



Article

The Assessment of the Association of Proton Pump Inhibitor Usage with Chronic Kidney Disease Progression through a Process Mining Approach

Kaile Chen ^{1,2,*}, Farhad Abtahi ^{1,2,3} , Hong Xu ⁴ , Carlos Fernandez-Llatas ^{1,5} , Juan-Jesus Carrero ⁶ and Fernando Seoane ^{1,3,7,8}

- ¹ Department of Clinical Science, Intervention and Technology, Karolinska Institutet, 17177 Stockholm, Sweden; farhad.abtahi@ki.se (F.A.); cflatas@itaca.upv.es (C.F.-L.); fernando.seoane@ki.se (F.S.)
- ² Department of Biomedical Engineering and Health Systems, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, 14157 Huddinge, Sweden
- ³ Department of Clinical Physiology, Karolinska University Hospital, 17176 Stockholm, Sweden
- ⁴ Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society (NVS), Karolinska Institutet, 17177 Stockholm, Sweden; hong.xu.2@ki.se
- ⁵ Institute of Information and Communication Technologies (SABIEN-ITACA), Universitat Politècnica de València, Camino de Vera S/N, 46022 Valencia, Spain
- ⁶ Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 17177 Stockholm, Sweden; juan.jesus.carrero@ki.se
- ⁷ Department of Medical Technology, Karolinska University Hospital, 17176 Stockholm, Sweden
- ⁸ Department of Textile Technology, University of Borås, 50190 Borås, Sweden
- * Correspondence: kaile@kth.se; Tel.: +46-762819715

Abstract: Previous studies have suggested an association between Proton Pump Inhibitors (PPIs) and the progression of chronic kidney disease (CKD). This study aims to assess the association between PPI use and CKD progression by analysing estimated glomerular filtration rate (eGFR) trajectories using a process mining approach. We conducted a retrospective cohort study from 1 January 2006 to 31 December 2011, utilising data from the Stockholm Creatinine Measurements (SCREAM). New users of PPIs and H2 blockers (H2Bs) with CKD (eGFR < 60) were identified using a new-user and active-comparator design. Process mining discovery is a technique that discovers patterns and sequences in events over time, making it suitable for studying longitudinal eGFR trajectories. We used this technique to construct eGFR trajectory models for both PPI and H2B users. Our analysis indicated that PPI users exhibited more complex and rapidly declining eGFR trajectories compared to H2B users, with a 75% increased risk (adjusted hazard ratio [HR] 1.75, 95% confidence interval [CI] 1.49 to 2.06) of transitioning from moderate eGFR stage (G3) to more severe stages (G4 or G5). These findings suggest that PPI use is associated with an increased risk of CKD progression, demonstrating the utility of process mining for longitudinal analysis in epidemiology, leading to an improved understanding of disease progression.

Keywords: eGFR trajectory; process mining; multistate model; proton pump inhibitors (PPIs); H2 blockers (H2Bs); chronic kidney disease (CKD); longitudinal data analysis



Citation: Chen, K.; Abtahi, F.; Xu, H.; Fernandez-Llatas, C.; Carrero, J.-J.; Seoane, F. The Assessment of the Association of Proton Pump Inhibitor Usage with Chronic Kidney Disease Progression through a Process Mining Approach. *Biomedicines* **2024**, *12*, 1362. <https://doi.org/10.3390/biomedicines12061362>

Academic Editor: Stephen Riordan

Received: 10 May 2024

Revised: 12 June 2024

Accepted: 13 June 2024

Published: 19 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chronic kidney disease (CKD) poses a significant global health challenge, affecting millions worldwide as it gradually impairs kidney function [1]. The estimated global prevalence of chronic kidney disease (CKD) was approximately 9% in 2017, representing a 29.3% increase since 1990 [2]. CKD also became the 12th leading cause of death in 2017, rising from 17th place in 1990 [2]. Glomerular filtration rate decline is a surrogate endpoint in evaluating the progression of kidney disease [3]. Empirical evidence underscores the important role of eGFR point estimate as a robust prognosticator for adverse CKD events,

notably end-stage renal disease (ESRD) [4]. Longitudinal analysis of eGFR trajectories and retrospective analysis of eGFR slopes are also important for understanding the dynamics of CKD progression and predicting patient outcomes [4–6]. Some studies have relied on the absolute change in eGFR as a surrogate [7,8] for kidney function progression, but they have failed to capture the rich insights that can be derived from analysing multiple eGFR measurements.

Proton pump inhibitors (PPIs) are widely prescribed for acid-related gastrointestinal disorders [9]. PPI usage greatly reduces stomach acid levels for longer periods [10–12] compared to H₂ blockers (H₂Bs). However, many observational investigations have indicated PPI usage with kidney adverse events, such as an increased risk of CKD, CKD progression and acute kidney injury (AKI) [8,13,14]. While most observational studies have shown associations between PPIs and CKD, some studies have reported no significant association [15] or only a weak link [16]. The underlying mechanism of PPI usage with adverse effects on CKD remains the subject of an ongoing inquiry. Plausible explanations include the induction of acute interstitial nephritis (AIN) [17,18] and hypomagnesemia [19] by the usage of PPIs. Delayed treatment or incomplete recovery from AIN can lead to acute kidney disease, potentially increasing the risk of developing CKD. H₂B [20] is another class of medication used to treat gastrointestinal conditions but has rarely been reported to be associated with nephrotoxicity.

Process discovery, a process mining technique for discovering and analysing sequences of events over time, has gained traction in healthcare for studying clinical pathways and disease trajectories [21]. For instance, one study successfully predicted in-hospital mortality using a process mining-based methodology [22]. Additionally, researchers in [23] modelled treatment pathways for patients with cancer using process mining techniques. In a different application, process mining facilitated the development of dynamic risk models for chronic conditions [24], resulting in simplified process maps that enhance the understanding and measurement of disease progression. Furthermore, the study [25] employed process mining to uncover disease trajectories in periodontitis patients, enabling researchers to monitor the progression of related multi-morbidities. However, its application in analysing the progression of chronic diseases within epidemiological studies remains scarce. Pharmacoepidemiology research often applies cohort designs and drug new-user designs to assess the association between exposures and outcomes [26]. These designs take the importance of time sequence and derive more robust evidence for causal inference compared to cross-sectional or case-control studies. Process mining excels at visualising such time-ordered events, making it a valuable tool for exploring disease trajectories. This capability enables us to use process mining to extract longitudinal eGFR data and compare the effects of PPIs and H₂Bs on the risk of CKD progression.

This study contributes to the fields of nephrology and epidemiology by applying a novel approach based on process mining. Firstly, it employs a process mining technique for the discovery, visualisation and comprehension of CKD patients' trajectories over time. Additionally, the study seeks to provide evidence regarding the association between the use of PPIs and the accelerated progression of CKD through an extensive longitudinal analysis.

In this retrospective cohort study, we aim to evaluate the usefulness of the process mining approach in epidemiological research by extracting eGFR-mediated longitudinal trajectories to assess the association between exposure to PPIs and CKD progression in comparison to exposure to H₂Bs.

2. Materials and Methods

Section 2.1 details the study design and resources utilised in this research. Following this, Section 2.2 outlines the population inclusion criteria adopted for the study. Section 2.3 describes the primary methodology of process mining discovery. Subsequently, Section 2.4 provides a comprehensive overview of the statistical methods employed in this study.

2.1. Study Design and Resources

This study employed a retrospective drug new-user cohort design utilising data from the Stockholm Creatinine Measurements (SCREAM) cohort, a population-based cohort in Sweden [27]. The study period was from 1 January 2006 to 31 December 2011. A drug new-user cohort was constructed, comprising PPI new users as the analysis group and H2B new users as the active-comparator group. The use of PPIs and H2Bs was assumed to be continuous from initiation to the study endpoints. Identification of PPIs, H2Bs and other concomitant medications was accomplished through Anatomical Therapeutic Classification (ATC) codes (Supplementary Table S1). Comorbidities were identified using ICD-10 codes (Supplementary Table S1). This study was carried out following the rules of the Declaration of Helsinki. The Swedish National Board of Welfare and the Stockholm Regional Ethics Review Board approved this study (Ethical Approval Code 2017/793-31).

