

## ARTICLE OPEN



# Optimizing drug selection from a prescription trajectory of one patient

Alejandro Aguayo-Orozco <sup>1,2</sup>, Amalie Dahl Haue <sup>1,3</sup>, Isabella Friis Jørgensen <sup>1</sup>, David Westergaard <sup>1,2,4</sup>, Pope Lloyd Moseley<sup>1</sup>, Laust Hvas Mortensen <sup>2,5</sup> and Søren Brunak <sup>1</sup>✉

It is unknown how sequential drug patterns convey information on a patient's health status and treatment guidelines rarely account for this. Drug-agnostic longitudinal analyses of prescription trajectories in a population-wide setting are needed. In this cohort study, we used 24 years of data (1.1 billion prescriptions) from the Danish prescription registry to model the risk of sequentially redeeming a drug after another. Drug pairs were used to build multistep longitudinal prescription trajectories. These were subsequently used to stratify patients and calculate survival hazard ratios between the stratified groups. The similarity between prescription histories was used to determine individuals' best treatment option. Over the course of 122 million person-years of observation, we identified 9 million common prescription trajectories and demonstrated their predictive power using hypertension as a case. Among patients treated with agents acting on the renin-angiotensin system we identified four groups: patients prescribed angiotensin converting enzyme (ACE) inhibitor without change, angiotensin receptor blockers (ARBs) without change, ACE with posterior change to ARB, and ARB posteriorly changed to ACE. In an adjusted time-to-event analysis, individuals treated with ACE compared to those treated with ARB had lower survival probability (hazard ratio, 0.73 [95% CI, 0.64–0.82];  $P < 1 \times 10^{-16}$ ). Replication in UK Biobank data showed the same trends. Prescription trajectories can provide novel insights into how individuals' drug use change over time, identify suboptimal or futile prescriptions and suggest initial treatments different from first line therapies. Observations of this kind may also be important when updating treatment guidelines.

npj Digital Medicine (2021)4:150; <https://doi.org/10.1038/s41746-021-00522-4>

## INTRODUCTION

High cost of prescription drugs has become a problem in most Western societies, where the expenditure of prescription is ever increasing and consumes a growing share of the total healthcare outlay<sup>1,2</sup>. A better understanding of variation in treatment response and tolerance would increase quality of life and improve treatment guidelines while also reducing costs of maintaining a healthy ageing population partly via prescribed treatment<sup>3</sup>. During the past 40 years, handwritten prescriptions have largely been replaced by electronic versions, and in most Western countries e-health infrastructures have been designed to manage them<sup>4</sup>. This transformation has facilitated secondary use of healthcare data, including studies of prescription databases. Since 1995, all prescriptions redeemed at a pharmacy in Denmark have been registered in the Danish National Prescription Registry (DNPR), now holding data on 7.2 million people<sup>5</sup>. DNPR is among the oldest and largest prescription registries in the world. Examples of other prescription data resources include the Finnish database on drug utilization that was established in 1994<sup>6</sup>, and the UK Clinical Practice Research Datalink (CPRD), which also holds prescription data<sup>7</sup>.

Research using these resources has been performed mostly in a hypothesis driven manner and highly targeted, either in terms of diseases (e.g., myocardial infarction)<sup>8</sup> or drugs (e.g., antibiotics and gastric acid inhibitors)<sup>9–11</sup>. In a large longitudinal study, prescriptions from a cohort of over 9 million people were used to conclude that 26% of all prescribed drugs influence cancer risk<sup>12</sup>. Other studies have addressed opioid use and the rapid increase in

opioid prescription over the last decades<sup>13</sup>. Prescription studies have also been designed to predict prescriptions based on previous, redeemed prescriptions. Data-mining prescriptions of anti-diabetic drugs have, for example, been used to predict future anti-diabetic prescription<sup>14</sup>. These studies involve less than one million people and focus on a subset of all drugs only<sup>15,16</sup>.

Numerous guidelines derived from randomized controlled trials (RCTs) and meta-analyses—for specific diseases and complications—have been published, e.g., hypertension, diabetes, and cardiovascular complications<sup>17</sup>. However, such guidelines rarely consider the full spectrum of diseases that are often being treated simultaneously; and typically, guidelines present an overview of state-of-the-art literature and evidence levels rather than answers to conflicting clinical considerations. This traditional limitation may, in turn, result in inadequate or unnecessary treatment for some patients<sup>18</sup>. Yet, the digital transformation, where entire health registries are being studied in a data-driven manner, can potentially compliment results from classical RCTs and meta-analyses in understanding the variation in treatment responses and tolerance<sup>19</sup>.

Previous longitudinal analyses of sequential drug prescriptions have been performed in a targeted fashion based on selected diseases or groups of drugs. However, it is unknown how sequential drugs patterns convey information about an individual's health status and treatment guidelines rarely account for this. This study is the first to perform a drug-agnostic longitudinal analysis of prescription trajectories in a population-wide setting.

<sup>1</sup>Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark. <sup>2</sup>Statistics Denmark, 2100 Copenhagen, Denmark. <sup>3</sup>The Heart Center, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. <sup>4</sup>Department of Obstetrics and Gynaecology, Amager Hvidovre Hospital, Copenhagen University Hospital, Hvidovre, Denmark. <sup>5</sup>Section of Epidemiology, Department of Public Health, University of Copenhagen, 1014 Copenhagen, Denmark. ✉email: [lhm@dst.dk](mailto:lhm@dst.dk); [soren.brunak@cpr.ku.dk](mailto:soren.brunak@cpr.ku.dk)

The objective of this cohort study was to identify and characterize the most prevalent prescription trajectories in the Danish population, hence providing a model to help fill gaps between disease registry data from hospitals, indications from general practitioners and dispensed prescriptions.

Mapping population-wide prescription trajectories can provide novel insight into how individuals' drug use change over time. This method can—at the level of individuals—identify potentially suboptimal prescription sequences, as well as futile prescriptions.

## RESULTS

### Condensing prescription data into trajectories

We performed a comprehensive cohort analysis of DNPR covering prescription patterns across all drugs dispensed at all pharmacies in a nation-wide, universal healthcare setting. We condensed the longitudinal prescription redemption data on each individual into prescription trajectories. Then we characterized over 9 million prescription trajectories of varying lengths among 7,255,919 individuals, 49.3% male and 50.7% female, (Supplementary Tables 1 and 2). For 48% of the population (3.5 million people) their presence in the registry start in 1995 and end in 2019 (Supplementary Figs. 1–3).

