Where next for diabetes research?

RESOURCES DOCUMENT
Prevalence of diabetes among Chinese patients with SARS-CoV-2

Preval (95% CI)                  % Weight
0.02  (0.00,  0.07)       7.4
0.06  (0.02,  0.11)       8.7
0.06  (0.02,  0.13)       8.2
0.07  (0.06,  0.09)       13.7
0.08  (0.00,  0.24)       4.2
0.10  (0.07,  0.13)       100.0
0.10  (0.06,  0.16)       10.0
0.10  (0.06,  0.16)       10.1
0.12  (0.06,  0.19)       9.0
0.19  (0.14,  0.25)       11.0
0.20  (0.09,  0.33)       5.9
0.11  (0.00,  0.42)       2.0
0.12  (0.06,  0.19)       9.0
0.12  (0.07,  0.18)       10.1
0.19  (0.14,  0.25)       11.0
0.10  (0.07,  0.13)       100.0

Q=35.76, p=0.00, I²=69%

Fadini et al. J Endocrinol Invest 2020
Prevalence and impact of diabetes among people infected with SARS-CoV-2

G. P. Fadini1 · M. L. Morieri1 · E. Longato2 · A. Avogaro1

Received: 15 March 2020 / Accepted: 20 March 2020 © Italian Society of Endocrinology (SIE) 2020

To the Editor

In December 2019, a new coronavirus causing severe acute respiratory syndrome (SARS-CoV-2) emerged in Wuhan, China. The virus spread rapidly to more than 150 countries and, by the time we are writing (March 19th 2020), it affected > 230,000 individuals causing almost 10,000 deaths [1].

It has been suggested that diabetes mellitus is one of the most common comorbidities in infected people, but its exact prevalence is unclear. We show results of a meta-analysis of studies reporting the prevalence of diabetes among people infected with the SARS-CoV-2 and its impact on disease severity or progression. We retrieved 12 studies reporting data from 2108 Chinese patients with confirmed SARS-CoV-2 infection (Data supplement). Mean age was 49.6 years and the prevalence of diabetes was 10.3% (Fig. 1a). For comparison, the nationwide prevalence of diabetes in China in 2013 was 10.9% overall and 12.3% among people aged 40–59 [2].

As of March 19th 2020, Italy was the second country most affected with SARS-Cov-2 (n = 41,035 individuals with confirmed SARS-CoV-2). At the University Hospital of Padova, located at the centre of an outbreak, among 146 hospitalized patients with confirmed SARS-CoV-2 infection and a mean age of 65.3 years, 13 had pre-existing diabetes, yielding a prevalence of 8.9% (95% CI 5.3–14.6). For comparison, the prevalence of diabetes in the same Region in 2018 was 6.2% overall and 11.0% among people aged 55–75 years (mean 65) [3]. A relatively low prevalence of diabetes among SARS-CoV-2 infected people could be due to under-reporting, chance, or a biological phenomenon. Dipeptidyl peptidase-4 (DPP-4) was discovered as a receptor for middle-east respiratory syndrome coronavirus while SARS-CoV-2 appears to use preferentially angiotensin-converting enzyme 2 to enter the cell [4]. However, since DPP-4 inhibitors are popular glucose-lowering medications worldwide, it will be of interest to explore whether they might protect against SARS-CoV-2 infection.

Six of the meta-analysed studies on Chinese patients reported the prevalence of diabetes according to disease severity or outcome (n = 1687 patients). The pooled rate ratio of diabetes among patients with adverse disease course as compared to those with the more favourable course was 2.26 (95% CI 1.47–3.49) (Fig. 1b and Supplemental Table 1).

As of March 17th, the median age of 2003 Italian patients who died while being infected with SARS-CoV-2 was 80.5 (IQR 31–103) and 70% were men. Among 355 deceased patients with available information on comorbidities, diabetes prevalence was 35.5% [5]. In 2018, diabetes prevalence among Italian citizens with the same age range and sex distribution was 20.3% [6]. Thus, the rate ratio of diabetes among patients who died with SARS-CoV-2 infection compared to the general population was 1.75.

