

OPEN ACCESS

EDITED BY

Chutian Zhang,
Northwest A&F University, China

REVIEWED BY

Li Li,
Southern Medical University, China
Feng Jin,
Shunyi Maternal and Children's Hospital
of Beijing Children's Hospital, China

*CORRESPONDENCE

Pärt Peterson
✉ part.peterson@ut.ee

RECEIVED 19 March 2026

REVISED 09 June 2026

ACCEPTED 18 June 2026

PUBLISHED 16 July 2026

CITATION

Sõnajalg H, Vanker M, Läll K, Möls M,
Alver M, Milani L, Kisand K and
Peterson P (2026) Bidirectional
associations between chronic diseases
and Long COVID: a population-based
cohort and interrupted time-series
analysis.

Front. Public Health 14:1834448.
doi: 10.3389/fpubh.2026.1834448

COPYRIGHT

© 2026 Sõnajalg, Vanker, Läll, Möls,
Alver, Milani, Kisand and Peterson. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Bidirectional associations between chronic diseases and Long COVID: a population-based cohort and interrupted time-series analysis

Hanna Sõnajalg^{1,2}, Martti Vanker³, Kristi Läll¹, Märt Möls^{1,2},
Maris Alver¹, Lili Milani¹, Kai Kisand³ and Pärt Peterson^{3*}, On
behalf of the Estonian Biobank Research Team

¹Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia, ²Institute of
Mathematics and Statistics, University of Tartu, Tartu, Estonia, ³Institute of Biomedicine and
Translational Medicine, University of Tartu, Tartu, Estonia

Background: Long COVID has been associated with an increased incidence of chronic diseases, including type 2 diabetes (T2D), which has been suggested to have a bidirectional relationship with COVID-19. We examined these associations in a population-based cohort.

Methods: Chronic non-communicable diseases (NCDs) diagnosed before (2010–2019) or after (2021–2023) the COVID-19 pandemic onset were tested for associations with Long COVID among Estonian Biobank participants ($N = 2,639$ with Long COVID, $N = 199,171$ controls) using logistic regression. Interrupted time-series analysis using a SARIMA model assessed changes in monthly T2D incidence after March 2020. A polygenic risk score (PRS) for severe COVID-19 was examined for association with new-onset T2D.

Results: Pre-existing NCDs were associated with Long COVID, and individuals with Long COVID had increased odds of first-recorded NCD diagnoses after 2020. Long COVID was associated with COPD (OR = 2.33), asthma (OR = 2.18), and chronic kidney disease (OR = 2.15). It was also associated with higher odds of incident asthma (OR = 5.12), COPD (OR = 4.08), and sleep apnea (OR = 3.05), while the association with T2D did not surpass Bonferroni correction. However, T2D incidence increased after March 2020 beyond expected trends ($p = 1.84 \times 10^{-80}$). Individuals in the highest PRS quantile for severe COVID-19 had increased odds of T2D (OR = 1.19).

Conclusion: The findings support bidirectional associations between COVID-19 and NCDs, including T2D. Pre-existing T2D was associated with increased odds of Long COVID, while the pandemic period showed increasing trends in T2D incidence.

KEYWORDS

asthma, COPD, Long COVID, polygenic risk score (PRS), type 2 diabetes

Introduction

Recent studies estimate that a substantial proportion of individuals, ranging from 10–30% of non-hospitalized and 50–70% of hospitalized cases, develop post-acute sequelae of SARS-CoV-2 infection, commonly referred to as Long COVID, after recovering from the initial illness (1–3). Among the various complications associated with Long COVID, cardiovascular diseases (4), pre-existing chronic respiratory diseases such as asthma and chronic obstructive pulmonary disease (COPD) (5) and chronic kidney disease (CKD) have been associated with a higher risk of Long COVID (6). Among these conditions, metabolic diseases, and particularly type 2 diabetes (T2D), have received increasing attention due to their bidirectional relationship with COVID-19 outcomes.

Pre-existing T2D is consistently associated with worse outcomes after SARS-CoV-2 infection, including increased hospitalization, intensive care unit admission, and mortality (7–12). Individuals with T2D also appear to have a higher risk of developing Long COVID; a UK cohort study reported a greater likelihood of persistent Long COVID symptoms among people with T2D compared to those without diabetes (13). Interestingly, SARS-CoV-2 infection has also been linked to a higher probability of T2D remission in individuals recently diagnosed with the disease (14). Beyond its impact on pre-existing diabetes, COVID-19 has been associated with increased incidence of new-onset T2D. In a cohort of 629,000 individuals tested for SARS-CoV-2, diabetes incidence showed a 3–5% excess burden over a median follow-up of 257 days (15). Meta-analyses report a significantly elevated diabetes risk ($RR = 1.62$ – 1.64) following COVID-19 compared to non-infected controls (7, 16, 17). Similarly, a Korean population-based study found a 30% higher risk of new-onset diabetes after COVID-19 (18), and a large English cohort of 16 million individuals demonstrated increased T2D incidence among those hospitalized with COVID-19 (19). These findings are consistent with a relationship between COVID-19 and T2D, however, most studies focus on short-term outcomes, and data on the long-term burden of T2D and its complications after COVID-19 remain limited.

