Genetic risk of type 2 diabetes and its relationship to severe Covid-19 disease

John Vytas Kushleika

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science

University of Washington

2024

Committee:

Burcu Darst

Sara Lindstrom

Program Authorized to Offer Degree:

Institute for Public Health Genetics

University of Washington

Abstract

Genetic risk of type 2 diabetes and its relationship to severe Covid-19 disease

John Vytas Kushleika

Chair of the Supervisory Committee:

Burcu Darst

Institute for Public Health Genetics

Type 2 diabetes (T2D) is a risk factor for severe cases of Covid-19. T2D is partially heritable, though few studies have examined the role of T2D genetic risk factors in severe Covid-19. We analyzed data from 459,493 participants in the UK Biobank for association of a T2D polygenic risk score (PRS) with severe Covid-19 outcomes diagnosed between February 1, 2020 and October 31, 2022. Participants with severe Covid-19 (SC, N=8,367) were those hospitalized for Covid-19 (ICD-10 codes U07.1 or U07.2) or who died of Covid-19. Logistic regression using SC (yes/no) as the outcome was employed, adjusting for age, sex and the first ten principal components to account for potential population stratification. We found positive and statistically significant relationships between the T2D PRS and SC. Among all participants, each SD unit increase in PRS was associated with an odds ratio of 1.09 (95% CI= 1.06 – 1.11; $p=1.7e^{-12}$) for risk of SC. Examined as a categorical variable, participants within the top T2D PRS quintile had 18% (95% CI= 11% – 27%; $p=9.1e^{-7}$) higher risks of SC in comparison to participants in the median (40%-60%) T2D PRS quintile. To address potential issues with selection bias, we repeated our analyses using three alternative control groups including those who (i) had never been diagnosed with Covid-19 at time of data collection, (ii) had non-severe Covid-19, or (iii) had undergone Covid-19-testing (regardless of their test results), obtaining very similar results. The T2D PRS was also predictive of severe Covid-19 among participants with no diagnosis of T2D. These results imply that genetic factors associated with T2D influence susceptibility to SC outcomes, providing support for the potential use of PRS in assessing patient risks and furthering our understanding of the mechanisms underlying the increased risk of SC among individuals with T2D.

Acknowledgement

I would like to express my gratitude and appreciation to my committee members/advisors, Burcu Darst and Sara Lindstrom, for the hours of encouragement and mentoring which they contributed to my progress through this program, as well as for their help with completion of this research project. I am grateful also for the collegiality of the Darst/Lindstrom lab group and the talented and dedicated faculty of the University of Washington.

Table of Contents

Introduction

Covid-19, caused by infection with the SARS-Cov2 virus[1], has been a significant worldwide public health concern since its identification as a cause of severe respiratory illness in late 2019. Diabetes was rapidly identified as one of the risk factors[2] for becoming severely ill during infection with SARS-Cov2, perhaps unsurprisingly since diabetics are known to be susceptible to infectious diseases[3] including respiratory illnesses[4] such as tuberculosis. Prevalent diabetes cases in the U.S. are mostly (~90%) type-2 diabetes (T2D), characterized by resistance to insulin in tissues with potential progression to pancreatic B-cell failure[5]. Explanations for the vulnerability of diabetics to infectious diseases include immune system changes[6] such as lymphocytopenia, alterations in cytokine signaling and elevated levels of glycated immune proteins. Regarding Covid-19 specifically, the functional immune changes in T2D affect the immune system's ability to counter the runaway secretion of inflammatory cytokines[7] known as "cytokine storm", which associates with severe outcomes in Covid-19 cases. There is also suggestive evidence for molecular links[8] between T2D and Covid-19 from gene network analyses[9]. One notable molecular player is the protein angiotensin-converting enzyme-2 (ACE2), which is the putative cellular receptor for the virus and has been reported to be upregulated in diabetic tissues[10]. Also, it has been observed that hospitalized Covid-19 patients often become hyperglycemic[2], leading to questions about cause and effect regarding these two conditions.

T2D has a heritable component which has been estimated to explain ~10-20% of disease variability[5]. There are numerous genetic variants associated with T2D which have been identified in genome-wide association studies, and combining these into polygenic risk scores (PRS) has shown promise for individualized risk prediction[11]. Genetic risk prediction could contribute to T2D prevention as the disease typically manifests in middle-aged adults and is known to respond favorably to lifestyle changes. Research using genetic risk prediction tools can also shed light on mechanistic questions. As few prior studies have addressed the intersection of genetic risk for T2D and Covid-19, this project aimed to examine the association of a T2D PRS with severe Covid-19 among participants in the UK Biobank.

