

Pre-diagnostic comorbidity pathways are associated with the timing of pulmonary hypertension outcomes: a case control study

Mohsin Masood · Hong Pan · Juan Delgado-SanMartin · Siying Jane Ong · Xiaohe Zhang · Naailah Bhamji · Roger Thompson · Allan Lawrie · Dennis Wang 

Received: 11 September 2025 / Accepted: 13 August 2026

© The Author(s) 2026

Abstract

Background Pulmonary hypertension (PH) is a life-threatening condition often diagnosed late due to non-specific symptoms and overlapping comorbidities with other cardiovascular and respiratory diseases. Although there have been advances in treatment, diagnostic delays are common, and patients are treated late in their disease course. Investigating the temporal ordering of recorded comorbidities may provide insight into diagnostic patterns associated with delayed PH recognition and adverse outcomes.

Methods We identified 2,727 PH patients from the UK Biobank cohort of 440,014 participants with diseases diagnoses, and sampled two demographics-matched cohorts of chronic obstructive pulmonary disease (COPD) and heart failure (HF) with the same number of patients as the PH cohort but were not diagnosed with PH. Employing process mining, a methodology originating from business process analysis to extract temporal patterns from event logs, the temporal diagnostic sequences of comorbidity combinations preceding PH diagnosis were compared to those sequences in patients who received a final diagnosis COPD or HF. We also assessed the relative risk of mortality given different combinations of comorbidities and the genetic correlation between comorbidities.

Results Patients diagnosed with PH had prior diagnoses of hypertension (58.1%), atrial fibrillation (31.2%), and heart failure (43.2%). Process mining revealed that the temporal order of cardiometabolic comorbidities influences the timing to diagnosis of PH, COPD and HF. Patients in which type 2 diabetes was diagnosed before systemic hypertension had received PH diagnosis 2 years and 9 months faster than the reverse sequence, where hypertension preceded diabetes. For patients with a final diagnosis of PH, HF and COPD, diagnostic sequences in which hypercholesterolemia was recorded before systemic hypertension and type 2 diabetes were associated with shorter median intervals to PH diagnosis. A sequence involving diabetes, hypertension, and hypercholesterolemia was associated with a significantly higher mortality risk ($HR = 1.97$; $P = 0.001$). Additional mortality risk factors included specific orderings of metabolic conditions and comorbidity clusters.

Conclusion Temporal diagnostic sequences of comorbidities were associated with the timing of PH diagnosis. Our findings suggest that the presentation and ordering of metabolic diseases may serve as hypothesis-generating

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Hong Pan, Juan Delgado-SanMartin, Siying Jane Ong, Xiaohe Zhang, Naailah Bhamji, Roger Thompson, Allan Lawrie, and Dennis Wang contributed equally to this work.



epidemiological indicators of PH risk. Given the limited PH subtype granularity available within UK Biobank, these findings should be interpreted within a heterogeneous PH population and require further validation in clinically characterized cohorts with detailed PH subtype information.

Keywords Pulmonary hypertension · Process mining · Comorbidity sequence · Mortality · Longitudinal analysis

Abbreviations

PH	Pulmonary hypertension
COPD	Chronic obstructive pulmonary disease
HF	Heart failure
ICD	International Classification of Diseases
HR	Hazard ratio
LDSC	Linkage disequilibrium score regression
GWAS	Genome-Wide Association Studies

Introduction

Pulmonary Hypertension (PH) is a chronic cardiovascular disease described by high blood pressure in the pulmonary arteries. PH affects the quality of life and reduces life expectancy [1, 2] and regularly develops in conjunction with heart failure and chronic lung conditions [3–5]. The initial symptoms of PH, including breathlessness, fatigue, and dizziness, may contribute to diagnostic uncertainty and alternative clinical coding prior to PH recognition. As a result, patients may initially receive diagnoses of other conditions such as asthma or heart failure (HF), which can be associated with delays in definitive PH diagnosis [6, 7]. Right heart catheterization to measure pulmonary artery pressure stands as the best method to confirm PH diagnosis, but its invasive and specialized nature prevents it from serving as an initial screening procedure [8]. Survival rates vary between different subgroups of PH [9], and previous studies have shown even mildly elevated mean pulmonary artery pressure of 19–24 mm Hg are associated with an increased risk of death. Falling within this borderline range has redefined earlier diagnostic guidelines to emphasize the importance of early detection and management [10–12]. Yet patients suspected of having PH may need multiple visits to primary healthcare providers and specialists to undergo numerous medical tests, before a diagnosis is reached, particularly in cases of pulmonary arterial hypertension (PAH), where symptoms are often non-specific and diagnosis comes from by exclusion [13]. A survey from the Pulmonary Hypertension Association UK revealed that one-third of PH patients waited over three years for diagnosis confirmation [14]. To reduce the diagnostic delay, medical professionals should enhance awareness and screening of patients so that the most at risk of PH receive diagnostic tests earlier [7, 8].

Prior studies have faced challenges in identifying PH subtypes due to limited coding granularity, which may increase the potential for misclassification [15, 16]. Late detection is also common, as PH frequently presents alongside cardiovascular and respiratory comorbidities. This highlights the need to evaluate whether prior diagnosis of comorbidity conditions can aid in early detection and risk stratification of PH [17]. Previous studies show PH often co-occurs with cardiovascular conditions (e.g., hypertension, atrial fibrillation, heart failure) and respiratory disorders (e.g., COPD, interstitial lung disease, pulmonary fibrosis) [18, 19]. However, identifying which comorbidities are associated with earlier or later recorded PH diagnosis remains a challenge. Importantly, the relative timing of these comorbidities relative to diagnosis and each other may offer deeper insights into disease risk. These patterns may help identify opportunities for earlier investigation and risk stratification [20–22]. Several studies have emphasized the importance of understanding the

temporal sequence and clustering of comorbidities. For instance, findings from the COMPERA registry showed increasing comorbidity rates over time in PH patients, underscoring the need for adaptive treatment approaches [23]. A 15-year longitudinal study by Santiago-Vacas et al. [24] examined temporal patterns in the recorded diagnosis of PH among patients with heart failure and right ventricular dysfunction. However, such studies did not specifically evaluate diagnostic sequences in which cardiovascular symptoms or comorbidities were recorded prior to PH diagnosis.

In this study, we analyzed the temporal ordering and timing of recorded comorbidity diagnoses preceding PH diagnosis using a data driven technique in business analytics, process mining, that has been successfully applied in different disease contexts [25–29]. We compared different temporal diagnostic sequences of comorbidities before PH diagnosis with demographic-matched cohorts of heart and lung diseases. To assess the molecular underpinnings of PH comorbidities, we incorporated publicly available GWAS summary statistics and applied linkage disequilibrium score regression (LDSC) to estimate genetic correlations between PH and its major comorbidities. Finally, we estimated the increased risk of mortality associated with the different comorbidity sequences prior to final diagnosis.

Methods

Identifying a PH cohort

From the 502,499 UK Biobank participants with electronic health records under our application ID 85457, we extracted the ICD-10 codes dataset, containing diagnostic information for 440,017 participants (covering the period from 1992 to 2021), and the all-cause mortality dataset, which included death records for 37,897 participants and their detailed causes of death for 37,731 (99.5%) of them. After removing duplicates, 3 participants were excluded due to missing key information (e.g., incomplete diagnostic records or inconsistent ICD-10 codes). This resulted in a final data set comprising 440,014 participants. From this, we identified a cohort of 2,727 PH patients. We subdivided the population into three primary classes: Primary PH (I27.0, $n=776$), Other Secondary PH (I27.2, $n=1,949$), and Secondary PH (I27.9, $n=395$). This PH cohort of 2,727 included 1,451 males (53.21%) and 1,276 females (46.79%), differing slightly from the broader UK Biobank population, where males comprised 45.17% and females 54.83%. A full breakdown of the cohort characteristics is provided in Supplementary Table S8. ICD-10 PH codes map onto the PH subgroups 1–5: Group 1 (PAH, I27.0) and Groups 2–5 covering secondary PH from heart disease, lung disease, thromboembolism, or multifactorial causes [1, 10].