Based on these data, we conclude that diabetes may not increase the risk of SARS-CoV-2 infection but can worsen the outcome of this new coronavirus disease. This finding is consistent with the association between diabetes and excess mortality from any acute and chronic condition, including infections [7]. The situation is rapidly evolving, and close monitoring of these data are important for a correct communication to patients and healthcare providers.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40618-020-01236-2) contains supplementary material, which is available to authorized users.

* G. P. Fadini
  gianpaolo.fadini@unipd.it

1 Department of Medicine, University of Padova, Via Giustiniani 2, 35128 Padua, Italy
2 Department of Information Engineering, University of Padova, Padua, Italy

Published online: 28 March 2020
**Fig. 1** a Forest plot of diabetes prevalence among SARS-CoV-2 infected patients. b Forest plot of diabetes rate ratio (RR) among patients with more severe versus those with less severe infection.

### Funding
None.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no competing interest.

**Ethical approval** The study has been conducted in compliance with ethical standards for research involving human participants. Protocol approval for evaluation of the healthcare status in diabetic patients was obtained from the Ethical Committee of the University Hospital of Padova.

**Informed consent** No informed consent was collected for this study because we used aggregated data or data that are publicly available.

### References


Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses

Michael Letko, Andrea Marzi and Vincent Munster

Over the past 20 years, several coronaviruses have crossed the species barrier into humans, causing outbreaks of severe, and often fatal, respiratory illness. Since SARS-CoV was first identified in animal markets, global viromics projects have discovered thousands of coronavirus sequences in diverse animals and geographic regions. Unfortunately, there are few tools available to functionally test these viruses for their ability to infect humans, which has severely hampered efforts to predict the next zoonotic viral outbreak. Here, we developed an approach to rapidly screen lineage B betacoronaviruses, such as SARS-CoV and the recent SARS-CoV-2, for receptor usage and their ability to infect cell types from different species. We show that host protease processing during viral entry is a significant barrier for several lineage B viruses and that bypassing this barrier allows several lineage B viruses to enter human cells through an unknown receptor. We also demonstrate how different lineage B viruses can recombine to gain entry into human cells, and confirm that human ACE2 is the receptor for the recently emerging SARS-CoV-2.

Severe acute respiratory syndrome-related coronavirus (SARS-CoV) first emerged in humans in 2003 after transmitting from animals in open-air markets in China. Shortly thereafter, several genetically related viruses were identified in Chinese horseshoe bats (Rhinolophus sinicus). At the same time, improvements in next-generation sequencing technology led to a boom of virus discovery, uncovering thousands of virus sequences in wild animal populations around the world. While most of these viruses have never been found in humans, many are genetically similar to known human viruses within the betacoronavirus genus. The betacoronaviruses are further divided into four lineages (that is, A–D). Lineage B, which includes SARS-CoV and the newly emerging SARS-CoV-2, has approximately 200 published virus sequences, whereas lineage C, which includes Middle East respiratory syndrome-related coronavirus (MERS-CoV), has over 500 viral sequences.

Every year, additional CoV sequences are discovered. However, there is a massive knowledge gap in the field, as very little work is performed after the viral sequences are published. Therefore, it is unknown whether these viruses have the potential to emerge in human populations.

Current methods for studying betacoronaviruses are technically demanding. Viral isolation from field samples is rarely successful and reverse genetics recovery of recombinant virus is labor intensive and expensive, as the synthesis of a single genome can cost upwards of US$15,000. These limitations are prohibitive to studying CoVs at the scale at which they are discovered.

Cell entry is an essential component of cross-species transmission, especially for the betacoronaviruses. All CoVs encode a surface glycoprotein, spike, which binds to the host-cell receptor and mediates viral entry. For betacoronaviruses, a single region of the spike protein called the receptor-binding domain (RBD) mediates the interaction with the host-cell receptor. After binding the receptor, a nearby host protease cleaves the spike, which releases the spike fusion peptide, facilitating virus entry.