In this study, we investigated the role of chronic diseases, including T2D, as risk factors for Long COVID using the Estonian Biobank (EstBB), comprising 210,000 individuals with linked health records and genotyping data. We further analyzed whether the incidence of chronic diseases associated with severe SARS-CoV-2 infection, such as T2D, increased during the pandemic period (2020–2023).

Methods

Study participants

The study was performed using data from EstBB, a volunteer-based biobank of the Estonian adult population (≥ 18 years old), with a current cohort size of more than 210,000 individuals and representing the Estonian adult population in terms of age, sex, and geographic distribution. EstBB is regularly linked to national registries, hospital databases, and the national Health Insurance Fund, which provide comprehensive information on reimbursed healthcare services at the national level. Further details about the EstBB cohort and genotyping procedures have been published elsewhere (20). The study focused on individuals with the Long COVID condition or post-viral fatigue

syndrome (ICD-10 codes U09/U09.9 or G93.3, respectively), defined between June 2020 and December 2023. The ICD-10 code G93.3 was included because it was frequently used to code chronic fatigue-like symptoms that many COVID-19 patients experienced during long-term recovery. The controls were EstBB individuals without a Long COVID diagnosis, however, the majority were likely exposed to SARS-CoV-2 during the pandemic period between March 2020 and 2023. These individuals were analyzed for 12 non-communicable diseases (NCD) represented by 28 ICD-10 codes documented in EstBB (Supplementary Tables S1, S2).

Statistical analysis

Association between pre-existing and new-onset NCDs and Long COVID

Logistic regression was used to test associations involving the 12 NCD groups separately, as well as any of the diseases combined, adjusting for birth year and sex as covariates. For analyses of pre-existing disease, NCD diagnoses recorded between 2010 and 2019 were used as exposures and Long COVID diagnosis during 2020–2023 as the outcome. For analyses of new-onset disease, Long COVID diagnosis was used as the exposure and first-recorded NCD diagnoses during 2021–2023 as outcomes, restricting the analysis to individuals without the respective diagnosis before 2020. Sensitivity analyses were conducted that additionally included BMI as a covariate. The BMI measurement closest prior to March 2020 was used. Because multiple disease groups were evaluated, statistical significance was assessed using a Bonferroni-corrected threshold of $\alpha = 0.0042$ ($0.05/12$).

Interrupted time series analysis

We conducted an interrupted time series analysis following Kurvits et al. (21) to examine the impact of the COVID-19 pandemic on monthly incidence rates of newly diagnosed T2D. Using a seasonal autoregressive integrated moving average (SARIMA) model with external covariates, we assessed longitudinal changes while accounting for autocorrelation and seasonal variation (22, 23). March 2020 marked the start of the pandemic period, coinciding with Estonia's national emergency declaration. We analyzed data from January 2012 to February 2020 (pre-pandemic) and from March 2020 to December 2023 (pandemic). To model the pandemic's effect, we included three intervention terms: (i) a pulse indicator for the sharp drop in cases during March–May 2020, (ii) a step term for a sustained shift in incidence, and (iii) a ramp function for gradual trend changes (23). Seasonality was modeled using calendar month, and a linear time trend accounted for long-term changes (Supplement Figures S1A,B). The optimal SARIMA model was selected using the `auto.arima` function (R package `forecast` [39]), prioritizing the lowest Akaike Information Criterion (AIC) and minimal residual variance. Model fit was confirmed with residual diagnostics and the Ljung–Box test. We model the interrupted time series using regression model: with autocorrelated errors. Specifically,

$$Z_t = \beta_0 + \sum_{m=2}^{12} \delta_m D_{m,t} + \beta_1 t + \beta_2 \text{pulse}_t + \beta_3 \text{ramp}_t + \beta_4 \text{step}_t + Y_t$$

Where Z_t denotes the number of incident T2D diagnoses at time t , $D_{m,t}$ are monthly dummy variables capturing deterministic

seasonal effects and t represents time in months. The intervention variables represent a temporary pulse effect (pulse), a gradual post-intervention trend change (ramp) and a permanent level change (step) associated with the COVID-19 pandemic. The regression errors Y_t follow a seasonal ARMA process, specifically a SARMA $(1,0,2) \times (1,0,1)_{12}$ model:

$$(1 - \phi_1 B) (1 - \Phi_1 B^{12}) Y_t = (1 + \theta_1 B + \theta_2 B^2) (1 + \Theta_1 B^{12}) \varepsilon_t$$

Where B denotes the lag (backshift) operator, defined by $B^k Z_t = Z_{t-k}$, and $\{\varepsilon_t\}$ is a white-noise process with mean zero, constant variance σ^2 and no serial correlation. To obtain the counterfactual trajectory in the absence of the COVID-19 pandemic, a separate regression model: excluding the COVID-related intervention variables was fitted using data up to February 2020. This baseline model includes monthly dummy variables and a linear time trend and is combined with a seasonal ARMA structure SARMA $(1,0,1) \times (1,0,1)_{12}$. To estimate the cumulative T2D burden attributable to the pandemic, we summed up the predicted contributions of the intervention terms and calculated a 95% confidence interval using standard errors, assuming parameter independence. Analyses were conducted in R 4.4.2 (RStudio), with significance set at a two-sided p -value < 0.05 .