2.2. Study Population

Based on the study aims and design, the study population was selected on five specified inclusion and exclusion criteria. (1) We included PPI or H2B new users. The new user was defined as the same as a prior classic epidemiology study by Klatte [8] (i.e., an individual who has received their first PPI or H2B dispensation between 1 January 2007 and 31 December 2010 was considered a new user; the data in 2006 were used to ascertain no prescription of the medication before, and the data in 2011 were used to make sure at least 1 year follow-up). (2) At least 1 time serum creatinine (Scr) test less than 6 months before the index date. (3) New users of PPIs or H2Bs with chronic kidney disease (CKD), defined by a baseline eGFR less than 60 mL/min/1.73 m². The baseline eGFR value was the latest measurement before or on the date of medication prescription. (4) People ≥ 18 years old. Exclusion: Individuals with a baseline eGFR of less than 15 mL/min/1.73 m² were excluded, as they were in the end stage of CKD.

The eGFR was calculated based on the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) 2009 creatinine equation [28] in adults aged 18 and older, which uses serum creatinine, age, ethnicity and gender. Although ethnicity data were unavailable due to legal restrictions, the misclassification of eGFR is expected to be minimal because most residents in the Stockholm region are of Caucasian origin. All creatinine measurements were conducted using either the enzymatic method or the corrected Jaffe method (alkaline picrate reaction) across three laboratory companies in the Stockholm region: Aleris, Unilabs and Karolinska [27]. Both inter- and intra-laboratory variations are considered minimal, as these laboratories undergo frequent quality audits and harmonisation checks by the national organisation EQUALIS (www.equalis.se accessed on 7 May 2024) [29].

2.3. Process Mining Discovery

The core methodology was based on the interactive process mining process [30]. The process mining discovery technique from the bupaR package [31] in R was used to derive process models and indicators [32]. Event logs were created for both PPI and H2B users.

The event was built on the eGFR stages, comprising five stages that align with KDIGO categories [33,34]: G1 (normal or high, eGFR ≥ 90 mL/min/1.73 m²), G2 (mildly decreased, eGFR 60–89 mL/min/1.73 m²), G3A (mildly to moderately decreased, eGFR 45–59 mL/min/1.73 m²), G3B (moderately to severely decreased, eGFR 30–44 mL/min/1.73 m²), CKD4 (severely decreased, eGFR 15–29 mL/min/1.73 m²) and CKD5 (kidney failure, eGFR < 15 mL/min/1.73 m²). The follow-up period started from the date of PPI or H2B dispensation until either emigration, death or the conclusion of the study on 31 December 2011.

2.4. Statistical Analysis

Baseline demographic and clinical characteristics of the patients were summarised using frequencies and percentages for categorical variables, means with standard deviations (SDs) for normally distributed continuous variables, and medians with interquartile ranges (IQR) for continuous variables that are not normally distributed.

The process indicator described the eGFR trajectories, including the frequency and transition time. Transition time is the duration between two events, also referred to as sojourn time in this analysis. It can be understood as the interval between two measurements or the observed time it takes for eGFR to shift from one stage to another.

Age, gender, comorbidities and concomitant medications were recorded on the index date. Comorbidities included gastrointestinal diseases, cardiovascular and cerebrovascular diseases, diabetes mellitus and chronic obstructive pulmonary disease, with detailed ICD-10 codes provided in Supplementary Table S1. Concomitant medications, such as NSAIDs, statins and antithrombotics, have detailed ATC codes available in Supplementary Table S1.

We employed a multistate model [35] to analyse the transition probabilities between different eGFR stages identified through the process indicator. This model incorporated PPI/H2B use, age, gender, comorbidities and concomitant medications as covariates to account for their potential influence on the transitions between eGFR stages. After adjusting for these covariates, the multistate model produced adjusted hazard ratios with corresponding 95% confidence intervals (CIs) to estimate the effect of PPI/H2B use on eGFR decline. The specific model structure incorporating these multiple covariates is presented in Figure 1. All statistical analyses and data visualisations were conducted using R 4.3.1.

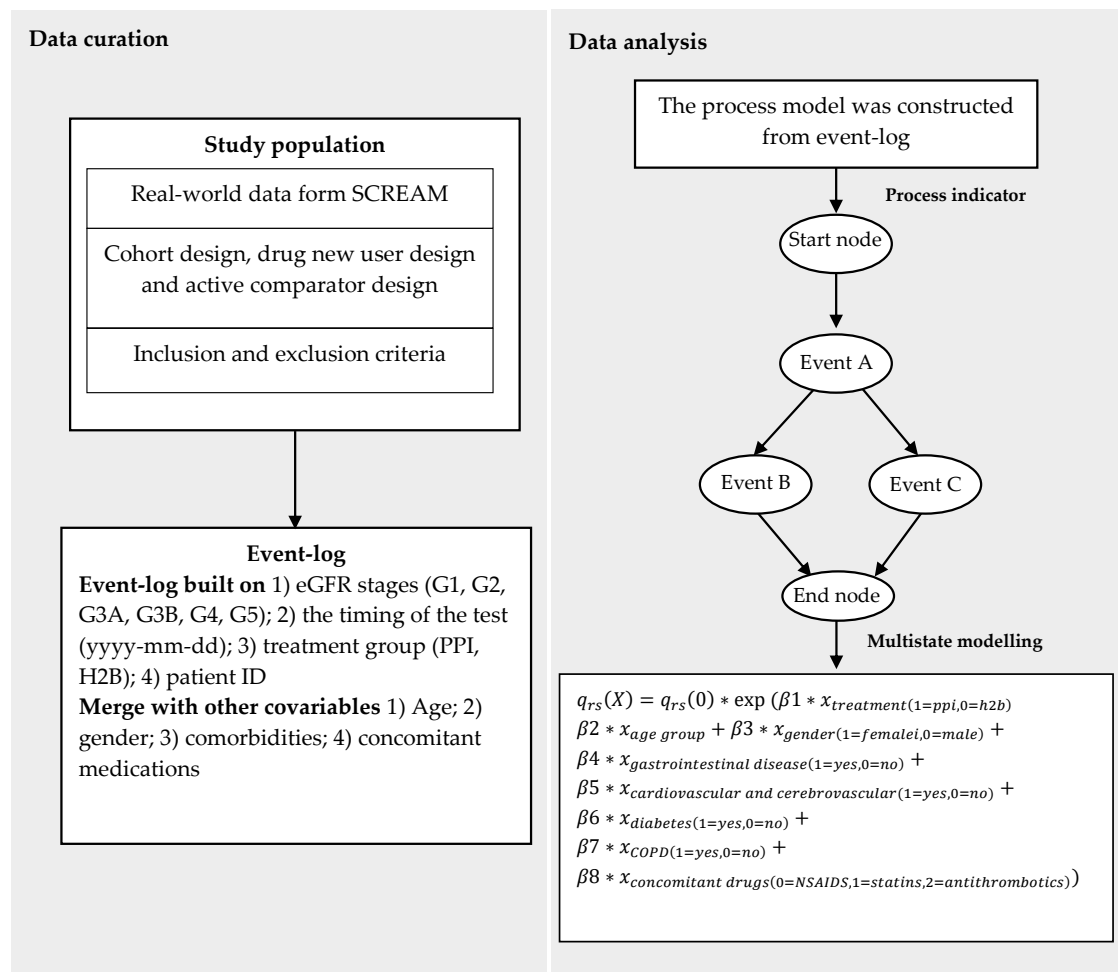


Figure 1. Diagram illustrating data curation and analysis in this study; data curation includes study design used for collection and event-log build; data analysis describes the process mining approach and the algorithm utilised in the multistate model. In the algorithm, $q_{rs}(X)$ represents the transition intensity or the instantaneous rate from event r to event s . The term $q_{rs}(0)$ denotes the baseline transition intensity, while $\exp(\beta_1)$ corresponds to the hazard ratio.

3. Results

A total of 12,043 newly initiated users of PPIs and H2Bs with CKD were included in this investigation, comprising 11,486 PPI users and 557 H2B users (Figure 2). The baseline characteristics showed a median age of 81 years for the PPI cohort and 78 years for the H2B cohort (Table 1), indicating more elderly individuals in the former. Moreover, both PPI and H2B cohorts exhibited a higher proportion of females than males (57% and 65%, respectively). Regarding baseline eGFR, the PPI cohort displayed a lower median compared to the H2B cohort. Additionally, the PPI group has a greater prevalence of comorbidities compared to the H2B group.