Known host receptors for betacoronaviruses include angiotensin-converting enzyme 2 (ACE2) for SARS-CoV and dipeptidyl peptidase-4 (DPP4) for MERS-CoV. Structural studies of coronaviruses have shown that the spike RBD is capable of folding independent of the rest of the spike protein and contains all of the structural information for host receptor binding. Additionally, a previous study showed that replacing the RBD of the lineage B bat virus Rp3 allowed the virus to enter cells expressing human ACE2. We therefore developed a method to functionally test the RBDs from lineage B betacoronaviruses in place of the SARS-CoV spike RBD (Fig. 1). Synthesizing just the RBD of spike is much faster and cost effective than conventional pseudotyping methods that rely on the synthesis of the full spike sequence (~4 kilobases) for CoVs—a process that can take weeks and is cost prohibitive for large panels of spike sequences. The short turnaround time for our approach allowed us to test the receptor usage of all published, unique RBD sequences in lineage B, and to rapidly confirm the ACE2 receptor usage of the SARS-CoV-2 spike, which emerged in China in January 2020 as our study was ongoing.

We show that lineage B RBDs divide into functionally distinct clades and that several previously unappreciated viruses exhibit compatibility with an unknown receptor on human cells. We also show that these clades are capable of recombining to impart human host-cell entry phenotypes, and that, beyond the RBD–receptor interaction, host protease processing is another species barrier encountered by lineage B betacoronaviruses during cell entry.

Results

ACE2 entry is lineage B clade 1 specific. The RBD of lineage B betacoronaviruses is a single, continuous domain that contains all of the structural information necessary to interact with the host receptor (Fig. 1a,b). We introduced silent mutations in the codon-optimized coding sequence for SARS-CoV to facilitate replacing the SARS RBD with the RBD from other lineage B viruses (Fig. 1b). All lineage B sequences were downloaded from online repositories...
Fig. 1 | Betacoronavirus lineage B entry with human ACE2 is clade specific. a, Betacoronaviruses, including SARS-CoV, interact with the host-cell receptor via the RBD in spike (Protein Data Bank ID: 5X5B; 2AJF). b, Engineered silent mutations in SARS spike facilitated replacement of the RBD sequence. SARS spike amino acid numbers are indicated in black for the silent cloning sites and orange for the RBD. c, Outline of the experimental workflow. d, Western blot of producer cell lysates and concentrated reporter particles. The labels along the top show the origin of the RBD in the SARS-CoV spike protein. e, Cladogram of the 29 spikes tested. Cells expressing either human ACE2 or empty vector were infected with pseudotyped VSV reporter particles, and luciferase was measured and normalized to no spike as a readout for cell entry. The data are representative of three technical replicates. Vertical bars indicate mean values of all three replicates and horizontal bars indicate s.d.

and parsed to 29 unique RBD sequences, representing all published variations of the lineage B RBD (Extended Data Fig. 1a,b). These 29 RBDs phylogenetically cluster into three clades, as previously described, but these RBD clades were not apparent in phylogenetic analysis of other viral sequences, such as the RNA-dependent RNA polymerase (Extended Data Fig. 1c). All 29 RBDs were codon optimized, synthesized and cloned in place of the SARS RBD, generating chimeric spike expression constructs. We then produced vesicular stomatitis virus (VSV) reporter particles pseudotyped with the chimeric spikes (Fig. 1c). We chose VSV over lentiviruses as our pseudotype platform because lentiviral pseudotypes have variable receptor specificity compared with the wild-type SARS spike (Fig. 1d). We then infected baby hamster kidney (BHK) cells expressing the receptor for SARS-CoV or empty vector (Fig. 1e) and observed that only clade 1, which includes SARS-CoV, could enter cells transfected with human ACE2 (Fig. 1e).

**Protease enhances clade 2 entry.** After binding the host receptor, host-cell protease cleaves spike, releasing the fusion peptide and allowing for host-cell entry. Previous studies have shown that an absence of the host protease or incompatibility between the host protease and viral spike can block viral entry. To circumvent host-cell protease incompatibility or absence, we protease treated our pseudotype panel and infected a wide variety of cell types from different host species (Fig. 2 and Extended Data Fig. 2a,b). In the absence of exogenous protease, only clade 1 infected cells from African green monkey kidney, human gastrointestinal tract, human liver and porcine kidney, in agreement with previous studies. Surprisingly, exogenous protease enhanced entry of a subset of clade 2 spike chimeras in nonhuman primate, bat and human cells (Fig. 2). Importantly, VSV-G-pseudotyped particles were able to produce luciferase signal in all cell lines tested in this study (Extended Data Fig. 2).