Association with the polygenic risk score for severe COVID-19

Genetic predisposition to severe COVID-19 was quantified using a polygenic risk score (PRS) derived from GWAS summary statistics for the B2 phenotype (hospitalized COVID-19 vs. population controls) from the COVID-19 Host Genetics Initiative (24). Summary statistics from the meta-analysis, excluding the 23andMe and EstBB cohorts to avoid sample overlap, were used. All EstBB participants were genotyped using Illumina GSAv1.0 and GSAv2.0 arrays, as well as GSAv2.0_EST and GSAv3.0_EST arrays with add-on SNV content from EstBB whole-genome sequencing data, aligned to the GRCh37 reference genome. Quality control followed best practices. Pre-phasing was carried out with Eagle v2.3 and imputation with Beagle v.28Sep18.79367 (25) using the Estonian population-specific reference panel based on 2,695 whole-genome sequencing samples (26). The COVID-19 PRS was calculated using a Bayesian polygenic prediction approach (PGS-CS) that models linkage disequilibrium and applies continuous shrinkage to single-variant effect sizes. Analyses were restricted to well-imputed autosomal single-nucleotide variants included among 1.1 million HapMap variants (27) and an external European sample-based LD reference from the 1,000 Genomes Project (28). The resulting PRS was standardized to a mean of 0 and standard deviation of 1, and individuals of European ancestry were subsequently stratified into low (0–20%), intermediate (20–80%), and high (80–100%) PRS quantiles for downstream analyses. Logistic regression was used for testing for an association between incident T2D and COVID-19 PGS, adjusting for birth year, sex, and 4 first genotype principal components. A sensitivity analysis was conducted that additionally adjusted for BMI.

Results

Characteristics of the study population

The study included 201,810 Estonian Biobank participants with available genotype and electronic health records, of whom 2,639 had a recorded Long COVID diagnosis and 199,171 served as controls. Individuals with Long COVID were predominantly female (75.3%) and had a mean age of 49.5 years and mean BMI of 27.5. Controls were 65% female, with a mean age of 46.4 years and mean BMI of 26.5. Long COVID cases were most common among individuals aged 40–59 years (42.7%) (Supplementary Tables S1, S2).

Association between pre-existing NCDs and Long COVID

We first assessed whether pre-existing non-communicable diseases (NCDs) diagnosed between 2010 and 2019 were associated with subsequent Long COVID diagnosis during 2020–2023. Individuals with any of the examined NCDs had higher odds of Long COVID (OR = 1.76, 95% CI 1.60–1.93, $p < 0.0001$). Several individual conditions were significantly associated with Long COVID (Figure 1; Supplementary Tables S3). The strongest associations were observed for chronic obstructive pulmonary disease (COPD) (OR = 2.33, 95% CI 1.94–2.80, $p < 0.0001$), asthma (OR = 2.18, 95% CI 1.95–2.44, $p < 0.0001$), and chronic kidney disease (OR = 2.15, 95% CI 1.57–2.95, $p < 0.0001$). Other significant associations included sleep apnea (OR = 1.45, 95% CI 1.29–1.61, $p < 0.0001$), thyroid disease (OR = 1.45, 95% CI 1.31–1.63, $p < 0.0001$), inflammatory polyarthropathies (OR = 1.43, 95% CI 1.21–1.69, $p < 0.0001$), hypertension (OR = 1.37, 95% CI 1.24–1.53, $p < 0.0001$), and T2D (OR = 1.33, 95% CI 1.13–1.58, $p = 0.0008$). When additionally adjusting for BMI, the overall pattern of associations remained largely unchanged, except for T2D (OR = 1.20, 95% CI 1.01–1.43, $p = 0.0391$), which no longer met the Bonferroni-corrected threshold (Supplementary Figure S1; Supplementary Table S4).