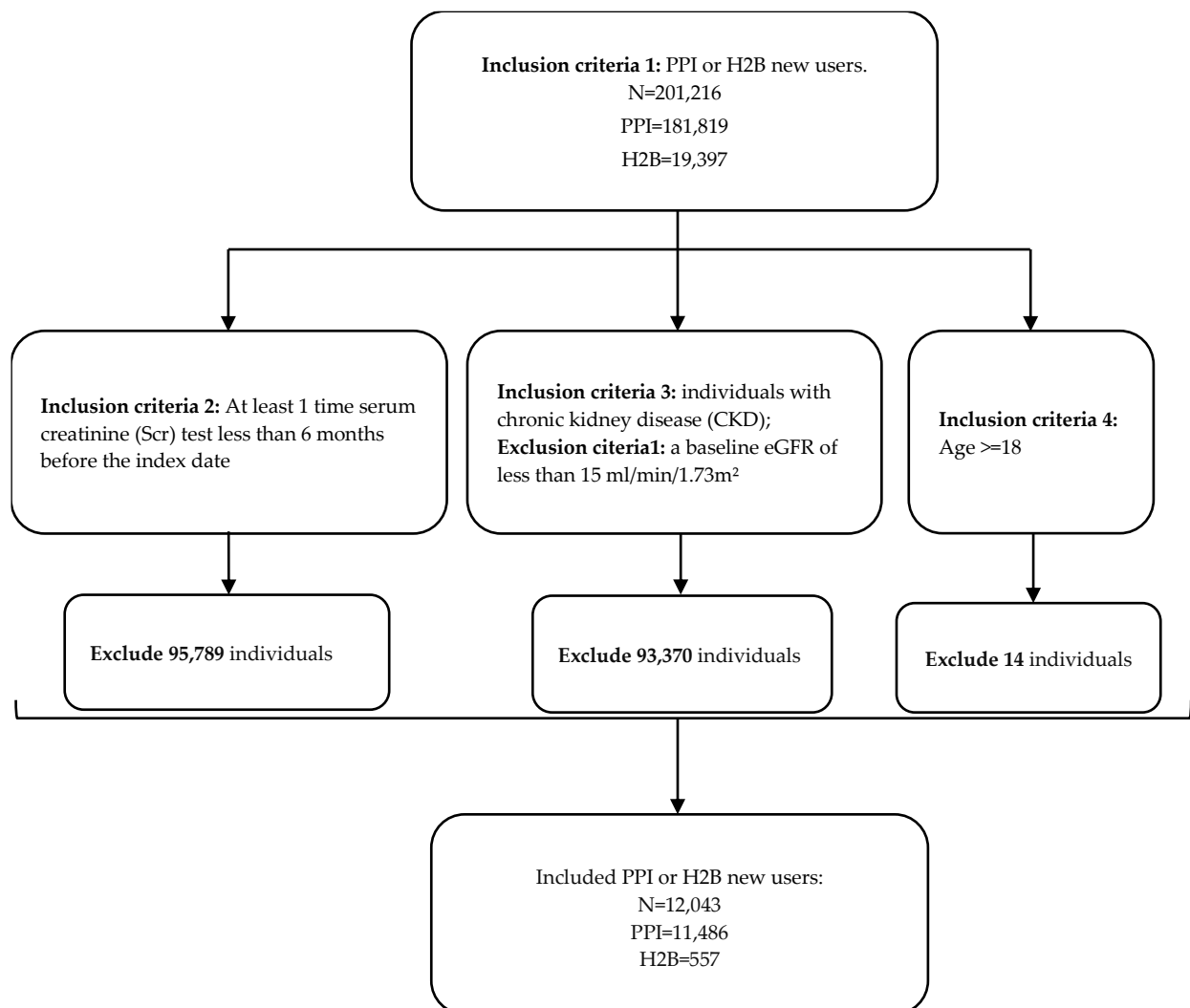


Figure 2. Flowchart of the selection process of study participants among new users of PPI and H2B, in accordance with the study's inclusion and exclusion criteria.

Table 1. Characteristics of the baseline between new users of PPIs and H2Bs with CKD in Stockholm during 2006–2011.

	PPI N = 11,486	H2B N = 557
Age, median (IQR)	81 (73, 86)	78 (70, 85)
Age group, n%		
18~45	160 (1.4)	10 (1.8)
46~64	1109 (9.7)	74 (13.3)
65~80	4262 (37.1)	251 (45.1)
≥81	5954 (51.8)	222 (39.9)
Gender		
Female	6603 (57)	361 (65)
Male	4883 (43)	196 (35)
Comorbidities, n%		
Gastrointestinal diseases		
Gastroesophageal reflux disease	547 (4.8)	19 (3.4)
Upper gastrointestinal tract bleeding	488 (4.2)	10 (1.8)
Ulcer disease	1097 (9.6)	18 (3.2)
<i>H. Pylori</i> infection		
Cardiovascular and cerebrovascular diseases		
Myocardial infarction	2164 (18.8)	95 (17.1)
Cerebrovascular disease	2527 (22)	104 (18.7)
Peripheral vascular disease	1290 (11.2)	58 (10.4)
Congestive heart failure	1976 (17.2)	80 (14.4)
Hypertension	10,005 (87.1)	469 (84.2)
Diabetes mellitus	2743 (23.9)	133 (23.9)
Chronic obstructive pulmonary disease (COPD)	2042 (17.8)	94 (16.9)
Concomitant medication, n%		
NSAIDs, aspirin	4941 (43)	257 (46.1)
Statins	4179 (36.4)	217 (39)
Antithrombotics	7553 (65.8)	321 (57.6)
Baseline eGFR, median (IQR)	47.8 (38.2, 54.6)	50.7 (42.4, 55.9)
G3A (45–59 mL/min/1.73 m ²)	6681 (58.2)	382 (68.6)
G3B (30–44 mL/min/1.73 m ²)	3506 (30.5)	139 (24.9)
G4 (15–29 mL/min/1.73 m ²)	1299 (11.3)	36 (6.5)

3.1. Process Indicator

The process indicators (Figures 3 and 4), which describe the transition of eGFR stages of the PPI and H2B groups over the follow-up period, were structured. These indicators are sequentially ordered from the node “start” to “end”; directional arrows denote the progression from antecedent to consequent events, with nodes representing eGFR stages. Node shading correlates with event frequency, displaying both relative and absolute frequencies. Additionally, arrow annotations depict relative frequencies and median transition times.

Figures 3 and 4 illustrate the transition process among five eGFR stages for PPI and H2B groups, respectively. Both groups initially presented at stage 3 (G3) or stage 4 (G4), aligning with our inclusion criteria of CKD patients (eGFR < 60 and ≥15 mL/min/1.73 m²). Three arrows emanating from the “start” node represent the baseline eGFR distribution within each group. For instance, in the PPI group, 58.2% started at G3A, 30.5% at G3B and 11.3% at G4. The H2B group had a higher proportion starting at G3A (68.6%) and lower proportions starting at G3B (25%) and G4 (6.5%) compared to the PPI group. The stage connected to the “end” represents the last observed event. Transitions among eGFR stages can be visually compared between the two groups in Figures 3 and 4. Notably, the PPI group exhibited a higher proportion of patients progressing to more severe stages (G4 or G5, e.g., G3A → G4 is 7% in the PPI group and 5% in the H2B group) and fewer instances of improvement, suggesting a potentially negative impact of PPI usage on CKD progression. Conversely, individuals in the H2B group relatively experienced more favourable

transitions, with a higher proportion transitioning back to better eGFR stages (indicating improved kidney function such as G1 and G2, e.g., G3A → G2 is 41% in the PPI group and 47% in H2B group) compared to the PPI group.

