honoraria or research support from AstraZeneca, Novartis, Servier, and BI-Lilly. The authors report no other conflicts of interest in this work.

References

1. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions version 5.0.1. 2008. Available from: http://handbook.cochrane.org. Accessed May 13, 2024.
2. International Society for Pharmacoepidemiology. DIVERSE project: protocol for the scoping review. 2021. Available from: https://catalogues.ema.europa.eu/sites/default/files/document_files/DIVERSE_protocol_v1.0.pdf. Accessed May 13, 2024.
3. Klungel OH, Kurz X, de Groot MC, et al. Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project. *Pharmacoepidemiol Drug Saf*. 2016;25(Suppl 1):156–165. doi:10.1002/pds.3968
4. Madigan D, Ryan PB, Schuemie M, et al. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol*. 2013;178(4):645–651. doi:10.1093/aje/kwt010
5. Nordon C, Rhodes K, Quint JK, et al. EXAcerbations of COPD and their OutcomeS on CardioVascular diseases (EXACOS-CV) Programme: protocol of multicountry observational cohort studies. *BMJ Open*. 2023;13(4):e070022. doi:10.1136/bmjopen-2022-070022
6. Müllerová H, Marshall J, de Nigris E, et al. Association of COPD exacerbations and acute cardiovascular events: a systematic review and meta-analysis. *Ther Adv Respir Dis*. 2022;16. doi:10.1177/17534666221113647
7. Hawkins NM, Nordon C, Rhodes K, et al. Heightened long-term cardiovascular risks after exacerbation of chronic obstructive pulmonary disease. *Heart*. 2024;110(10):702–709. doi:10.1136/heartjnl-2023-323487

8. Vogelmeier CF, Rhodes K, Garbe E, et al. Elucidating the risk of cardiopulmonary consequences of an exacerbation of COPD: results of the EXACOS-CV study in Germany. *BMJ Open Respir Res*. 2024;11(1):e002153. doi:10.1136/bmjresp-2023-002153
9. Santos S, Manito N, Sanchez-Covisa J, et al. Risk of severe cardiovascular events following COPD exacerbations: results from the EXACOS-CV study in Spain. *Rev Esp Cardiol*. 2024;78:138–150. doi:10.1016/j.rec.2024.1006.1003
10. Swart KMA, Baak BN, Lemmens L, et al. Risk of cardiovascular events after an exacerbation of chronic obstructive pulmonary disease: results from the EXACOS-CV cohort study using the PHARMO Data Network in the Netherlands. *Respir Res*. 2023;24(1):293. doi:10.1186/s12931-023-02601-4
11. Nordon C, Battin C, Verdoux H, et al. The use of random-effects models to identify health care center-related characteristics modifying the effect of antipsychotic drugs. *Clin Epidemiol*. 2017;9:689–698. doi:10.2147/CLEP.S145353
12. Thompson S, Ekelund U, Jebb S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. *Int J Epidemiol*. 2011;40(3):765–777. doi:10.1093/ije/dyq248
13. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573–577. doi:10.7326/0003-4819-147-8-200710160-00010
14. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. *J R Stat Soc Ser a Stat Soc*. 2009;172(1):21–47. doi:10.1111/j.1467-985X.2008.00547.x
15. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–560. doi:10.1136/bmj.327.7414.557
16. Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res Synth Methods*. 2019;10(1):83–98. doi:10.1002/jrsm.1316
17. Viechtbauer W. Package 'metafor'. 2015. Available from: https://cran.r-project.org/web/packages/metafor/metafor.pdf. Accessed May 13, 2024.
18. Mannino DM, Higuchi K, Yu T-C, et al. Economic burden of COPD in the presence of comorbidities. *Chest*. 2015;148(1):138–150. doi:10.1378/chest.14-2434
19. Chetty U, McLean G, Morrison D, Agur K, Guthrie B, Mercer SW. Chronic obstructive pulmonary disease and comorbidities: a large cross-sectional study in primary care. *Br J Gen Pract*. 2017;67(658):e321–e328. doi:10.3399/bjgp17X690605
20. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD*. 2009;6(5):388–394. doi:10.1080/15412550903140865
21. Greulich T, Weist BJD, Koczulla AR, et al. Prevalence of comorbidities in COPD patients by disease severity in a German population. *Respir Med*. 2017;132:132–138. doi:10.1016/j.rmed.2017.10.007
22. Karch A, Vogelmeier C, Welte T, et al. The German COPD cohort COSYCONET: aims, methods and descriptive analysis of the study population at baseline. *Respir Med*. 2016;114:27–37. doi:10.1016/j.rmed.2016.03.008
23. Roberto G, Leal I, Sattar N, et al. Identifying cases of type 2 diabetes in heterogeneous data sources: strategy from the EMIF project. *PLoS One*. 2016;11(8):e0160648. doi:10.1371/journal.pone.0160648
24. Simons SO, Heptinstall AB, Marjenberg Z, et al. Temporal dynamics of cardiovascular risk in patients with chronic obstructive pulmonary disease during stable disease and exacerbations: review of the mechanisms and implications. *Int J Chron Obstruct Pulmon Dis*. 2024;19:2259–2271. doi:10.2147/COPD.S466280
25. Engelkes M, de Ridder MA, Svensson E, et al. Multinational cohort study of mortality in patients with asthma and severe asthma. *Respir Med*. 2020;165:105919. doi:10.1016/j.rmed.2020.105919