Identifying comorbidities in the PH cohort, and in matched COPD and heart failure cohorts

We further extracted the most frequently co-occurring ICD-10 diagnoses to identify comorbidities within this population before categorizing them at the subchapter level. In the PH patient group consisting of 2,727 participants, essential hypertension was also present in 1,611 (59.1%), atrial fibrillation in 868 (31.8%), COPD in 1,822 (66.8%), heart failure in 1,179 (43.2%), hypercholesterolemia in 830 (30.4%), and diabetes mellitus in 579 (21.2%) patients. Since COPD and HF are thought to have related pathology to PH, we examined their comorbidities as comparable controls to the comorbidities observed for PH. We extracted a COPD-matched cohort and a heart failure (HF)-matched cohort, each consisting of 2,441 participants who had never been diagnosed with PH. Participants in the COPD and HF comparison cohorts were restricted to individuals with no recorded PH-related ICD-10 codes (I27.0, I27.2, I27.9) at any time during follow-up. A detailed description of the control cohort selection, ICD-10 codes and matching methodology is provided in the Supplementary Materials (Sect. 1.1–1.2, Figure S1 and S3).

Assessing genetic overlap between PH and comorbidities using GWAS Data

We used linkage disequilibrium score regression (LDSC v1.0.1) to evaluate whether PH and its comorbidities share common genetic risk factors by estimating SNP-based heritability and pairwise genetic correlations. We extracted common genetic variants called SNPs from publicly available GWAS summary statistics while concentrating on European-ancestry groups and high-quality HapMap3 SNPs (Supplementary material: Table S9 – S10). We excluded comorbid traits that showed very low heritability levels below 1% and those with non-significant correlation results exceeding a p-value of 0.05. Supplementary Materials contain detailed descriptions of dataset sources, quality control filters, and LDSC parameters used in our methodology.

Process mining of comorbidity progression over time

We extracted dates of diagnostic events from hospital inpatient records and primary care data from the UK Biobank which links each patient's ICD-10 diagnosis to the date they were diagnosed or hospitalized. The initial occurrence date was used to timestamp all ICD-10-coded diagnoses for each participant. We further identified the PH diagnosis date by selecting the first recorded date among all ICD-10 codes related to PH which include I27.0, I27.2, and I27.9. Comorbidities are sorted into chronological order based on the PH diagnosis date, then separated into pre-PH (diagnosed before PH) and post-PH (diagnosed after PH) categories. Diagnoses recorded on the same date were treated as concurrent events, and no directional ordering was assumed for such events in the temporal analyses. We utilized process mining techniques in the bupaR package in R to create sequences of comorbidities that describe the order and time between diagnoses. The ordered diagnostic sequences were visualized as event logs to illustrate temporal patterns of recorded comorbidities preceding PH, COPD, and HF diagnoses. For longitudinal trajectory analyses, follow-up began at each participant's earliest recorded diagnosis across linked primary care and inpatient records to reconstruct comorbidity sequences preceding PH. For survival analyses, analytic time zero was defined as the first recorded PH diagnosis. Sensitivity analyses using incident PH definitions were performed to minimize potential left truncation and immortal-time bias due to differences in data capture across sources. Participants were followed until death or administrative censoring (30 November 2021), whichever occurred first.

Estimating mortality risk associated with different comorbidity sequences

To evaluate mortality risk associated with each identified comorbidity sequence, we used linked mortality data from the UK Biobank. We performed time-to-event analysis using Cox proportional hazards models, estimating hazard ratios (HRs) to quantify the risk of death associated with specific comorbidity trajectories. Models were adjusted for potential confounders including age at PH diagnosis, smoking status, body mass index category, socioeconomic deprivation (IMD quintiles), and index year where data availability permitted (details in Supplementary material 1.6). We pre-specified primary hypotheses focusing on major metabolic-cardiovascular diagnostic sequences preceding PH diagnosis. All additional sequence contrasts were treated as exploratory. To account for multiple comparisons, both Bonferroni correction and false discovery rate (FDR) adjustment were applied. Hazard ratios are reported with 95% confidence intervals together with adjusted p-values (details in Supplementary material 1.8). Cause-specific Cox models were additionally fitted to evaluate cardiovascular mortality while accounting for competing risks. The proportional hazards assumption was assessed using Schoenfeld residual-based tests, with detailed diagnostic results provided in Supplementary Material Sect. 1.7.

Software and code availability

The data cleaning and process mining stages of the analysis were implemented in Python (v 3.9.12) and R (v 4.3.2). The complete set of experiments and associated code is freely available on <https://doi.org/10.5281/zenodo.15747766>.

Results

Pulmonary hypertension has a similar profile of pre-diagnosis comorbidities as heart failure

There were 2,727 individuals in UK Biobank who received a PH diagnosis from 1992 to 2021 as described by their ICD-10 codes. We classified comorbid conditions into three types based on their diagnosis timing relative to the assignment of a PH diagnosis code (I27.0, I27.2, I27.9): pre-PH comorbidities (diagnosed only before PH diagnosis), post-PH comorbidities (diagnosed only after PH diagnosis), and common PH comorbidities (diagnosed both before and after PH diagnosis). To eliminate duplication, we included only the first recorded instance of each comorbidity. The distribution of comorbidities across PH, COPD, and HF was consistent with patterns reported in specialized PH centers [30] (Fig. 1). For the COPD matched control cohort, the most prevalent comorbidities included chronic ischemic heart disease with 1,377 cases representing 56.4%, hypertension with 1,138 cases equating to 46.6%, along with hypercholesterolemia which occurred in 543 cases or 22.2%. A total of 1,746 individuals in the HF matched cohort suffered from hypertension (71.5%), followed by hypercholesterolemia in 1,042 individuals (42.7%) and atrial fibrillation in 803 individuals (32.9%).

All comorbidities are diagnosed before their respective index conditions, which indicates that these conditions were frequently recorded before the index diagnosis rather than effects of those diseases. The most common comorbidity found before diagnosis was essential hypertension with a prevalence of 58.1% in PH patients, 38.5% in COPD patients, and 69.7% in HF patients (Fig. 2). Atrial fibrillation and flutter showed high prevalence in PH patients (31.2%) and HF patients (31.3%) while remaining significantly less common in COPD patients (6.5%, $\chi^2(2)=568.07, p<0.0001$). Asthma was most prevalent in PH patients (15.8%), compared to both COPD (1.6%) and HF (9.8%). PH patients exhibited a pattern of cardiovascular comorbidities, including hypertension, atrial fibrillation, atherosclerosis, and hypercholesterolemia, that was more similar to that of HF patients than to COPD patients. Unlike COPD patients, who showed the highest prevalence of chronic ischemic heart disease (56.4%), PH patients had a lower prevalence (14.9%). The pre-diagnosis timing of comorbidities in PH patients followed a pattern that was more comparable to HF patients than to COPD patients.

Strong genetic overlap between PH and cardiovascular comorbidities

We observed the strongest genetic correlation between PH and heart failure, with an estimated $r_g=0.65$ ($SE=0.18, P=3.08 \times 10^{-4}$) (Supplementary Figure S4). Notable correlations were also found with hypertension ($r_g=0.62, SE=0.14, P=1.82 \times 10^{-5}$), hypercholesterolemia ($r_g=0.52, SE=0.14, P=3.26 \times 10^{-4}$), atrial fibrillation ($r_g=0.64, SE=0.17, P=1.10 \times 10^{-4}$), and coronary atherosclerosis ($r_g=0.42, SE=0.11, P=1.38 \times 10^{-4}$) (Supplementary Table S10). A smaller but statistically significant correlation was identified between PH and type 2 diabetes ($r_g=0.26, SE=0.08, P=1.92 \times 10^{-3}$). In contrast, no significant genetic correlation was observed between PH and either COPD or asthma in the GWAS datasets analyzed.