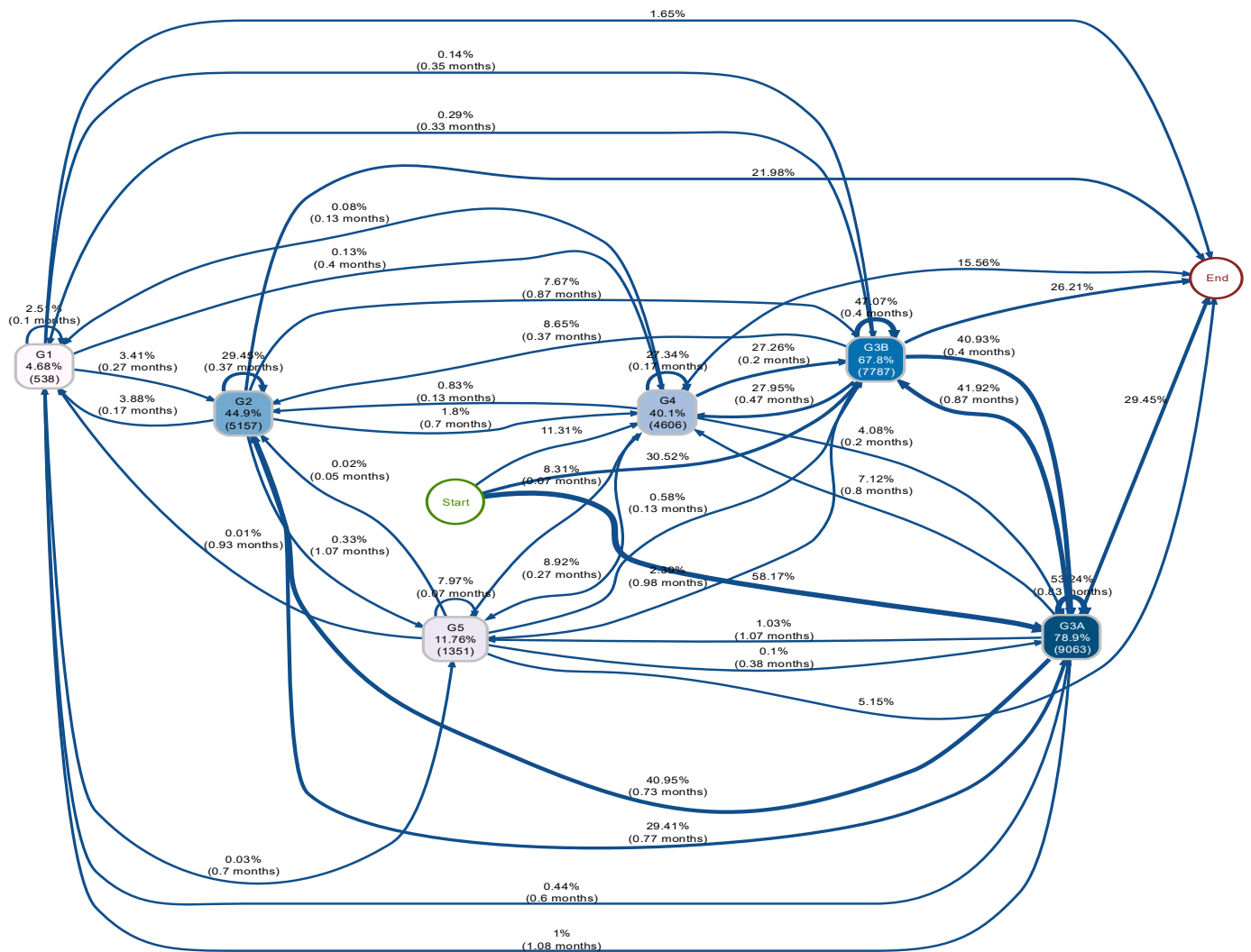


Figure 3. Initial process indicator of eGFR stage transitions over time in the PPI group. This figure illustrates the transition of eGFR stages in the PPI group over the follow-up period. Nodes represent event relative and absolute frequencies, with darker shades indicating higher event frequencies. Transition relative frequencies and median times are annotated on the arrows. Arrows signify time-ordered sequences of events, with varying thicknesses reflecting the frequency of transitions.

Due to the unstructured nature and complexity of Figures 3 and 4 and to facilitate the comparison of eGFR trajectories between PPI and H2B users, eGFR stages were integrated to simplify the process indicators, as illustrated in Figure 5. Specifically, G1 and G2 were merged into a singular event denoted as “G1 or G2”, G3A and G3B were combined into “G3”, and G4 and G5 were consolidated into “G4 or G5”.

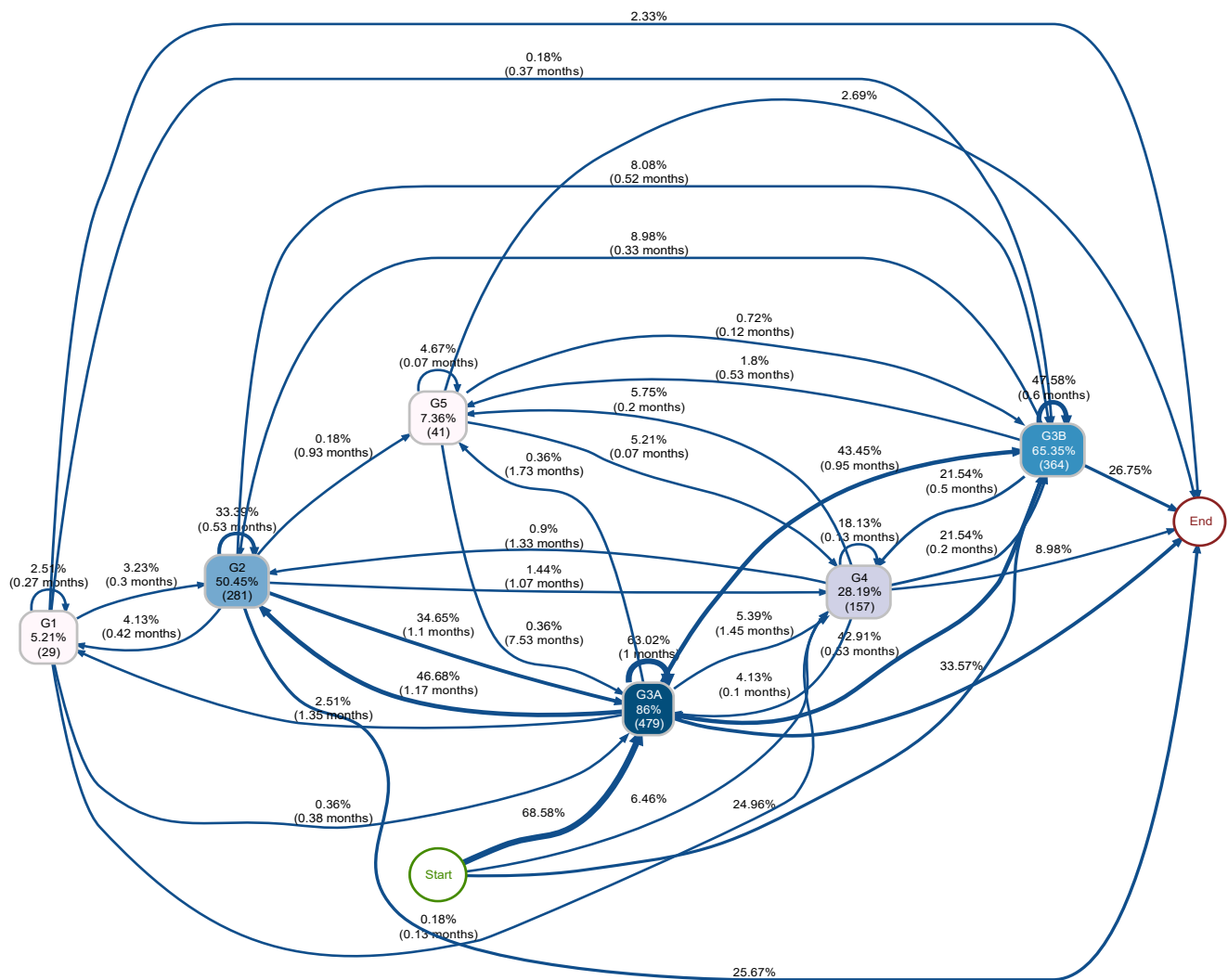


Figure 4. Initial process indicator of eGFR stage transitions over time in the H2B group. This figure illustrates the transition of eGFR stages in the H2B group over the follow-up period. Nodes represent event relative and absolute frequencies, with darker shades indicating higher event frequencies. Transition relative frequencies and median times are annotated on the arrows. Arrows signify time-ordered sequences of events, with varying thicknesses reflecting the frequency of transitions.