In Fig. 1a females display an increase in the proportion of anatomical group G (genitourinary system and sex hormones) during the teenage period. Furthermore, for both sexes in the age interval 5–15 years there is an increase in the proportion of anatomical groups H and P (systemic hormonal preparations, excluding sex hormones and insulins; and antiparasitic products, insecticides and repellents), possibly related to growth and the infections of childhood. At older ages, the proportions of drugs related to anatomical group B and C (blood forming organs and cardiovascular system) increase, representing medication for the most common diseases of the age group such as hypertension and atherosclerosis (Fig. 1a). Figure 1b shows a spike at young age for both sexes, primarily related to anatomical groups J, R, and S (anti-infectives for systemic use, respiratory system, and sensory organs). There is also a continuous increase from young adults with a peak value at the age of 65–70 years for males and 75 years in females, reflecting that the prevalence of polypharmacy increases with age and that life expectancy is higher for females than for males (Supplementary Fig. 4).

### Temporal drug prescription association analysis

From the full data set, 229,803 sequential drug pairs ( $P1 \rightarrow P2$ ) were identified, where all chemical subgroups of each anatomical group were represented. We then excluded all prescription pairs that were redeemed by less than 100 patients, those with  $Q$ -value above  $10^{-03}$  and those with  $RR < 1$  (Supplementary Fig. 4a).

Figure 2a (Supplementary Table 3) presents the most frequent, statistically significant directional prescription pairs (classified by chemical ATC class). All pairs were prescribed to more than 100,000 people and interestingly all had an  $RR > 2$ . The pair with shortest average intermediate time is formed by two drugs that belong to the nervous system group  $N02AA \rightarrow N02AB$ , natural opium alkaloids and phenylpiperidine derivatives, respectively ( $RR = 2.76$ , 95% confidence interval (95% CI): 2.74–4.13,  $P = 2.70 \times 10^{-8}$ ), with an average intermediate time of 2 years. Beta blocking agents redemption increases the risk for subsequent redemption of digitalis glycoside ( $C07AB \rightarrow C01AA$ ;  $RR = 3.55$ , 95% CI: 2.30–5.47,  $P = 9.07 \times 10^{-9}$ ). Both chemical subgroups are used as antiarrhythmic drugs. The pair with the highest  $RR$  (~4) is  $G03AA \rightarrow G01AF$ , progestogens and estrogens, which include hormonal contraceptives, to imidazole derivatives, respectively ( $RR = 3.98$ , 95% CI: 2.42–6.55,  $P = 5.27 \times 10^{-8}$ ), while the pair with the highest number of patients is  $G03AA \rightarrow J02AC$ , triazole derivatives ( $RR = 2.82$ , 95% CI: 1.76–4.53,  $P = 1.66 \times 10^{-5}$ ). Both

imidazole and triazole derivatives are used to treat fungal infections<sup>20</sup>.

The directional prescription pair of drugs from different anatomical groups having the lowest  $RR$  is  $D07BB \rightarrow AB10K$ , corticosteroids, moderately potent, combination with antiseptics, and sodium-glucose co-transporter 2 (SGLT2) inhibitors, respectively ( $RR = 0.002$ , 95% CI: 0.001–0.003,  $P = 4.44 \times 10^{-142}$ ). An increased risk of redemption of opioids ( $N02$ ) is observed after prescription of diabetic treatment, particularly blood glucose lowering drugs ( $A10B$ ). Diabetic patients often use opioids to treat neuropathic pain derived from diabetes<sup>21</sup>. The use of direct factor Xa inhibitors ( $B01AF$ ), which is a group of antithrombotic agents associated with a lower risk of subsequent redemption of other anti-dementia drugs ( $N06DX$ ) ( $RR = 0.41$ , 95% CI: 0.37–0.46,  $P = 1.38 \times 10^{-62}$ ) (Fig. 2b). In Fig. 2c pairs that belong to the same anatomical group show that there is a high risk of taking vitamin B12 ( $B03BA$ ) after redemption of iron ( $B03AE$ ) ( $RR = 9.87$ , 95% CI: 6.78–14.37,  $P = 6.95 \times 10^{-33}$ ). Patients redeeming GLP-1 prescriptions ( $A10BH$ ) are at lower risk of redeeming SGLT-2 ( $A10BJ$ ) prescriptions subsequently ( $RR = 0.45$ , 95% CI: 0.37–0.55,  $P = 1.34 \times 10^{-14}$ ). Both types of medication are effective antihyperglycemics and have also been associated with cardiovascular protection<sup>22</sup>.

### Prescription trajectories across chemical subgroups

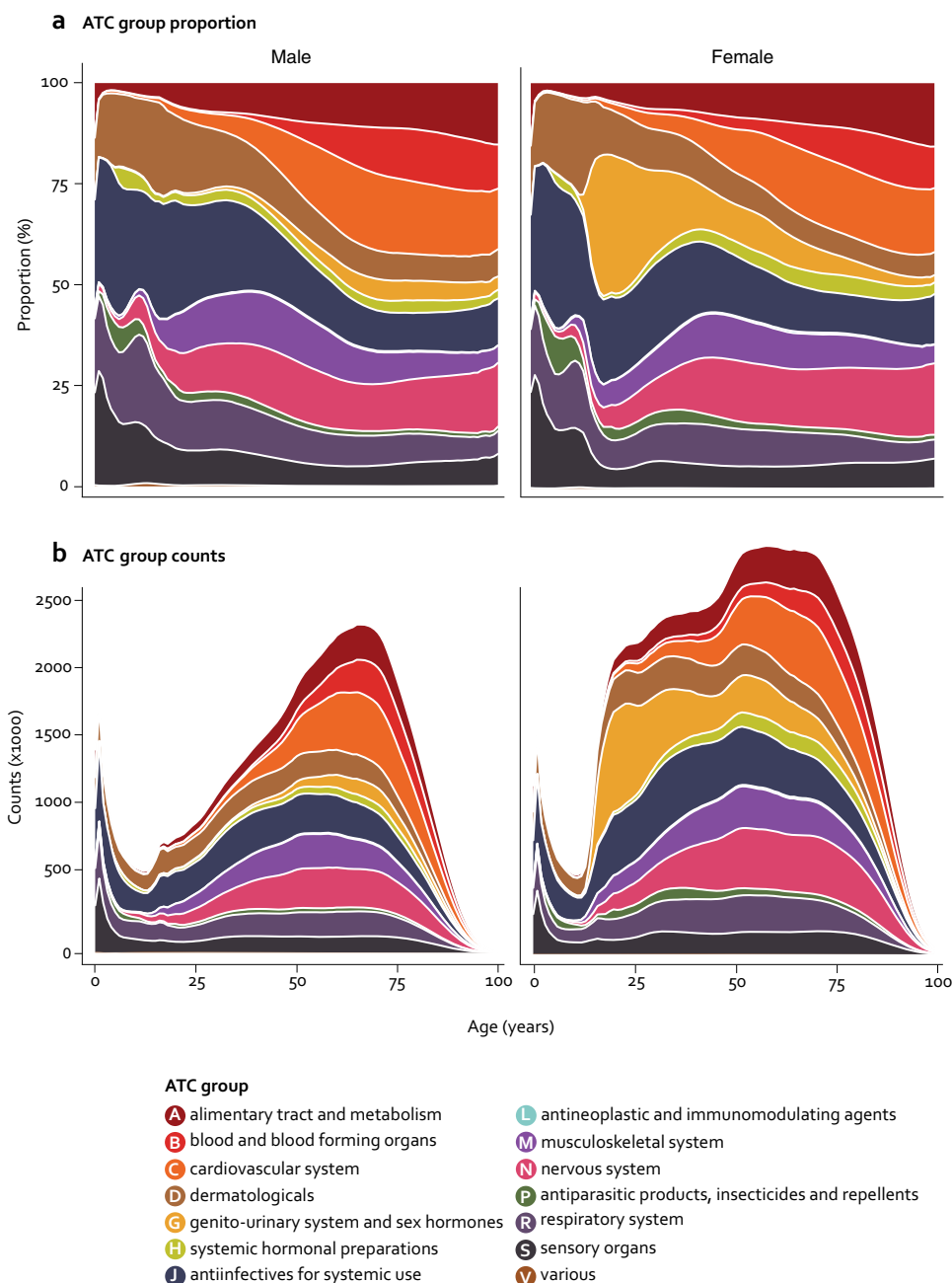
The directional prescription pairs were then combined into prescription trajectories defined by recurrent prescriptions of three or more chemical subgroups, which were followed by a minimum of 1000 patients (Table 1). Trajectories containing prescriptions of four different chemical subgroups exceeded 3.2 million. All trajectories form a heterogeneous network with intertwined connections across anatomical subgroups (Fig. 3a), attesting to the complexity of prescription patterns, as well as the fine-grained level of patient stratification these analyses can support. Supplementary Fig. 4a displays trajectories where all prescriptions are classified in the same anatomical subgroup. The largest number of trajectories is formed by ATC groups J, D, and N (anti-infectives, dermatologicals, and nervous system, respectively).