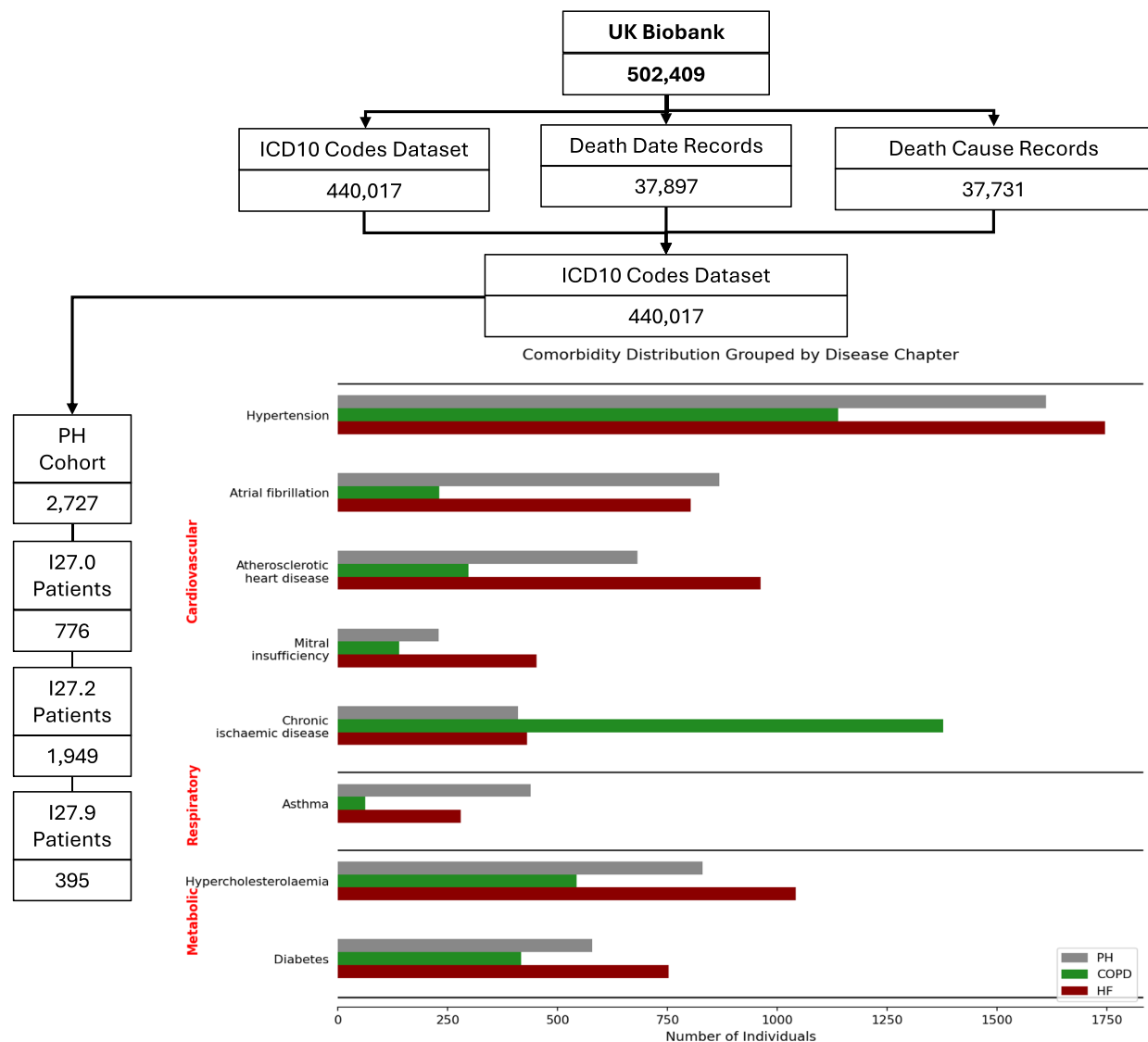


Fig. 1 Study cohorts with extracted diagnoses data: The selection of participants in the UK Biobank ($n=502,409$) towards the identification of 2,727 PH patients, stratified by ICD-10 codes into I27.0, I27.2, and I27.9 classes. The bars show the distribution of common comorbidities among the PH cohort, COPD-matched controls, and HF-matched controls, grouped by the ICD-10 codes of cardiovascular, respiratory, and metabolic comorbidities. The control cohorts with a final diagnosis of HF or COPD were the same size as the PH cohort and matched by demographics

Process mining reveals temporal comorbidity sequences preceding pulmonary hypertension diagnosis

Across participants diagnosed with PH, we applied process mining to examine comorbidity sequences leading to PH (Fig. 3). Hypertension was the most common pre-PH comorbidity, with a direct diagnostic trajectory preceding PH in 565 patients over 3.90 median years. Two main temporal diagnostic sequences were identified: (1) hypertension → atrial fibrillation → PH, where 235 patients developed atrial fibrillation 2.30 years after hypertension onset, persisting 4.31 years in 294 patients before PH; (2) hypertension → atherosclerotic heart disease → PH, where 228 patients developed atherosclerosis 2.66 years after hypertension, progressing to PH over 1.94 years in 297 patients. Asthma persisted 4.52 years in 156 patients, and hypercholesterolemia 2.92 years in 265 patients before PH. After PH diagnosis, mitral valve insufficiency

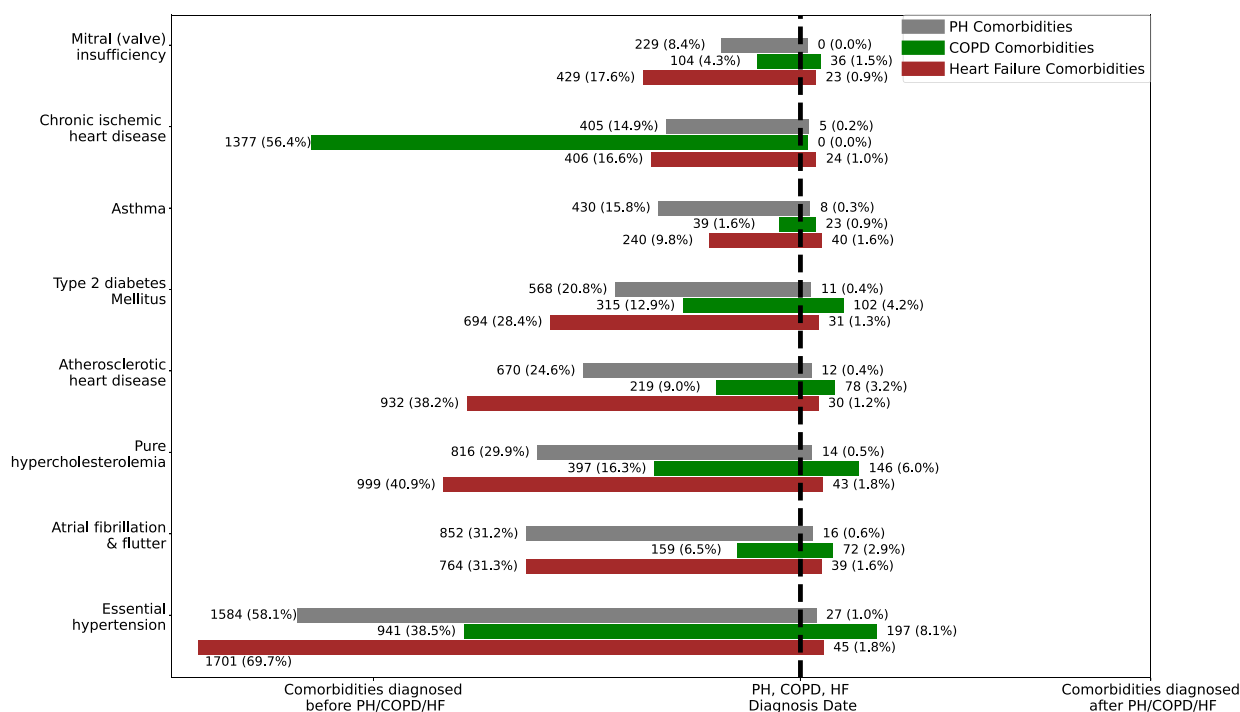


Fig. 2 Timing and frequency of comorbidities in PH, and matched cohorts of COPD and HF: The frequency of cohort individuals with the comorbidity diagnosed before, after, and on the same date as the final diagnosis of either PH, HF or COPD. Each bar is labeled with the number of individuals and as a percentage relative to the size of the cohort

was common, diagnosed in 151 patients within 0.09 years, likely reflecting concurrent detection during echocardiography. Mortality was high, with 900 of 2,727 patients dying within 2.13 years of PH diagnosis.