Figure 5 illustrates the distribution of transitions between eGFR stages between PPI and H2B groups that exhibited notable disparities. Notably, 33% of PPI users experienced a reduced eGFR from “G3” to “G4 or G5”, compared to 25% of H2B users who underwent this transition within the same median duration. After declining to “G4 or G5”, relatively more H2B users transitioned back to “G3” than PPI users. Similarly, 2.19% of PPI users worsened from “G1 or G2” to “G4 or G5”, while only 1.8% of H2B users experienced this decline. Furthermore, in the improvement transition from “G3” to “G1 or G2”, 45% of PPI users underwent this change, whereas 51% of H2B users experienced the transition. Moreover, in the H2B group, a larger percentage of individuals remained at “G1 or G2” (28%) and “G3” (60.3%) compared to PPI users, where the corresponding figures were slightly lower, at 23.6% and 55.7%, respectively. Conversely, in the PPI group, a higher proportion of individuals stayed at “G4 or G5” (20.7%) compared to the H2B group (11.7%).

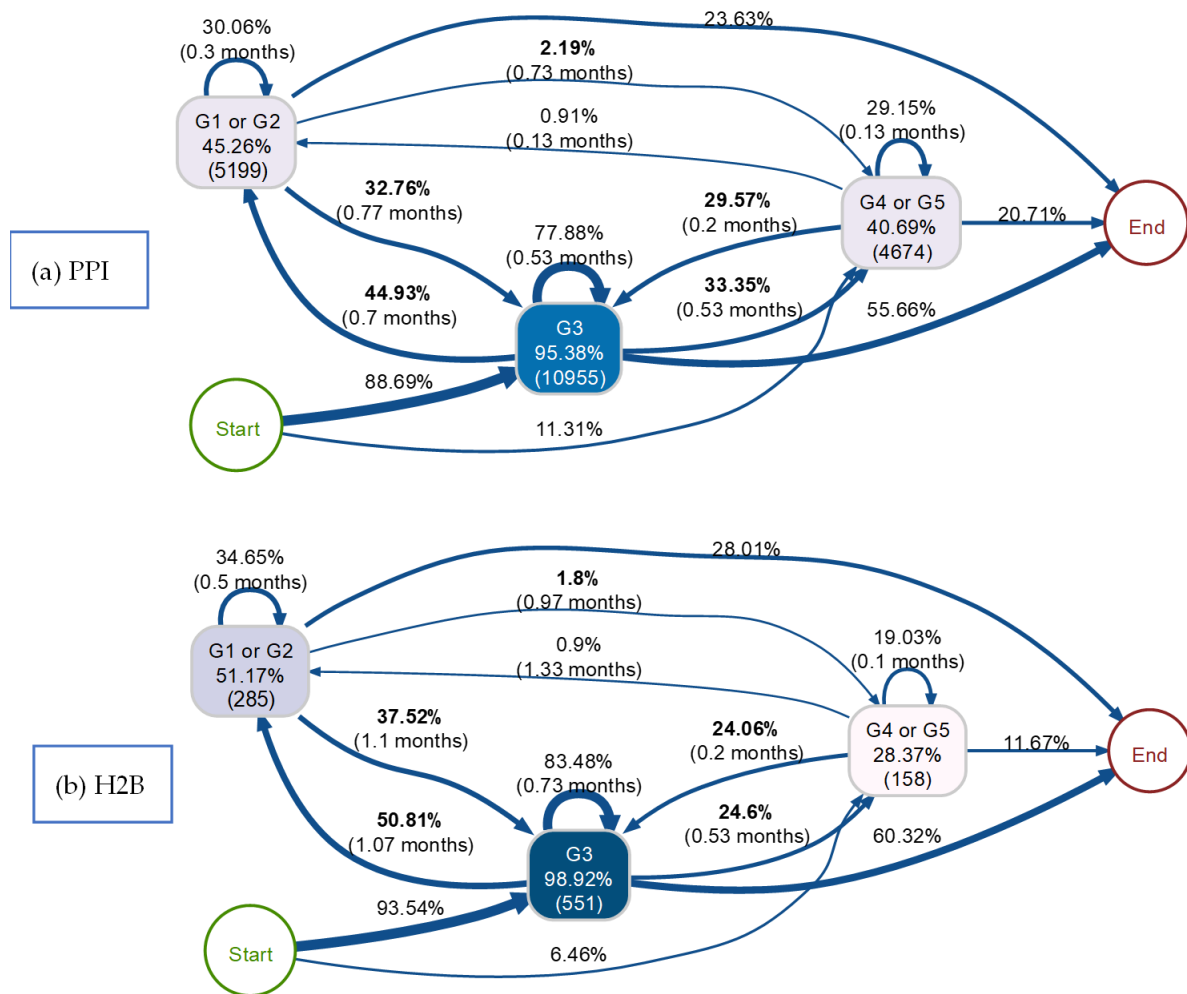


Figure 5. Process indicators of integrated eGFR stage transition over time between the PPI and H2B groups. (a) represents the PPI group; (b) represents the H2B group. Nodes display event relative and absolute frequencies, with darker shades indicating higher event frequencies. Transition relative frequencies and median times are depicted on the arrows. Arrows signify time-ordered sequences of events, with varying thicknesses reflecting the frequency of transitions. Transitions with proportions showing potential significant differences between the PPI and H2B groups are highlighted in bold.

3.2. The Patterns of eGFR Trajectories

Each trajectory signifies the changes in individual eGFR stages throughout follow-up. The length of each trajectory corresponds to the number of eGFR stages observed during follow-up. Notably, individuals utilising PPI exhibit longer eGFR trajectory lengths compared to those in the H2B group. Additionally, PPI users have a shorter median follow-up duration in contrast to H2B users (Table 2).

Compared to the H2B group, the PPI group shows a 75% higher hazard (HR: 1.75, 95% CI: 1.49, 1.52) for the progression from eGFR stage “G3” to stages “G4 or G5” (G3 → G4/5, as shown in Table 2). There was no statistically significant distinction in the transition from “G3” to “G1/2”, “G1/2” to “G3” and “G4/5” to “G3” between patients being administered PPIs versus those receiving H2Bs.

Subgroup analysis to compare the effects of PPIs versus H2Bs on eGFR trajectories across different subgroups is presented in Supplementary Table S2.

Table 2. The comparison of eGFR trajectory patterns between PPI and H2B groups.

	PPI	H2B
Trajectory length, median (IQR)	11 (5, 21)	10 (5, 21)
Follow-up time(month), median (IQR)	23.3 (11.1, 37.7)	33.3 (17.7, 46.3)
eGFR stages transition:		
G3 → G1/2		
Number, %	5161 (44.9)	283 (50.8)
Median time (IQR), months	0.70 (0.13, 2.97)	1.07 (0.13, 4.08)
^a Hazard ratio, 95% CI	0.95 (0.78, 1.16)	Reference
G1/2 → G3		
Number, %	3763 (32.76)	209 (37.52)
Median time (IQR), months	0.77 (0.17, 2.70)	1.10 (0.20, 2.80)
^a Hazard ratio, 95% CI	0.92 (0.75, 1.12)	Reference
G3 → G4/5		
Number, %	3831 (33.35)	137 (24.60)
Median time (IQR), months	0.53 (0.13, 2.0)	0.53 (0.10, 2.13)
^a Hazard ratio, 95% CI	1.75 (1.49, 2.06)	Reference
G4/5 → G3		
Number, %	3396 (29.57)	134 (24.06)
Median time (IQR), months	0.20 (0.07, 0.83)	0.20 (0.07, 0.89)
^a Hazard ratio, 95% CI	1.13 (0.97, 1.31)	Reference

^a. The hazard ratio and 95% CI were derived from multistate modelling after adjusting age, sex, comorbidities, concomitant medication and baseline eGFR.

3.3. Integrating Process Mining with Epidemiological Design versus Classic Epidemiology Methods

A detailed comparison with the study by Klatte et al. is presented in Table 3. Our study's main finding aligns with Klatte's research, providing evidence that the process mining approach can yield results comparable to those of classical epidemiological methods. We applied process mining techniques to achieve a more comprehensive understanding of disease progression. Unlike the previous study that focused on the association between exposure (PPI) and outcome (eGFR decline of 30%), we constructed eGFR trajectories by capturing all observed eGFR stages. This approach allows for a different perspective analysis by comparing the properties of these trajectories between individuals exposed to PPIs and those exposed to H2Bs.