We found that shifts between different cardiovascular drugs are heterogeneous and complex (Fig. 3B). Patients starting with low-ceiling diuretics ( $C03AB$ ) continue to another prescription within this ATC group in 77% of the cases, whereas 26% of patients starting with high-ceiling diuretics ( $C03CA$ ) continue to another prescription in this ATC group (Supplementary Fig. 4b). The use of plain sulfonamides significantly increases the risk that the patient with this prescription later will be prescribed potassium-sparing agents ( $C03CA \rightarrow C03DB$ ;  $RR = 6.43$ , 95% CI: 4.16–9.93,  $P = 5.40 \times 10^{-17}$ ). The use of thiazides and potassium in combination ( $C03AB$ ) increases the risk of redeeming a prescription subsequently that comprises several other drug subgroups, such as beta-blocking agents ( $C07$ ), dihydropyridine derivatives ( $C08CA$ ), angiotensin II antagonists ( $C09C$  and  $C09D$ ), among others (Fig. 3B).

### Using prescription trajectories in risk stratification

In Fig. 3b, the chemical subgroup agents acting on the renin-angiotensin system ( $C09$ ) show a large number of trajectories interconnected across all chemical subgroups, following prescriptions within the same group subsequently. More than 50% of those who ever redeemed  $C09A$  (ACE inhibitors, plain) ( $n = 1,074,196$ ) also eventually redeemed a second drug from the chemical subgroup  $C09$  ( $n = 569,212$ ). More than six thousand individuals ( $n = 6379$ ) follow a trajectory that starts with  $C09A$ , followed by  $C09B$  (ACE inhibitors, combinations),  $C09C$  (angiotensin II receptor blockers (ARBs), plain) and finally  $C09D$  (angiotensin II receptor blockers (ARBs), combinations) in a short span of time (average 7 years between first and last redemption).