Temporal sequences of cardiometabolic comorbidities identify accelerated diagnostic trajectories to PH and reveal order-specific risks

To evaluate whether temporal diagnostic ordering of comorbidities was associated with differences in the timing of recorded PH and other cardiopulmonary diagnoses, we compared matched sequences with alternative temporal arrangements (Fig. 4). Analyses focused on temporal diagnostic sequences demonstrating a ≥ 2 -year difference in median inter-diagnosis intervals and consistent patterns across PH, COPD, and HF cohorts. For the sequence diabetes $\rightarrow$ hypertension $\rightarrow$ PH/COPD/HF compared with hypertension $\rightarrow$ diabetes $\rightarrow$ PH/COPD/HF, the median interval to recorded PH diagnosis was 5y11m versus 8y8m, respectively. Similar patterns were observed in COPD (3 m vs. 9y7m) and HF (5y5m vs. 10y3m), suggesting that differences in diagnostic ordering are associated with variation in the timing of subsequent disease diagnoses. Another pair, diabetes $\rightarrow$ hypertension $\rightarrow$ hypercholesterolemia $\rightarrow$ PH/COPD/HF versus diabetes $\rightarrow$ hypercholesterolemia $\rightarrow$ hypertension $\rightarrow$ PH/COPD/HF, showed PH progression in 9y2m vs. 6y9m; similar results occurred in COPD (10y8m vs. 5 m) and HF (8y9m vs. 10y4m), emphasizing hypercholesterolemia's role. The shortest temporal diagnostic sequence was hypertension $\rightarrow$ atherosclerotic heart disease $\rightarrow$ PH, with a median interval of 4y11m compared with 5y11m for the reversed sequence. Comparable differences were observed when COPD (2y2m vs. 1y5m) or HF (4y4m vs. 10y1m) represented the final recorded diagnosis. Detailed median inter-event interval statistics for all diagnostic sequences (Fig. 4) are provided in the Supplementary Material (Tables S11–S13).

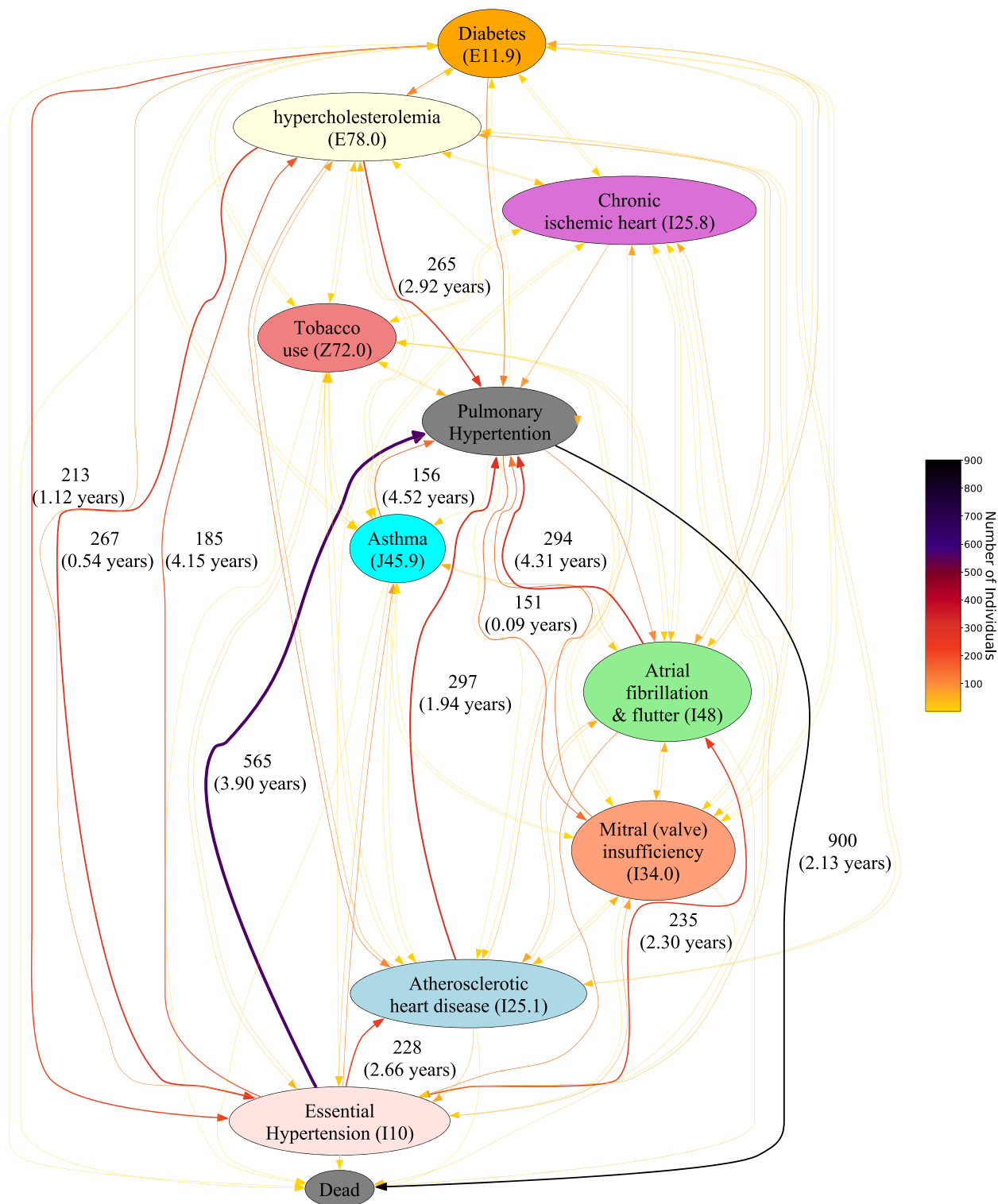


Fig. 3 Process mining of longitudinal comorbidity trajectories in PH cases: Nodes represent individual comorbid conditions, and directed edges indicate the temporal sequence of disease occurrence. Edge thickness and color intensity correspond to the number of individuals within each diagnostic sequence, with darker and thicker edges representing more frequently observed transitions between diagnoses. Numerical labels adjacent to edges indicate the number of patients and the median time (years) between successive diagnoses. The direction of arrows reflects disease progression over time. The pulmonary hypertension node is highlighted to denote the index diagnosis event in trajectory analysis

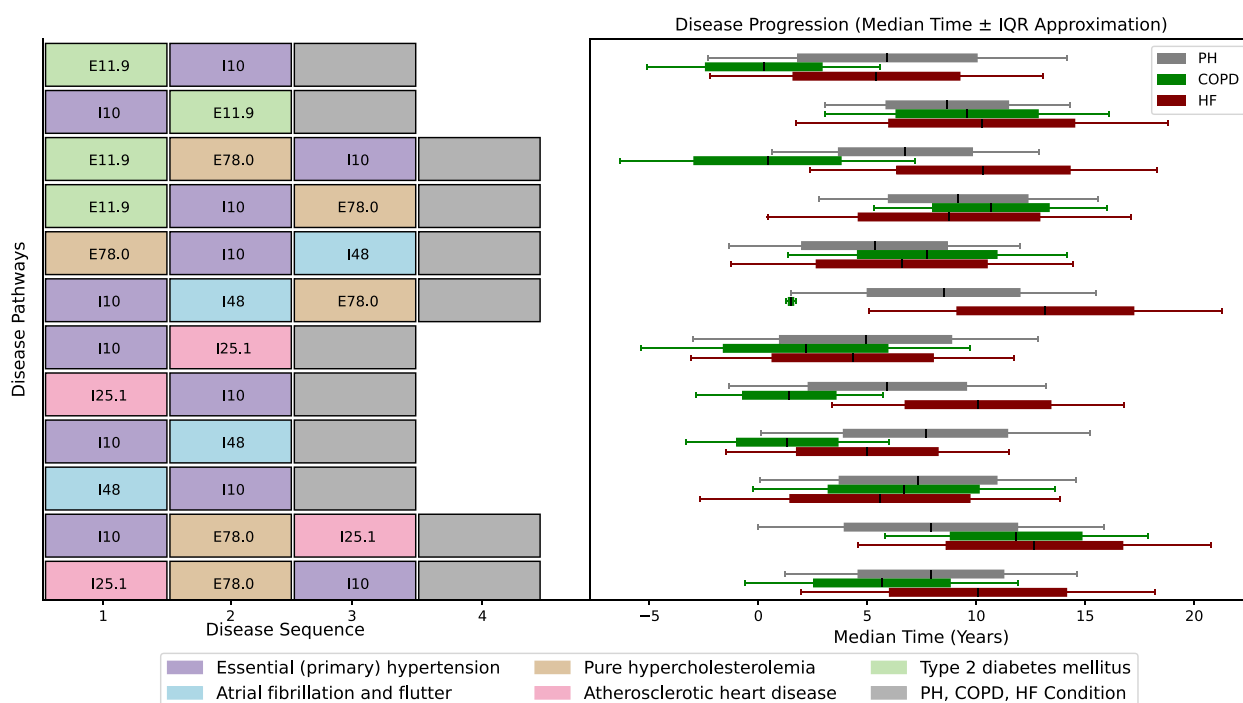


Fig. 4 Comorbidity sequences and median time preceding diagnosis in PH, COPD, and HF cohorts: The cohorts were matched on demographic characteristics and excluded individuals with the corresponding index condition of the comparison cohorts. The left panel visualizes temporal diagnostic sequences preceding pulmonary hypertension (PH), chronic obstructive pulmonary disease (COPD), and heart failure (HF) diagnoses, with each node representing a sequentially recorded ICD-coded condition. Sequences are grouped according to the initial recorded comorbidity. The right panel presents horizontal boxplots showing the median time (years ± IQR approximation) from the first recorded condition in each sequence to the corresponding diagnosis across PH, COPD, and HF cohorts, enabling comparison of temporal diagnostic patterns between cohorts