Association between Long COVID and new-onset NCDs diagnosed after 2020

We next examined whether individuals with Long COVID had increased odds of receiving first-recorded NCD diagnoses during the pandemic period (2021–2023). Long COVID diagnosis was associated with higher odds of several NCDs compared with individuals without Long COVID (Figure 2; Supplementary Tables S5). The strongest associations were observed for asthma (OR = 5.12, 95% CI 4.19–6.25, $p < 0.0001$) and COPD (OR = 4.08, 95% CI 3.01–5.53, $p < 0.0001$). Elevated odds were also observed for sleep apnea (OR = 3.05, 2.46–3.86, $p < 0.0001$), atrial fibrillation (OR = 1.86, 95% CI 1.36–2.54, $p = 0.0001$), thyroid diseases (OR = 1.83, 95% CI 1.44–2.32, $p < 0.0001$) and hypertension (OR = 1.80, 95% CI 1.49–2.17, $p < 0.0001$), and any of the diseases combined (OR = 2.17, 95% CI 1.80–2.62, $p < 0.0001$). The association with T2D was nominally significant but did not survive Bonferroni correction (OR = 1.44, 95% CI 1.06–1.96, $p = 0.0201$). In the BMI-adjusted model, results were unchanged overall, except for inflammatory polyarthropathies, which reached

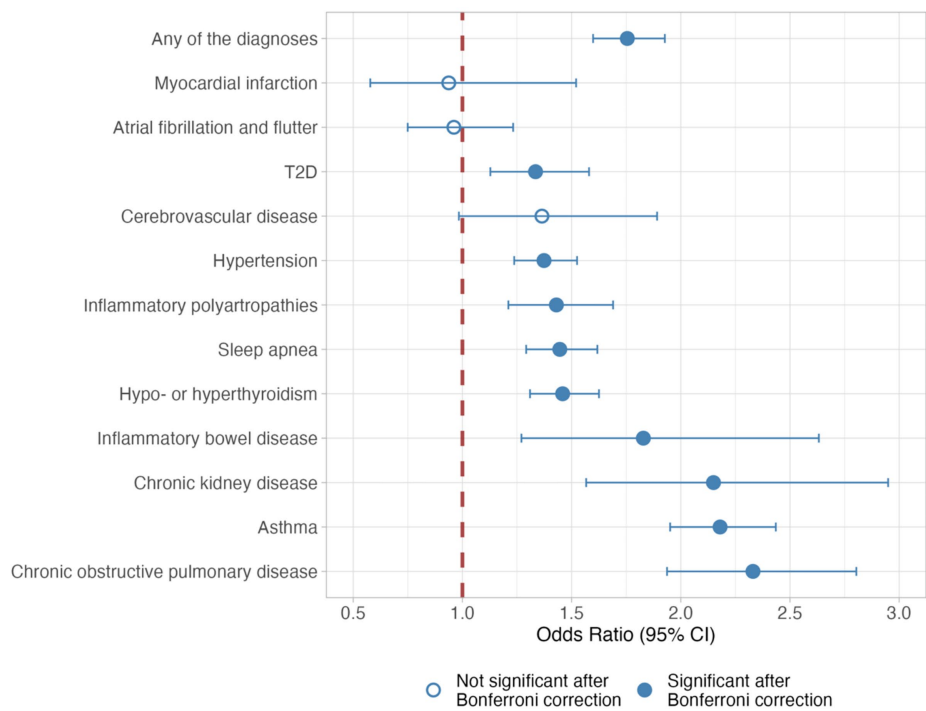


FIGURE 1
Association of pre-pandemic NCDs (2010–2019) with subsequent long COVID diagnosis (2020–2023). Odds ratios (95% confidence intervals) from logistic regression models estimating the association between NCD diagnoses recorded between January 1, 2010 and December 31, 2019 and subsequent Long COVID diagnosis between June 2020 and December 2023. The dashed vertical line indicates the null value (OR = 1).