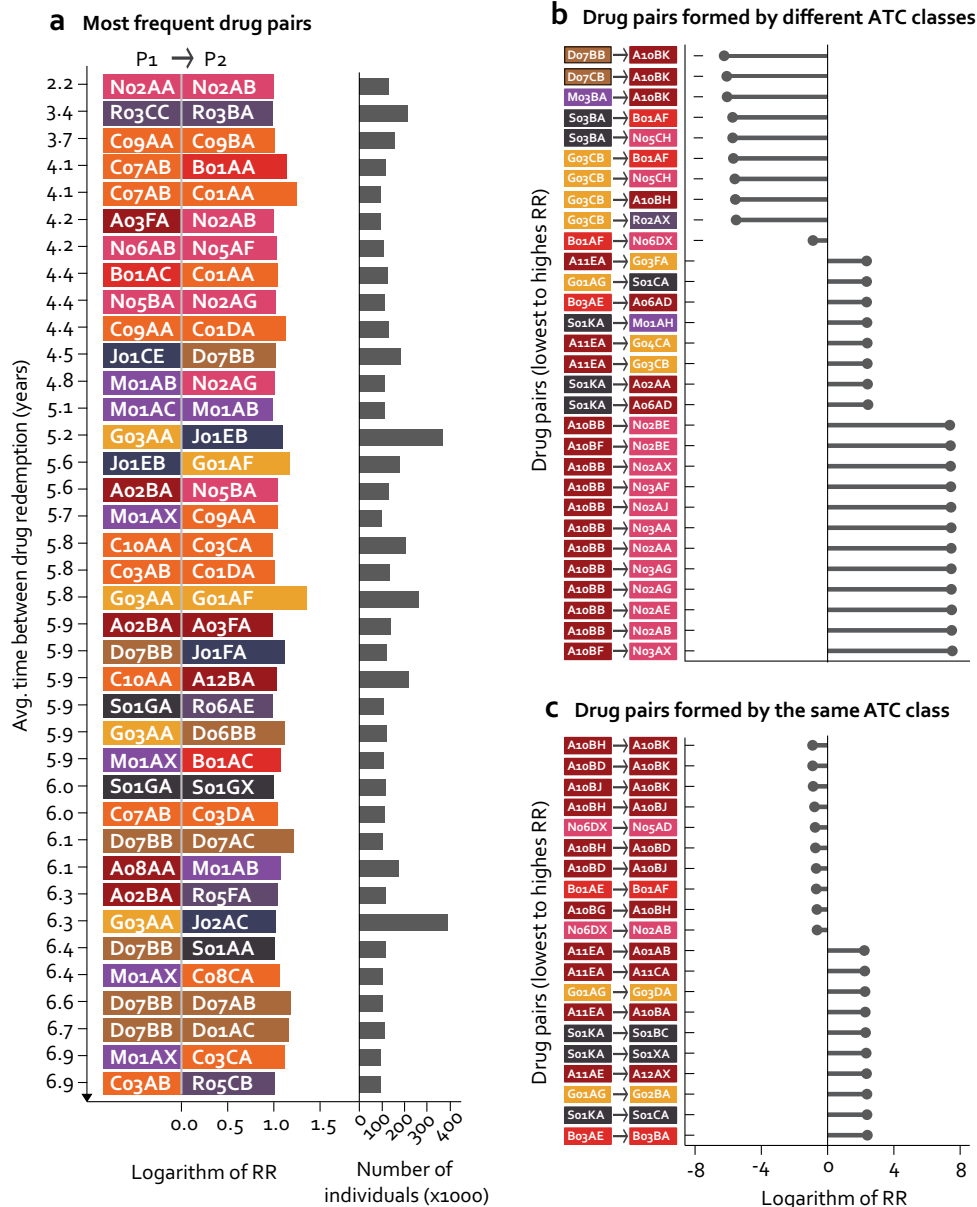


**Fig. 1** Prescriptions redeemed at all Danish pharmacies in the period 1995–2019 stratified by the 14 anatomical ATC groups for males (left) and females (right). **a** Proportion of different ATC groups against age at redemption. **b** Total counts of prescriptions against age at redemption. Color key according to anatomical ATC group.

Next, we tested if the differences in proportion of patients starting with ACE inhibitors and ARBs, and subsequently redeeming another cardiovascular drug or not, would correlate with the 24-year survival probability of these patients. The demographics of patients treated with ACE inhibitors (C09A and C09B) and having no earlier C09 redemptions, and patients who posterior to ACE got prescribed ARBs (C09C and C09D) were similar (Supplementary Table 4). However, we identified four groups of patients with different survival probabilities (Fig. 4a). A Cox proportional hazard model showed that individuals treated solely with ACE are associated with an increased hazard of mortality. Individuals treated with ARBs as first line treatment and with no posterior change in prescription vs individuals treated only with ACE inhibitors have an HR < 1 (HR = 0.73, 95% CI: 0.64–0.82,  $P < 1 \times 10^{-16}$ ).

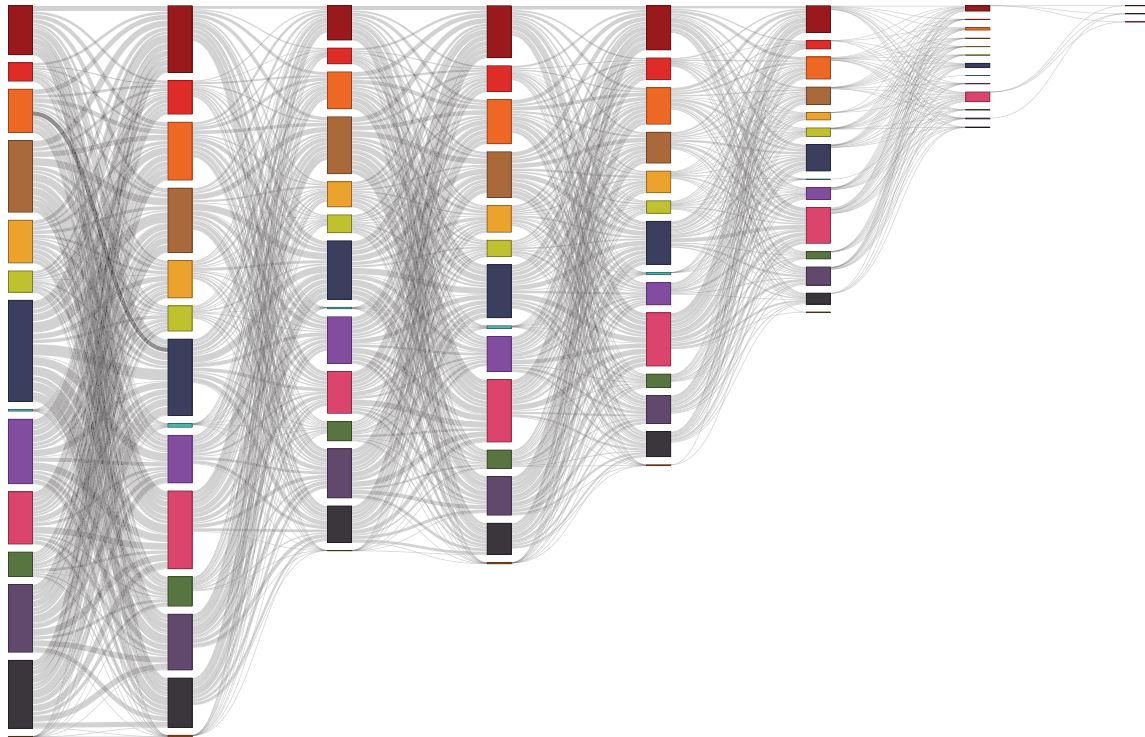
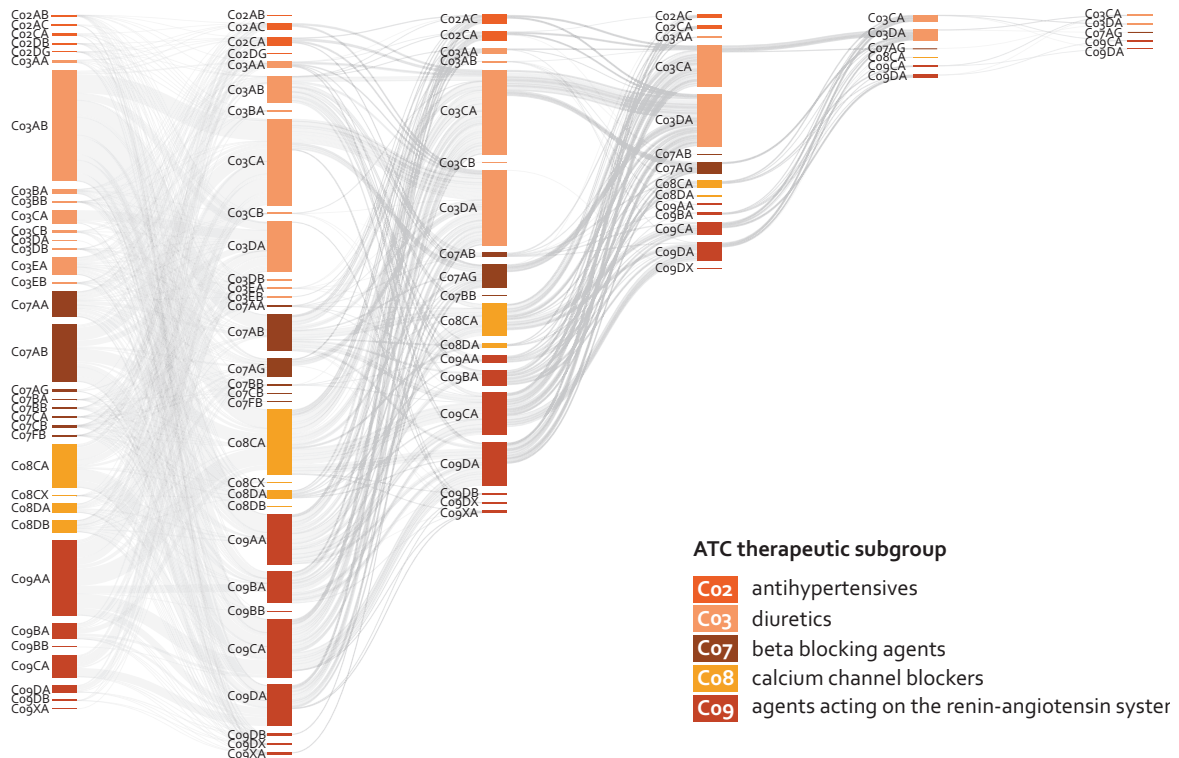
Individuals treated with ACE inhibitors as first line treatment with posterior change to ARBs vs individuals treated with ACE inhibitors, and individuals treated with ARBs with posterior change to ACE vs. individuals treated with ACE and no change, also have HR < 1 (HR = 0.47, 95% CI: 0.41–0.54,  $P < 1 \times 10^{-16}$ ; and HR = 0.96, 95% CI: 0.77–1.19,  $P < 1 \times 10^{-16}$ , respectively).

