# Socio-demographic and genetic risk factors for drug adherence and persistence: a retrospective nation-wide and biobank study across 5 medication classes and 1 814 521 individuals

M. Cordioli[1,#], A. Corbetta[1,2,3,#], H.M. Kariis[4], S. Jukarainen[1], P. Vartiainen[1], T. Kiiskinen[1], M. Ferro[1], FinnGen, Estonian Biobank research team, K. Lehto[4], M. Niemi[5,6,7], S. Ripatti[1,8,9], L. Milani[4], A. Ganna[1,10,11]

[1] Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[2] CHDS - Health Data Science Center, Human Technopole, Milan, Italy
[3] MOX - Laboratory for Modeling and Scientific Computing, Department of Mathematics, Politecnico di Milano, Milan, Italy
[4] Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
[5] Department of Clinical Pharmacology, University of Helsinki, Helsinki, Finland
[6] Individualized Drug Therapy Research Program, University of Helsinki, Helsinki, Finland
[7] Department of Clinical Pharmacology, HUS Diagnostic Center, Helsinki University Hospital, Helsinki, Finland
[8] Broad Institute of MIT and Harvard, Cambridge, MA, USA
[9] Department of Public Health, University of Helsinki, Helsinki, Finland
[10] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
[11] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[#] These authors contributed equally to this work
Correspondence to Andrea Ganna (andrea.ganna@helsinki.fi)

## Abstract

Low drug adherence is a major obstacle to the benefits of pharmacotherapies, and it is therefore important to identify factors associated with discontinuing or being poorly adherent to a prescribed treatment regimen. Using high-quality nationwide health registry data and genome-wide genotyping, we evaluate the impact of socio-demographic and genetic risk factors on adherence and persistence for 5 common medication classes that require long-term, regular therapy (N = 1 814 591 individuals, 185 931 with genetic data).
Need for social assistance and immigration status showed a notable negative effect on persistence and adherence across the examined medications (odd ratios between 0.48 and 0.82 for persistence and between 1.1% to 4.3% decrease in adherence) while demographic (age and sex), health (comorbidity status), and genetic factors showed comparably modest or inconsistent effects. Some pharmacogenes were modestly associated with persistence, but not with adherence. We observed significant genetic correlations between medication adherence and participation in research studies. Overall, our findings suggest that socio-economically disadvantaged groups would benefit from targeted interventions to improve the dispensing and uptake of pharmacological treatments.

## Introduction

Drug adherence is defined as the extent to which patients follow the prescribed therapeutic regimen when taking medications [1]. High adherence is considered to be crucial for the effectiveness of pharmacological treatments [2], and improving adherence has been considered as one of the leading challenges to healthcare professionals [3]. A better understanding of the risk factors driving adherence is fundamental to help clinicians improve their clinical practice and identify patients who are at risk of poor adherence at an earlier stage.

While there are multiple dimensions of adherence, some related to the healthcare system and the mode of therapy delivery [4,5], studies have shown that individuals with low adherence in placebo arms of randomized controlled trials have higher mortality[6]. This suggests that adherence to pharmacotherapy may be a surrogate marker for overall health behavior ("healthy adherer" effect) and that risk factors for adherence might be shared across different drugs.

Accurately estimating patients drug usage and adherence is not a trivial task. Traditionally questionnaires have been used, but with the increasing availability of electronic health records, studies have started investigating determinants of drug adherence and persistence using pharmacy administrative databases or insurance claims data [7–10]. However, most studies focus on a single medication or condition and consider adherence over a limited follow-up available due to data limitations. Few studies have investigated determinants of adherence across different medications [11,12], but have only considered a limited number of demographic and socioeconomic risk factors.

Beyond socio-demographic factors, genetics might play a role in drug adherence. Pharmacogenomic studies have identified several genetic variants relevant to drug response that can increase an individual's risk for adverse drug reactions [13] and a recent randomized trial suggested that a pre-emptive testing with a 12-gene pharmacogenetic panel can significantly reduce the incidence of clinically relevant adverse drug reactions [14]. Because an adverse drug reaction decreases the likelihood of a patient continuing to take a medication, genetics might impact adherence via this mechanism. Nonetheless, the impact of clinically relevant pharmacogenetic variants on drug adherence has not yet been studied. Moreover, identifying novel genetic variants associated with adherence might help to discover new aspects of drug response. An alternative, but not mutually exclusive hypothesis, is that genetics might capture underlying health and behavioral aspects that are associated with adherence. In this scenario, polygenic scores (PGS) for certain diseases or behavioral traits might help predicting drug adherence.

Here, we conducted a comprehensive and systematic investigation of adherence and persistence to drug treatments across multiple patient-related dimensions encompassing demographic, socioeconomic and genetic factors. We used high-quality nation-wide electronic health record data on prescription medication purchases made at pharmacies to accurately estimate two established drug-taking phenotypes [15–18] (**Methods**): (a) *persistence*, defined as

purchasing medications for at least one year versus early discontinuation of the medication (i.e. after one purchase) and (b) *adherence*, measured as Medication Possession Ratio (MPR), which is the "proportion (or percentage) of days supply obtained during a specified time period" (**Figure 1, Methods**). We considered five commonly prescribed medications that are typically used for long-term, continuous, regular therapy, and do not have directly observable effects on symptoms.

We first explored the role of multiple health, demographics and socioeconomic factors in FinRegistry [19], a nation-wide cohort study. Next, we considered the role of genetics in a subset of the Finnish population from the FinnGen study [20], where genome-wide genotype information is available. In FinnGen, we focused on well-known pharmacogenes as well as on genome-wide effects. We performed a genome-wide association study (GWAS) to identify potential novel genetic associations with adherence and persistence and we evaluated if these results can be used for risk stratification. Finally, we used genetic correlations analysis to examine the overlap between persistence and adherence and other diseases or behavioral traits, and whether PGSs capturing these traits can be used to predict the two drug-taking phenotypes of interest. The results were replicated in an independent cohort from the Estonian Biobank.

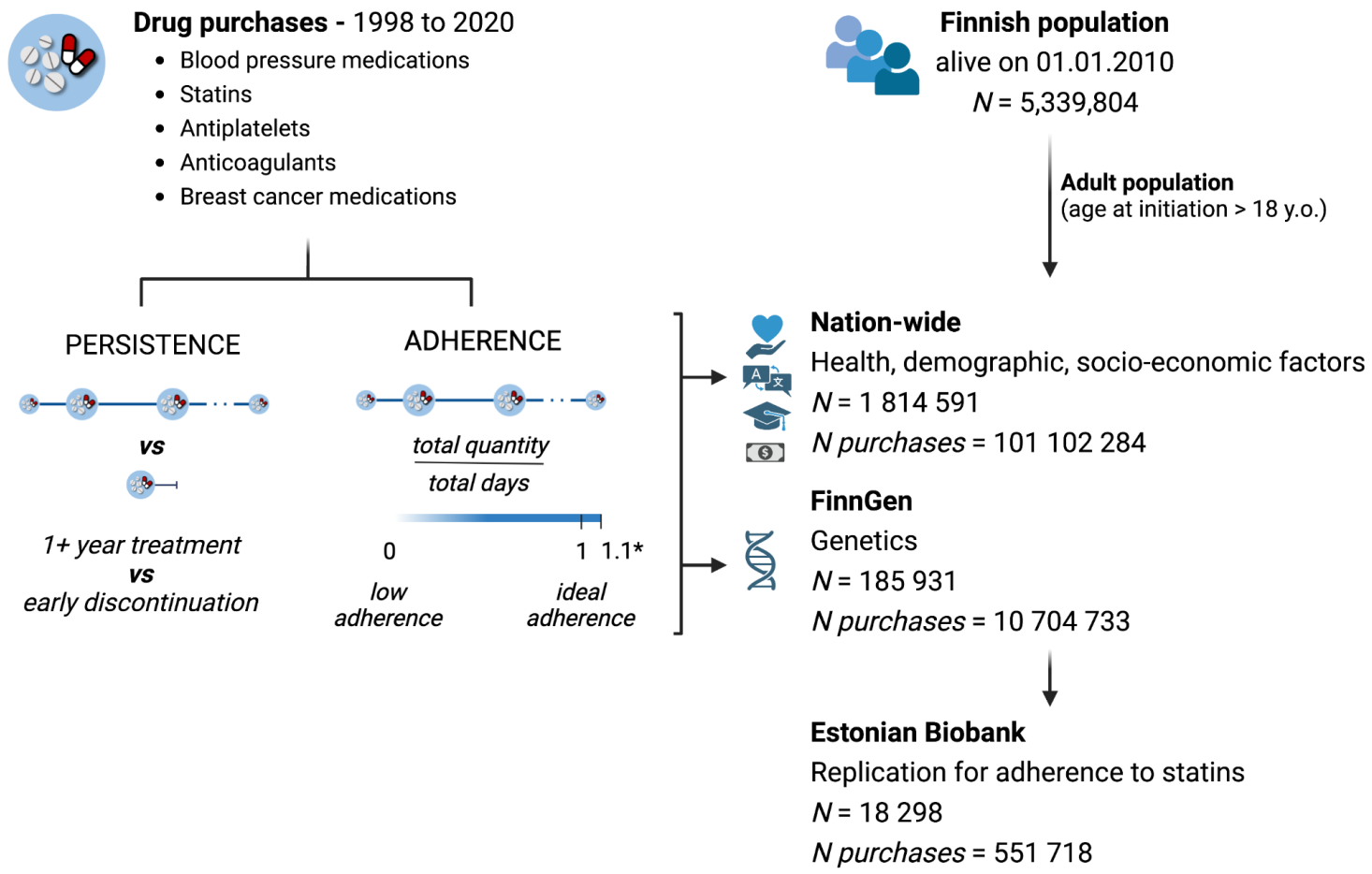



**Figure 1 | Schematic overview of the study.** We included all individuals resident in Finland and alive on 1.1.2010 (N=5 339 804) and used information from the Finnish drug purchase register to identify individual with at least one purchase for the following medication medication classes: blood pressure medications, statins, antiplatelets, direct oral anticoagulants and breast cancer medications. We considered only the adult population, starting the treatment after age 18, and reconstructed individual medication purchase trajectories as a proxy to define the medication usage phenotypes: persistence and adherence. Persistence was defined as continuing the treatment for at least one year versus discontinuing it after the first purchase. We defined adherence as the Medication Possession Ratio (ratio of the quantity purchased in the observation period to the days in the observation period), for patients continuing the treatment for at least one year. Adherence is a continuous trait ranging from 0 (complete lack of adherence) to 1 (ideal adherence). *We excluded patients with adherence above 1.1, to exclude excessive stockpiling or overbuying behaviors, but still allowing for some variability. We then studied the association of these two medication use phenotypes with health, demographic and socio-economic factors from nation-wide registries (N=1 814 591 with complete information) and with genetic information for a subset of the population from the FinnGen study (N=185 931). Created with Biorender.com.