Mortality risk associated with pre-diagnostic comorbidity sequences

Given the different timelines observed across comorbidity sequences preceding diagnosis, we hypothesized that these temporal diagnostic patterns may also be associated with differences in mortality risk among PH, COPD, and HF cohorts. We assessed overall survival in patients grouped according to their comorbidity sequences preceding PH and COPD diagnoses (Fig. 5). For each comorbidity sequence, Cox proportional hazards models were used to estimate mortality risk, comparing patients within that sequence to all other patients within the same disease group. The resulting hazard ratios also allowed us to compare which comorbidity sequences were associated with higher mortality risk in PH versus COPD populations. With hazard ratios of 1.88 ($p=0.001$) for PH and 1.99 ($p=0.010$) for COPD, the *diabetes* → *hypertension* → *hypercholesterolemia* → PH/COPD sequence had similar mortality risk for both conditions. The sequence: *hypercholesterolemia* → *hypertension* → *atrial fibrillation* → COPD was associated with increased mortality risk in the COPD cohort (HR=2.51, $p=0.001$). In contrast, this same sequence does not show elevated risk in the PH cohort. The sequence associated with the highest mortality risk was *hypertension* → *diabetes* → *hypercholesterolemia* → COPD (HR=3.67, $p=0.001$). In the PH cohort, the same sequence was also associated with increased mortality risk, although the magnitude of association was lower (HR=1.67, $p=0.001$). A considerable mortality risk is shown by the sequence: *atherosclerotic heart disease* → *hypertension* → COPD (HR=2.27, $p=0.010$), while the corresponding sequence preceding PH diagnosis does not show the same mortality risk (HR=1.39, $p=0.16$).

This same comparison was performed to evaluate the mortality risks associated with various comorbidity sequences preceding PH and HF diagnoses (Fig. 6). The sequence: *diabetes* → *hypertension* → *hypercholesterolemia* → PH

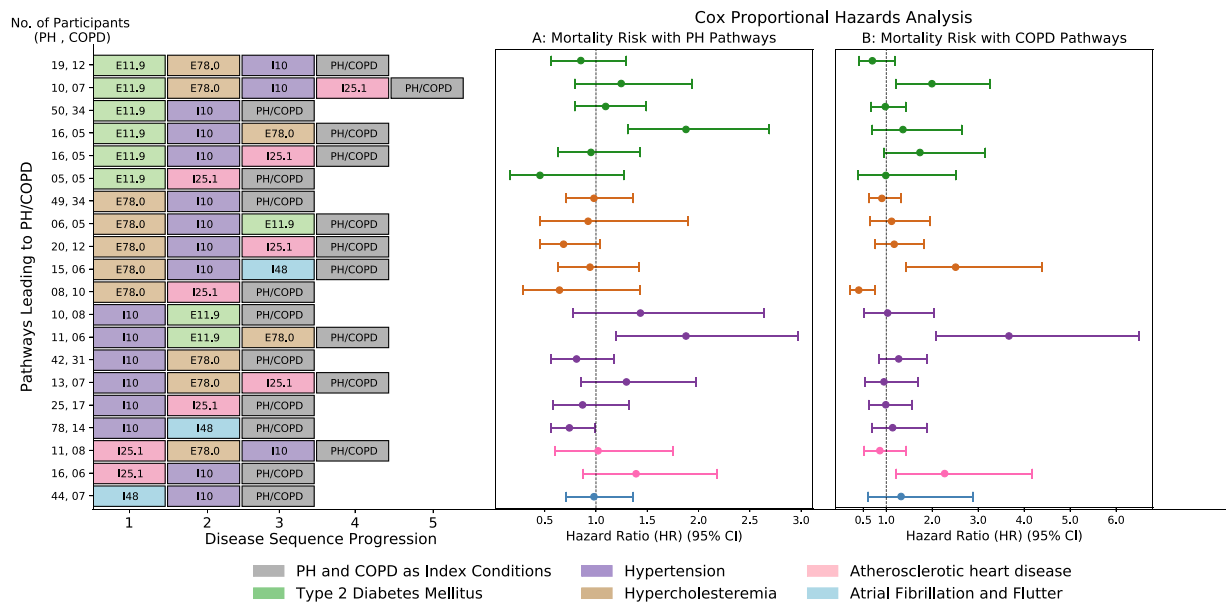


Fig. 5 Diagnostic Sequences Preceding PH and COPD with Mortality Risk Assessment Using Cox Proportional Hazards Analysis: The left panel illustrates temporal diagnostic sequences preceding PH and COPD diagnoses, with each node representing a sequentially recorded ICD-coded condition. The number of participants in each cohort (PH, COPD) and within each diagnostic sequence is shown along the vertical axis. The right panels present hazard ratios with 95% confidence intervals for mortality risk associated with diagnostic sequences preceding PH diagnosis (**A**) and COPD diagnosis (**B**)

presents the highest mortality risk in PH (HR = 1.97, $p = 0.001$). The same sequence of comorbidities also had a poor prognosis in HF patients with an HR of 1.34 and p -value at 0.010. There is a considerable mortality risk for HF patients with a prior comorbidity sequence of *hypercholesterolemia* → *hypertension* → *atrial fibrillation* → *HF* (HR = 1.60, $p = 0.005$). In contrast, the sequence *hypertension* → *diabetes* → *hypercholesterolemia* → PH was associated with the highest observed mortality risk for PH (HR = 2.05, $p = 0.001$). These findings emphasize distinct high-risk multimorbidity sequences across disease cohorts. Notably, no single comorbidity sequence was found to be significantly associated with high mortality risk in both PH and HF cohorts.

Discussion

This study provides insights into the temporal ordering of diagnoses and mortality risks of comorbidities in PH, HF, and COPD, leveraging UK Biobank data. The analysis identified clinically relevant temporal sequences of comorbidities through a dynamic analytical approach that surpasses traditional cross-sectional studies [1, 29]. Hypertension affected 58.1% of patients, with atrial fibrillation (31.2%) and heart failure (43.2%) also highly prevalent. The distribution of the most common comorbidities across PH patients was similar to what was observed in a PH-specific disease cohort [30]. The similarity between the datasets supports the robustness of the observed comorbidity patterns across these populations; however, further validation in independent and clinically characterized cohorts remains warranted. These findings suggest that cardiometabolic multimorbidity represents an important feature of the PH population studied, especially in elderly and high-risk groups. Furthermore, our analysis shows that PH patients are frequently diagnosed with other cardiovascular conditions, hypertension (58.1%) and atrial fibrillation (31.2%) being most common [1–3, 30]. Overlapping metabolic conditions such as diabetes, hypercholesterolemia, and atherosclerotic heart disease were prevalent, supporting the presence of cardiometabolic multimorbidity in PH patients. While these comorbidities are associated with worse outcomes,