**Clade 2 entry is receptor dependent.** Next, we tested human variants of known betacoronavirus receptors for their ability to mediate cell entry of clade 2 and 3 spike chimeras. We also tested human aminopeptidase N (APN)—a common receptor for alphacoronaviruses (Fig. 3a). Protease treatment only enhanced entry of clade 1 RBDs on cells expressing human ACE2, but not human DPP4 or APN. No entry was observed with clade 2 or 3 spikes, regardless of receptor or protease addition. Human DPP4—a receptor for the lineage C betacoronaviruses—only mediated entry of MERS-CoV, whereas APN only mediated entry of the alphacoronavirus HCoV-229E
(Fig. 2b). In the absence of receptor, no entry was observed for any of the pseudotypes, suggesting that protease-mediated entry is receptor dependent (Fig. 2b).

**Receptor usage of SARS-CoV-2.** While our study was ongoing, a lineage B virus tentatively named SARS-CoV-2 was identified as the cause of a pneumonia outbreak in Hubei, China. Once the sequence was publicly available, we synthesized, cloned and tested the RBD from SARS-CoV-2 in our assay with human variants of known coronavirus receptors. The chimeric SARS–SARS-CoV-2 spike protein expressed was incorporated into particles similarly to other clade 1 chimeric spikes and was capable of entering cells expressing human ACE2, but not any of the other receptors tested (Fig. 3c,d and Extended Data Fig. 3).

**Clade determinants for ACE2 usage.** Consensus sequences of the three lineage B clades showed several key differences between these groups. Only clade 1 RBDs contain all 14 residues that have been shown through crystallography to interact with human ACE2 (Fig. 4a and Extended Data Fig. 4). The majority of these residues are absent from clades 2 and 3, which contain additional deletions in surface-exposed loops that cluster at the interface with ACE2 (Fig. 4a,b). We generated a series of clade consensus RBD variants to determine the minimum number of mutations needed to impart ACE2 function on clade 2 and 3 RBDs (Fig. 4c). Introducing the two loop deletions from clade 1 in clade 2 resulted in reduced spike expression, impaired pseudotype incorporation and loss of cell entry (Fig. 4c,d). Restoring these loops in clades 2 and 3 from the loops found in clade 1 did not enhance entry with ACE2.
Full spike and RBD chimeras are comparable. Next, we synthesized full-length clade 2 and 3 spikes to compare with our RBD chimera.

**Fig. 3 | Lineage B entry into cells with known CoV receptors.**

**a.** Schematic of known coronavirus spikes and their receptors. N-terminal RBD found in spike S1 of lineage A betacoronaviruses and gammacoronaviruses is indicated in orange. C-terminal RBD found in spike S1 of alphacoronaviruses and lineage B betacoronaviruses is indicated in blue. CECA1, carcinoembryonic antigen-related cell adhesion molecule 1; HCoV, human coronavirus; IBV, infectious bronchitis virus; MHV, mouse hepatitis virus; TGEV, transmissible gastroenteritis coronavirus.

**b, c.** Pseudotyped particles were either left untreated or treated with trypsin and subsequently used to infect BHK cells expressing the indicated coronavirus receptors. WT, wild type.

**d.** Pseudotypes were used to infect BHK cells expressing known receptors without protease treatment. The x axis labels indicate the origin of the RBD in the SARS spike protein. Data for all panels represent three technical replicates. Vertical bars indicate mean values of all three replicates and horizontal bars indicate s.d.

(Fig. 4c) 2 → 1 and 3 → 1 (version 1)). Introducing all 14 ACE2 contact points in clade 2 or 3 also failed to restore ACE2 entry (Fig. 4c; 2 → 1 and 3 → 1 (version 2)). Only replacing all 14 contact points and the surrounding amino acids (known as the receptor-binding motif (RBM)) led to increased ACE2 entry with clade 2 and 3 RBDs (Fig. 4c; 2 → 1 (version 3) = clade 2 residues 322–400 + clade 1 residues 400–501; 3 → 1 (version 3) = clade 3 residues 322–385 + clade 1 residues 386–501).