## Results

### Adherence and persistence for 5 medications in 1 814 591 individuals in the general population

We considered five classes of medications: statins, blood pressure (BP) medications, antiplatelets, breast cancer medications, direct oral anticoagulants (DOAC) (ATC codes included are described in **Table 1** and **Methods**). For breast cancer medications and DOACs, we included individuals with, respectively, a breast cancer or atrial fibrillation, pulmonary embolism or venous thromboembolism diagnosis before treatment initiation (**Methods**). For antiplatelets, we did not analyze persistence, as this class of medications is frequently prescribed for temporary treatments of 6 to 12 months. Information about the date, quantity, and type of reimbursable medications that were purchased in pharmacies were available between 1998 and 2020 and were used to calculate adherence and persistence.

Our initial study population included every individual who was a resident in Finland on 1/1/2010 (N=5 339 804). 2 289 455 individuals (43% of the whole population) purchased at least one of the 5 medications after 18 years of age, with a total of 162 231 752 purchases.
Most individuals purchased the 5 medications for at least 12 months (ranging from 89.7% for breast cancer medications to 64.7% for anticoagulants, with the exception of antiplatelets (44.2%), **Supplementary Table 1**) and were eligible for the adherence analysis (N = 1 876 212). We further excluded 141 452 individuals who had a medication possession ratio (MPR) > 1.1, to account for stockpiling of medications or overbuying behaviors (**Methods**), and 1 270 individuals with incomplete registry information for the 8 health and socioeconomic factors considered in the main analysis (**Supplementary Table 1**). In the persistence analysis, we considered people discontinuing the treatment after one purchase, and compared them to those continuing for at least 12 months and were eligible for the adherence analysis. Our final study population thus included 1 814 591 individuals, for a total of 101 102 284 purchases.

BP medications were the most purchased (N of individuals = 1 316 889) and had a mean treatment length of over 10 years (**Table 1**). Breast cancer medications were purchased by the fewest individuals (N=55 270) and most individuals stopped after 5 years.
Persistence was above 90% for all drugs (**Figure 2**). Adherence was 0.94 averaging across all medications, highest in breast cancer medications (0.97) and lowest in statins (0.89; **Table 1**). Adherence values were similar to those reported in the literature in comparable studies or meta-analyses [11,21] with all medication classes having at least 80% of individuals considered as “good adherers” (**Figure 2**). **Extended Figure 1** provides an overview of adherence values by showing the relationship between expected and observed consumption across the 5 medication classes.
For the subset of 185 931 individuals participating in FinnGen, persistence rates and adherence distributions were consistent with the general population. (**Table 1**).

| Medication class | ATC codes | Cohort | N individuals | Total purchases | Proportion of persistent individuals | Mean adherence (SD) | Mean treatment length in years |
|---|---|---|---|---|---|---|---|
| Blood Pressure medications | C02*, C03*, C08*, C09* | FinRegistry | 1,316,889 | 61,213,707 | 0.947 | 0.93 (0.123) | 10.755 |
| | | FinnGen | 134,615 | 6,399,494 | 0.937 | 0.932 (0.121) | 11.172 |
| Statins | C10AA* | FinRegistry | 1,061,926 | 33,864,971 | 0.952 | 0.892 (0.123) | 9.168 |
| | | FinnGen | 116,439 | 3,828,359 | 0.956 | 0.901 (0.118) | 9.454 |
| Antiplatelets | B01AC04, B01AC30 | FinRegistry | 138,007 | 3,120,237 | - | 0.931 (0.114) | 5.528 |
| | | FinnGen | 12,618 | 285,902 | - | 0.933 (0.112) | 5.572 |
| Anticoagulants | B01AF01, B01AF02, B01AF03, B01AF07 | FinRegistry | 105,113 | 2,003,698 | 0.975 | 0.957 (0.094) | 3.127 |
| | | FinnGen | 2,792 | 45,394 | 0.889 | 0.952 (0.101) | 2.984 |
| Breast Cancer medications | L02BA01, L02BG04, L02BG06, L02BG03 | FinRegistry | 55,270 | 899,671 | 0.984 | 0.969 (0.067) | 4.556 |
| | | FinnGen | 9,857 | 145,584 | 0.985 | 0.971 (0.067) | 4.128 |

**Table 1 | Descriptive statistics for the 5 medication classes included in the analysis.** For each medication class, we report the ATC codes used to define the medication purchases of interest and other descriptive statistics in the general population (FinRegistry) and in the subset of people with genetic data (FinnGen). Number of individuals is the total number of individuals included in the analysis, meaning the number of individuals considered for both the persistence and adherence analysis. Total purchases is the total number of purchases per each medication and each cohort considered in both analyses. Mean adherence (and its standard deviation) and the mean treatment length are reported for the individuals with at least one year long treatment (individuals considered in the adherence analysis).

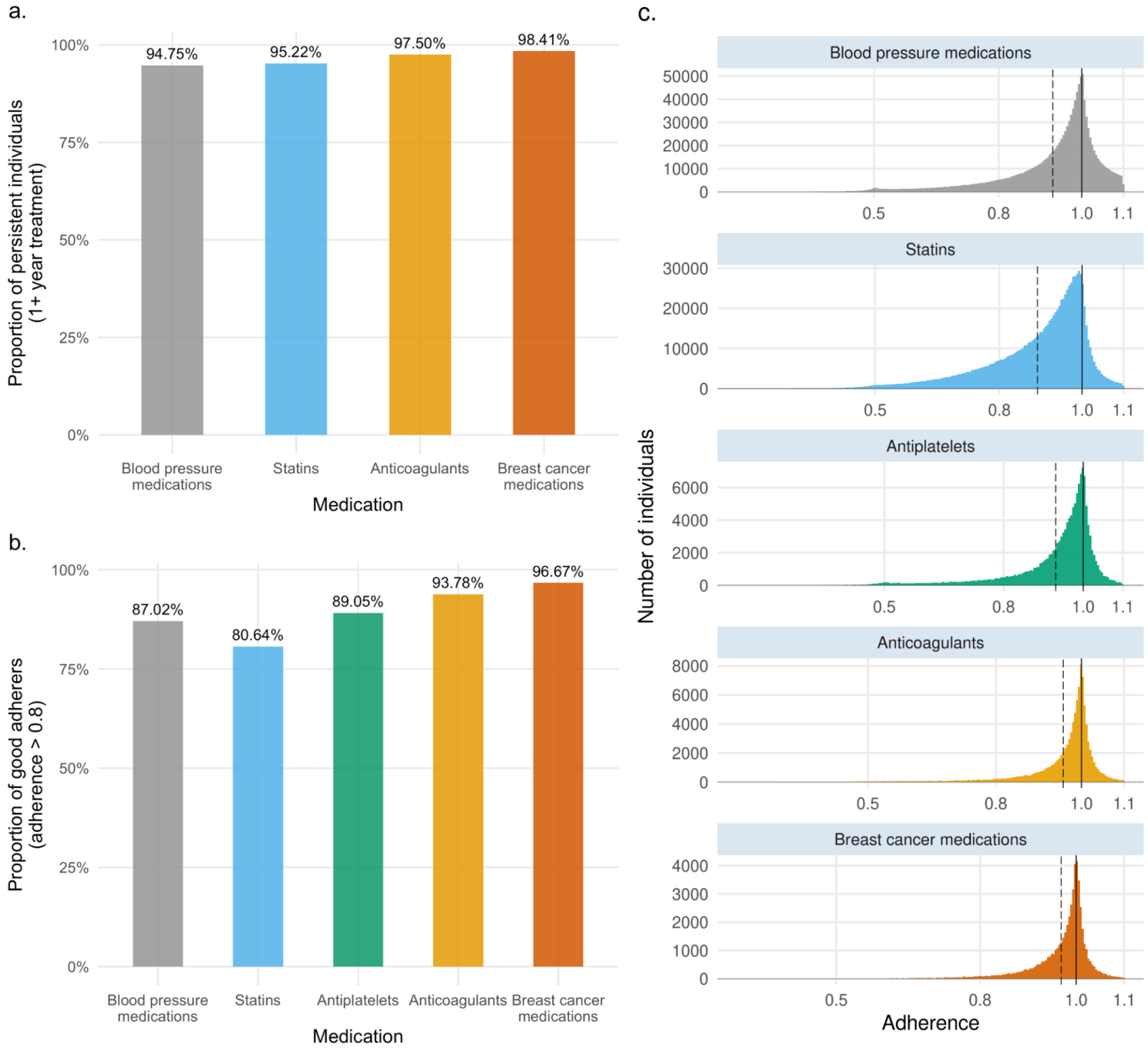



**Figure 2 | Persistence and adherence to 5 medications in the general population. a.,** Proportion of persistent individuals for each medication class considered in the analysis. **b.**, Proportion of "good adherers" (individuals with adherence greater than 0.8) for each medication class. **c.,** Adherence distribution and mean adherence (dotted line) for the 5 medication classes considered in the analysis.