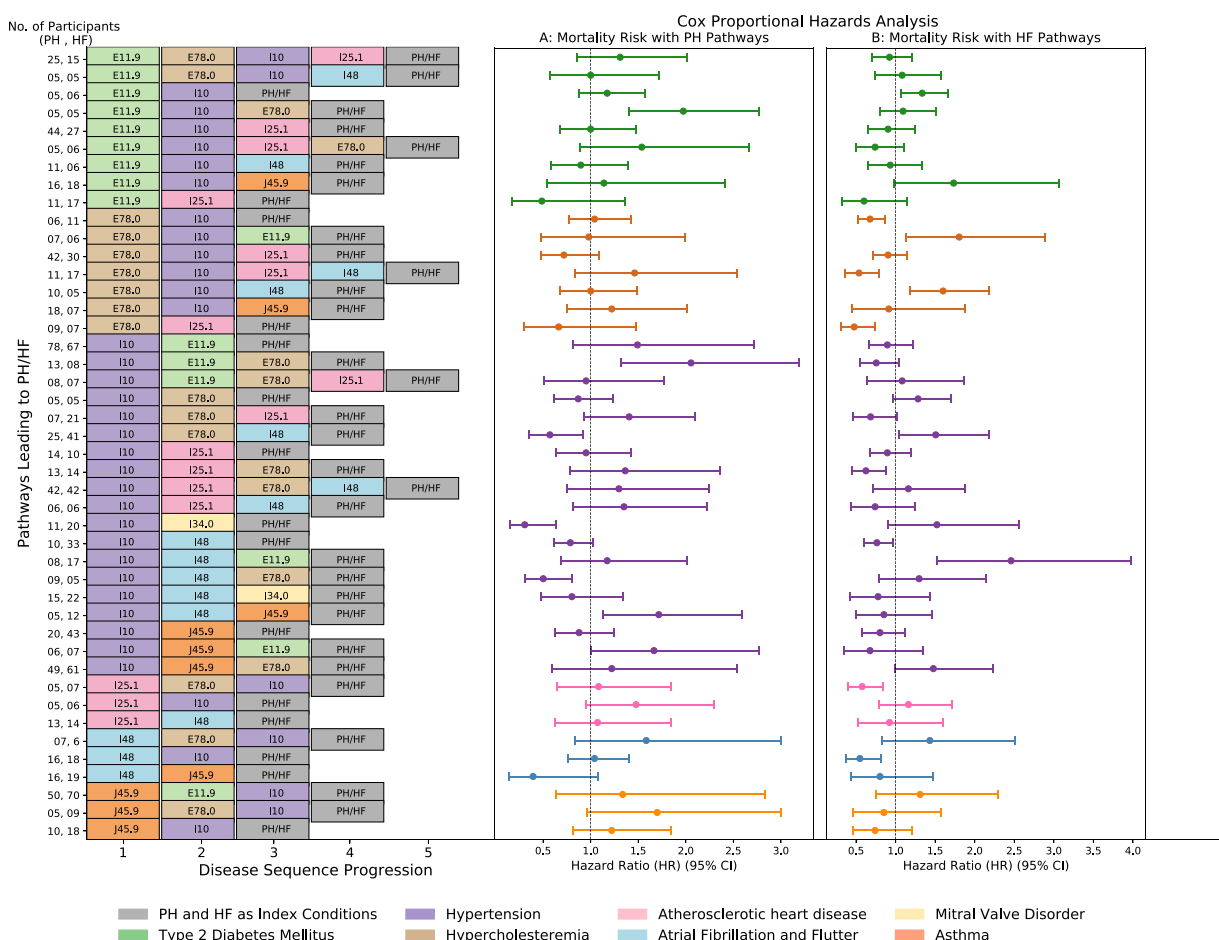


Fig. 6 Diagnostic sequences preceding PH and HF with mortality risk assessment using cox proportional hazards analysis: The left panel illustrates temporal diagnostic sequences preceding PH and HF diagnoses, with nodes representing sequentially recorded ICD-coded conditions and their temporal ordering. The number of participants in each cohort (PH, HF) and within each diagnostic sequence is shown along the vertical axis. The right panels present hazard ratios with 95% confidence intervals for mortality risk associated with diagnostic sequences preceding PH diagnosis (**A**) and HF diagnosis (**B**)

their mechanistic role in PH progression remains unclear and warrants further investigation. Although we did not directly measure systemic inflammation or vascular resistance, the observed comorbidity patterns, including high prevalence of metabolic conditions such as diabetes and hypercholesterolemia, may be associated with PH and its recorded diagnostic patterns [29–31]. Supporting this is the genetic correlation between PH and several other cardiometabolic conditions, most notably heart failure, hypertension, atrial fibrillation, hypercholesterolemia, and type 2 diabetes, suggesting shared genetic liability across these related traits. However, these correlations should be interpreted cautiously, as genetic correlation does not imply causality and may also be influenced by heterogeneity in PH subtypes.

Process mining revealed temporal diagnostic patterns preceding PH diagnosis, often preceded by cardiovascular conditions such as hypertension and atrial fibrillation, and metabolic disorders including diabetes and hypercholesterolemia. These conditions frequently appear in diagnostic sequences before PH, suggesting potential shared risk factors or diagnostic complexity rather than direct causality [32]. The sometimes-fast decline in patient health after PH diagnosis highlights the critical need for prompt intervention and specialized treatment approaches to prevent negative health results [33]. Diagnostic sequences in which diabetes was recorded before hypertension were associated with shorter intervals to PH diagnosis, suggesting a temporal diagnostic pattern that

warrants further investigation. For example, hypertension → atherosclerotic heart disease → PH was associated with the shortest median interval to PH diagnosis (4y11m), suggesting that hypertension frequently appears early within recorded diagnostic sequences preceding PH. These findings align with earlier reports linking metabolic syndrome components to PH pathogenesis [34], but our results provide a longitudinal lens, demonstrating that sequence, not just the presence of comorbidities, carries important epidemiological implications.

For PH, the sequence *diabetes* → *hypertension* → *hypercholesterolemia* → PH was associated with the highest mortality risk. This pattern reflects the frequent co-occurrence of diabetes, hypertension, and hypercholesterolemia within PH diagnostic sequences associated with increased mortality risk. In HF, the sequence *hypertension* → *atrial fibrillation* → *diabetes* → HF was associated with the highest mortality risk [35]. These mortality-associated sequences highlight recurring patterns of metabolic and cardiovascular comorbidities, but further work is needed to determine whether sequence or individual conditions are key contributors to risk. Some comorbidity sequences are also associated with better outcomes, such as reduced mortality in sequences involving valvular diseases (*hypertension* → *mitral valve disorder* → PH) and chronic ischemic heart disease (*atherosclerotic heart disease* → *hypercholesterolemia* → HF). These findings should be interpreted cautiously, as the observed associations do not establish causality and may reflect differences in clinical management or underlying disease severity [36, 37]. The occurrence rates of hypertension and atrial fibrillation within PH diagnostic sequences are consistent with findings reported by Simonneau et al. [1], which identified left heart disease as the defining cause of Group 2 PH. Similarly, the observed HF diagnostic sequences are consistent with previous studies highlighting the importance of cardiovascular comorbidities in HF outcomes [38]. This analysis supports previous observations regarding the importance of cardiometabolic comorbidities while identifying distinct temporal diagnostic patterns between PH and HF. These findings provide epidemiological insights that may help generate hypotheses for future research but should not be interpreted as evidence for disease-specific mechanisms or direct changes to clinical management. While these temporal diagnostic patterns provide important insight into clinical pathways preceding PH recognition, they should be interpreted within the context of healthcare utilization patterns, referral pathways, diagnostic surveillance intensity, and administrative coding practices. Differences in the timing and ordering of recorded diagnoses may reflect variations in clinical contact frequency or investigation of timing rather than true biological acceleration of disease processes. Therefore, the identified comorbidity trajectories are best understood as epidemiological risk-flagging patterns within routinely collected healthcare data rather than causal disease pathways. In estimating mortality risk, proportional hazards assumptions may have been violated for several key covariates with potential time-varying effects. While sensitivity analyses and graphical diagnostics suggested that these deviations did not materially alter the overall interpretation of hazard ratio estimates, the presence of non-proportional hazards may influence the precision of effect estimates. Therefore, hazard ratios should be interpreted with caution, particularly when comparing risks across heterogeneous patient subgroups. Alternative modelling approaches, including time-varying coefficient models, were considered during the interpretation of the survival analyses. However, given that the primary objective of this study was to identify epidemiological patterns and temporal diagnostic sequences rather than to develop a fully specified causal survival model, these analyses were not formally pursued. Future studies using clinically characterized cohorts with richer longitudinal data may further explore time-dependent effects and dynamic risk relationships.