Table 3. Summary of current study vs. a prior study that employed classic epidemiological methods.

	Klatte et al. [8]	Present Study
Study type	Classic epidemiology study	Process mining integrated with epidemiology design
Study design	Cohort; active-comparator design; new-user design	Cohort; active-comparator design; new-user design
Data source	SCREAM	SCREAM
Outcome	A doubling of serum creatinine or >30% decline in eGFR	Multiple eGFR stages
Measure	eGFR absolute changes; once a 30% decline or more in eGFR is observed, the follow-up will be discontinued	eGFR stages; collecting all observed eGFR stages until the end of the study
Analysis methods	Descriptive statistics; Cox proportional hazard regression model	Descriptive statistics; process indicators; multistate model
Study period	2006–2011	2006–2011
Study population	PPI or H2B new users with baseline eGFR ≥ 15 mL/min/1.73 m ²	PPI or H2B new users with CKD (15 mL/min/1.73 m ² $\leq$ eGFR < 60 mL/min/1.73 m ²)
Key results	PPI users had an increased risk for eGFR decline of 30% or more compared with H2B users	eGFR trajectory distribution differently between PPI users and H2B users; PPI users had an increased risk from eGFR 'G3' transit to 'G4 or G5'

4. Discussion

We applied a process mining approach to extract the eGFR trajectory patterns, followed by comparing the patterns between PPI and H2B groups. The results showed that the pattern of eGFR trajectory is more complex in PPI users than in H2B users. In addition, the PPI group has a significantly higher risk of undergoing from eGFR stage G3 to the eGFR stage G4 or G5, representing a faster decline in kidney function than the H2B group.

4.1. Process Mining in Chronic Disease Research

Process mining techniques have been applied in healthcare since 2001 [35], providing three main techniques: process discovery, conformance checking and enhancement [36]. These techniques are predominantly utilised to analyse data from electronic health records pertaining to clinical or treatment pathways to identify bottlenecks within hospitals and enhance overall efficiency [37]. While some previous studies have attempted to use process mining to trace disease trajectories in diabetes, sepsis and general diseases [38–41], our study initiated the attempt to synthesise process mining with epidemiological cohort design to assess the association between PPI usage and kidney function decline by comparing the eGFR trajectory patterns among CKD patients.

4.2. Investigating the Association between PPI Use and CKD Progression through eGFR Trajectories

We used eGFR stages as the measurement events to investigate the patterns of eGFR in PPI users compared to H2B users. Specifically, the eGFR stages transit to higher eGFR stages signify the progression of CKD. Previous research primarily defines CKD progression by quantifying absolute eGFR changes, operating under the assumption of its linear progression [5]. However, the decline in kidney function in the population might have different patterns due to various factors that do not follow a steady linear decline. Some studies used the mixed model of linear quadratic, cubic and probabilistic clustering to produce trajectories to cope with the property of non-linearity [6] or the Bayesian smoothing method to define trajectories [42]. Applying process mining, in this case, provides dynamic changes with multiple measurements in eGFR over time since we observed the manifest difference in eGFR trajectories, which also represent the complexity of an individual. Notably, individuals in the PPI group exhibited more frequent serum creatinine measurements within a considerably shorter follow-up duration, indicative of heightened healthcare utilisation and worse baseline conditions, which could be a higher burden of comorbidities and older age, as corroborated by baseline characteristics. In the Stockholm region, PPI is generally the first-line treatment for gastrointestinal conditions like gastroesophageal reflux disease and ulcer disease [43].

Following adjustment for potential confounders such as baseline conditions of comorbidities, concomitant medications, baseline kidney function (baseline eGFR), sex and age, PPI users demonstrated a significantly higher rate of transition from G3 to G4 or G5 stages compared to H2B users, suggesting that PPI usage is associated with increased risk of CKD progression. These findings align with previous research indicating an association between PPI use and an increased risk of CKD progression [8,14,18]. Furthermore, we construed transitions to lower eGFR stages (i.e., larger eGFR values) as improvements or fluctuations, notwithstanding prevailing beliefs regarding the irreversibility of kidney function decline [44]. Contrary to the conventional notion of steady GFR progression, many CKD patients exhibit non-linear GFR trajectories or protracted periods of non-progression [42]. Our findings align with existing evidence regarding the non-linear eGFR changes, revealing transitions to improved kidney function (e.g., G3 to G1/G2 or G4/G5 to G3). Notably, a higher proportion of the H2B group experienced such improvements compared to the PPI group, although this difference did not reach statistical significance. Similarly, intermittent transitions from G1 or G2 to severe G3 showed no significant disparity between PPI and H2B cohorts. These transitions may signify eGFR fluctuations from G3 to G1/2, subsequently reverting to G3 or higher stages G4/G5. Other studies [5,45] have also posited

that short-term increases in eGFR may stem from muscle mass loss or recovery from acute kidney injury.

4.3. The Process Mining Approach in This Study and the Comparison with Classic Epidemiological Research and Artificial Intelligence

A comparison was conducted between the current study and the prior investigation to ascertain the utility and reproducibility of the process mining-assisted approach. This study also employed the cohort design [46], drug new-user design and active-comparator design [47] as Klatte's study [8] in order to reduce selection bias. This study goes beyond replicating previous findings concerning the association between PPI usage and the increased risk of CKD progression. Combining process mining with these epidemiological designs introduces a new approach to epidemiological studies. This integration allows for a more comprehensive analysis and understanding of the data. To the best of our knowledge, this is the first application of interactive process mining to analyse longitudinal renal function trajectories in real-world epidemiological data.

Process indicators were derived from the serum creatine test records and the timing of these tests, resulting in divergent patient patterns, including testing frequency variations indicative of baseline conditions. Consequently, these process indicators depicted the disease trajectories of patients from a distant point, with greater complexity in the process map potentially signalling severe conditions in reality. Compared to the property of classic epidemiology studies, which directly test the hypothesis of exposure and outcome, incorporating process mining into epidemiological settings may be less efficient than classical methods. However, process mining can provide valuable insights into the data landscape and illuminate associations from a broader perspective (in a process indicator/map). This underscores the usefulness of integrating process mining into epidemiological studies.

We additionally constructed a multistate model based on the process indicator. The process indicator reflects real-world eGFR trajectories, offering a descriptive patient trajectory. Therefore, the initial relationship represented on process indicators might potentially be biased by confounding factors. Thus, multistate modelling allows for comparing the transition intensities between PPI and H2B groups with the adjustment of potential confounders. In healthcare research utilising process mining, it is common practice to integrate other statistical methods to test whether differences between two groups are attributable to factors beyond potential confounding variables or random errors [38].

Process mining, when combined with multistate models, presents a valuable tool in healthcare for patient journey modelling, disease progression tracking, and risk stratification. This hybrid approach, incorporating rule-based logic and statistical modelling elements, offers a high degree of interpretability, allowing for clear insights into patient transitions between distinct health states over time. However, unlike machine learning and deep learning methodologies, which excel at handling complex, high-dimensional data and uncovering intricate patterns, multistate models are inherently limited to scenarios with well-defined states and transitions. While deep learning techniques can process extensive amounts of unstructured data to predict outcomes, they often do so at the expense of transparency and interpretability. Consequently, multistate models are particularly suited for applications where understanding the underlying processes and maintaining model transparency are critical. In scenarios where complex data and predictive accuracy are paramount, machine learning and deep learning can be combined with multistate models to leverage the strengths of both approaches. For instance, machine learning can be used to enhance the predictive power and adaptability of multistate models.

4.4. Limitations, Challenges and Future Work

A limitation of this study is the high median age of the population. This is because the SCREAM dataset can only capture individuals who have accessed healthcare services, and younger individuals, who are generally healthier, are less likely to utilise healthcare resources. Consequently, our findings are representative of the real world, which includes

people who take PPIs/H2Bs with CKD and have accessed healthcare utilisation within the Stockholm region. Additionally, analysing complex medical histories with multiple events challenges deriving meaningful associations. Domain knowledge is crucial to simplify this process and extract valuable insights.

Future work should focus on standardising the application of process mining in epidemiological research. There is potential for using process mining to investigate disease interactions, such as mediation and collider effects. Additionally, expanding the application of process mining to other diseases presents a promising avenue for future research. Fully leveraging process mining's strengths in visualisation can automatically generate associations between diseases, diagnoses and medications, enabling comprehensive data-driven hypotheses.