We then used the UKB self-reported medication information to stratify patients according to RAS treatment, as described above. Similarly to the results obtained using the stratification by RAS of patients in DNPR, a Cox proportional hazard model showed that individuals treated solely with ACE inhibitors had a higher HR compared with the other three groups of patients (ARB vs. ACE: HR = 0.90, 95% CI: 0.84–0.95,  $P < 1 \times 10^{-3}$ ; ACE–ARB vs. ACE: HR =

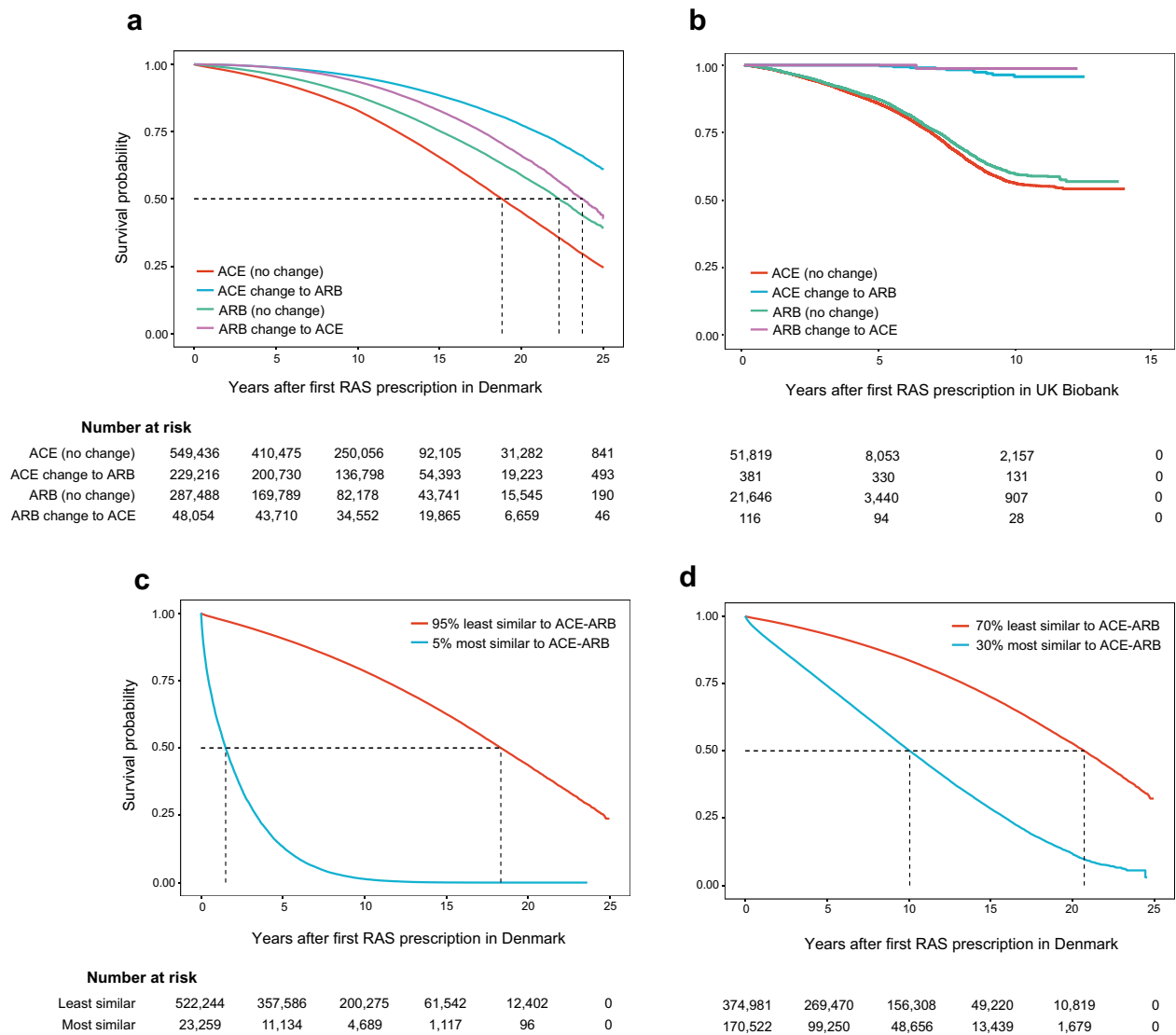


**Fig. 2 Selected directional prescription pairs.** **a** Attrition of prescription pairs following temporal statistical modeling. This includes the drug pair steps (P1 → P2) that were taken sequentially by over 100,000 individuals, had a relative risk ratio (RR) > 2 and a Q-value < 10<sup>-03</sup>. Drug pairs were ordered vertically by time between P1 and P2 from shortest to longest time (top to bottom). Sizes of bars for each prescription (P1 and P2) are equal and correspond to the absolute number of the logarithm of its RR. The pair with the largest RR is G03AA→G01AF. The unannotated bars to the right indicate the number of individuals following each of the directional prescription pairs. **b** Drug pairs (P1→P2) with the lowest and highest RR where P1 and P2 belong different ATC classes. **c** Drug pairs (P1→P2) with the lowest and highest RR where P1 and P2 belong to the same ATC class. Pairs are ordered by RR, lowest to highest (top to bottom). \*Refer to Fig. 1 for ATC group legend.