Methods

Participants

The UK Biobank holds information on genetics, lifestyle, diet, and medical history for more than 500,000 individuals in the United Kingdom[12]. Participant information is updated regularly via follow-up visits and linkage to National Health Service (NHS) records. Data was obtained from the UK Biobank under application #95770, which provided access to participant records through October 2022 and allowed us to identify Covid-19 and T2D status. Participants were excluded for poor heterozygosity or high missingness (data field 22010, "Recommended genomic analysis exclusions"), or if self-reported sex did not match genetically determined sex. Participants who died on or before January 31, 2020 (N=28,947) were also excluded. The remaining participants used for analysis numbered 459,493. The study was reviewed by the Human Subjects Division of the University of Washington which determined that formal human subjects review was not necessary for this project.

Definitions

As the UK Biobank is not a random sample of the U.K. population, and being tested for Covid-19 (or even receiving a positive test result) might depend upon a variety of factors that could produce biased results, we examined a variety of participant subgroups for use as alternative controls. Individuals who tested for Covid-19 (TESTED) were those who had a record of any Covid-19 test during the period between February 1, 2020 and October 31, 2022, regardless of the test results. Covid-positive (Covid(+)) participants were defined as those having one or more Covid-19 tests with a positive Covid-19 result. We defined severe Covid-19 (SC) cases as those Covid(+) participants who had a hospital inpatient diagnosis of Covid-19 (ICD-10 diagnostic codes U07.1 or U07.2), or those who died of Covid-19 and had a Covid-19 diagnostic code listed as a primary or secondary cause of death. The remaining Covid(+) participants (who did not also have an inpatient diagnosis of Covid-19) were assumed to have non-severe Covid-19 (N-SC). Covid-negative (Covid(-)) participants were those without a record of a positive Covid-19 test (including those without any Covid-19 test on record). T2D was defined as participants meeting the following two criteria: 1) an inpatient diagnosis of T2D (ICD-10 code E11) after the age of 40, and 2) the absence of any diagnostic codes for types of diabetes other than type 2 (codes E10, E12, E13 or E14). For comparative purposes, some analyses were also done using a broader definition of T2D (T2D-ALT), defined as participants with either 1) an inpatient ICD-10 code of E11, or 2) those who self-reported having diabetes but did not subsequently answer "yes" to the follow-up survey

question “Did you only have diabetes during pregnancy?” Participant age was calculated by subtracting birth year from 2021, the approximate mid-point of the study period. Body-mass index (BMI) was defined as the calculated variable from a participant’s initial study visit (data field 21001.0). Values of the Multiple Deprivation Index (data fields 26410, 26426 & 26427) were used as a measure of socioeconomic status. The R package “Comorbidity[14]” was used to calculate Swiss-weighted Elixhauser comorbidity scores for each participant, using as input the participant’s complete list of inpatient ICD-10 diagnostic codes.

PRS Construction

A T2D PRS developed by Thompson[13] et.al. was utilized in analyses. This was generated from meta-analyses of large genome-wide association studies including individuals of European, East Asian and South Asian ancestries (approximately 212K cases and 1.14M controls; data external to the UK Biobank) and includes 13,628,692 variants. Individual PRS values were calculated as the genome-wide per-variant sum of effect sizes multiplied by the allele dosage, and were zero-centered and standardized.

Statistical Analyses

Multivariable logistic regression was employed to examine the utility of the T2D PRS for prediction of T2D in the cohort, with adjustment for age, sex and the first ten principal components. The T2D PRS and selected covariates (age, sex, BMI and the first ten principal components) were also tested in univariable analyses to assess their individual effects on SC. Then, multivariable logistic regression was used to examine the association of the T2D PRS with SC. Analyses each included the T2D PRS as predictor, with the additional covariates age, sex and first ten principal components to account for potential population stratification.

Analyses conducted were as follows: 1) SC versus the absence of SC (i.e., vs N-SC and Covid(-)); 2) SC versus N-SC, 3) SC versus Covid(-), and 4) SC vs TESTED. We repeated comparisons (1), (2), and (3) after stratifying by T2D status. Additionally, these relationships were examined after partitioning the T2D PRS into categories (quintiles), with PRS category cutoffs defined based on PRS distributions in the appropriate control group. For these categorical analyses the middle (40%-60%) quintile was used as reference.