A key limitation of this study is the reliance on ICD-10 coded diagnoses without haemodynamic confirmation, which may lead to misclassification across PH subgroups. Given the high number of comorbidities with heart conditions in UK Biobank, ICD-defined PH is likely to predominantly represent Group 2 PH, with some contribution from Group 3 PH rather than the full spectrum of PH subtypes. Therefore, the observed comorbidity trajectories should be interpreted as aggregated epidemiological patterns within a heterogeneous PH population, rather than subtype-specific biological mechanisms, and should not be assumed to apply uniformly across all PH groups. Consequently, these findings do not provide direct insight into earlier diagnosis or intervention in other PH subgroups, particularly Groups 1 and 4, which have distinct pathophysiological mechanisms and clinical management pathways that are not captured in population datasets. Furthermore, participants in the

COPD and HF comparison cohorts were restricted to individuals without recorded PH-related ICD-10 diagnoses codes (I27.0, I27.2, or I27.9) throughout the available follow-up period. However, systematic echocardiographic screening was not available within UK Biobank and therefore the presence of undiagnosed or subclinical PH cannot be completely excluded. Any resulting misclassification would be expected to reduce differences between cohorts rather than artificially inflate observed associations, suggesting that the reported associations may represent conservative estimates of between-group differences. Furthermore, although propensity score matching was applied to construct comparable COPD and HF cohorts, some residual imbalance remained for certain covariates, including sex and mortality status. While most variables achieved acceptable balance, these residual differences may introduce bias in comparative analyses and should be considered when interpreting sequence-specific associations and mortality risk estimates. In addition, the temporal ordering of diagnoses derived from routinely collected administrative healthcare data may reflect variations in healthcare utilization, referral timing, diagnostic surveillance intensity, or coding practices rather than true biological disease progression. Therefore, observed diagnostic sequences should be interpreted as epidemiological associations within recorded care pathways rather than definitive mechanistic trajectories. When multiple diagnoses were recorded on the same calendar date, they were treated as concurrent events rather than being assigned a directional order. Consequently, these events should not be interpreted as reflecting the true chronological sequence of clinical events occurring during routine clinical care. To address potential misclassification inherent to ICD-10-based PH identification, we conducted sensitivity analyses using stricter PH definitions excluding unspecified PH (I27.9) and COPD diagnoses reduced the cohort by 10.2%, 34.8%, and 37.2% when applied jointly, but the temporal comorbidity trajectories and mortality associations remained consistent; detailed cohort derivation and analyses are provided in the Supplementary Material (Sect. 1.2.4). Treatment effects were not considered due to missing records, which may influence trajectories. Future work should integrate treatment, lifestyle, genetic, and biomarker data, and apply machine learning to improve disease classification and survival prediction. These findings highlight the potential of temporal comorbidity patterns as epidemiological indicators of PH risk. However, these findings should be interpreted as hypothesis-generating insights rather than direct tools for clinical decision-making. Further validation in clinically characterized cohorts with detailed haemodynamic and subtype-specific data, as well as in ethnically diverse populations and different healthcare systems, is required before these epidemiological observations can be generalized or translated into clinical applications.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1186/s12890-026-04610-7>.

Acknowledgements This research was carried out using the UK Biobank Resource under Application Number 85457. We gratefully acknowledge the researchers and collaborators who provided support and guidance during this study.

Code availability The full analysis codebase for data processing, comorbidity trajectory mining, and figure generation is permanently archived and available at: <https://doi.org/10.5281/zenodo.15747766>.

Authors' contributions All authors contributed to the study conception and design. Material preparation, data extraction, coding, analysis, and development of supplementary material were performed by Mohsin Masood. Hong Pan, and Naailah Bhamji supported coding and supplementary material. Genetic correlation analysis was carried out by Hong Pan. The first draft of the manuscript was written by Mohsin Masood. Juan Delgado-SanMartin, Siying Jane Ong, and Xiaohe Zhang contributed to manuscript refinement and editing. All authors read and approved the final manuscript. Dennis Wang, Roger Thompson, and Allan Lawrie provided critical input on the research concept and contributed to manuscript review and refinement as co-senior authors.

Funding This study is supported by funding from the Agency for Science, Technology and Research (A*STAR) Use-Inspired Basic Research Award, the Academy of Medical Sciences Professorship (APR7.1002), and the UK Research and Innovation (MR/Z505468/1).

Data availability The data used in this study are available from the <https://www.ukbiobank.ac.uk> and <https://doi.org/10.5281/zenodo.15747766> under approved access (Application Number: 85457). Access requires formal application and institut

ional approval.

Declarations

Ethics approval and consent to participate UK Biobank has obtained ethics approval from the Northwest Multi-center Research Ethics Committee (REC reference: 11/NW/0382). All participants provided informed consent at recruitment.

Competing interests The authors declare no conflict of interest/Competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Simonneau G, Montani D, Celermajer DS, Denton CP, Gatzoulis MA, Krowka M, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J*. 2019;53(1):1801913. <https://doi.org/10.1183/13993003.01913-2018>.
2. Armstrong I, Billings C, Kiely DG, Yorke J, Harries C, Clayton S, et al. The patient experience of pulmonary hypertension: a large cross-sectional study of UK patients. *BMC Pulm Med*. 2019. <https://doi.org/10.1186/s12890-019-0827-5>.
3. Rosenkranz S, Howard LS, Gomberg-Maitland M, Hoeper MM. Systemic consequences of pulmonary hypertension and right-sided heart failure. *Circulation*. 2020;141(8):678–93.
4. Lin Y, Pang L, Huang S, Shen J, Wu W, Tang F, et al. The prevalence and survival of pulmonary hypertension due to left heart failure: A retrospective analysis of a multicentre prospective cohort study. *Front Cardiovasc Med*. 2022;9(9):908215. <https://doi.org/10.3389/fcvm.2022.908215>.
5. Zangiabadi A, De Pasquale CG, Sajkov D. Pulmonary hypertension and right heart dysfunction in chronic lung disease. *BioMed Res Int*. 2014. <https://doi.org/10.1155/2014/739674>.
6. Hoeper MM, Bogaard HJ, Condliffe R, Frantz R, Khanna D, Kurzyna M, et al. Definitions and diagnosis of pulmonary hypertension. *J Am Coll Cardiol*. 2013;62(25 Supplement):D42–50.
7. Hewes JL, Lee JY, Fagan KA, Bauer NN. The changing face of pulmonary hypertension diagnosis: a historical perspective on the influence of diagnostics and biomarkers. *Pulm Circ*. 2020. <https://doi.org/10.1177/2045894019892801>.
8. Balakrishnan B, Owens B, Hayes R, Wen S. A systematic review of echocardiographic parameters for the screening of pulmonary hypertension: what are the odds? *Cureus*. 2022. <https://doi.org/10.7759/cureus.32185>.
9. Hurdman J, Condliffe R, Elliot CA, Davies C, Hill C, Wild JM, et al. ASPIRE registry: assessing the spectrum of pulmonary hypertension identified at a referral centre. *Eur Respir J*. 2011;39(4):945–55. <https://doi.org/10.1183/09031936.00078411>.
10. Maron BA, Hess E, Maddox TM, Opatowsky AR, Tedford RJ, Lahm T, et al. Association of borderline pulmonary hypertension with mortality and hospitalization in a large patient cohort: insights from the Veterans Affairs Clinical Assessment, Reporting, and Tracking Program. *Circulation*. 2016;133(13):1240–8. <https://doi.org/10.1161/CIRCULATIONAHA.115.020207>.
11. Karia N, Howard L, Johnson M, Kiely DG, Lordan J, McCabe C, et al. Predictors of outcomes in mild pulmonary hypertension according to 2022 ESC/ERS Guidelines: the EVIDENCE-PAH UK study. *Eur Heart J*. 2023;44(44):4678–91. Available from: <https://pubmed.ncbi.nlm.nih.gov/37619574/>.
12. Kolte D, Lakshmanan S, Jankowich MD, Brittain EL, Maron BA, Choudhary G. Mild pulmonary hypertension is associated with increased mortality: a systematic review and meta-analysis. *J Am Heart Assoc*. 2018. <https://doi.org/10.1161/JAHA.118.009729>.
13. Gal   N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Heart J*. 2016;37(1):67–119. <https://www.ncbi.nlm.nih.gov/pubmed/26320113>.
14. Armstrong I. Breathless, not voiceless: What it means to live with PH today | PHA. PHA. 2023. <https://www.phauk.org/research/pha-uk-led-research/breathless-not-voiceless-what-it-means-to-live-with-ph-today/>.