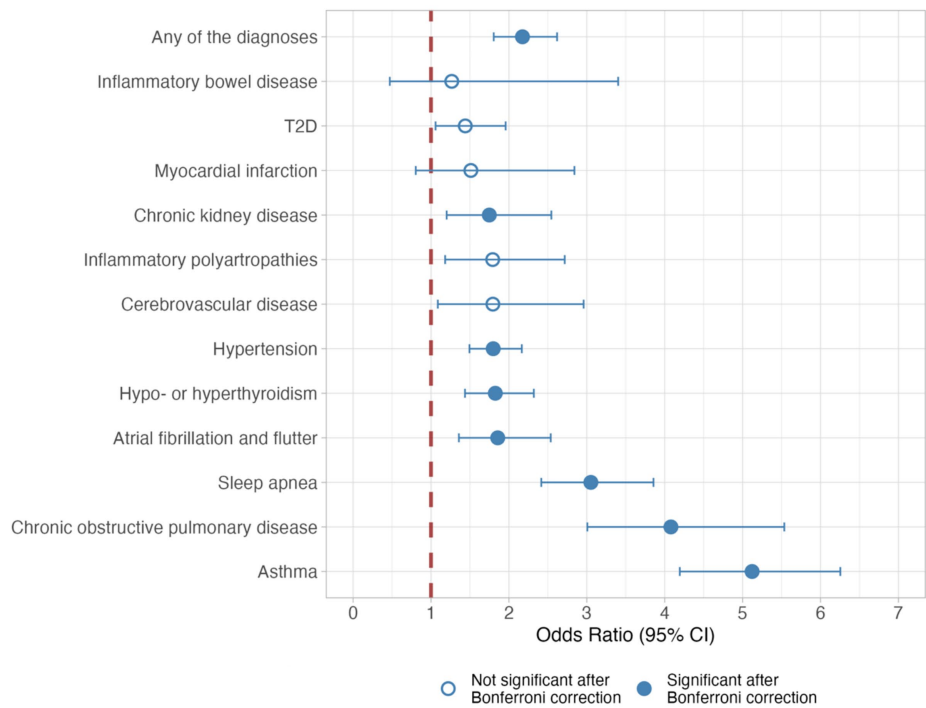


FIGURE 2
Association between Long COVID diagnosis and first-recorded NCDs during 2021–2023. Odds ratios (95% confidence intervals) for first-recorded NCD diagnoses between January 1, 2021 and December 31, 2023 among individuals with Long COVID compared to participants without a Long COVID diagnosis. The dashed vertical line represents the null value (OR = 1).

Bonferroni-corrected significance (Supplementary Figure S2; Supplementary Tables S6). Higher age, female sex, and being overweight or obese were also associated with increased odds of Long COVID (Supplementary Figure S3).

Increase in new-onset T2D diagnoses following the COVID-19 pandemic

To evaluate population-level changes in T2D incidence, we performed an interrupted time-series analysis using monthly counts of newly diagnosed T2D cases from 2010 to 2023. The SARIMA model was fitted to pre-pandemic data (January 2012–February 2020) to estimate the expected counterfactual trajectory. Observed T2D diagnoses during the pandemic period exceeded model-based projections, indicating a significant increase in incidence beyond expected temporal and seasonal trends ($p = 1.84 \times 10^{-80}$) (Figure 3; Supplementary Figure S4). These results suggest an excess burden of newly diagnosed T2D during the COVID-19 pandemic period.

Association between COVID-19 polygenic risk score and incident T2D

Finally, we assessed whether genetic susceptibility to severe COVID-19 was associated with incident T2D during the pandemic period. Participants were stratified by polygenic risk score (PRS) for severe COVID-19. Individuals in the highest PRS quantile (top 20%) had higher odds of incident T2D than those in the middle PRS group (20–80%), when adjusted for birth year and sex (OR = 1.17, 95% CI 1.05–1.31, $p = 0.0048$) (Figure 4). The association remained significant in the sensitivity analysis that additionally included BMI as a covariate (OR = 1.13, 95% CI 1.01–1.26, $p = 0.0385$) (Supplementary Figure S5). These results suggest that

genetic predisposition to severe COVID-19 is linked with incident T2D.

Discussion

In this population-based study of more than 210,000 EstBB participants with linked electronic health records, we examined clinical and epidemiological associations between Long COVID and NCDs. We observed three key findings. First, several pre-existing NCDs, such as chronic respiratory diseases, chronic kidney disease, hypertension, and T2D, were associated with increased odds of Long COVID diagnosis. Second, individuals with Long COVID had higher odds of receiving first-recorded diagnoses of several NCDs during the pandemic period. Third, population-level incidence of newly diagnosed T2D exceeded expected trends during the COVID-19 pandemic, and genetic susceptibility to severe COVID-19 was associated with increased risk of incident T2D.

Our findings are consistent with previous reports showing that individuals with pre-existing cardiometabolic and respiratory diseases have an increased risk of Long COVID. Conditions such as COPD, asthma, chronic kidney disease, and hypertension may predispose individuals to persistent post-infectious symptoms through mechanisms including systemic inflammation, immune dysregulation, and endothelial dysfunction (3, 29–31). We also observed increased odds of several NCD diagnoses among individuals with Long COVID during the pandemic period, with the strongest associations for respiratory diseases such as asthma and COPD. These patterns are consistent with evidence that COVID-19 may contribute to persistent respiratory and cardiometabolic complications. Although pre-existing T2D was associated with increased odds of Long COVID, the association between Long COVID and incident T2D did not remain significant after multiple testing correction. This discrepancy may reflect limited statistical power for incident T2D cases, a short follow-up