**Discussion**

Despite significant advances in next-generation sequencing technologies, which have facilitated the discovery of thousands of animal-derived viruses, tools for downstream functional assessment of
these sequences are lacking. To gain traction on this ever-growing problem, we took a reductionist approach to coronavirus entry and developed a scalable, biosafety level 2-compatible method for testing only the minimal region of the virus essential for interacting with the host receptor (Fig. 1 and Extended Data Fig. 1a). Because most of these viruses have never been isolated, we resorted to synthetic biology and molecular engineering to reduce the burden of gene synthesis to just a small fragment. Thus, the cost and synthesis production time for testing several spikes for entry in our system is dramatically reduced (Extended Data Fig. 1d). In theory, this approach to functional viromics should be applicable to a wide variety of virus–host proteins and interactions.

Coronavirus entry is a multi-step process involving several distinct domains in spike that mediate virus attachment to the cell surface, receptor engagement, protease processing and membrane fusion. While the RBD–receptor interaction is the most studied in the absence of exogenous protease, only clade 1 RBDs entered nonhuman primate, human and porcine cell lines (Fig. 2a,b). These findings are in agreement with previous studies that have either isolated virus (WIV1) or rescued recombinant chimeric viruses (SHC014, Rs4231 and Rs7237)\(^1\). However, with trypsin, a subset of genetically similar clade 2 RBDs also observed cell rounding during their trypsin infections. Therefore, further studies are needed to assess where trypsin is enhancing entry of coronaviruses: at the level of spike, the receptor, or both.

In the absence of exogenous protease, only clade 1 RBVs entered nonhuman primate, human and porcine cell lines (Fig. 2a,b). These findings are in agreement with previous studies that have either isolated virus (WIV1) or rescued recombinant chimeric viruses (SHC014, Rs4231 and Rs7237)\(^1\). However, with trypsin, a subset of genetically similar clade 2 RBVs gained entry in these cells, suggesting that their barrier is at the level of protease processing (Fig. 2a,b). The other spikes from clade 2 and 3 did not enter the cells we tested, regardless of protease addition, suggesting an absent or incompatible receptor. Surprisingly, the protease-dependent entry phenotype was consistent in the reverse spike chimeras in which we replaced the RBD in clade 2 or 3 spike
with a clade 1 RBD (Fig. 5d), suggesting that either the protease site between S1 and S2 is not compatible with the chimeric spike backbone or the protease is not expressed in these cells (Extended Data Fig. 5). Because clade 1 spikes enter cells expressing human ACE2 without the addition of protease, but clade 2–clade 1 chimeras require protease, our data show that the spike protease cleavage site may be adapted to the protease environment of the receptor-bound RBD (Extended Data Fig. 5).

None of the spike pseudotypes efficiently entered *Rhinolophus* cells, as was observed in previous studies using these cells29,30 (Fig. 2c). Surprisingly, Aji cells were selectively permissive for only clade 2 entry following protease treatment, which suggests that clade 2 spikes enter cells expressing human ACE2 without the addition of protease, but clade 2–clade 1 chimeras require protease, our data show that the spike protease cleavage site may be adapted to the protease environment of the receptor-bound RBD (Extended Data Fig. 5).

None of the spike pseudotypes efficiently entered *Rhinolophus* cells, as was observed in previous studies using these cells29,30 (Fig. 2c). Surprisingly, Aji cells were selectively permissive for only clade 2 entry following protease treatment, which suggests that clade 2 spikes enter cells expressing human ACE2 without the addition of protease, but clade 2–clade 1 chimeras require protease, our data show that the spike protease cleavage site may be adapted to the protease environment of the receptor-bound RBD (Extended Data Fig. 5).

None of the spike pseudotypes efficiently entered *Rhinolophus* cells, as was observed in previous studies using these cells29,30 (Fig. 2c). Surprisingly, Aji cells were selectively permissive for only clade 2 entry following protease treatment, which suggests that clade 2 spikes enter cells expressing human ACE2 without the addition of protease, but clade 2–clade 1 chimeras require protease, our data show that the spike protease cleavage site may be adapted to the protease environment of the receptor-bound RBD (Extended Data Fig. 5).