15. Link J, Glazer C, Torres F, Chin K. International classification of diseases coding changes lead to profound declines in reported idiopathic pulmonary arterial hypertension mortality and hospitalizations. *Chest*. 2011;139(3):497–504. <https://doi.org/10.1378/chest.10-0837>.
16. Mathai SC, Hemnes AR, Manaker S, Anguiano RH, Dean BB, Saundankar V, et al. Identifying Patients with Pulmonary Arterial Hypertension Using Administrative Claims Algorithms. *Ann Am Thorac Soc*. 2019;16(7):797–806. <https://doi.org/10.1513/AnnalsATS.201810-672CMEPMCID:PMC6600840>.
17. Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Heart J*. 2022;43(38). <https://doi.org/10.1093/eurheartj/ehac237>.
18. Gialamas I, Arvanitaki A, Rosenkranz S, Wort SJ, Göran Rådegran, Badagliacca R, et al. The impact of cardiovascular and lung comorbidities in patients with pulmonary arterial hypertension: A systematic review and meta-analysis. *J Heart Lung Transplant*. 2024;43(9):1383–94. <https://doi.org/10.1016/j.healun.2024.04.066>.
19. Reinders S, Didden EM, Ong R. Survival, morbidity, and quality of life in pulmonary arterial hypertension patients: a systematic review of outcomes reported by population-based observational studies. *Respir Res*. 2024;25(1). <https://doi.org/10.1186/s12931-024-02994-w>.
20. Ling Y, Johnson MK, Kiely DG, Condliffe R, Elliot CA, Gibbs JSR, et al. Changing demographics, epidemiology, and survival of incident pulmonary arterial hypertension: results from the pulmonary hypertension registry of the United Kingdom and Ireland. *Am J Respir Crit Care Med*. 2012;186(8):790–6.
21. Anderson JJ, Lau EM. Pulmonary hypertension definition, classification and epidemiology in Asia. *JACC Asia*. 2022. <https://doi.org/10.1016/j.jacasi.2022.04.008>.
22. Virsinskaite R, Karia N, Kotecha T, Schreiber BE, Coghlan JG, Knight DS. Pulmonary hypertension – the latest updates for physicians. *Clin Med*. 2023;23(5):449–54. <https://doi.org/10.7861/clinmed.2023-23.5.Cardio4>.
23. Hoeper MM, Pausch C, Ekkehard Grünig, Staehler G, Huscher D, Pittrow D, et al. Temporal trends in pulmonary arterial hypertension: results from the COMPERA registry. *The Eur Respir J*. 2021;59(6):2102024–4. <https://doi.org/10.1183/13993003.02024-2021>.
24. Santiago-Vacas E, Josep Lupón, Giovana Gavidia-Bovadilla, Gual-Capllonch F, Antonio M de, Domingo M, et al. Pulmonary hypertension and right ventricular dysfunction in heart failure: prognosis and 15-year prospective longitudinal trajectories in survivors. *Eur J Heart Failure*. 2020;22(7):1214–25. <https://doi.org/10.1002/ehf.1862>.
25. Pishgar M, Harford S, Theis J, Galanter W, Rodriguez-Fernández JM, Chaisson LH, et al. A process mining- deep learning approach to predict survival in a cohort of hospitalized COVID-19 patients. *BMC Med Inform Decis Mak*. 2022. <https://doi.org/10.1186/s12911-022-01934-2>.
26. Shafei I, Karnon J, Crotty M. Process mining and customer journey mapping in healthcare: enhancing patient-centred care in stroke rehabilitation. *Digit Health*. 2024. <https://doi.org/10.1177/20552076241249264>.
27. Beyel H, Verket M, Peeva V, Rennert C, Pegoraro M, Schütt K, et al. Process-Aware Analysis of Treatment Paths in Heart Failure Patients: A Case Study. *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies*. 2024;506–15. <https://arxiv.org/abs/2403.10544>.
28. de Boer TR, Arntzen RJ, Bekker R, Buurman BM, Willems HC, van der Mei RD. Process mining on national health care data for the discovery of patient journeys of older adults. *J Am Med Dir Assoc*. 2025;26(1):105333. <https://doi.org/10.1016/j.jamda.2024.105333>.
29. Launders N, Hayes JF, Price G, Osborn DP. Clustering of physical health multimorbidity in people with severe mental illness: An accumulated prevalence analysis of United Kingdom primary care data. Wodchis WP, editor. *PLOS Med*. 2022;19(4): e1003976. <https://doi.org/10.1371/journal.pmed.1003976>.
30. Lawrie A, Hamilton N, Wood S, Exposto F, Muzwidzwa R, Raiteri L, Beaudet A, Muller A, Sauter R, Pillai N, Kiely DG; ASPIRE Consortium. Association of risk assessment at diagnosis with healthcare resource utilization and health-related quality of life outcomes in pulmonary arterial hypertension. *Pulm Circ*. 2024;14(3): e12399. <https://doi.org/10.1002/pul.2.12399>.
31. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–5.
32. Choi J, Lee SR, Choi EK, Lee H, Han M, Ahn HJ, et al. Accumulated hypertension burden on atrial fibrillation risk in diabetes mellitus: a nationwide population study. *Cardiovasc Diabetol*. 2023;22:12. <https://doi.org/10.1186/s12933-023-01736-4>.
33. Hachulla E, Denton CP. Early intervention in pulmonary arterial hypertension associated with systemic sclerosis: an essential component of disease management. *Eur Respir Rev*. 2010;19(118):314–20. <https://doi.org/10.1183/09059180.00007810>.
34. Ussavarungsi K, Thomas CS, Burger CD. Prevalence of metabolic syndrome in patients with pulmonary hypertension. *Clin Respir J*. 2017;11(6):721–6. <https://doi.org/10.1111/crj.12406>.
35. Chatterjee NA, Chae CU, Kim E, Moorthy MV, Conen D, Sandhu RK, et al. Modifiable risk factors for incident heart failure in atrial fibrillation. *JACC Heart Fail*. 2017;5(8):552–60. <https://doi.org/10.1016/j.jchf.2017.04.004>.
36. Bermejo J, et al. Persistent Pulmonary Hypertension in Corrected Valvular Heart Disease: Hemodynamic Insights and Long-Term Survival. *J Am Heart Assoc*. 2021;10(2):e018816. <https://doi.org/10.1161/JAHA.120.01994>.

37. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 1994;344(8934):1383–9. [https://doi.org/10.1016/S0140-6736\(94\)90566-5](https://doi.org/10.1016/S0140-6736(94)90566-5).
38. Younis A, Tabaja C, Santangeli P, Nakagawa H, Sipko J, Madden R, Bouscher P, and Hussein AA. Outcomes of Atrial Fibrillation Ablation in Heart Failure Subtypes. *Circ Arrhythm Electrophysiol*. 2024; 17(9): 1 <https://doi.org/10.1161/CIRCEP.124.01292>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Mohsin Masood^{1,6,7} · Hong Pan^{2,3} · Juan Delgado-SanMartin¹ · Siying Jane Ong² · Xiaohe Zhang⁵ · Naailah Bhamji¹ · Roger Thompson⁴ · Allan Lawrie¹ · Dennis Wang^{1,2,3} 

✉ Mohsin Masood
mohsin.masood@kcl.ac.uk

✉ Dennis Wang
dennis.wang@imperial.ac.uk

Hong Pan
pan_hong@a-star.edu.sg

Juan Delgado-SanMartin
j.delgado-san-martin@imperial.ac.uk

Siying Jane Ong
siyingjaneong@gmail.com

Xiaohe Zhang
haleyxiaohe.zhang@yale.edu

Naailah Bhamji
naailah.bhamji21@imperial.ac.uk

Roger Thompson
r.thompson@sheffield.ac.uk

Allan Lawrie
a.lawrie@imperial.ac.uk

¹ National Heart & Lung Institute, Faculty of Medicine, Imperial College London, England, UK

² Institute for Human Development and Potential (IHDP), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore

³ Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore

⁴ School of Medicine and Population Health, Medical School, University of Sheffield, England, UK

⁵ Yale University, New Haven, CT, USA

⁶ Department of Population Health Sciences, Faculty of Life Sciences & Medicine, King's College London, London, England, UK

⁷ NIHR Nottingham Biomedical Research Centre, School of Medicine, Translational Medical Sciences, University of Nottingham, England, UK