In sensitivity analyses, the effect of including body-mass index (BMI, data field 21001.0) as a regression covariate in addition to age, sex and first ten principal components was also briefly examined by repeating analysis (1) above. All analyses were conducted in R, version 4.4.0

Results

Participant characteristics

Table 1 and **Supplementary Table 1** present participant characteristics for the full cohort as well as for groups defined by presence (+) or absence (-) of any recorded Covid-19 infection occurring during the study period. Participants with SC numbered 8,367 (~1.8% of the cohort). Characteristics between the Covid(-) and Covid(+) subgroups (Table 1) were generally similar. While Covid(+) participants tended to be somewhat younger than Covid(-) and small differences were noted in smoking status, other characteristics (sex, BMI, comorbidity index, deprivation index & prevalence of T2D) did not differ notably. However, those with SC were distinguished markedly from N-SC or Covid(-) participants (**Supplementary Table 2**). Those with SC tended to be male, older, and have higher values /rates for BMI, smoking, comorbidity & deprivation indices compared with N-SC or Covid(-). Additionally, the prevalence of T2D was roughly three times higher among those with SC compared with N-SC (~20.8% vs. ~7%). There were 30,663 participants with T2D and 39,630 with T2D-ALT (see Methods section above). Compared to all participants, those with T2D were more likely to be male, older and have higher values /rates for BMI, smoking, comorbidity & deprivation indices (**Supplementary Table 1**). Participant characteristics did not differ notably between the T2D and T2D-ALT groups.

Validation of the T2D PRS and covariates

We confirmed that the T2D PRS was associated with T2D in our cohort (**Supplementary Table 2**). Each SD unit increase in the T2D PRS was associated with an estimated odds ratio of 2.02 for T2D (95% CI: 2.0 – 2.05; p<2e-16), implying an approximate doubling of risk of T2D per SD unit increase. We also examined the association between preselected covariates (age, sex and the first ten principal components) and SC in individual logistic regression analyses, confirming that the variables age and sex in particular were strongly associated with SC, while the principal components had in general smaller (albeit sometimes statistically significant) effects.

Logistic regression analyses for effects of the T2D PRS on severe Covid outcomes

The T2D PRS was positively associated with SC (**Table 2**), with an odds ratio of 1.09 (95% CI=1.06 – 1.11; $p=1.6e^{-12}$), implying an approximately +9% elevation of risk for each +1 SD change in PRS value. These associations were highly statistically significant ($p<0.001$; **Table 2**). Sensitivity analyses across different control groups (see Methods) showed qualitatively the same results.

We repeated the analyses stratified by T2D status. Results are presented in **Table 3**. Among T2D controls (participants without T2D), the T2D PRS was positively associated with risk of SC (OR: 1.03; 95% CI: 1.00 - 1.05; p=0.043). Among T2D cases on the other hand, a protective effect of the T2D PRS was observed for SC (OR: 0.89, 95% CI, 0.84 – 0.94; $p=9.2e^{-6}$). Using different control groups as part of sensitivity analyses did not alter these results (**Table 3**).

Results from categorical analyses of the T2D PRS among all subjects are presented in **Table 4**. Odds ratio estimates implied a higher risk of SC with higher PRS categories, with the clearest evidence of an effect seen at the extremes of PRS categories (the first and fifth quintiles). Participants in the highest (80%-100%) PRS quintile had an approximately 18% higher risk (95% CI= 11% - 27%; $p = 9.1e^{-7}$) of SC compared with the median (40% - 60%) PRS quintile (**Table 4**). Conversely subjects in the bottom quintile of risk scores (0-20%) were estimated to have approximately 8% lower risk (95% CI= 1% - 14%; p = 0.031) of SC than the median. Stratifying by Covid infection status resulted in very similar odds ratios for these categories (**Supplementary Table 3**).

The effect of BMI on the association of the T2D PRS with severe Covid-19 outcomes was also explored. When BMI was included as a covariate in comparison (1) in addition to our standard covariates age, sex and 10 PCs, the OR estimate was substantially attenuated (OR: 1.04; 95% CI= 1.02 – 1.06) although still significant ($p= 7.3e^{-4)}$; **Supplementary Table 4**.