Trajectory length	Number of combinatorial possibilities (×10 <sup>9</sup> )	Number of trajectories (followed by >1000 patients)	RR > 1*	Number of patients	Avg. time of trajectories (days)
2	0.00016	38,607	26,018	6,971,152	2,903
3	0.031	86,715	570,374	6,092,274	4,436
4	4.52	4,595,060	3,267,979	5,970,284	5,327
5	519.24	4,314,951	3,186,683	4,547,832	5,880
6	49,414.73	3,779,496	2,411,796	3,593,486	7,708
7	4,023,771.39	183,213	81,417	1,475,953	8,594
8	286,190,740.36	203	62	53,512	9,060

**a** Trajectory network**b** Antihypertensives trajectory network

**Fig. 3 Prescription trajectory space.** **a** All statistically significant prescription trajectories followed by more than 1000 patients are represented in this interconnected network. The order of the trajectories and the organization of the network is arbitrary. The length of the trajectory is depicted by the number of nodes it contains (from left to right, ranging from 2 to 8). The heterogeneity of prescription trajectories is depicted here. **b** Alluvial plot of the chemical subgroups used in treatment for hypertension, including C02, C03, C07, C08, C09 (antihypertensives, diuretics, beta blocking agents, calcium channel blockers, and agents acting on the renin-angiotensin system, respectively). The plot is ordered by length from left to right, and arbitrarily by chemical subgroup from top to bottom. The edge height represents the number of people that move from one node to the next (going left to right). Note that patients can follow more than one trajectory at the same time. \*Refer to Fig. 1 for ATC group legend in.



**Fig. 4** Survival curves for patients treated with different RAS drugs as different lines of treatment. **a** Survival curves for different risk groups stratified by prescription trajectories. Age, sex, and CCI adjusted Cox proportional hazard survival probability by groups of RAS treatment: i) redeeming solely ACE inhibitors with no change to ARB; ii) redeeming ACE inhibitors with posterior change to ARB; iii) redeeming uniquely ARB(s) with no change to ACE; and iv) redeeming ARB(s) with posterior change to ACE. The hazard ratio (HR) is lower for all individuals treated solely with ARB or treated with ARB before or after ACE when compared to individuals treated with ACE with no change to second line treatment. **b** Survival curves for the same groups of patients in the UKBB. **c** 95th percentile most similar ACE patients to ACE-ARB trajectories survival vs. ACE patients less similar to ACE-ARB. **d** Seventieth percentile most similar ACE patients to ACE-ARB trajectories survival vs. ACE patients less similar to ACE-ARB.

0.06, 95% CI: 0.03–0.11,  $P < 1 \times 10^{-16}$ ; ARB-ACE vs. ACE: HR = 0.02, 95% CI: 0.003–0.16,  $P < 1 \times 10^{-4}$ ). Interestingly, these groups had different survival probabilities (Fig. 4b).

Next, we calculated the trajectory similarity of patients treated with ACE with no change (ACE) to patients treated with ACE with posterior change to ARB (ACE-ARB), based on their length four trajectories leading up to RAS. Using the adjusted Tanimoto similarity score we ranked the patients treated with ACE based on similarity to each frequent trajectory (followed by more than 1000 patients) from ACE-ARB patients. The survival probability of those individuals, whose trajectory history leading to RAS was most similar to the trajectories of ACE-ARB was significantly lower. The top 5% (95th percentile) most similar patients had a survival probability loss of almost 15 years in comparison to the rest of the ACE patients; and the top 30% (70th percentile) most similar

patients to ACE-ARB had a survival probability of approximately 10 years less compared to the rest of ACE patients (Fig. 4c, d).

## DISCUSSION

This study sheds light on the immense heterogeneity in the prescription trajectory space in a population covered by a one-payer health model. As population-wide health research with registries and EHRs evolves rapidly, it grows closer to account for patient dissimilarities, i.e., advance precision medicine initiatives<sup>23–27</sup>.

Treatment decisions can be supported by these longitudinal analyses, as, for instance, the findings for increased risk after C03AB of different diuretic drug redemptions. Thus, prescribing both drugs at the same time, or changing the first line treatment



for some groups of patients might improve the outcome and reduce the risk of multiple unnecessary treatment shifts. This finding is of special importance for anti-hypertensive drugs as more than 70% of the patients in this study that have a prescription for an anti-hypertensive will be prescribed more than one<sup>28,29</sup>.

Patients with hypertension are very heterogeneous and display different tolerance to drugs. Prescription trajectories might in many cases represent the existing treatment guidelines, which indicate the sequence of drugs to administer given specific diseases and symptoms, and only in case of side effects or incapacity to tackle the diseases can the doctor prescribe the subsequent drugs (primary, secondary, and third prophylaxis)<sup>30</sup>. As exemplified above, in more than 50% of cases, the use of ACE inhibitors (C90A and C09B) as a first prescribed drug is followed by a series of different drugs within the same therapeutic group (C09). Often, drugs redeemed after ACE inhibitors include C09C and C09D. These trajectories suggest that the treatment is more effective when reaching the level of ARB therapy in the guideline. International guidelines advise to consider ARBs as first-line antihypertensive therapy only if there is a compelling indication<sup>31</sup>; the present trajectory-based analysis offers an insight that perhaps most patients receive no substantial benefits from the first line of treatment and only when reaching later treatments, such as C09C or C09D are they are treated effectively. Individuals' trajectory similarity can be used to support decision treatment in patients that require treatment with RAS drugs.

Our study had several strengths and limitations. Using registry data from the entire Danish population ensured a large sample size preventing problems related to inclusion and exclusion, as well as cohort grouping during the drug pair modeling. Furthermore, using a nation-wide registry with structured information on prescription redemptions, where redemptions are always registered, guaranteed that we did not miss many cases in our study. Nevertheless, codes will be missing in the cases where drugs are sold over the counter or dispensed directly by health care providers. Second, the prescription of drugs will be limited to the decision made by the physician, making observations confounded by indication. Third, with respect to the number of people that follow a trajectory, those formed by drugs commonly prescribed will be over-represented.