**Health and socio-demographic risk factors for persistence and adherence**

We studied the effects of the following risk factors, measured at the time of first purchase, on persistence and adherence: age, sex, primary *vs* secondary prevention, Charlson comorbidity index, years spent in education, living area, mother tongue (a proxy for first-generation immigration status), and receiving any social assistance benefit in the year before starting the treatment. We chose these 8 factors because they broadly capture overall health and socio-demographic aspects, were previously examined in the literature [4] and were available from nation-wide registers. Basic descriptive statistics for each factor and each medication are reported in **Supplementary Table 2**. Some socio-demographic factors were consistently associated with lower persistence and adherence across all medications in a multivariable analysis (**Figure 3** considers dichotomized factors while **Supplementary Tables 5 and 6** report similar analysis when continuous factors are treated as such). For example, not speaking Finnish or Swedish and having received social assistance benefits were associated with an average 3.2% [1.9%-4.3%] and 2% [1.1%-3.1%] lower adherence across medications, respectively. Starting a treatment after a related disease event (secondary prevention) was also consistently associated with higher odds of persistence for statins (odds ratio (OR) = 1.53, 95% confidence interval (CI) = 1.49-1.58, $P = 4.9 \times 10^{-198}$) and increased adherence to statins and antiplatelets (1.7% increase, 95% CI = 1.64-1.76% $P < 2 \times 10^{-308}$ for statins; 0.47% increase, 95% CI = 0.35-0.59%, $P = 7.4 \times 10^{-15}$ for antiplatelets). Age and sex showed discordant effects across different medications after adjustment for the remaining factors. Being female showed a consistent negative effect on persistence for all drugs but showed heterogeneous effects on adherence. Age also showed heterogeneous effects across different medications (**Supplementary Tables 3-6**).

Overall, the considered health and socio-demographic factors showed a higher effect on persistence than on adherence as indicated by overall higher variance explained (**Supplementary Table 7**).

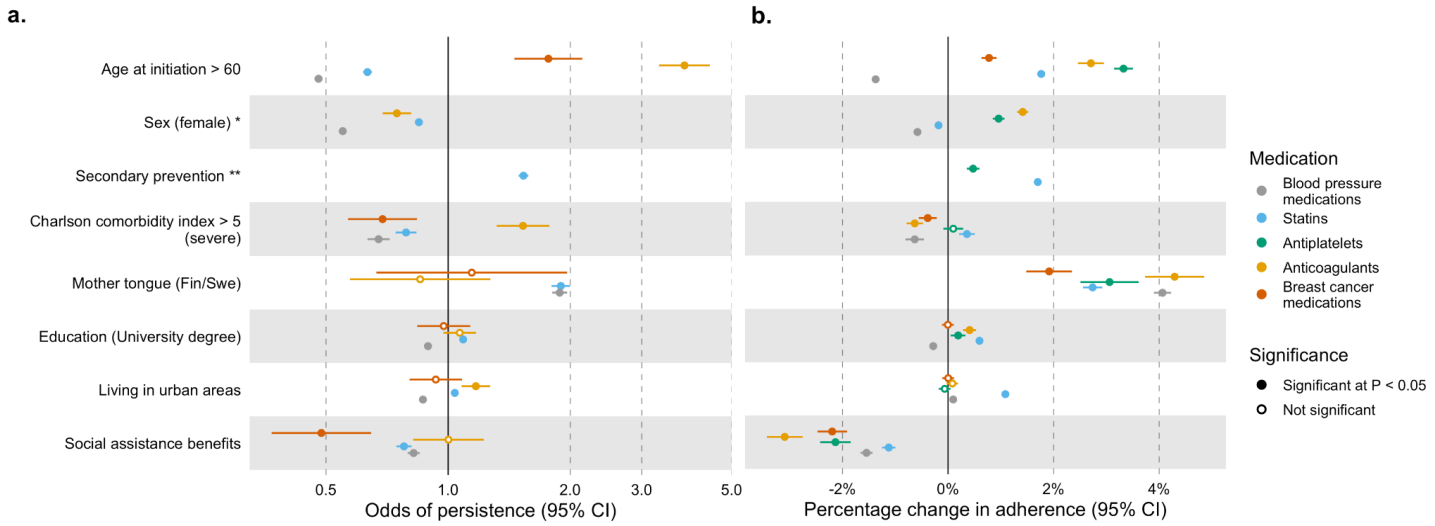



**Figure 3 | Effect of 8 health, demographic and socioeconomic factors on persistence and adherence. a.**, Odds of persistence given each factor presence. ORs are from a multivariable logistic regression model, error bars represent the 95% confidence interval for the estimates. **b.**, Coefficients are estimated from a multivariable linear regression model and were rescaled in order to represent the percentage change in adherence given each factor presence. Error bars represent the 95% confidence interval for the estimate. Descriptive statistics for each risk factor are reported in Supplementary Table 2. In this figure we binarized the continuous risk factors to make effect sizes comparable, however results for non-binarized risk factors are available in Supplementary Table 5-6.
*****sex not included for breast cancer medications as all individuals included are females
******secondary prevention defined only for statins and antiplatelets (**Methods**)

### Established pharmacogenes do not associate with persistence and adherence

We tested the effects of 5 relevant pharmacogenes on persistent and adherence. We selected the following pharmacogenes based on the guidelines from the Clinical Pharmacogenetics Implementation Consortium [22–25], filtering for the drug/gene pairs which were reported with recommended prescribing actions (e.g. alternative therapies or dosing): ABCG2 for rosuvastatins, CYP2C9 for fluvastatins, SLCO1B1 for all type of statins, CYP2C19 for clopidogrel, and CYP2D6 for tamoxifen. For each gene, we defined individual diplotypes and grouped them in phenotypes defined based on each diplotype activity score. We then examined the effect of each phenotype on persistence and adherence. The CYP2C9 intermediate metabolizer and CYP2D6 poor metabolizer profiles were nominally significantly associated with lower odds of persistence for, respectively, rosuvastatins and tamoxifen (OR = 0.51, 95% CI = 0.29 - 0.91, P = 0.02 for CYP2C9, OR = 0.4, 95% CI = 0.18 - 0.87, P = 0.02 for CYP2D6), but were not significant after multiple testing correction (P < 0.002, Bonferroni correction for 19 phenotypes tested).

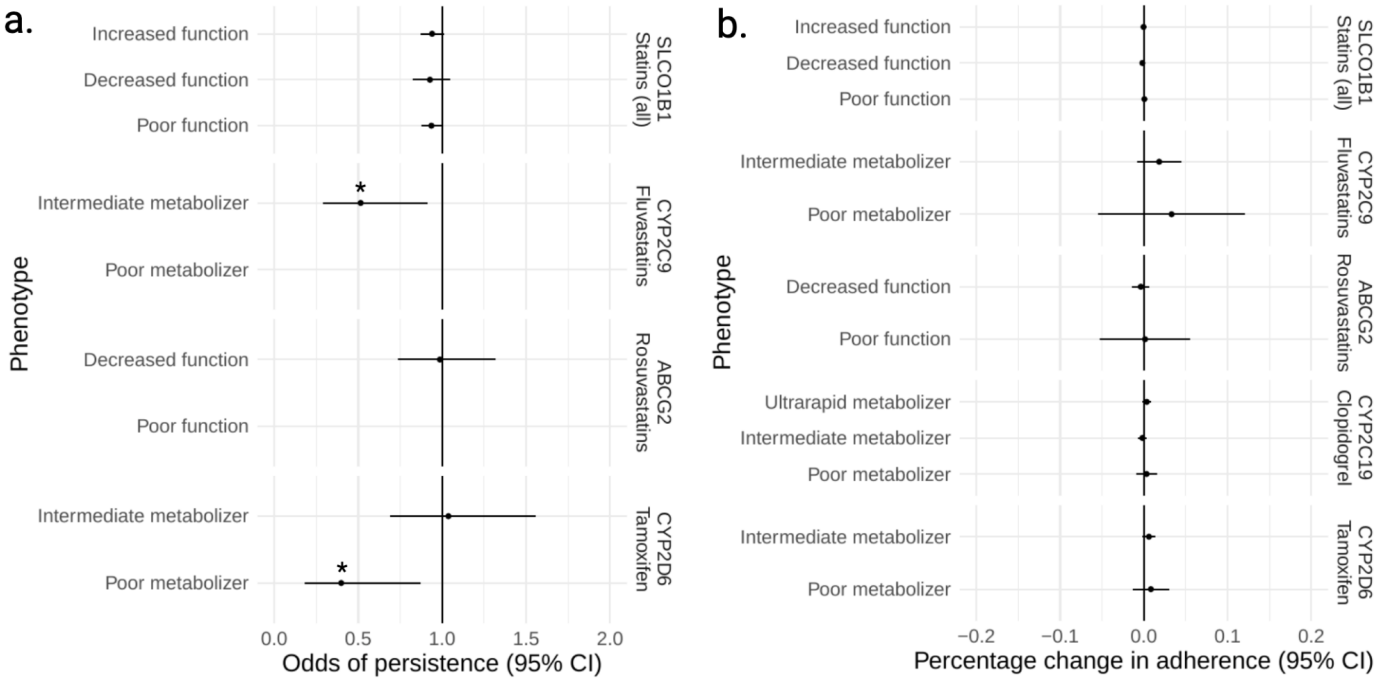



**Figure 4 | Forest plot for associations between clinically relevant pharmacogenes metabolic phenotypes and persistence and adherence**. a. Odds of persistence given each phenotype level. ORs are estimated from a logistic regression model, separately for each gene/drug pair. Error bars represent the 95% confidence interval for the estimate. Stars represent nominal (P < 0.05) statistical significance. ORs are not reported for phenotypes with N < 5. **b**. Percentage change in adherence given each phenotype level. Coefficients are estimated from a linear regression model, separately for each gene/drug pair. Error bars represent the 95% confidence interval for the estimate.