Discussion

We report that a measure of genetic risk for T2D, a PRS, has a moderately positive association with SC among UK Biobank participants. Associations persisted after statistical adjustment for age, sex, genetic ancestry and BMI, after stratification by SARS-Cov2 infection status or Covid-19 test status, and when assessed within diabetes-free participants. Our results imply that genetic factors associated with T2D influence susceptibility to SC, potentially furthering our understanding of the mechanisms underlying

the increased risk of SC among individuals with T2D. The findings can also more generally inform discussions about the causal role of genetic factors in infectious respiratory diseases.

Our results complement those of a recent study[15] by Lee et.al., which similarly reported association of genetic risk for T2D with SC in the UK Biobank. Lee et. al used a PRS generated from data internal to the UK Biobank via 10-fold cross-validation, while our study relied on a PRS developed using data external to the UK Biobank. Nonetheless the magnitude and direction of effects on SC were similar in both studies. It is also noteworthy that both studies demonstrated positive associations of T2D genetic risk with SC even within participants without any diagnosis of T2D. This finding might reflect the presence of some incipient T2D among 'healthy' participants or alternatively could point to the influence of T2D genetics on other risk factors; in either case, it underlines the possibility of using such risk scores to identify patients with an increased susceptibility to SC.

Upon restricting our analyses to participants with T2D, we found that higher T2D PRS scores were protective against SC. This result may appear anomalous, but in fact it is reasonable to expect an attenuation or reversal of effect for analyses which are limited to participants with a particular outcome (in this case those with T2D) and which employ a predictor (T2D PRS) associated with that outcome. In examining the characteristics of the 30,663 participants with T2D in our study after stratifying them by SC status (**Supplementary Table 1**), it is apparent that those who became severely ill with Covid-19 (N=1,738) had markedly higher values for all auxiliary risk factors examined (male sex, age, BMI, smoking, comorbidity & deprivation indices) compared to participants with T2D who did not experience SC (N=28,925). This observation highlights that a variety of sociodemographic, environmental and lifestyle factors may readily outweigh risks due to genetics in the context of T2D and is consistent with the notion that modifying these factors can improve overall health for individuals with diabetes.

Including BMI as a covariate in regression analyses resulted in attenuating the OR for SC by approximately 50% (**Supplementary Table 4**) though the association remained statistically significant ($p = 7.3e^{-4}$). Obesity (as measured alternatively by BMI, waist/hip ratio or related metrics) is well-established as a risk factor for T2D[16], and it is also theorized that the two conditions cooperate bi-directionally in support of one another. A higher PRS for BMI was also recently found to be predictive of increased risk of SC in the UK Biobank[17]. As BMI is the manifestation of environmental, genetic and other factors, the

moderating effect of including BMI as a regression covariate supports the idea that genetic risk for T2D overlaps with or interacts with causes of obesity.

Because the UK Biobank is not a random sample of the British population[18] and because testing positive for Covid-19 (or even receiving a test for Covid-19) may be subject to various influencing factors, spurious associations could be induced in the context of Covid-19 research in the UK Biobank, harming the generalizability of research findings[19]. We addressed this potential issue through sensitivity analyses, namely examining the association of the T2D PRS on Covid-19 outcomes within a variety of participant strata. Though individuals in the UK Biobank who were tested one or more times for Covid-19 differ in many characteristics from those who were not tested[19], we found similar odds ratios for T2D genetic risk in tested participants compared with the full cohort (**Table 2**). And as testing positive (or negative) for Covid-19 might also be influenced by factors (for example occupation or socioeconomic status), we stratified subjects into Covid(-) and Covid(+) groups but again failed to find any notable differential effect.

Our outcome category (severe Covid-19) captures a significant and consequential portion of the public health burden of Covid-19 disease, and the rigorous nature of data collection in the UK Biobank provides a large degree of confidence in measuring this outcome. However it is reasonable to assume that our study does not avoid all misclassification bias, due to factors such as variation in the tendency to seek medical care, limits on access to care or hospital capacity, and varying approaches to recording causes of death. Additionally, it is notable that comorbidities are commonly seen among individuals with SC, thus complicating diagnoses. The UK Biobank data could be probed for further insights into whether particular medical conditions, or combinations of conditions, help to explain severe outcomes in Covid-19 patients.