5. Conclusions

This study evaluated the feasibility of applying process mining techniques to analyse and provide useful insights into real-world data. Process mining discovered the disease progression transitions in real-world data, which were used to construct a multistate model. By quantifying transition intensity using the multistate model, we identified a significantly higher risk of CKD progression in the PPI group compared to the H2B group. This is supported by a greater likelihood of eGFR trajectories advancing to higher stages in the PPI group. These findings confirm the prior evidence that PPI usage increases the risk for CKD progression and indicate that the process mining approach is beneficial for analysing longitudinal epidemiological data.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines12061362/s1>, Table S1: The ICD-10 codes of comorbidities and ATC codes of PPI/H2B; Table S2: Adjusted hazard ratios and 95% confidence intervals (CI) for the effect of PPI versus H2B on eGFR trajectories in different subgroups.

Author Contributions: Conceptualization, J.-J.C. and F.S.; Formal analysis, K.C.; Methodology, K.C., C.F.-L., J.-J.C. and F.S.; Resources, J.-J.C.; Supervision, F.A., H.X., C.F.-L., J.-J.C. and F.S.; Validation, F.A.; Writing—original draft, K.C.; Writing—review and editing, F.A., H.X., C.F.-L., J.-J.C. and F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research has received funding from EIT Health, ID 220649, the innovation community on Health of the European Institute of Innovation and Technology (EIT), a body of the European Union, under Horizon 2020, the EU Framework Programme for Research and Innovation. KC has received funding from the China Scholarship Council (CSC, NO.202108610086). HX has received funding from the Center for Innovative Medicine Foundation (CIMED, FoUI-963369), Åke Wibergs stiftelse (#M22-0170), U&L Angeby stiftelse (#2022-02882) and the Swedish research council grant (#2022-01428). JJC has received funding from the Swedish Research Council, the Swedish Heart and Lung Foundation and ALF Medicin. The APC was funded by the KTH Royal Institute of Technology.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki. Approval for the study was obtained from the Swedish National Board of Welfare and the Stockholm Regional Ethics Review Board (Ethical Approval Code 2017/793-31), who deemed that informed consent was not required and provided de-identified data.

Informed Consent Statement: Patient consent was waived by the Swedish National Board of Welfare and the Stockholm Regional Ethics Review Board, who deemed that informed consent was not required and provided de-identified data.

Data Availability Statement: The SCREAM contains sensitive personal data that cannot be publicly shared due to GDPR regulations. We welcome collaboration project proposals that adhere to GDPR and national and institutional regulations concerning data sharing and access. For inquiries, please contact juan.jesus.carrero@ki.se.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Kovesdy, C.P. Epidemiology of chronic kidney disease: An update 2022. *Kidney Int. Suppl.* **2022**, *12*, 7–11. [\[CrossRef\]](#)
2. Cockwell, P.; Fisher, L.-A. The global burden of chronic kidney disease. *Lancet* **2020**, *395*, 662–664. [\[CrossRef\]](#)
3. Badve, S.V.; Palmer, S.C.; Hawley, C.M.; Pascoe, E.M.; Strippoli, G.F.; Johnson, D.W. Glomerular filtration rate decline as a surrogate end point in kidney disease progression trials. *Nephrol. Dial. Transpl.* **2016**, *31*, 1425–1436. [\[CrossRef\]](#)
4. Kovesdy, C.P.; Coresh, J.; Ballew, S.H.; Woodward, M.; Levin, A.; Naimark, D.M.; Nally, J.; Rothenbacher, D.; Stengel, B.; Iseki, K.; et al. Past Decline Versus Current eGFR and Subsequent ESRD Risk. *J. Am. Soc. Nephrol.* **2016**, *27*, 2447–2455. [\[CrossRef\]](#)
5. Sumida, K.; Kovesdy, C.P. Disease Trajectories Before ESRD: Implications for Clinical Management. *Semin. Nephrol.* **2017**, *37*, 132–143. [\[CrossRef\]](#)
6. Santos, J.; Oliveira, P.; Severo, M.; Lobato, L.; Cabrita, A.; Fonseca, I. Different kidney function trajectory patterns before dialysis in elderly patients: Clinical implications and outcomes. *Ren. Fail.* **2021**, *43*, 1049–1059. [\[CrossRef\]](#)
7. Coresh, J.; Turin, T.C.; Matsushita, K.; Sang, Y.; Ballew, S.H.; Appel, L.J.; Arima, H.; Chadban, S.J.; Cirillo, M.; Djurdjev, O.; et al. Decline in Estimated Glomerular Filtration Rate and Subsequent Risk of End-Stage Renal Disease and Mortality. *JAMA* **2014**, *311*, 2518. [\[CrossRef\]](#)
8. Klatte, D.C.F.; Gasparini, A.; Xu, H.; de Deco, P.; Trevisan, M.; Johansson, A.L.V.; Wettermark, B.; Arnlov, J.; Janmaat, C.J.; Lindholm, B.; et al. Association Between Proton Pump Inhibitor Use and Risk of Progression of Chronic Kidney Disease. *Gastroenterology* **2017**, *153*, 702–710. [\[CrossRef\]](#)
9. Scarpignato, C.; Gatta, L.; Zullo, A.; Blandizzi, C.; Scarpignato, C.; Blandizzi, C.; Gatta, L.; Zullo, A.; Kohn, A.; Leandro, G.; et al. Effective and safe proton pump inhibitor therapy in acid-related diseases—A position paper addressing benefits and potential harms of acid suppression. *BMC Med.* **2016**, *14*, 179. [\[CrossRef\]](#)
10. Cederberg, C.; Röhss, K.; Lundborg, P.; Olbe, L. Effect of once daily intravenous and oral omeprazole on 24-hour intragastric acidity in healthy subjects. *Scand. J. Gastroenterol.* **1993**, *28*, 179–184. [\[CrossRef\]](#)
11. Wilder-Smith, C.; Halter, F.; Ernst, T.; Gennoni, M.; Zeyen, B.; Varga, L.; Roehmel, J.J.; Merki, H.S. Loss of acid suppression during dosing with H₂-receptor antagonists. *Aliment. Pharmacol. Ther.* **1990**, *4* (Suppl. S1), 15–27.
12. Maclaren, R.; Reynolds, P.M.; Allen, R.R. Histamine-2 Receptor Antagonists vs Proton Pump Inhibitors on Gastrointestinal Tract Hemorrhage and Infectious Complications in the Intensive Care Unit. *JAMA Intern. Med.* **2014**, *174*, 564. [\[CrossRef\]](#)
13. Han, C.T.; Islam, M.M.; Poly, T.N.; Lu, Y.C.; Lin, M.C. A Meta-Analysis of Proton Pump Inhibitor Use and the Risk of Acute Kidney Injury: Geographical Differences and Associated Factors. *J. Clin. Med.* **2023**, *12*, 2467. [\[CrossRef\]](#)
14. Dos Santos, A.S.; de Menezes, S.T.; Silva, I.R.; Oliveira, W.N.; Pereira, M.L.; Mill, J.G.; Barreto, S.M.; Figueiredo, R.C. Kidney function decline associated with proton pump inhibitors: Results from the ELSA-Brasil cohort. *BMC Nephrol.* **2023**, *24*, 285. [\[CrossRef\]](#)
15. Kweon, T.; Kim, Y.; Lee, K.J.; Seo, W.W.; Seo, S.I.; Shin, W.G.; Shin, D.H. Proton pump inhibitors and chronic kidney disease risk: A comparative study with histamine-2 receptor antagonists. *Sci. Rep.* **2023**, *13*, 21169. [\[CrossRef\]](#)
16. Cheema, E. Investigating the association of proton pump inhibitors with chronic kidney disease and its impact on clinical practice and future research: A review. *J. Pharm. Policy Pract.* **2019**, *12*, 6. [\[CrossRef\]](#)
17. Morschel, C.F.; Mafra, D.; Eduardo, J.C.C. The relationship between proton pump inhibitors and renal disease. *J. Bras. Nefrol.* **2018**, *40*, 301–306. [\[CrossRef\]](#)
18. Pauksakon, P.; Fogo, A.B. Do Proton-Pump Inhibitors Cause CKD and Progression of CKD?: Commentary. *Kidney360* **2022**, *3*, 1141–1143. [\[CrossRef\]](#)
19. Florentin, M.; Elisaf, M.S. Proton pump inhibitor-induced hypomagnesemia: A new challenge. *World J. Nephrol.* **2012**, *1*, 151–154. [\[CrossRef\]](#)
20. Nugent, C.C.; Falkson, S.R.; Terrell, J.M. *H₂ Blockers*; StatPearls Publishing: Treasure Island, FL, USA, 2018.
21. Chen, K.; Abtahi, F.; Carrero, J.-J.; Fernandez-Llatas, C.; Seoane, F. Process mining and data mining applications in the domain of chronic diseases: A systematic review. *Artif. Intell. Med.* **2023**, *144*, 102645. [\[CrossRef\]](#)
22. Theis, J.; Galanter, W.L.; Boyd, A.D.; Darabi, H. Improving the In-Hospital Mortality Prediction of Diabetes ICU Patients Using a Process Mining/Deep Learning Architecture. *IEEE J. Biomed. Health Inf.* **2022**, *26*, 388–399. [\[CrossRef\]](#)
23. Baker, K.; Dunwoodie, E.; Jones, R.G.; Newsham, A.; Johnson, O.; Price, C.P.; Wolstenholme, J.; Leal, J.; McGinley, P.; Twelves, C.; et al. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. *Int. J. Med. Inf.* **2017**, *103*, 32–41. [\[CrossRef\]](#)
24. Valero-Ramon, Z.; Fernandez-Llatas, C.; Valdivieso, B.; Traver, V. Dynamic Models Supporting Personalised Chronic Disease Management through Healthcare Sensors with Interactive Process Mining. *Sensors* **2020**, *20*, 5330. [\[CrossRef\]](#)
25. Larvin, H.; Kang, J.; Aggarwal, V.R.; Pavitt, S.; Wu, J. Multimorbid disease trajectories for people with periodontitis. *J. Clin. Periodontol.* **2021**, *48*, 1587–1596. [\[CrossRef\]](#)
26. Luijken, K.; Spekrijse, J.J.; van Smeden, M.; Gardarsdottir, H.; Groenwold, R.H.H. New-user and prevalent-user designs and the definition of study time origin in pharmacoepidemiology: A review of reporting practices. *Pharmacoepidemiol. Drug Saf.* **2021**, *30*, 960–974. [\[CrossRef\]](#)
27. Carrero, J.J.; Elinder, C.G. The Stockholm Creatinine Measurements (SCREAM) project: Fostering improvements in chronic kidney disease care. *J. Intern. Med.* **2022**, *291*, 254–268. [\[CrossRef\]](#)