### Genome-wide association study of drug adherence and persistence

We conducted a GWAS of persistence and adherence to each medication, adjusting for age at initiation, year of birth, sex, genotyping batch, and first 10 genetic principal components (**Extended Figure 2**). No variant showed significant association with either phenotypes at a significance level of $P < 1x10^{-8}$ (standard genome-wide significance corrected for multiple testing across the 5 medications). 5 SNPs, however, were genome-wide significant at the standard $P < 5 \times 10^{-8}$ threshold (**Supplementary Table 10**). *rs1339882991* was positively correlated with both persistence and adherence to BP medications and was previously reported to be positively associated with hypertension [26]. *rs111349244* was negatively associated with persistence to BP medication, and, interestingly, associated with higher risk of prolonged pregnancy in FinnGen (OR = 1.12, $P = 8.2x10^{-9}$). The remaining variants did not have significant reported associations in Open Target Genetics [27].

Considering the sample size observed for statins, we have at least 80% power to identify genome-wide significant variants with an effect of greater than 2.7% on adherence, assuming a minor allele frequency of 0.3. Such an effect is in line with what is observed for the strongest

epidemiological risk factor for statins adherence (**Figure 3**). As we expect individual genetic effects for common variants to be smaller than socio-environmental, our results indicate that at current sample size, power to identify genetic associations for adherence and persistence is limited.

**Polygenic overlap between persistence, adherence and other health and behavioral traits**

We explored the genetic overlap between persistence and adherence and 33 major diseases, disease risk factors and behavioral traits (**Figure 5, Supplementary Tables 11,12**). Higher systolic blood pressure was genetically correlated with higher persistence to BP medications (*rg* = 0.77, P = $3.1x10^{-23}$) and with higher adherence to both statins and BP medications (*rg* = 0.17, P = $3.4x10^{-06}$ and *rg* = 0.4, P= $3.5x10^{-30}$, respectively). On the contrary, higher LDL cholesterol was genetically correlated with lower statins adherence (*rg* = -0.14, P = 0.006), although when considering a GWAS of LDL cholesterol adjusted by statins usage the effect was not anymore statistically significant. Notably, both adherence to statins and BP medications were positively genetically correlated with participating in one or more of the optional components of UK Biobank [28] (*rg* between 0.11 and 0.31, P < $8.2x10^{-3}$), indicating some shared effect between drug adherence and continued participation in scientific studies.

We wondered if existing polygenic scores (PGS) could be used to predict adherence and persistence and how their effect compared to that of the health and socio-demographic risk factors examined before. Among the 33 PGS examined (**Extended Figure 3,4** and **Supplementary Table 13,14**) the strongest effects were observed for: systolic blood pressure (SBP) PGS on BP medications persistence (OR = 1.64 for 1 standard deviation (SD) in PGS, 95% CI = 1.60-1.68, P < $2x10^{-308}$ ) and adherence (0.8% increase per 1-SD change in PGS, 95% CI = 0.74-0.86%, P = $2.6x10^{-126}$); LDL cholesterol (adjusted for statins usage) PGS on statins persistence (OR = 1.24, 95% CI = 1.20-1.28, P = $6.2x10^{-47}$) and SBP PGS on statins adherence (0.28% increase, 95% CI = 0.22-0.34%, P = $2x10^{-16}$); coronary artery disease PGS on antiplatelets adherence (0.47% increase, 95% CI = 0.29-0.67%, P = $4.3x10^{-7}$) and PGS for schizophrenia and adherence to breast cancer medications (0.22% decrease, 95% CI = 0.1%-0.34%, P = $4.2x10^{-4}$ ). No significant association was observed between any of the PGS and adherence or persistence to DOAC.

Overall, these PGSs could explain a lower fraction of the variance in adherence than the combined effect of the health and socio-demographic risk factors examined above, after accounting for age and sex (**Supplementary Table 15**).

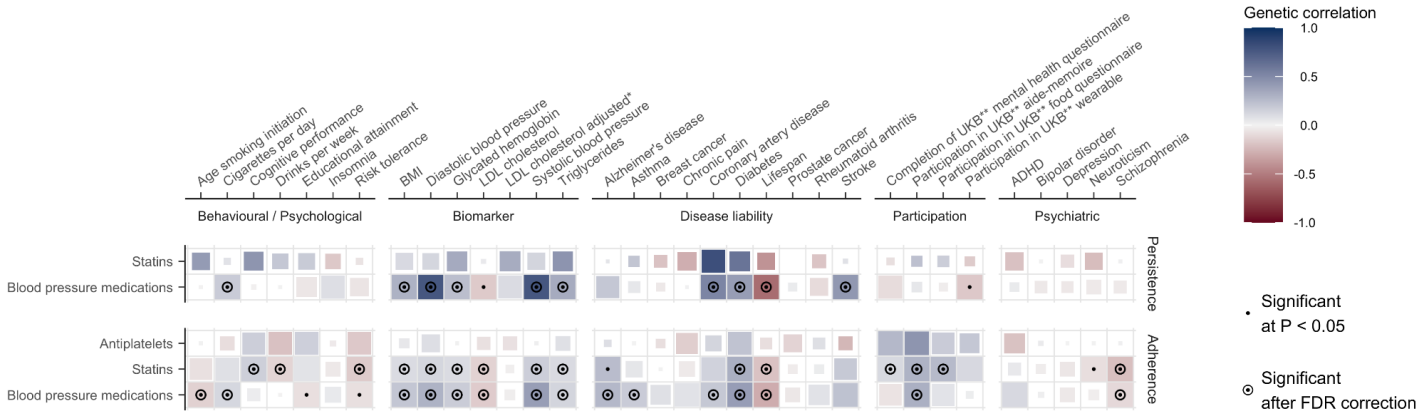



**Figure 5 | Genetic correlates of persistence and adherence.** Genetic correlations between 33 clinically relevant traits and persistence and adherence. Medications for which genetic correlations could not be computed are not displayed. Blue represents positive correlations, red negative. The size of each tile is proportional to the statistical significance (the bigger the tile, the lower the P-value). Correlations nominally significant at P < 0.05 are marked with a dot, while correlations which were significant after FDR correction (within each medication) are additionally marked with a circle.
***** adjusted by the use of cholesterol lowering medication
****** UKB = UK Biobank

### Replication and PGS prediction in Estonian Biobank and sensitivity analysis on persistence definition

We further replicated our adherence analysis on statins in Estonian Biobank [29], a cohort with similar drug purchase data. A total of 18 298 individuals (6.7% of the study) were on statins and eligible for the adherence analysis. The mean adherence was considerably lower (0.54, SD = 0.16) than in the Finnish data (0.89, SD = 0.12) (**Supplementary Table 16**).

First, we evaluated if the association between health and socio-demographic risk factors available in both studies (sex, age at initiation, prevention type, years of education) and adherence was consistent. The direction and magnitude of effects was similar between the two studies, although only age treatment initiation was statistically significant (**Supplementary Table 17**).
Second, we tested whether the GWAS results from FinnGen could be used to build a PGS for statin adherence in Estonian Biobank. We found the PGS to be significantly associated with observed adherence (percentage change in adherence for 1 SD change in PGS = 0.39%, P = $3x10^{-4}$).

Our definition of persistence resulted in higher proportions of persistent individuals compared to what observed in literature [30]. Thus, as sensitivity analysis, we consider a different definition of persistence, looking at the proportion of individuals that stop medication within one year among those with at least one purchase of the medication. This definition includes a broader population (i.e. does not exclude overbuying behaviors) and consider as not persistence those individuals that stop after two and three purchases, as opposed to only one. With this definition,

persistence for statins was 75.6%, consistent with the literature [30]. When looking at the effect of PGS on persistence, we observed directionally consistent effects, but an overall reduction in effect sizes despite the increased number of statistically significant associations (**Supplementary Table 18**, **Supplementary Figure 1**).

## Discussion

Drug adherence remains a major obstacle to the benefits of pharmacotherapies. Although many external factors (e.g. healthcare- and system-related factors) contribute to drug adherence [4,5], focusing on the patient-specific determinants might help in identifying patients at risk of low adherence. In this study we provide a comprehensive exploration of patient-specific factors affecting drug persistence and adherence across 5 different classes of medications, using nation-wide high quality registry data and, for the first time, also exploring genetic factors.

Overall, our results suggest that the need for social assistance and being first-generation immigrants have a moderate to large effect on drug adherence and persistence which is consistent across medication classes. In comparison, demographic characteristics (age and sex), comorbidity burden, and genetic factors have modest or heterogeneous effects. The only health factor with a robust effect was the presence of a disease related to the medication, which substantially increased adherence among those using drugs for secondary prevention.
In line with previous reports [21], the effects of age and sex were heterogeneous across different medications. Nevertheless, females had consistently lower odds of persistence across all medications. We speculate that a higher rate of adverse drug reactions in women compared to men [31,32], or lower perceived risk for cardiometabolic diseases among women can explain these findings.

We performed three main genetic analyses. First, we evaluated if established pharmacogenes for statins, clopidogrel and tamoxifen were associated with persistence and adherence of these drugs. Our results indicate that being a CYP2C9 intermediate metabolizer or a CYP2D6 poor metabolizer reduced by half the odds of persistence. However, as these results were not significant after multiple-testing correction, larger studies should specifically test this hypothesis. Because pharmacogenomics testing is not widespread in Finland, we do not believe these results are biased by genetically-informed drug prescription. Overall, the lack of association between most pharmacogenes and adherence/persistence suggests that widespread pharmacogenetic testing before prescribing the medications included in this study is not going to substantially improve adherence in the population.