None of the spike pseudotypes efficiently entered *Rhinolophus* cells, as was observed in previous studies using these cells29,30 (Fig. 2c). Surprisingly, Aji cells were selectively permissive for only clade 2 entry following protease treatment, which suggests that clade 2 spikes enter cells expressing human ACE2 without the addition of protease, but clade 2–clade 1 chimeras require protease, our data show that the spike protease cleavage site may be adapted to the protease environment of the receptor-bound RBD (Extended Data Fig. 5).

Our results show that, despite all of them being classified as the same virus species, most lineage B betacoronaviruses do not use currently known coronavirus receptors (Figs. 1e and 3a,b). We did not observe any pseudotype entry in the presence of protease and the absence of receptor, suggesting that lineage B cell entry is still receptor dependent following protease treatment (Fig. 3b). While our study was ongoing, a lineage B betacoronavirus was identified as the etiological agent behind an outbreak of pneumonia in Wuhan, Hubei, China (SARS-CoV-2). The RBD for SARS-CoV-2 has residues and motifs found in all three clades but forms a distinct clade, so we tested it for receptor usage and observed entry of human ACE2 but not other known coronavirus receptors (Fig. 3d). Interestingly, within the backbone of SARS-CoV spike, cell entry of SARS-CoV-2 was similar to the other clade 1 spikes tested, including SARS-CoV. These findings suggest that SARS-CoV-2 is capable of using human ACE2 as efficiently, if not more, as SARS-CoV, which may help to explain the human-to-human transmissibility of this virus. Our finding that SARS-CoV-2 uses human ACE2 has now been confirmed with the full virus and a computer modelling-based approach32,33.

The RBM is a region in the carboxy-terminal half of the RBD that contains all of the residues that interface with the host receptor (Fig. 3a)35. The 14 contact points in the co-structure of the SARS
These engineered spike sequences were synthesized and cloned into pcDNA3.1+ (GenScript).

Spike RBDs were first codon optimized for human cells, appended with regions of the target spike backbone to facilitate In-Fusion cloning and synthesized as double-stranded DNA fragments (Integrated DNA Technologies). SARS-CoV, As6526 or BM48-31 engineered spike plasmids were digested with their corresponding restriction enzymes and gel purified. RBD inserts were resuspended in water and In-Fusion cloned into gel-purified, digested spike backbone vectors (Takara).

For Human ACE2 (Q8BYF1), DPP4 (XM_005246731.3) or APN (NP_001141.2) were synthesized and cloned into pcDNA3.1+ (GenScript). All DNA constructs were verified by Sanger sequencing (ACGT).

Receptor transfection. BHK cells were seeded in black 96-well plates and transfected the next day with 100 ng plasmid DNA encoding human ACE2, DPP4, APN or empty vector, using polyethylenimine (Polysciences). All downstream experiments were performed 24 h post-transfection.

Pseudotype production. To produce seed particles (VSVΔG-luc/GFP + VSV-G), 293T cells were seeded in six-well plates and transfected 24 h later with 2 μg VSV-luc/GFP + 2 μg T7 polymerase, 0.3 μg VSVN, 0.25 μg VSVL, 1.25 μg VSV-P and 1 μg VSV-G. VSV seed particles were harvested 48 h post-transfection. Cell supernatants were collected, cleared from cell debris by centrifugation, aliquoted and stored at −80°C.

CoV spike pseudotypes were produced as previously described. 293T cells were seeded onto six-well plates pre-coated with poly-L-lysine (Sigma–Aldrich) and transfected the next day with 1,200 ng of empty plasmid and 400 ng of plasmid encoding coronavirus spike or green fluorescent protein (GFP) as a no-spike control. After 24 h, transfected cells were infected with VSVΔG particles pseudotyped with VSV-G, as previously described. After 1 h of incubating at 37°C, cells were washed three times and incubated in 2 ml DMEM supplemented with 1% FBS, penicillin/streptomycin and L-glutamine for 48 h. Supernatants were collected, centrifuged at 500g for 5 min, aliquoted and stored at −80°C.