In conclusion, we found a moderately positive effect of a T2D genetic risk score on the likelihood of becoming severely ill or dying from Covid-19 in UK biobank participants. These results add to the body of research on Covid-19 and could contribute to diagnostic and preventative efforts for this serious respiratory disease.

References

1. Centers for Disease Control and Prevention. Accessed April 17, 2021. https://www.cdc.gov/

2. Muniyappa R, Gubbi S. COVID-19 pandemic, coronaviruses, and diabetes mellitus. *Am J Physiol Endocrinol Metab*. 2020;318(5):E736-E741. doi:10.1152/ajpendo.00124.2020

3. Visca D, Pignatti P, Spanevello A, Lucini E, La Rocca E. Relationship between diabetes and respiratory diseases—Clinical and therapeutic aspects. *Pharmacological Research*. 2018;137:230-235. doi:10.1016/j.phrs.2018.10.008

4. Jensen AV, Faurholt-Jepsen D, Egelund GB, et al. Undiagnosed Diabetes Mellitus in Community-Acquired Pneumonia: A Prospective Cohort Study. *Clinical Infectious Diseases*. 2017;65(12):2091-2098. doi:10.1093/cid/cix703

5. DeFronzo RA, Ferrannini E, Groop L, et al. Type 2 diabetes mellitus. *Nat Rev Dis Primers*. 2015;1:15019. doi:10.1038/nrdp.2015.19

6. Geerlings SE, Hoepelman AIM. Immune dysfunction in patients with diabetes mellitus (DM). *FEMS Immunology & Medical Microbiology*. 1999;26(3-4):259-265. doi:10.1111/j.1574-695X.1999.tb01397.x

7. Ramasamy S, Subbian S. Critical Determinants of Cytokine Storm and Type I Interferon Response in COVID-19 Pathogenesis. *Clin Microbiol Rev*. 2021;34(3):e00299-20. doi:10.1128/CMR.00299-20

8. Gao Q, Zhang W, Li T, et al. Interrelationship between 2019-nCov receptor DPP4 and diabetes mellitus targets based on protein interaction network. *Sci Rep*. 2022;12(1):188. doi:10.1038/s41598-021-03912-6

9. Saik OV, Klimontov VV. Gene Networks of Hyperglycemia, Diabetic Complications, and Human Proteins Targeted by SARS-CoV-2: What Is the Molecular Basis for Comorbidity? *Int J Mol Sci*. 2022;23(13):7247. doi:10.3390/ijms23137247

10. Subbarayan K, Ulagappan K, Wickenhauser C, Bachmann M, Seliger B. Immune Interaction Map of Human SARS-CoV-2 Target Genes: Implications for Therapeutic Avenues. *Front Immunol*. 2021;12:597399. doi:10.3389/fimmu.2021.597399

11. Polfus LM, Darst BF, Highland H, et al. Genetic discovery and risk characterization in type 2 diabetes across diverse populations. *Human Genetics and Genomics Advances*. 2021;2(2):100029. doi:10.1016/j.xhgg.2021.100029

12. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z

13. Thompson DJ, Wells D, Selzam S, et al. UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. *medRxiv*. Published online January 1, 2022:2022.06.16.22276246. doi:10.1101/2022.06.16.22276246

14. Gasparini A. comorbidity: An R package for computing comorbidity scores. *JOSS*. 2018;3(23):648. doi:10.21105/joss.00648

15. Lee A, Seo J, Park S, 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(1):122. doi:10.1038/s42003-024-05799-1

16. Ruze R, Liu T, Zou X, et al. Obesity and type 2 diabetes mellitus: connections in epidemiology, pathogenesis, and treatments. *Front Endocrinol*. 2023;14:1161521. doi:10.3389/fendo.2023.1161521

17. Zhu Z, Hasegawa K, Ma B, Fujiogi M, Camargo CA, Liang L. Association of obesity and its genetic predisposition with the risk of severe COVID-19: Analysis of population-based cohort data. *Metabolism*. 2020;112:154345. doi:10.1016/j.metabol.2020.154345

18. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

19. Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun*. 2020;11(1):5749. doi:10.1038/s41467-020-19478-2

Tables

Table 1: Participant characteristics: full cohort; Covid(-) participants; Covid(+) participants.