Second, we performed a GWAS of drug adherence and persistence and used these results to understand the overlap with other traits. We did not identify robust individual genetic signals but we observed genome-wide overlaps with a proxy for study participation indicating some shared effect between adherence to drug therapy and participation in scientific research, similarly to what we recently observe for COVID-19 vaccination uptake [33]. Our power calculations indicate that even if this is the largest study of this kind, power is still limited to identify genome-wide

significant associations of small effects. A previous study using FinnGen data compared cardiometabolic-related medication user vs non-user or changes in medication use and identified > 40 loci reflecting underlying cardiometabolic risk [34]. The differences in the magnitude of genetic signals and genetic correlation profiles in the current study compared to our previous study indicate that drug adherence and persistence have lower genetic contribution and a stronger behavioral component, which does not necessarily reflect the underlying disease risk.

Third, we explored if existing PGS could predict adherence and persistence and found the strongest association between systolic blood pressure PGS and adherence/persistence to BP medications. This association probably reflects individuals with higher BP being more adherent due to higher perceived risk [35]. Finally, while our results could be used to generate a PGS for statin adherence, which significantly replicated in Estonia Biobank, the variance explained by this PGS was low.

A strength of this study is the use of Finnish registries covering all drug purchases since 1998 which allows an in-depth longitudinal study of drug purchasing. For example, the average treatment length for BP medications was over 10 years. Moreover, by using nation-wide data, we could confirm that adherence and persistence measures in FinnGen were similar to what was observed in the general population. Importantly, this study considers both socio-economic and genetic risk factors, helping to better contextualize the results and bridging the divide between social science and human genetics, as advocated by others [36]. Furthermore, we derived adherence and persistence measures from 5 different medication classes, which allowed us to capture adherence-related factors beyond single disease- and drug-related factors. Finally, both epidemiological and genetic results were replicated in an external study.

A limitation of the study is that our definitions of persistence and adherence can be biased in different ways. First, our persistence definition resulted in higher proportions of persistent individuals compared to what observed in literature[30]. Nonetheless, we chose this definition to better capture early discontinuation behaviors (i.e. after the first purchase). Moreover, we deemed it fair to compare these individuals only to those included in the adherence analysis, so as not to include overbuying behaviors. Sensitivity analyses suggest that our results are robust to a different definition of persistence and that our stricter inclusion criteria resulted in larger effect sizes at the expenses of decreased statistical power.

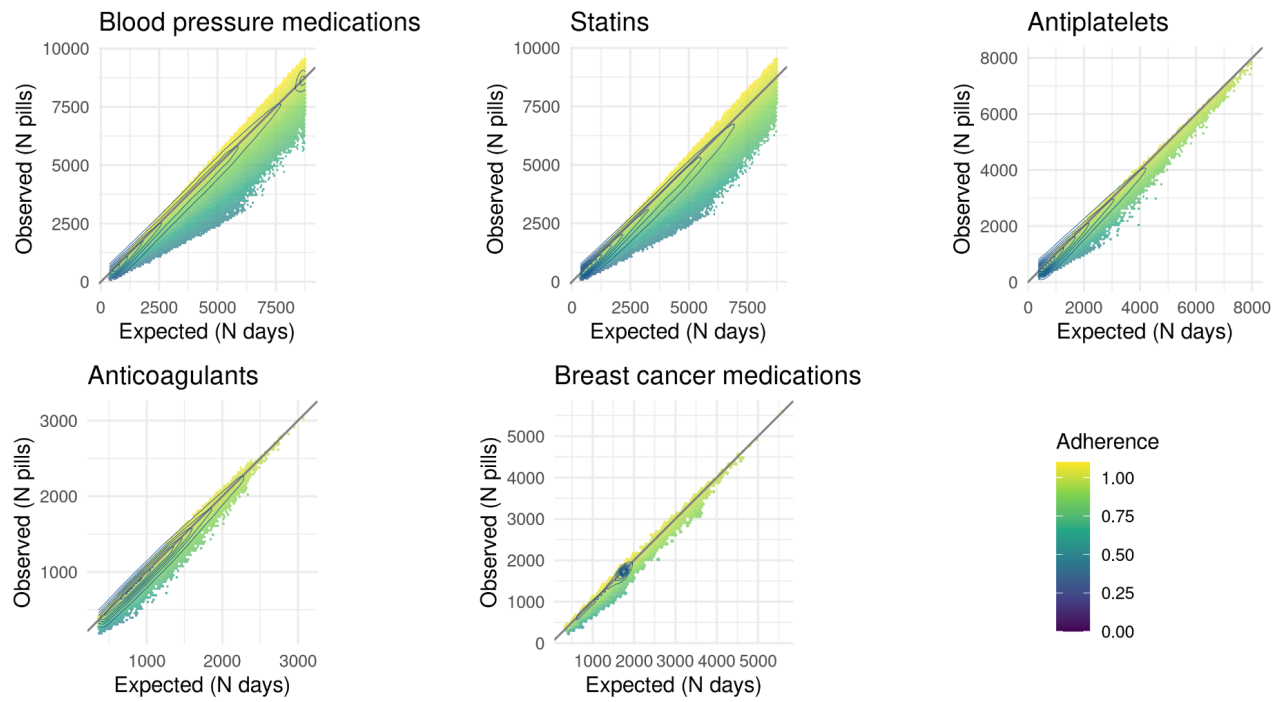
As for the adherence definition, first we cannot know if individuals are really taking medications and instead rely on the expected and observed consumption. Second, we don't have information on the dose prescribed by the doctor for the entire period and rely on the suggested dosage of the package. For example, individuals which are suggested by the doctor to take half-tablet of statins/day despite package indication, will result in low adherence. Two results indicate that this behavior is not common. First, we do not see enrichment in adherence values that would capture a systematic indication of non-standard dose of the medication. Second, since 2015 prescription information is available as text. Using regular expressions to extract this information (**Methods**) we found that in 2019, 96% of the prescriptions were indicating 1 statin tablet/day. Finally, the observed adherence is higher than what was reported from a previous

Finnish study [37]. Considering only individuals that had taken medication for at least 12 months and excluding long inactive periods might have contributed to higher adherence values. Nonetheless, early discontinuation is captured by our persistence measure, while long breaks in drug usage might reflect hospital admission, partial emigration, or other external factors. Overall, we used a definition of adherence that is more robust to large external shocks and better suited to capture smaller and consistent variations in drug purchases over longer time periods.

In conclusion, our results suggest a limited role of genetic risk factors, including clinically relevant pharmacogenes, in explaining persistence and adherence for 5 common medication classes that require long-term, regular therapy. While larger GWASs are needed to identify smaller genetic effects, the observed genetic overlap with behavioral traits, might complicate the interpretation of the results. The robust associations with measures of social deprivation suggest that we need better understanding of the drivers of poor adherence in these socio-economically disadvantaged groups. Identifying these drivers might offer points of intervention in clinical practice, or help recognizing system-wide barriers for drug adherence.