Some unaddressed challenges remain. For example, neither treatment compliance nor treatment duration is considered in the current version of the model. However, we argue that prevalent diseases like hypertension are ripe for analysis using this strategy given the large size of the data set and plethora of treatment options. We exemplify this using the prescription trajectories for ACE inhibitors and ARBs (Fig. 3). This model was developed based on the Danish population and thus a population with a homogenous demography and generally a low level of emigration, with 92.5% being European and the next most prevalent region of origin being Asia (2.9%) (Supplementary Table 5)<sup>32</sup>. The prescription trajectories might be confounded by patient characteristics. In order to reduce bias due to sex, age at redemption, or year of prescription, these were introduced as covariates in the model. For specific cases other covariates, such as specific diseases or other prescriptions, could be used. However, we did not include it in the general analysis as these are case-dependent, and we aimed for presenting a generic approach. Future work will be focused on integration of additional data sources, e.g., data regarding socio-economic status.

We identified the most prevalent prescription patterns in the Danish population, hence providing a model to help fill gaps between disease registry data from hospitals, indications from general practitioners and the dispensed prescription. Ultimately, the implementation of prescription trajectories in clinical decision support may aid in better patient stratification, further advancing modern healthcare towards personalized medicine.

## METHODS

### Study design

Using the Danish Central Person Registry (CPR) number, a unique identifier to each individual implemented in 1968, all residents of Denmark can be tracked and followed over time<sup>32</sup>.

We identified a cohort of all individuals in the Danish population recorded in CPR between January 1st, 1995 and June 30th, 2019 ( $n = 7,873,901$ ), who were at risk of redeeming one or more prescriptions. Prescription redemption is recorded in the population-based Danish National Prescription Registry (DNPR), and we found that 91% ( $n = 7,255,919$ ) of the population at risk redeemed at least one prescription in the study period.

From the DNPR, we extracted features from the Anatomical Therapeutic Chemical (ATC) Classification System at the chemical subgroup level (i.e., the 4th ATC level). For each individual and each ATC chemical subgroup (e.g., A10BA, biguanides), only the first redemption was included in the analysis.

Information regarding diseases was obtained from the Danish National Patient Registry, which contains administrative information related to hospital admissions such as primary and secondary diagnoses, which uses the International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) during the period 1994 to December 2017.

The UK Biobank (UKB) self-reported medication data were used to replicate patient stratification and survival analysis results.

### Statistical analysis

We examined all sequential redeemed prescriptions, i.e. two prescriptions redeemed at two different times by the same individual, for all patients across a 24-year period. The relative risk ratio of redeeming P2 after P1 relative to those who did not redeem P1 was modeled using a Poisson regression model, where the patients' counts were treated as the dependent variable. The analyses were adjusted for sex, age, calendar year, and time at risk as offset parameter. As patients are often observed over time, a single patient can contribute to multiple redemption, age and calendar year categories. Consequently, prescription redemption, age and calendar year were treated as time-varying factors in the analysis (see Supplementary Fig. 5 for an illustration).

To establish whether the parameters in the model are significant, a Wald Chi-squared test was used for each Poisson model to obtain associated  $p$ -values.  $Q$ -values were calculated from the  $p$ -values using the Benjamini–Hochberg (BH) procedure. Moreover, the event counts may have a variation greater than that predicted by a Poisson distribution, thus leading to overdispersion. Hence, some pairs will have too small  $p$ -values.

To determine directionality of the pairs (P1→P2 or P2→P1) we compare the RR derived from the model for each prescription co-occurrence. We used the  $p$ -value of both to select the more significant direction and we previously run a binomial test to validate them.

Pairs with a significant directionality (P1→P2 or P2→P1) were iteratively joined in prescription trajectories to identify longitudinal trajectories containing three or more prescriptions. These were obtained by combining pairs with overlapping prescriptions (P1→P2 and P2→P3 combined to P1→P2→P3). They were subsequently extended with more overlapping prescription pairs to obtain longer trajectories. In order to add robustness to the trajectories, we selected only those followed by 1000 or more patients. The average time difference between each prescription pair in the trajectory was computed.

### Survival analysis

We evaluated the association between prescription pairs with a significant directionality and death for the individuals who had been exposed to renin-angiotensin agents (C09). Patients were included if they had been prescribed any C09 twice or more, at any point in their recorded history. Patients were followed up until death or end of data period, which was up to 24 years. This was modeled using a left-truncated right-censored multivariable Cox regression model with individuals contributing to the risk group from the time they redeem the first prescription of interest. The risk group was compared to the individuals who did not redeem any prescription in the C09 subgroup twice or more. The model included age, sex, and Charlson comorbidity index (CCI)<sup>33</sup> as covariates (Supplementary Fig. 6).

## Trajectory similarity analysis

The similarity between redeemed, individual prescription histories was calculated using length four trajectories (followed by more than 1000 patients and up to 5 years before the first redeemed RAS drug). Patients treated with ACE inhibitors and no posterior change to ARB, where the last prescription was the RAS drug, were compared at the trajectory level to patients treated with ACE with posterior change to ARB (ACE-ARB). To quantify the similarity, we used an adjusted Tanimoto similarity score<sup>34</sup>. Each patient's prescription history in terms of statistically significant drug pairs in their length four trajectories leading to the first RAS treatment were compared against each frequent trajectory followed by ACE-ARB patients. The intersection of prescriptions was divided by the total number,  $PP$ , of elements in the ACE-ARB trajectory,

$$\text{sim}(\text{patient, ACE} - \text{ARB}) = \frac{\sum_{k=0}^n \frac{|\text{PP in patient history} \cap \text{PP in ACE-ARB trajectory}_k|}{|\text{PP in ACE-ARB trajectory}_k|}}{n} \quad (1)$$

Similarity analysis was performed across all ACE-ARB trajectories for each patient treated with ACE and no switching to other RAS drugs. The closer to 1 the more similar the patient is to the frequent trajectories followed by ACE-ARB patients. To quantify the overall similarity, we calculated the mean similarity for each patient across the set of trajectories.

## Data and material approval

This study has been approved by the Danish Data Protection Agency (ref: SUND-2016-83) and the Danish Health Authority (refs: FSEID-00001627, FSEID-00003092, and FSEID-00003096).

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

Aggregate data created in this study are available from the corresponding author on reasonable request. Permission to access and analyse the underlying person-sensitive data can be obtained following approval from the Danish Data Protection Agency and the Danish Health Authority. Due to privacy concerns the data has to be analysed in closed analysis environments.

## CODE AVAILABILITY

In order to analyse the data, the following software tool was used: R v. 3.6.1.