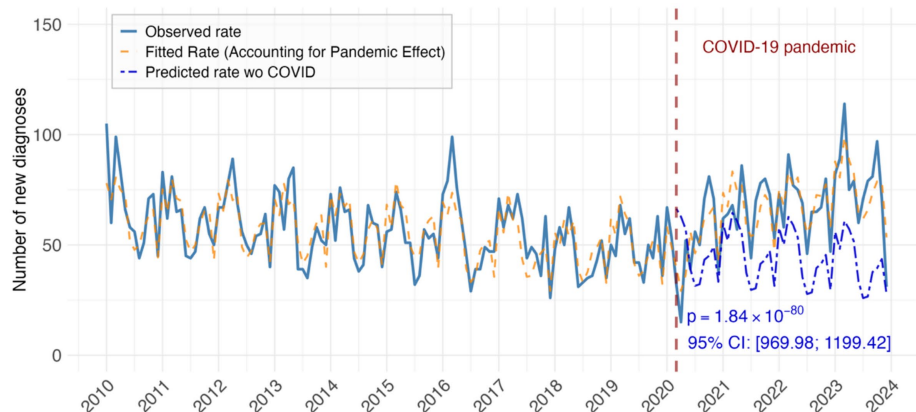


FIGURE 3

Interrupted time-series analysis of monthly incident T2D diagnoses before and during the COVID-19 pandemic (2012–2023). Observed and model-predicted monthly counts of newly diagnosed T2D. The dashed red vertical line indicates March 2020, marking the beginning of the COVID-19 pandemic period in Estonia. The solid blue line represents observed monthly diagnoses. The orange line represents the SARIMA model fitted to the observed data including pandemic effects. The dashed blue line represents the counterfactual prediction of T2D diagnoses in the absence of the pandemic. The p value corresponds to a test of whether the cumulative contribution of the COVID-19 intervention terms (pulse, step, and ramp) from March 2020 to December 2023 differs from zero, whereas the 95% confidence interval quantifies the plausible range for the total cumulative increase in incident T2D diagnoses attributable to these effects over the same period.

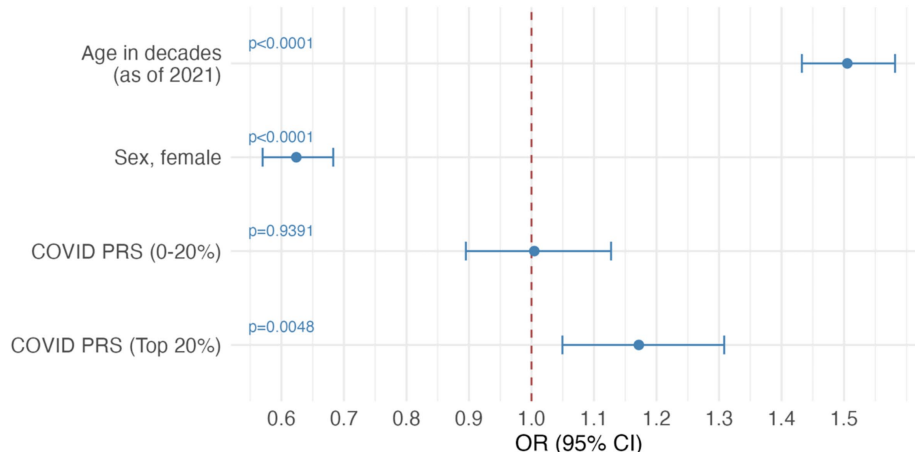


FIGURE 4

Association between COVID-19 polygenic risk score and incident T2D. Odds ratios (95% confidence intervals) for incident T2D according to COVID-19 PRS quantiles. Models were adjusted for birth year and sex. The dashed vertical line represents the null value (OR = 1). The intermediate PRS group (20–80%) was used as the reference category.

period for incident T2D, changes in healthcare accessibility, or that the development of T2D following COVID-19 may be more strongly influenced by broader population-level or pandemic-related factors rather than Long COVID itself. Also, the observed associations may partly reflect increased healthcare use or diagnostic surveillance among individuals with ongoing symptoms.

Our interrupted time-series analysis demonstrated a significant increase in newly diagnosed T2D after the onset of the COVID-19 pandemic compared with expected trends based on pre-pandemic data. Similar increases in T2D incidence following COVID-19 have been reported in large population-based studies and meta-analyses (7–10, 13). Several mechanisms may contribute to this phenomenon, including direct viral effects on pancreatic β -cells, systemic inflammation, and metabolic disturbances following infection. However, broader pandemic-related factors may also have played a role, including lifestyle changes, reduced physical activity, psychological stress, weight gain, and altered access to healthcare services (32).

We also found that the higher PRS for severe COVID-19 is associated with an increased likelihood of receiving a first diagnosis of T2D during the post-pandemic period. Previous research has shown the potential of T2D PRS to estimate COVID-19 severity and mortality (33). Our study extends this knowledge by demonstrating that individuals with higher genetic risk for severe COVID-19 are also more likely to develop T2D post-infection. However, this association should be interpreted with caution, as it may reflect horizontal pleiotropy rather than a causal relationship between severe COVID-19 susceptibility and T2D risk.

This study has several strengths. The EstBB provides a large population-based cohort with comprehensive linkage to national healthcare registries, enabling long-term follow-up and reliable ascertainment of clinical diagnoses. The integration of epidemiological and genetic data allowed us to explore potential biological links between COVID-19 severity and metabolic disease risk. In addition, the interrupted time-series design enabled evaluation of population-level changes in T2D incidence while accounting for underlying temporal and seasonal trends.