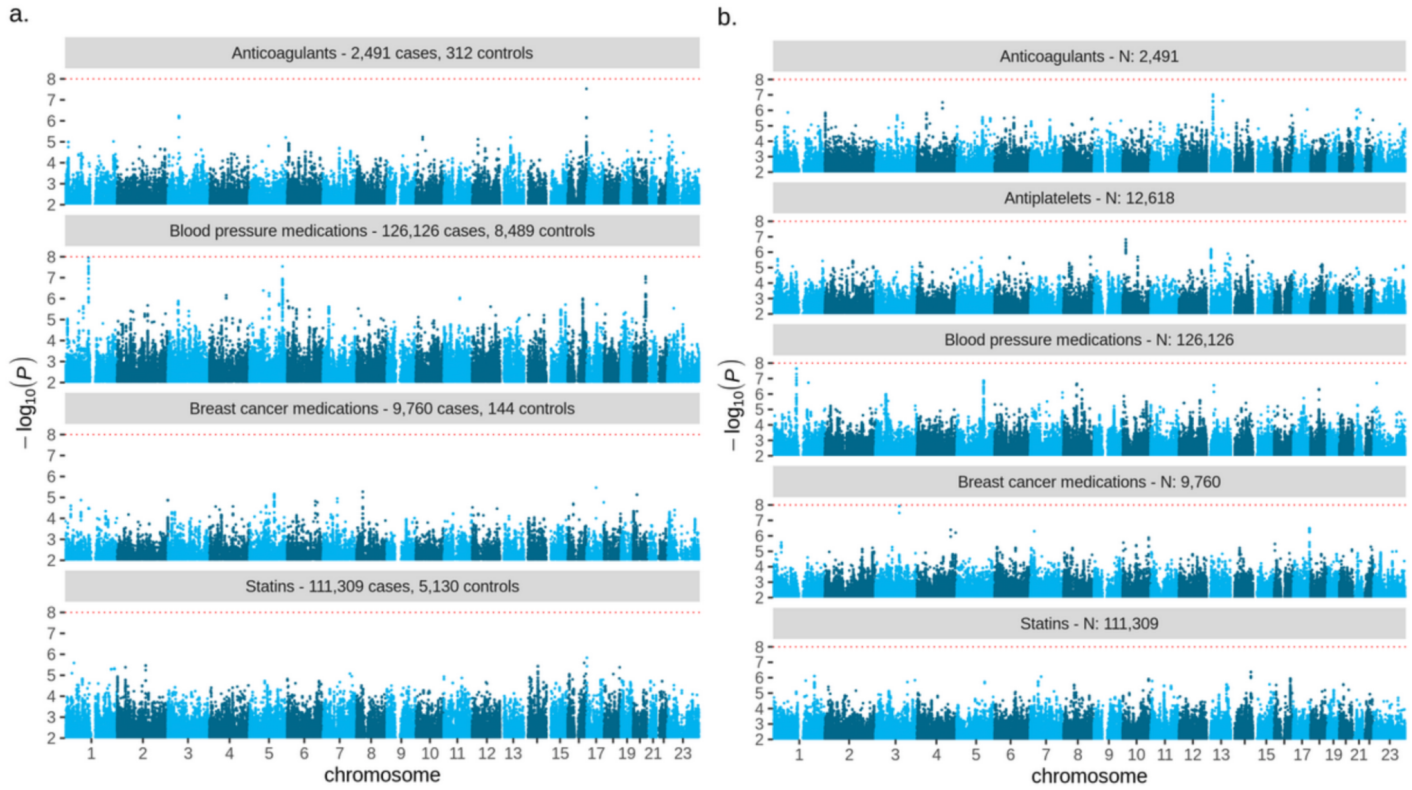
## Extended Figures

### Extended Figure 1



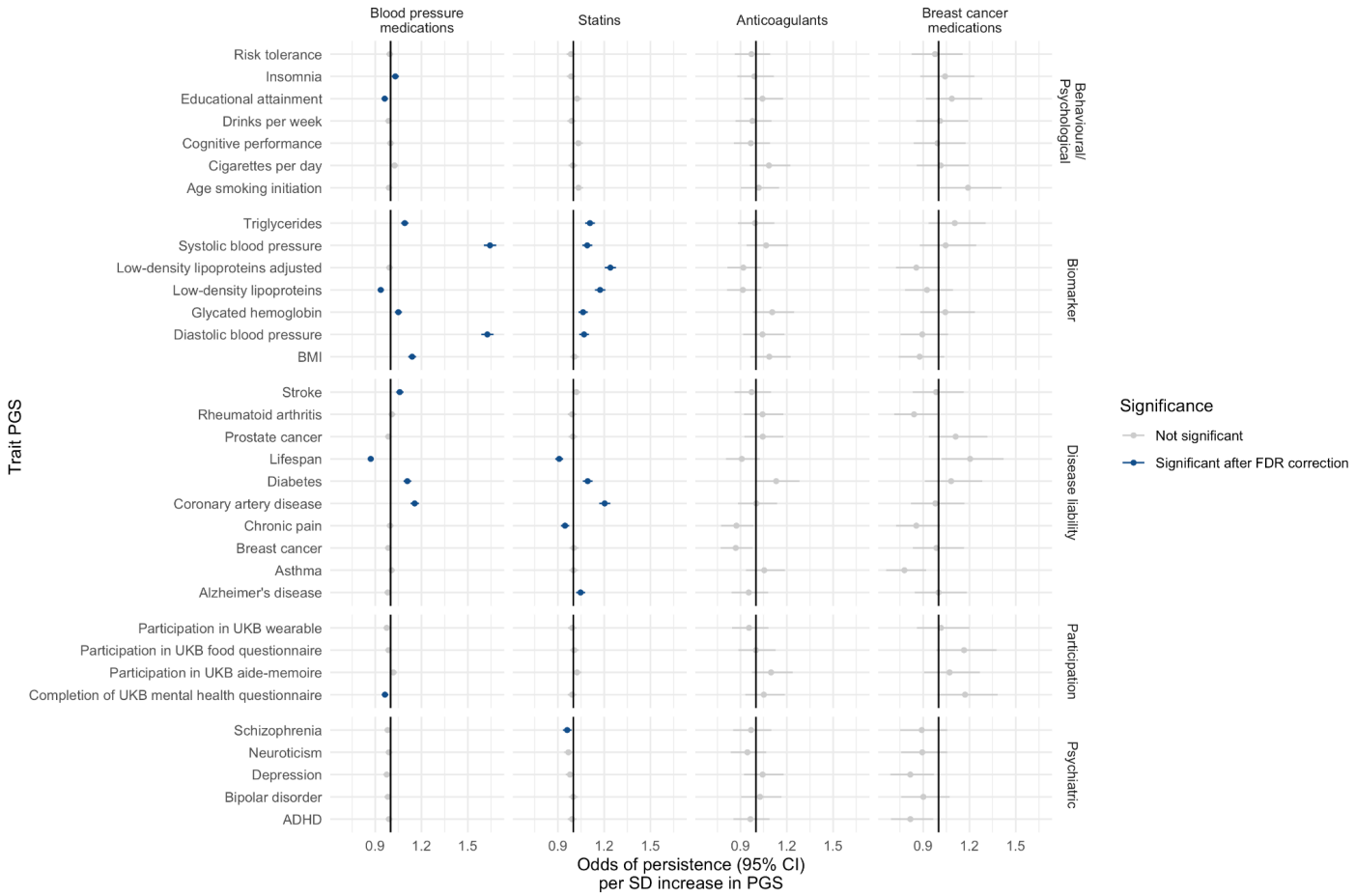

**Extended Figure 1 | Adherence as function of expected and observed consumption.** Hexagonal heatmap representing mean adherence levels at the varying of expected (number of days, x-axis) versus observed consumption (number of tablets, y-axis). Bins with less than 5 individuals are not displayed. The level lines represent the 2-dimensional density, each line including 10% of the individuals in the data.

## Extended Figure 2



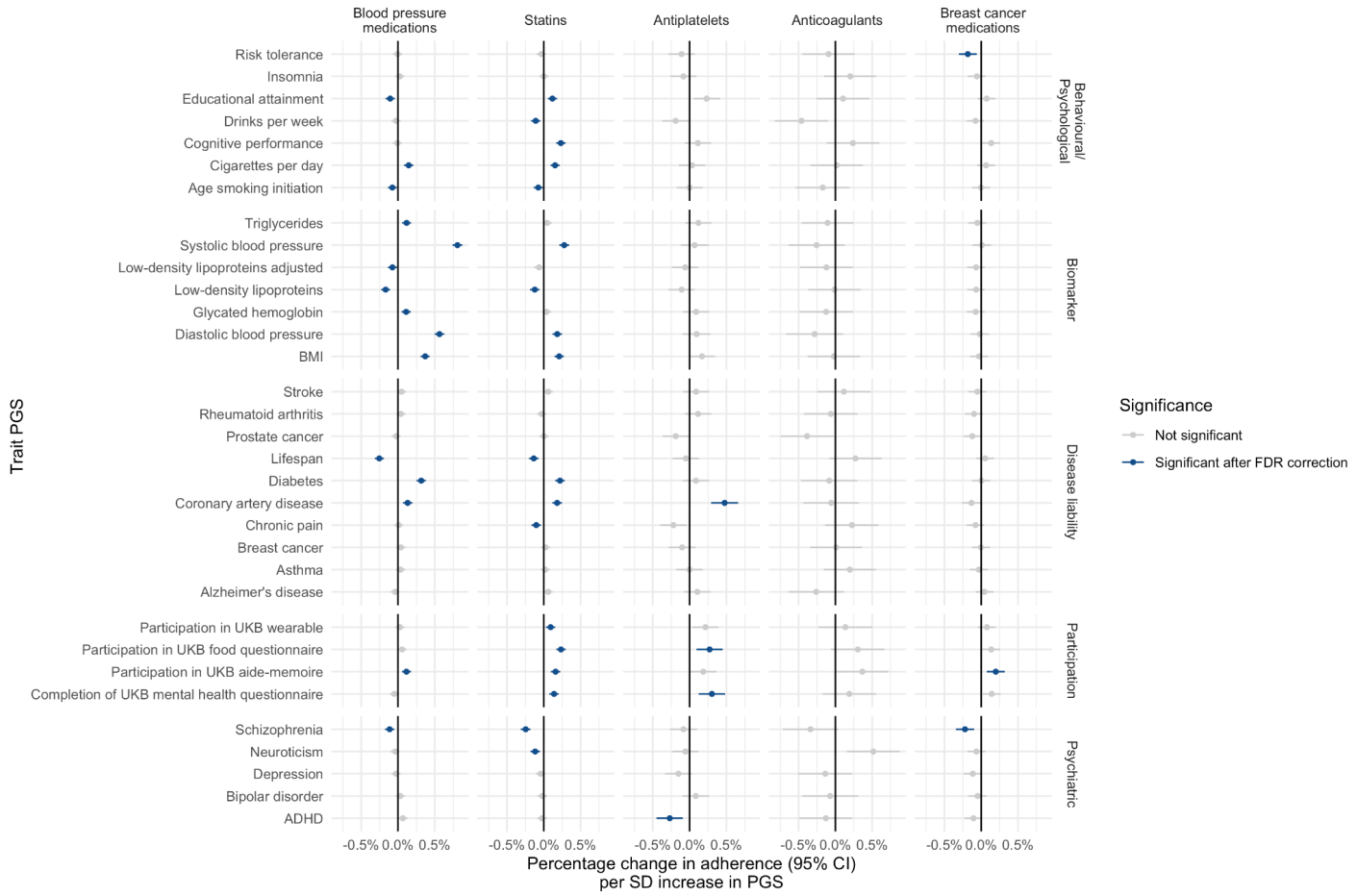

**Extended Figure 2 | Genome-wide association results for persistence and adherence. a.** results of genome-wide association study of persistence to each medication. Number of cases and controls are reported next to each drug name. Red dotted line represents genome-wide significance at P-value = 1x10-8 (Bonferroni corrected for 5 medications). b. results of genome-wide association study of adherence to each medication. Total sample size is reported next to each drug name. Red dotted line represents genome-wide significance at P-value = 1x10-8 (Bonferroni corrected for 5 medications). Results for 5 variants with significance < 5x10-8 are reported in Supplementary Table 10.

## Extended Figure 3




**Extended Figure 3 | Associations between persistence and PGS for 33 clinically relevant traits.** Odds of persistence per 1-SD increase in trait PGS. ORs are from a logistic regression model adjusted for sex, age at initiation, first 10 genetic principal components. Error bars represent the 95% confidence interval for the estimates. Results are reported in Supplementary Table 13.

## Extended Figure 4




**Extended Figure 4 | Associations between adherence and PGS for 33 clinically relevant traits.** Percentage change in adherence per 1-SD increase in trait PGS. Coefficients are from a linear regression model adjusted for sex, age at initiation, first 10 genetic principal components. Error bars represent the 95% confidence interval for the estimates. Results are reported in Supplementary Table 14.