Luciferase-based cell entry assay. Target cells were seeded in black 96-well plates and inoculated, in triplicate, with equivalent volumes of pseudotype stocks. For the tetrans experiments, pseudotype stocks were diluted 1:1 in DMEM without FBS, tetrans was added to a final concentration of 2,500 μg/ml−1 and the samples were incubated at 37°C for 15 min. The samples were then diluted again 1:1 in cold DMEM supplemented with 2% FBS, then added to cells. Inoculated plates were centrifuged at 1,200g at 4°C for 1 h and incubated overnight at 37°C. Approximately 18–20 h post-infection, Bright-Glo luciferase reagent (Promega) was added to each well, 1:1, without removing the culture media, and luciferase was measured. Relative entry was calculated as the fold-entry over the negative control, by normalizing the relative light unit for spike pseudotypes to the plate relative light unit average for the no-spike control.

Western blot. Producer cells (spike-transfected 293T) were lysed in 1% sodium dodecyl sulfate, 150 mM NaCl, 50 mM Tris-HCl and 5 mM EDTA and clarified by centrifugation at 14,000g for 20 min.

Pseudotyped particles were concentrated from producer cell supernatants that were overlaid on a 10% OptiPrep cushion in PBS (Sigma–Aldrich) and centrifuged at 22,000g for 2 h at 4°C. lysates and concentrated particles were analysed by FLAG (Sigma–Aldrich; A8592; 1:10,000), GADPH (Sigma–Aldrich; G8795; 1:10,000) and/or VSV-M (Kerafast; 23H12; 1:5,000) expression on 10% Bis–Tris PAGE gel (Thermo Fisher Scientific).

Statistics and reproducibility. Each figure shows the data for three technical replicates, with the average of these replicates indicated. During the course of this study, these experiments were performed multiple times with different batches of pseudotypes and cells, and performed at different times. Thus, the results shown are representative of the biological replicates we observed. Four separate clones were recovered in parallel for the SARS–SARS-CoV-2 construct and were tested simultaneously, in triplicate. Because all four clones were identical by Sanger sequencing and behaved similarly in our entry assays, representative results with one clone are shown in Fig. 3c,d.

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

Data availability. Unprocessed gel images for Figs. 1 and 3–5 and tables of luciferase assay data for Figs. 1–5 and Extended Data Fig. 2 are included as source data. Any other data that support the findings of this study are available from the corresponding author upon request. Accession codes for all of the spike sequences used here can be found in Extended Data Fig. 1.

Received: 3 February 2020; Accepted: 11 February 2020; Published online: 24 February 2020
References


Acknowledgements

We thank A. Schountz for generously sharing primary Artibeus cells, as well as M. Muller for providing the Rhinolophus cells used in this study. We also acknowledge F. Wu, S. Zhao, R. Yu, Y.-M. Chen, W. Wang, Y. Hu, Z.-G. Song, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-Z. Zheng, L. Xu, E. C. Holmes and Y.-Z. Zhang, as well as the Wuhan Institute of Virology, Chinese Academy of Medical Sciences and Peking Union Medical College, Chinese Center for Disease Control and Prevention, and Wuhan Jinyintan Hospital for rapidly providing the genome sequence for SARS-CoV-2 and other essential outbreak information to the scientific community. This work was supported by the Intramural Research program of the National Institute of Allergy and Infectious Diseases.

Author contributions

M.L. conceived of and designed the study, generated the RaKSM-2.5 and AJ cell lines, performed all of the experiments, analysed all of the data, assembled the figures and wrote the manuscript. A.M. generated the dual-reporter VSV system used for the pseudotype assays. V.M. secured funding and supervised the study. All authors contributed to the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-020-0688-y. Extended data is available for this paper at https://doi.org/10.1038/s41564-020-0688-y. Correspondence and requests for materials should be addressed to M.L. or V.M.