Several limitations should be considered. Long COVID was identified using ICD-10 codes recorded in routine clinical practice, and

diagnostic practices may have varied during the pandemic. Associations between Long COVID and subsequent NCD diagnoses may partly reflect increased healthcare utilization or surveillance bias. The exclusion of 2020 as a washout period to reduce potential misclassification may have led to underestimation of short-term incident NCD cases occurring early in the pandemic. The interrupted time-series analysis captures population-level changes during the pandemic but cannot isolate the causal effects of SARS-CoV-2 infection. Finally, although the PRS findings suggest shared genetic susceptibility, pleiotropy and residual confounding cannot be excluded.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Genetic predisposition to severe COVID-19 was quantified using a polygenic risk score (PRS) derived from GWAS summary statistics for the B2 phenotype from the COVID-19 Host Genetics Initiative. Summary statistics from the meta-analysis were used. All EstBB participants were genotyped using Illumina GSAv1.0 and GSAv2.0 arrays, as well as GSAv2.0_EST and GSAv3.0_EST arrays with add-on SNV content from EstBB whole-genome sequencing data, aligned to the GRCh37 reference genome. Individual-level data at EstBB can only be accessed through EstBB. Requests to access these datasets should be directed to <https://genomics.ut.ee/en/content/Estonian-biobank>.

Ethics statement

The current study was approved by the Estonian Committee on Bioethics and Human Research at the Estonian Ministry of Social Affairs (no. 1.1-12/2706 and no.1.1-12/624). All EstBB participants have signed a broad informed consent. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

HS: Visualization, Formal analysis, Methodology, Data curation, Writing – review & editing. MV: Formal analysis, Methodology, Writing – review & editing, Data curation, Visualization. MM: Methodology, Writing – review & editing. MA: Conceptualization, Writing – review & editing, Supervision. LM: Supervision, Writing – review & editing, Conceptualization. KK: Conceptualization, Supervision, Writing – review & editing. PP: Writing – original draft, Conceptualization, Resources, Funding acquisition, Supervision, Writing – review & editing. KL: Methodology, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This project received funding from the European Union's Horizon Europe Research and Innovation Programme (grant no. 101137196-POINT). The research was conducted using the Estonian Center of Genomics/Roadmap II funded by the Estonian Research Council (project number TT17) and PRG2011.

Group member of Estonian Biobank Research Team

Andres Metspalu, Lili Milani, Tõnu Esko, Reedik Mägi, Mait Metspalu, Mari Nelis and Georgi Hudjashov.

Acknowledgments

We would like to thank all biobank participants for enabling this research. Data analysis was carried out in part in the High-Performance Computing Center of the University of Tartu.

References

- Altmann DM, Whettlock EM, Liu S, Arachchillage DJ, Boyton RJ. The immunology of long COVID. *Nat Rev Immunol.* (2023) 23:618–34. doi: 10.1038/s41577-023-00904-7
- Cai M, Xie Y, Topol EJ, Al-Aly Z. Three-year outcomes of post-acute sequelae of COVID-19. *Nat Med.* (2024) 30:1564–73. doi: 10.1038/s41591-024-02987-8
- Al-Aly Z, Davis H, McCorkell L, Soares L, Wulf-Hanson S, Iwasaki A, et al. Long COVID science, research and policy. *Nat Med.* (2024) 30:2148–64. doi: 10.1038/s41591-024-03173-6
- Rezel-Potts E, Douiri A, Sun X, Chowieńczyk PJ, Shah AM, Gulliford MC. Cardiometabolic outcomes up to 12 months after COVID-19 infection. A matched cohort study in the UK. *PLoS Med.* (2022) 19:e1004052. doi: 10.1371/journal.pmed.1004052
- Terry P, Heidel RE, Wilson AQ, Dhand R. Risk of long COVID in patients with pre-existing chronic respiratory diseases: a systematic review and meta-analysis. *BMJ Open Respir Res.* (2025) 12:2528. doi: 10.1136/bmjresp-2024-002528
- Anzalone AJ, Krichevsky S, Yoo YJ, Wilkins KJ, Alakwaa F, Liu F, et al. The association of long COVID and CKD: findings from the National Clinical Cohort Collaborative. *Clin J Am Soc Nephrol.* (2025) 20:1323–40. doi: 10.2215/CJN.0000000773
- Lai H, Yang M, Sun M, Pan B, Wang Q, Wang J, et al. Risk of incident diabetes after COVID-19 infection: a systematic review and meta-analysis. *Metabolism.* (2022) 137:155330. doi: 10.1016/j.metabol.2022.155330
- Nazari P, Pozzilli P. Type 2 diabetes and COVID-19: lessons learnt, unanswered questions and hints for the future. *Diabetes Res Clin Pract.* (2023) 204:110896. doi: 10.1016/j.diabres.2023.110896
- Kim SH, Arora I, Hsia DS, Knowler WC, LeBlanc E, Mylonakis E, et al. New-onset diabetes after COVID-19. *J Clin Endocrinol Metab.* (2023) 108:e1164–74. doi: 10.1210/clinem/dgad284
- Xie Y, Al-Aly Z. Risks and burdens of incident diabetes in long COVID: a cohort study. *Lancet Diabetes Endocrinol.* (2022) 10:311–21. doi: 10.1016/S2213-8587(22)00044-4
- Raveendran AV, Misra A. Post COVID-19 syndrome ("long COVID") and diabetes: challenges in diagnosis and management. *Diabetes Metab Syndr.* (2021) 15:102235. doi: 10.1016/j.dsx.2021.102235
- Harding JL, Oviedo SA, Ali MK, Oforokun I, Gander JC, Patel SA, et al. The bidirectional association between diabetes and long-COVID-19 - a systematic review. *Diabetes Res Clin Pract.* (2023) 195:110202. doi: 10.1016/j.diabres.2022.110202
- Heald AH, Williams R, Jenkins DA, Stewart S, Bakerly ND, Mccay K, et al. The prevalence of long COVID in people with diabetes mellitus-evidence from a UK cohort. *EClinicalMedicine.* (2024) 71:102607. doi: 10.1016/j.edinnm.2024.102607
- Wander PL, Lowy E, Korpak A, Beste LA, Kahn SE, Boyko EJ. SARS-CoV-2 infection is associated with higher chance of diabetes remission among veterans with incident diabetes. *PLoS One.* (2025) 20:e0317348. doi: 10.1371/journal.pone.0317348
- Naveed Z, Velásquez García HA, Wong S, Wilton J, McKee G, Mahmood B, et al. Association of COVID-19 infection with incident diabetes. *JAMA Netw Open.* (2023) 6:e238866. doi: 10.1001/jamanetworkopen.2023.8866
- Hirani D, Salem V, Khunti K, Misra S. Newly detected diabetes during the COVID-19 pandemic: what have we learnt? *Best Pract Res Clin Endocrinol Metab.* (2023) 37:101793. doi: 10.1016/j.beem.2023.101793