	Full Cohort	%	Covid(-)	%	Covid(+)	%
All	459493		346653	75.4%	112840	24.6%
Male	206421	44.9%	155129	44.8%	51292	45.5%
Female	253072	55.1%	191524	55.2%	61548	54.5%
Age: median (IQR)	70. (63, 76)		71 (64, 76)		67 (60, 74)	
<60	74594	16.2%	49270	14.2%	25324	22.4%
60 - 75	243002	52.9%	181534	52.4%	61468	54.5%
75 +	141897	30.9%	115849	33.4%	26048	23.1%
BMI: median (IQR)	26.7 (24.1, 29.8)		26.68 (24.1 , 29.8)		26.8. (24.2, 29.9)	
Smoking status						
current	45701	9.9%	35748	10.3%	9953	8.8%
former	156596	34.1%	117779	34.0%	38817	34.4%
never	254932	55.5%	191322	55.2%	63610	56.4%
Mean comorbidity score (SD) [A]	2.78 (9.0)		2.75 (9.0)		2.89 (9.3)	
Mean Townsend deprivation index (SD) [B]	17.10 (13.9)		17.1 (14.0)		17.1 (13.7)	
Type 2 diabetes diagnosis	30663	6.7%	23139	6.7%	7524	6.7%

A: Elixhauser comorbidity score with Swiss weightings

B: Townsend deprivation index at recruitment (Data field 22189)

Table 2. Estimated effects of the T2D PRS on severe Covid-19 (SC) among all participants, among Covid(+) participants, versus Covid(-) participants, and among TESTED participants.

Comparison	Cases	Controls	OR (95% CI)	p-value
		N-SC + Covid(-)		
SC / all participants	8,367	451126	1.09 (1.06 - 1.11)	1.70e−12
		Covid(+)		
SC / N-SC	8,367	104473	1.09 (1.06 - 1.12)	1.30e−12
		Covid(-)		
SC / Covid(-)	8,367	346653	1.08 (1.06 - 1.11)	2.99e−12
		TESTED		
SC / TESTED	8,042	259245	1.08 (1.05 - 1.10)	2.29e−10

From multivariable logistic regression.

Predictor: T2D PRS. Covariates: age, sex, 1st 10 PCs

Table 3. Estimated effects of the T2D PRS on SC and Covid-DTH among T2D controls (left) and among T2D cases (right).

	T2D Controls (T2D(-)), N=428830				T2D Cases (T2D(+)), N=30663			
	Cases	Controls	OR (95% CI)	p-value	Cases	Controls	OR (95% CI)	p-value
Strata: Subjects without T2D (N=428830)					**Strata: T2D cases (N= 30663)**			
	6629.0	422,201	1.03 (1.00 - 1.05)	4.34e−2	1738	28,925	0.89 (0.84 - 0.94)	9.22e−6
Strata: T2D(-) & Covid(+)					**Strata: T2D cases & Covid(+)**			
	6629.0	98,687	1.03 (1.00 - 1.06)	2.48e−2	1738	5,786	0.87 (0.82 - 0.93)	7.65e−6
Strata: T2D(-) &Covid(-)					**Strata: T2D cases & Covid(-)**			
	6629.0	323,514	1.03 (1.00 - 1.05)	5.06e−2	1738	23,139	0.89 (0.84 - 0.94)	1.95e−5

Predictor, T2D PRS. Covariates: age, sex, 1st 10 PCs

Table 4. Estimated effect of the T2D PRS categories (quintiles) on severe Covid-19 (SC) among all subjects.

Quintile	OR (95% CI)	p-value
1st quintile (0 - 20%)	0.92 (0.86 - 0.99)	3.08e−2
2nd quintile (20 - 40%)	0.99 (0.893 - 1.07)	8.49e−1
3rd quintile (40 - 60%)	reference	
4th quintile (60 -80%)	1.09 (1.02 - 1.17)	1.03e−2
5th quintile (80 - 100%)	1.18 (1.11 - 1.27)	9.08e−7

Predictor, T2D PRS. Covariates: age, sex, 1st 10 PCs

Reference group for categorical analysis: middle (40%- 60%) quintile

Supplementary Tables

Supplementary Table 1: Participant characteristics (percentages), including SC, Covid-DTH, T2D, T2D/SC and T2D-ALT groups.