28. Levey, A.S.; Stevens, L.A.; Schmid, C.H.; Zhang, Y.L.; Castro, A.F.; Feldman, H.I.; Kusek, J.W.; Eggers, P.; Van Lente, F.; Greene, T.; et al. A New Equation to Estimate Glomerular Filtration Rate. *Ann. Intern. Med.* **2009**, *150*, 604. [\[CrossRef\]](#)
29. Runesson, B.; Gasparini, A.; Qureshi, A.R.; Norin, O.; Evans, M.; Barany, P.; Wettermark, B.; Elinder, C.G.; Carrero, J.J. The Stockholm CREAtinine Measurements (SCREAM) project: Protocol overview and regional representativeness. *Clin. Kidney J.* **2016**, *9*, 119–127. [\[CrossRef\]](#)
30. Fernandez-Llatas, C. Applying Interactive Process Mining Paradigm in Healthcare Domain. In *Interactive Process Mining in Healthcare*; Fernandez-Llatas, C., Ed.; Springer International Publishing: Cham, Switzerland, 2021; pp. 103–117.
31. Janssenswillen, G.; Depaire, B.; Swennen, M.; Jans, M.; Vanhoof, K. bupaR: Enabling reproducible business process analysis. *Knowl.-Based Syst.* **2019**, *163*, 927–930. [\[CrossRef\]](#)
32. Fernandez-Llatas, C. Interactive Process Mining in Practice: Interactive Process Indicators. In *Interactive Process Mining in Healthcare*; Fernandez-Llatas, C., Ed.; Springer International Publishing: Cham, Switzerland, 2021; pp. 141–162.
33. Stevens, P.E.; Ahmed, S.B.; Carrero, J.J.; Foster, B.; Francis, A.; Hall, R.K.; Levin, A. KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int* **2024**, *105*, S117–S314. [\[CrossRef\]](#)
34. Bikbov, B.; Purcell, C.A.; Levey, A.S.; Smith, M.; Abdoli, A.; Abebe, M.; Adebayo, O.M.; Afarideh, M.; Agarwal, S.K.; Agudelo-Botero, M.; et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2020**, *395*, 709–733. [\[CrossRef\]](#)
35. Maruster, L.; van der Aalst, W.; Weijters, T.; van den Bosch, A.; Daelemans, W. Automated discovery of workflow models from hospital data. In Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2001), Amsterdam, The Netherlands, 25–26 October 2001; pp. 183–190.
36. Fernandez-Llatas, C. Interactive Process Mining in Healthcare: An Introduction. In *Interactive Process Mining in Healthcare*; Fernandez-Llatas, C., Ed.; Springer International Publishing: Cham, Switzerland, 2021; pp. 1–9.
37. Munoz-Gama, J.; Martin, N.; Fernandez-Llatas, C.; Johnson, O.A.; Sepúlveda, M.; Helm, E.; Galvez-Yanjari, V.; Rojas, E.; Martinez-Millana, A.; Aloini, D.; et al. Process mining for healthcare: Characteristics and challenges. *J. Biomed. Inform.* **2022**, *127*, 103994. [\[CrossRef\]](#)
38. Kusuma, G.P.; Kurniati, A.P.; Rojas, E.; McInerney, C.D.; Gale, C.P.; Johnson, O.A. Process Mining of Disease Trajectories: A Literature Review. *Stud. Health Technol. Inf.* **2021**, *281*, 457–461. [\[CrossRef\]](#)
39. Kusuma, G.; Kurniati, A.; McInerney, C.D.; Hall, M.; Gale, C.P.; Johnson, O. Process Mining of Disease Trajectories in MIMIC-III: A Case Study. In *Lecture Notes in Business Information Processing*; Springer International Publishing: Cham, Switzerland, 2021; pp. 305–316.
40. De Oliveira, H.; Prodel, M.; Lamarsalle, L.; Inada-Kim, M.; Ajayi, K.; Wilkins, J.; Sekelj, S.; Beecroft, S.; Snow, S.; Slater, R.; et al. “Bow-tie” optimal pathway discovery analysis of sepsis hospital admissions using the Hospital Episode Statistics database in England. *JAMIA Open* **2020**, *3*, 439–448. [\[CrossRef\]](#)
41. De Toledo, P.; Joppien, C.; Sesmero, M.P.; Drews, P. Mining Disease Courses across Organizations: A Methodology Based on Process Mining of Diagnosis Events Datasets. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019.
42. Li, L.; Astor, B.C.; Lewis, J.; Hu, B.; Appel, L.J.; Lipkowitz, M.S.; Toto, R.D.; Wang, X.; Wright, J.T., Jr.; Greene, T.H. Longitudinal progression trajectory of GFR among patients with CKD. *Am. J. Kidney Dis.* **2012**, *59*, 504–512. [\[CrossRef\]](#)
43. Klokistan.se. Available online: <https://klokistan.se/terapiomrade/matmaltningssorgan.html> (accessed on 7 May 2024).
44. Yan, M.T.; Chao, C.T.; Lin, S.H. Chronic Kidney Disease: Strategies to Retard Progression. *Int. J. Mol. Sci.* **2021**, *22*, 84. [\[CrossRef\]](#)
45. Rosansky, S.J. Renal Function Trajectory Is More Important than Chronic Kidney Disease Stage for Managing Patients with Chronic Kidney Disease. *Am. J. Nephrol.* **2012**, *36*, 1–10. [\[CrossRef\]](#)
46. Setia, M.S. Methodology Series Module 1: Cohort Studies. *Indian J. Dermatol.* **2016**, *61*, 21–25. [\[CrossRef\]](#)
47. Yoshida, K.; Solomon, D.H.; Kim, S.C. Active-comparator design and new-user design in observational studies. *Nat. Rev. Rheumatol.* **2015**, *11*, 437–441. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.