Received: 18 May 2021; Accepted: 27 September 2021;  
Published online: 20 October 2021

## REFERENCES

1. Organisation for Economic Co-operation and Development. OECD Health Statistics. 2019. <https://doi.org/10.1787/health-data-en>. Accessed June 2020.
2. Carroll, A. E. The high costs of unnecessary care. *JAMA* **318**, 1748–1749 (2017).
3. Mandl, K. D. & Kohane, I. S. Federalist principles for healthcare data networks. *Nat. Biotechnol.* **33**, 360–363 (2015).
4. Aanestad, M., Grisot, M., Hanseth, O. & Vassilakopoulou, P. *Information Infrastructures within European Health Care: Working with the Installed Base* (Springer, 2017).
5. Kildemoes, H. W., Sorensen, H. T. & Hallas, J. The Danish National Prescription Registry. *Scand. J. Public Health* **39**, 38–41 (2011).
6. Klaukka, T. The Finnish database on drug utilisation. *Nor. J. Epidemiol.* **11**, 19–22 (2001).
7. Wolf, A. et al. Data resource profile: clinical practice research datalink (CPRD) Aurum. *Int. J. Epidemiol.* **48**, 1740–174 (2011).
8. Sen, A. et al. Systematic assessment of prescribed medications and short-term risk of myocardial infarction—a pharmacoepi-wide association study from Norway and Sweden. *Sci. Rep.* **9**, 8257 (2019).
9. Yelin, I. et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. *Nat. Med.* **25**, 1143–1152 (2019).
10. Jess, T. et al. Antibiotic use during pregnancy and childhood overweight: a population-based nationwide cohort study. *Sci. Rep.* **9**, 11528 (2019).
11. Jordakieva, G. et al. Country-wide medical records infer increased allergy risk of gastric acid inhibition. *Nat. Commun.* **10**, 3298 (2019).
12. Patel, C. J., Ji, J., Sundquist, J., Ioannidis, J. P. A. & Sundquist, K. Systematic assessment of pharmaceutical prescriptions in association with cancer risk: a method to conduct a population-wide medication-wide longitudinal study. *Sci. Rep.* **6**, 31308 (2016).
13. Sigurdsson, M. I. et al. Association between preoperative opioid and benzodiazepine prescription patterns and mortality after noncardiac surgery. *JAMA Surg.* **154**, e191652 (2019).
14. Wright, A. P., Wright, A. T., McCoy, A. B. & Sittig, D. F. The use of sequential pattern mining to predict next prescribed medications. *J. Biomed. Inform.* **53**, 73–80 (2015).
15. Fredheim, O. M. et al. Chronic pain and use of opioids: a population-based pharmacoepidemiological study from the Norwegian prescription database and the Nord-Trøndelag health study. *Pain* **155**, 1213–1221 (2014).
16. Rod, N. H. et al. Trajectories of childhood adversity and mortality in early adulthood: a population-based cohort study. *Lancet* **396**, 489–497 (2020).
17. Unger, T. et al. 2020 International Society of Hypertension Global Hypertension Practice Guidelines. *Hypertension* **75**, 1334–13357 (2020).
18. Kessing, L. V., Vradi, E. & Andersen, P. K. Nationwide and population-based prescription patterns in bipolar disorder. *Bipolar Disord.* **18**, 174–182 (2016).
19. Webb, A. J., Fischer, U., Mehta, Z. & Rothwell, P. M. Effects of antihypertensive-drug class on interindividual variation in blood pressure and risk of stroke: a systematic review and meta-analysis. *Lancet* **375**, 906–915 (2010).
20. Aminzadeh, A., Sabeti Sanat, A. & Nik Akhtar, S. Frequency of candidiasis and colonization of *Candida albicans* in relation to oral contraceptive pills. *Iran. Red. Crescent Med. J.* **18**, e38909 (2016).
21. Vinik, A. I. Clinical practice. Diabetic sensory and motor neuropathy. *N. Engl. J. Med.* **374**, 1455–1464 (2016).
22. Scheen, A. J. SGLT2 inhibitor or GLP-1 receptor agonist in type 2 diabetes? *Lancet Diabetes Endocrinol.* **7**, 818–820 (2019).
23. Nielsen, A. B. et al. Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the Danish National Patient Registry and electronic patient records. *Lancet Digital Health* **1**, e78–e89 (2019).
24. Jensen, P. B., Jensen, L. J. & Brunak, S. Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* **13**, 395–405 (2012).
25. Schork, N. J. Personalized medicine: time for one-person trials. *Nature* **520**, 609–611 (2015).
26. Jorgensen, I. F., Aguayo-Orozco, A., Lademann, M. & Brunak, S. Age-stratified longitudinal study of Alzheimer's and vascular dementia patients. *Alzheimers Dement.* **16**, 908–917 (2020).
27. Siggaard, T. et al. Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million Danish patients. *Nat. Commun.* **11**, 4952 (2020).
28. Gradman, A. H., Basile, J. N., Carter, B. L. & Bakris, G. L., American Society of Hypertension Writing Group. Combination therapy in hypertension. *J. Am. Soc. Hypertens.* **4**, 90–98 (2010).
29. Yoon, E. Y. et al. Use of antihypertensive medications and diagnostic tests among privately insured adolescents and young adults with primary versus secondary hypertension. *J. Adolesc. Health* **55**, 73–78 (2014).
30. Mancía, G., Rea, F., Corra, G. & Grassi, G. Two-drug combinations as first-step antihypertensive treatment. *Circ. Res.* **124**, 1113–1123 (2019).
31. WHO. *Prevention of Cardiovascular Disease. Guidelines for Assessment and Management of Cardiovascular Risk* (WHO, 2007).
32. Pedersen, C. B. The Danish civil registration system. *Scand. J. Public Health* **39**, 22–25 (2011).
33. Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* **40**, 373–383 (1987).
34. Jorgensen, I. F. & Brunak, S. Time-ordered comorbidity correlations identify patients at risk of mis- and overdiagnoses. *npj Digit. Med.* **4**, 1–10 (2021).

## ACKNOWLEDGEMENTS

We would like to acknowledge funding from the Novo Nordisk Foundation (grant agreements NNF14CC0001, NNF17OC0027594, and NNF18SA0034956). This study has been approved by the Danish Data Protection Agency (ref: SUND-2016-83) and the Danish Health Authority (refs: FSEID-00001627, FSEID-00003092, and FSEID-00003096).

## AUTHOR CONTRIBUTIONS

A.A.O. and S.B. conceived the idea and wrote the first draft. A.D.H. and I.F.J. reviewed early drafts L.V.M. had oversight of the data analysis, which was carried out by A.A.O.

L.V.M., P.M., and D.W. contributed to later drafts. All authors approved the final submission.

### COMPETING INTERESTS

S.B. reports ownerships in Intomics A/S, Hoba Therapeutics Aps, Novo Nordisk A/S, Lundbeck A/S, and managing board memberships in Proscion A/S and Intomics A/S outside the submitted work. A.A.O., A.D.H., I.F.J., D.W., P.M., and L.V.M. declare no competing interests.

### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41746-021-00522-4>.

**Correspondence** and requests for materials should be addressed to Laust Hvas Mortensen or Søren Brunak.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021