## Methods

### Study population.

The FinRegistry dataset [19] contains data from Finnish nationwide health, demographic, and socioeconomic registries, for a total of 7 166 416 individuals. To ensure completeness of information for the variables considered in the epidemiological analysis, we included only individuals who were residents in Finland and alive on 1 January 2010 (5 339 804 individuals in total). The FinnGen dataset [20] additionally contains genetic data for a subset of the Finnish population (N = 430 885 for data freeze 10). We studied the two drug usage phenotypes in adults only, meaning individuals needed to be at least 18 years old at treatment initiation.

### Kela drug purchase register.

The Social Insurance Institution of Finland (Kela) is a government agency that provides basic economic security by financial support for everyone living in Finland. The Kela drug purchase register contains information about any reimbursable prescription medication bought from pharmacies in Finland between 1995 and 2020. We included only purchases from 1998 on to ensure better homogeneity with the data from the other registries. Each data entry contains information about the date of purchase, ATC code of the medication, number of purchased packages, and the Nordic Article Number (VNR) associated with the package, which gives information about the quantity and dosage of the medication. When the information was available, we additionally excluded those purchases made using pharmacies dose distribution services, which regularly provides patients with 2 weeks worth of medications at a time.

### Medications included.

We included medications commonly prescribed and purchased, that are typically used for a prolonged time and do not have a significant direct effect on wellbeing. The classes of medication we included are as follows: statins (all types of statins without differentiation), blood pressure medications (antihypertensives, diuretics, calcium channels blockers, agents acting on the renin-angiotensin system), antiplatelets (clopidogrel and acetylsalicylic acid in combination with dipyridamole), breast cancer medications (tamoxifen, anastrozole, letrozole, exemestane), direct oral anticoagulants (DOAC) (rivaroxaban, apixaban, edoxaban, caplacizumab). For breast cancer medications we included only females, starting the treatment after a breast cancer diagnosis. For DOAC, only patients starting the treatment after an atrial fibrillation diagnosis were included.

### Persistence.

We define persistence [17] as a binary phenotype, comparing individuals taking the medication for at least 12 months (i.e. all individuals included in the adherence analysis) versus patients early-discontinuing the drug after one purchase. To avoid right censoring, we considered as early discontinuation only purchases made at least 2 years before the end of follow-up. End of follow-

up date was either death date, moving abroad date or the last date for which the data are available (01.01.2020).

**Adherence.**

For each individual, we considered all registered purchases for each class of medication and reconstructed individual drug purchasing trajectories. We then estimated adherence using the Medication Possession Ratio measure (MPR) [18]:

$$MPR = \frac{days\ of\ supply\ during\ observation\ period}{days\ in\ observation\ period}$$

where

- *days of supply during observation period* is the total number of tablets purchased in the trajectory, normalized assuming drug-specific daily doses (e.g. statins: 1 tablet/day)
- *days in observation period* is the sum of the intervals (in days) between all purchases considered in the trajectory

When reconstructing the trajectories, we excluded the last purchase and purchases for which the interval to the next one was of at least 150 days without tablets available. We considered this as a break in the treatment which should not be taken into account into the adherence calculation. When patients were switching to different formulations (within the same medication class) and the interval between purchases was shorter than the drug supply, we discarded the leftover tablets of the previous formulation. We included only individuals with at least 1 year of purchases. We further excluded individuals with adherence greater than 1.1, to exclude overbuying behaviors.

**Health, demographic and socioeconomic risk factors for persistence and adherence.**

We assessed the effect of the following 8 epidemiological risk factors on persistence and adherence:

- sex (not considered for breast cancer medications)
- age at treatment initiation: age when the first purchase was recorded
- secondary prevention: whether the treatment was started after a major event related to the medication considered. The following diagnosis (ICD-10 codes) were considered:
    - statins: major coronary heart disease (I20.0,I21,I22), cerebrovascular diseases (I60-I69), atherosclerosis (excluding cerebral, coronary and PAD) (I70)
    - antiplatelets: stroke (I60-I64), transient ischemic attack (G45), myocardial infarction (I21,I22)
    - BP medications: no specific outcome was considered, as this is a broader class of medications
    - breast cancer medications: breast cancer (C50) - only secondary prevention patients were included

    - DOAC: atrial fibrillation (I48), pulmonary embolism or venous thromboembolism (I26,I80,I87.1) - only secondary prevention patients were included
- Charlson comorbidity index (at treatment initiation): we considered patients comorbidity status at treatment initiation by computing the Charlson comorbidity index (CCI) based on all the diagnoses registered previous to the first purchase. This was computed using the R package ICCI (https://github.com/dsgelab/ICCI), which uses the R package comorbidity (https://cran.r-project.org/web/packages/comorbidity/index.html).
- years of education (at treatment initiation): years expected to complete the recorded education level. When the information at treatment baseline was not available, the first available record was considered.
- living area (at treatment initiation): urban *vs* rural, as classified by the Finnish digital and population data agency (DVV).
- mother tongue: Finnish or Swedish (the two official languages of Finland) *vs* any other language
- social assistance benefits: receiving (or not) any amount of social assistance benefit in the year before treatment initiation

The effect sizes of each factor on persistence and adherence were derived by using a multivariate logistic/linear regression model including all the factors, adjusting for year of birth.

For persistence, the estimated effects are reported as OR from the logistic regression model, while for adherence we reported the percentage change in adherence by rescaling the log(OR) from the linear model by a factor of 1.1 (maximum adherence value). To provide a better interpretation of the results, in Figure 3 we report the effect sizes for a model where continuous variables were dichotomized according to the following thresholds: age at initiation > 60; CCI > 5; education, university degree or higher. The same effect sizes keeping the continuous variables continuous are reported in Supplementary Table 5-6.

**Pharmacogenes selection and phenotypes definition.**

We selected the pharmacogenes to test based on the guidelines provided by the Clinical Pharmacogenetics Implementation Consortium [22]. The CPIC assigns levels to gene/drug pairs based on PharmGKB [38] annotation levels, recommendations by external groups (e.g. FDA, EMA) or by other professional societies. We selected gene/drug pairs to match all the ATC codes included in our analyses, and with a CPIC level A or B, meaning that prescribing actions (e.g. alternative therapies or dosing) are recommended when the pharmacogene is present. The gene/drug pairs we selected are: *ABCG2 for rosuvastatins, CYP2C9 for fluvastatins, SLCO1B1* for statins (all types), *CYP2C19* for clopidogrel, and *CYP2D6* for tamoxifen.
For each gene and for each individual included in the analyses, we defined their star alleles diplotypes to test starting from the imputed phased genotypes available in FinnGen and using Stargazer [39,40]. Having genotype data available, we could define only diplotypes captured by SNP and not by structural variations (gene deletions or copy number variations). We then grouped individuals based on the phenotypes also provided as output by Stargazer and reflecting the metabolic activity score associated to each specific diplotype (poor/decreased/increased function for SLCO1B1 and ABCG2, poor/intermediate/ultrarapid

metabolizer for CYP2D6, CYP2C9, CYP2C19). We used a logistic/linear regression model to estimate the association of each phenotype with persistence/adherence, adjusting by sex, age at initiation and first 10 principal genetic components. For *SLCO1B1* and statins we considered all statins users, independent of the type of statins, but only rosuvastatins and fluvastatins users for ABCG2 and CYP2D9. Similarly, for *CYP2C19* we considered the subset of antiplatelets users taking clopidogrel, and for *CYP2D6* only breast cancer medications users taking tamoxifen. We only tested phenotypes when defined for at least 5 individuals in the adherence analysis, or for at least 5 persistent/non-persistent individuals in the persistence analysis.

**Genetic risk factors of persistence and adherence.**

We used REGENIE [41] to run a GWAS of persistence and adherence to each medication (see **Supplementary Methods** for details about genotyping and imputation in FinnGen). Persistence was treated as a binary phenotype, while adherence was kept as continuous trait, but scaled so to have mean 0 and variance equal to 1. For both analyses we adjusted for sex, genotyping batch, first 10 principal genetic components, age at first purchase and year of birth. The resulting summary statistics were post-processed to keep only SNPs with INFO score ≥ 0.8 and MAF ≥ 1%.

Genetic correlations between persistence and adherence and 33 publicly available traits (listed in **Supplementary Table 11**) were calculated using linkage disequilibrium score regression [42].

PGSs for the same traits were derived using PRS-CS [43]. Their association with persistence and adherence was assessed by running a logistic/linear regression model for each PGS individually, adjusting for sex and age at initiation.

**Replication and PGS prediction in Estonian Biobank.**

The study population in the Estonian Biobank (**Supplementary Methods**) cohorts consisted of individuals that initiated their statin treatment between 2007-2020. The patients were included if they were over the age of 18 at the start of treatment initiation and if they had used statins for at least 1.5 years. The 1.5 year treatment length was based on either having at least 2 purchases of statins within the first calendar year of the treatment and at least one purchase in the following year, or having one purchase at the end of the first year (180 days before end of the calendar year) and at least 2 purchases in the second year. Individual purchase trajectories were reconstructed and adherence was calculated using the same criteria specified above.

Among the health, demographic and socioeconomic factors considered in FinRegistry, the following were available in Estonian Biobank too: sex, age at treatment initiation, type of prevention, years of education. We assessed the association between adherence and these

factors by using a multivariate linear regression model, only considering individuals with complete information for all variables (N = 12 096).

We derived a PGS for adherence to statins from the summary statistics of the GWAS in FinnGen, using PRS-CS. We then tested the association between this PGS and adherence observed in Estonian Biobank using a linear regression model, adjusting for sex, age at initiation and first 10 genetic principal components.

**Acknowledgements**

We are grateful to Finnish individuals, whose data made this study possible. We would also like to thank the entire FinRegistry team for making the data available for the study.