	All Participants	Covid(-)	Covid(+)	SC	T2D	T2D / SC	T2D-ALT
All	459,493	346,653	112,840	8,367	30,663	1,738	39,858
Male	44.9%	44.8%	45.5%	56.7%	58.0%	64.2%	59.1%
Female	55.1%	55.2%	54.5%	43.3%	42.0%	35.8%	40.9%
Age: median (IQR)	70.0	71.0	67.0	75.0	74.0	75.0	73.0
<60	16.2%	14.2%	22.4%	8.6%	7.4%	7.2%	8.1%
60 - 74	52.9%	52.4%	54.5%	41.1%	48.2%	40.7%	49.0%
75 +	30.9%	33.4%	23.1%	50.4%	44.4%	52.1%	42.9%
BMI: median (IQR)	26.7	26.7	26.8	28.4	30.7	31.3	30.5
Smoking status							
current	9.9%	10.3%	8.8%	14.2%	12.6%	14.0%	12.1%
former	34.1%	34.0%	34.4%	41.1%	40.7%	45.8%	40.3%
never	55.5%	55.2%	56.4%	43.8%	45.9%	38.5%	46.8%
Comorbidity score	2.78	2.75	2.89	13.6	6.3	15.3	6.0
Deprivation index	17.1	17.1	17.1	21.8	21.7	24.7	21.5
Type-2 diabetes diagnosis	6.7%	6.7%	6.7%	20.8%	100.0%	100.0%	100.0%

SC = Severe Covid; Covid-DTH= death due to Covid-19; T2D/SC = participants with both T2D and SC; T2D-ALT= alternative definintion of T2D (see Methods)

Supplementary Table 2: Validation of the T2D PRS and covariates

Predictor	Outcome	OR (95% CI)	p-value
Effect of T2D PRS on T2D1: [A]			
T2D PRS	T2D	2.02 (2.0 - 2.05)	2e-16
Effect of covariates on SC: [B]			
sex	SC	1.62 (1.55 - 1.69)	2e-16
age	SC	1.064 (1.061. - 1.068)	2e-16
BMI	SC	1.07 (1.068 - 1.076)	2e-16
PC1	SC	1.0017 (1.0013 - 1.002)	2e-16
PC2	SC	1.000. (0.999 - 1.001)	0.981
PC3	SC	1.004 (1.002 - 1.005)	3.82e-08
PC4	SC	1.00 (0.99 - 1.00)	0.319
PC5	SC	1.00 (0.99 - 1.00)	0.0819
PC6	SC	1.00 (0.99 - 1.00)	0.414
PC7	SC	0.99 (0.99 - 1.00)	0.62
PC8	SC	1.00 (0.99 - 1.00)	0.759
PC9	SC	1.02 (1.01 - 1.021)	5.43e-09
PC10	SC	1.01 (1.01 - 1.014)	0.000425

(A): multivariable logistic regression; covariates include age, sex & 10 PCs

(B): univariable logistic regression. SC = Severe Covid; PC = Principal Component

Supplementary Table 3. Effects of categories (quintiles) of the T2D PRS on SC within Covid(+) strata and vs. Covid(-) controls.

Quintile	OR (95% CI)	p-value
within Covid(+)	(N=112840)	
1st quintile (0 - 20%)	0.92 (0.86 - 0.99)	2.20e−2
2nd quintile (20 - 40%)	1.00 (0.92 - 1.06)	8.26e−1
3rd quintile (40 - 60%)	reference	
4th quintile (60 -80%)	1.09 (1.02 - 1.17)	1.39e−2
5th quintile (80 - 100%)	1.18 (1.10 - 1.26)	1.90e−6
vs. Covid(-) controls	(N=346653)	
1st quintile (0 - 20%)	0.92 (0.86 - 0.99)	3.18e−2
2nd quintile (20 - 40%)	1.00 (0.93 - 1.07)	8.49e−1
3rd quintile (40 - 60%)		
4th quintile (60 -80%)	1.09 (1.02 - 1.17)	1.07e−2
5th quintile (80 - 100%)	1.18 (1.10 - 1.27)	8.12e−7

Predictor, T2D PRS. Covariates: age, sex, 1st 10 PCs

Reference group for categorical analysis: middle (40%- 60%) quintile

Supplementary Table 4. Effect of BMI: Without BMI (left) or including BMI as a covariate in regression analyses among all subjects.

		without BMI		plus BMI	
Cases	Controls	OR (95% CI)	p-value	OR (95% CI)	p-value
8367	451126	1.09 (1.06 - 1.11)	1.7e-12	1.04 (1.02 - 1.06)	0.00073

Predictor, T2D PRS. Covariates: Age, Sex & 1st 10 PCs & +/- BMI