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2026.1834448/full#supplementary-material>

17. Zhang T, Mei Q, Zhang Z, Walline JH, Liu Y, Zhu H, et al. Risk for newly diagnosed diabetes after COVID-19: a systematic review and meta-analysis. *BMC Med.* (2022) 20:444. doi: 10.1186/s12916-022-02656-y
18. Choi JH, Kim KM, Song K, Seo GH. Risk for newly diagnosed type 2 diabetes mellitus after COVID-19 among Korean adults: a Nationwide matched Cohort study. *Endocrinol Metab (Seoul).* (2023) 38:245–52. doi: 10.3803/EnM.2023.1662
19. Taylor K, Eastwood S, Walker V, Cezard G, Knight R, Al Arab M, et al. Incidence of diabetes after SARS-CoV-2 infection in England and the implications of COVID-19 vaccination: a retrospective cohort study of 16 million people. *Lancet Diabetes Endocrinol.* (2024) 12:558–68. doi: 10.1016/S2213-8587(24)00159-1
20. Milani L, Alver M, Laur S, Reisberg S, Haller T, Aasmets O, et al. The Estonian biobank's journey from biobanking to personalized medicine. *Nat Commun.* (2025) 16:3270. doi: 10.1038/s41467-025-58465-3
21. Kurvits K, Toompere K, Jaanson P, Uusküla A. The COVID-19 pandemic and the use of benzodiazepines and benzodiazepine-related drugs in Estonia: an interrupted time-series analysis. *Child Adolesc Psychiatry Ment Health.* (2024) 18:66. doi: 10.1186/s13034-024-00757-5
22. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol.* (2017) 46:dyw098–355. doi: 10.1093/ije/dyw098
23. Schaffer AL, Dobbins TA, Pearson SA. Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Med Res Methodol.* (2021) 21:58. doi: 10.1186/s12874-021-01235-8
24. C.-H. G. Initiative. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet.* (2020) 28:715–8. doi: 10.1038/s41431-020-0636-6
25. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* (2007) 81:1084–97. doi: 10.1086/521987
26. Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet.* (2017) 25:869–76. doi: 10.1038/ejhg.2017.51
27. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. *Nature.* (2010) 467:52–8. doi: 10.1038/nature09298
28. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature.* (2015) 526:68–74. doi: 10.1038/nature15393
29. Al-Aly Z, Topol E. Solving the puzzle of long Covid. *Science.* (2024) 383:830–2. doi: 10.1126/science.adl0867
30. Alwan NA, Al-Aly Z. Learning from long COVID: conceptualising action to tackle infection-associated non-communicable diseases. *Lancet.* (2025) 405:360–2. doi: 10.1016/S0140-6736(24)02256-6
31. Bowe B, Xie Y, Al-Aly Z. Postacute sequelae of COVID-19 at 2 years. *Nat Med.* (2023) 29:2347–57. doi: 10.1038/s41591-023-02521-2
32. Shitaye G, Getie M, Mekonnen Z, D'Abrosca G, Fattorusso R, Isernia C, et al. Molecular analysis of long COVID and new-onset diabetes mellitus: pathobiological relationships and current mechanistic views. *Front Endocrinol (Lausanne).* (2025) 16:1737894. doi: 10.3389/fendo.2025.1737894
33. Lee A, Seo J, Park S, Cho Y, Kim G, Li J, et al. Type 2 diabetes and its genetic susceptibility are associated with increased severity and mortality of COVID-19 in UK biobank. *Commun Biol.* (2024) 7:122. doi: 10.1038/s42003-024-05799-1