We want to acknowledge the participants and investigators of FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta), Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/) and Arctic Biobank (https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi). Finnish Biobank Cooperative -FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (https://site.fingenious.fi/en/) managed by FINBB.

This study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016775, from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant number 945733) and from Academy of Finland fellowship grant N. 323116.

**Data and code availability**

Data dictionaries for FinRegistry are publicly available on the FinRegistry website (www.finregistry.fi/finnish-registry-data). Access to FinRegistry data can be obtained by submitting a data permit application for individual-level data for the Finnish social and health data permit authority Findata ( https://asiointi.findata.fi/). The application includes information on the purpose of data use; the requested data, including the variables, definitions for the target and control groups, and external datasets to be combined with FinRegistry data; the dates of the data needed; and a data utilization plan. The requests are evaluated on a case-by-case basis. Once approved, the data are sent to a secure computing environment Kapseli and can be accessed within the European Economic Area (EEA) and countries with an adequacy decision from the European Commission.

Analysis code used to produce the results is available on GitHub at: https://github.com/dsgelab/drugs-persistence-adherence

**Ethics declarations**

FinRegistry is a collaboration project of the Finnish Institute for Health and Welfare (THL) and the Data Science Genetic Epidemiology research group at the Institute for Molecular Medicine Finland (FIMM), University of Helsinki. The FinRegistry project has received the following approvals for data access from the National Institute of Health and Welfare (THL/1776/6.02.00/2019 and subsequent amendments), DVV (VRK/5722/2019-2), Finnish Center for Pension (ETK/SUTI 22003) and Statistics Finland (TK-53-1451-19). The FinRegistry project has received IRB approval from the National Institute of Health and Welfare (Kokous 7/2019).

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.

The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020, THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/2021, Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20)

TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4th July 2019.

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 10 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, BB2021_65, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, HUS/150/2022 § 12, §13, §14, §15, §16, §17, §18, and §23, Auria Biobank AB17-5154 and amendment #1 (August 17 2020) and amendments BB_2021-0140, BB_2021-0156 (August 26 2021, Feb 2 2022), BB_2021-0169, BB_2021-0179, BB_2021-0161, AB20-5926 and amendment #1 (April 23 2020)and it´s modification (Sep 22 2021), Biobank Borealis of Northern Finland_2017_1013, 2021_5010, 2021_5018, 2021_5015, 2021_5023, 2021_5017, 2022_6001, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, 53§/2021, 13§/2022, 14§/2022, 15§/2022, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), §8/2021, §9/2022, §10/2022, §12/2022, §20/2022, §21/2022, §22/2022, §23/2022, Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020, Finnish Hematological Registry and Clinical Biobank decision 18th June 2021, Arctic biobank P0844: ARC_2021_1001.

## Authors contribution

Study design: MC, AC, SJ, KL, MN, LM and AG. Data analysis: MC, AC and HMK. Results interpretation: MC, AC, HMK, SJ, PK, TK, MF, SR, KL, MN, LM, AG. Writing – original draft: MC, AC and AG. All authors were involved in further drafts of the manuscript and revised it critically for content. All authors gave final approval of the version to be published. MC and AC contributed equally to this work.

**Conflicts of interest:** None declared.

## References


1. Osterberg, L. & Blaschke, T. Adherence to medication. *N. Engl. J. Med.* **353**, 487–497 (2005).
2. Simpson, R. J., Jr & Mendys, P. The effects of adherence and persistence on clinical outcomes in patients treated with statins: a systematic review. *J. Clin. Lipidol.* **4**, 462–471 (2010).
3. Miller, N. H., Hill, M., Kottke, T. & Ockene, I. S. The multilevel compliance challenge: recommendations for a call to action. A statement for healthcare professionals. *Circulation* **95**, 1085–1090 (1997).
4. Kardas, P., Lewek, P. & Matyjaszczyk, M. Determinants of patient adherence: a review of systematic reviews. *Front. Pharmacol.* **4**, 91 (2013).
5. Sabaté, E. *Adherence to Long-term Therapies: Evidence for Action*. (World Health Organization, 2003).

6. Simpson, S. H. *et al.* A meta-analysis of the association between adherence to drug therapy and mortality. *BMJ* **333**, 15 (2006).
7. Chan, D. C. *et al.* Patient, physician, and payment predictors of statin adherence. *Med. Care* **48**, 196–202 (2010).
8. Citarella, A. *et al.* Family history of cardiovascular disease and influence on statin therapy persistence. *Eur. J. Clin. Pharmacol.* **70**, 701–707 (2014).
9. Shin, D. W. *et al.* Antihypertensive medication adherence in cancer survivors and its affecting factors: results of a Korean population-based study. *Support. Care Cancer* **19**, 211–220 (2010).
10. Yang, Y. *et al.* Predictors of medication nonadherence among patients with diabetes in Medicare Part D programs: a retrospective cohort study. *Clin. Ther.* **31**, 2178–88; discussion 2150–1 (2009).
11. Rolnick, S. J., Pawloski, P. A., Hedblom, B. D., Asche, S. E. & Bruzek, R. J. Patient characteristics associated with medication adherence. *Clin. Med. Res.* **11**, 54–65 (2013).
12. Yeaw, J., Benner, J. S., Walt, J. G., Sian, S. & Smith, D. B. Comparing adherence and persistence across 6 chronic medication classes. *J. Manag. Care Pharm.* **15**, 728–740 (2009).
13. Pirmohamed, M. Pharmacogenomics: current status and future perspectives. *Nat. Rev. Genet.* **24**, 350–362 (2023).
14. Swen, J. J. *et al.* A 12-gene pharmacogenetic panel to prevent adverse drug reactions: an open-label, multicentre, controlled, cluster-randomised crossover implementation study. *Lancet* **401**, 347–356 (2023).
15. Hess, L. M., Raebel, M. A., Conner, D. A. & Malone, D. C. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. *Ann. Pharmacother.* **40**, 1280–1288 (2006).
16. Vrijens, B. *et al.* A new taxonomy for describing and defining adherence to medications. *Br. J. Clin. Pharmacol.* **73**, 691–705 (2012).
17. Cramer, J. A. *et al.* Medication compliance and persistence: terminology and definitions. *Value Health* **11**, 44–47 (2008).
18. Andrade, S. E., Kahler, K. H., Frech, F. & Chan, K. A. Methods for evaluation of medication adherence and persistence using automated databases. *Pharmacoepidemiol. Drug Saf.* **15**, 565–74;

discussion 575–7 (2006).

19. Viippola, E. *et al.* Data Resource Profile: Nationwide registry data for high-throughput epidemiology and machine learning (FinRegistry). *Int. J. Epidemiol.* (2023) doi:10.1093/ije/dyad091.
20. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
21. DiMatteo, M. R. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Med. Care* **42**, 200–209 (2004).
22. Relling, M. V. & Klein, T. E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin. Pharmacol. Ther.* **89**, 464–467 (2011).
23. Cooper-DeHoff, R. M. *et al.* The Clinical Pharmacogenetics Implementation Consortium Guideline for SLCO1B1, ABCG2, and CYP2C9 genotypes and Statin-Associated Musculoskeletal Symptoms. *Clin. Pharmacol. Ther.* **111**, 1007–1021 (2022).
24. Lee, C. R. *et al.* Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2C19 Genotype and Clopidogrel Therapy: 2022 Update. *Clin. Pharmacol. Ther.* **112**, 959–967 (2022).
25. Goetz, M. P. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and Tamoxifen Therapy. *Clin. Pharmacol. Ther.* **103**, 770–777 (2018).
26. Dönertaş, H. M., Fabian, D. K., Valenzuela, M. F., Partridge, L. & Thornton, J. M. Common genetic associations between age-related diseases. *Nat Aging* **1**, 400–412 (2021).
27. Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
28. Tyrrell, J. *et al.* Genetic predictors of participation in optional components of UK Biobank. *Nat. Commun.* **12**, 886 (2021).
29. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).
30. Helin-Salmivaara, A. *et al.* Long-term persistence with statin therapy: a nationwide register study in Finland. *Clin. Ther.* **30 Pt 2**, 2228–2240 (2008).
31. Zucker, I. & Prendergast, B. J. Sex differences in pharmacokinetics predict adverse drug reactions in women. *Biol. Sex Differ.* **11**, 32 (2020).

32. Brabete, A. C. *et al.* A Sex- and Gender-Based Analysis of Adverse Drug Reactions: A Scoping Review of Pharmacovigilance Databases. *Pharmaceuticals* **15**, (2022).
33. Hartonen, T. *et al.* Nationwide health, socio-economic and genetic predictors of COVID-19 vaccination status in Finland. *Nat Hum Behav* **7**, 1069–1083 (2023).
34. Kiiskinen, T. *et al.* Genetic predictors of lifelong medication-use patterns in cardiometabolic diseases. *Nat. Med.* **29**, 209–218 (2023).
35. Shiraly, R., Khani Jeihooni, A. & Bakhshizadeh Shirazi, R. Perception of risk of hypertension related complications and adherence to antihypertensive drugs: a primary healthcare based cross-sectional study. *BMC Prim Care* **23**, 303 (2022).
36. Harden, K. P. & Koellinger, P. D. Using genetics for social science. *Nat Hum Behav* **4**, 567–576 (2020).
37. Aarnio, E. J. *et al.* Register-based predictors of adherence among new statin users in Finland. *J. Clin. Lipidol.* **8**, 117–125 (2014).
38. Whirl-Carrillo, M. *et al.* An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin. Pharmacol. Ther.* **110**, 563–572 (2021).
39. Lee, S.-B. *et al.* Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genet. Med.* **21**, 361–372 (2019).
40. Lee, S.-B., Wheeler, M. M., Thummel, K. E. & Nickerson, D. A. Calling Star Alleles With Stargazer in 28 Pharmacogenes With Whole Genome Sequences. *Clin. Pharmacol. Ther.* **106**, 1328–1337 (2019).
41. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
42. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
43. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).