Reprints and permissions information is available at www.nature.com/reprints.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2020.
Extended Data Fig. 1 | Lineage B RBD panel assembly and phylogenetic analysis. a, Coronavirus sequences were downloaded from NCBI and further parsed to 29 unique RBD variants. b, Virus isolate name, accession number, clade, host species and location of identification listed for the 29 unique lineage B RBDs used in this study. c, Cladograms for the spike RBD and coronavirus RNA-dependent RNA polymerase (nsp12). d, Overview of experimental timeline.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Additional cell lines tested without protease. a, Additional human and cell lines or b, bat-derived cell lines derived from other species were infected with VSV-reporter particles pseudotyped with chimeric spikes and luciferase was measured a readout for cell entry. Trypsin was not used in these infections. c, All cell lines tested in this study supported entry and reporter expression of VSV-g pseudotyped particles. For all panels shown are the data for 3 technical replicates, horizontal bars indicate the s.d. and vertical bars indicate the mean value of all 3 replicates.
Extended Data Fig. 3 | 2019-nCoV uses human ACE2 to enter cells. VSVΔG-luciferase/GFP particles were pseudotyped with the indicated spikes and used to infect BHKs transfected with known coronavirus receptors. Microscopy images were taken 20 hours post-infection. Scale bar indicates 1000 μm.
Extended Data Fig. 4 | Lineage B panel RBD sequence features. a. Amino acid sequences corresponding to SARS-spike residues 317 through 500 were aligned with ClustalW. Contact points between SARS-spike and human ACE2 are indicated with an (*). Clade 2 sequences are shown as compared to clade 2 As6526, with identical residues indicated with a (.) and sites that vary between clade 2 viruses highlighted in purple. Loop deletions are highlighted in orange. b. Amino acid alignment of 2019-nCoV RBD and consensus RBD sequences for clade 1 and 2 and BM48-31 (clade 3). Loop deletions are highlighted in orange.
Extended Data Fig. 5 | Model of lineage B entry. a, SARS spike-clade 1 RBD enters cells expressing ACE2 and a host protease capable of cleaving SARS spike (left panel). While clade 2 RBDs can bind an unknown host receptor, the SARS spike backbone is incompatible with the receptor-associated protease, resulting in a lack of cleavage and entry (middle panel). The addition of exogenous protease may overcome the lack of endogenous protease cleavage of spike, resulting in receptor-dependent entry. Alternatively, the addition of exogenous protease may activate the receptor to facilitate entry. b, Replacing the RBD of As6526 spike with the clade 1 RBD allows for ACE2 interaction, but the As6526 spike backbone is incompatible with the ACE2-associated protease (left panel). Addition of exogenous protease overcomes protease incompatibility, allowing for ACE2-mediated entry (right panel).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated.

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
Raw luciferase data was exported to Microsoft Excel (v. 16.33).

Data analysis
Luciferase data was analyzed in Microsoft Excel (v. 16.33). Graphs of luciferase data were generated in Prism 8. Protein structures were rendered in PyMol (v.2.3.2) using existing, published structural data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Accession numbers for all viral sequences used in this study are provided in extended data figure 1b. Accession numbers for human ACE2, APN, and DPP4 are provided in the methods section, under “plasmids.” Unprocessed western blot images and graphed values are provided as source data. All reagents are freely available upon request.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample-size calculations were not performed. For each pseudotype in our study, we infected cells in triplicate to demonstrate magnitude and consistency of measurable differences.

Data exclusions

We did not exclude any data for the experiments presented.

Replication

The figures in our manuscript are “final” experiments after many months of preliminary data collection. All viral pseudotypes in our study were generated and tested multiple times before we produced large batches of pseudotypes for the “final” experiments. These results were extremely consistent between different batches of pseudotypes produced at different times and in cells of different passage numbers.

Randomization

No specific steps were taken to randomize experimental groups. This is because the generally binary outcome of the assay performed left results to little interpretation.

Blinding

All pseudotypes were numbered 1-29 for ease during our experiments. No further steps were taken to blind the investigator, because the data was generally binary - there either was or was not cell entry, leaving little room for investigator bias during data collection.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology
- Animals and other organisms
- Human research participants
- Clinical data

Methods

n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

Antibodies

Antibodies used

FLAG antibody for western blot - Sigma - A8592
GAPDH antibody for western blot - Sigma - GB795
anti-VSV M - Kerafast Inc., Boston, MA - 23H12

Validation

These FLAG and GAPDH antibodies are well established, commercially available, have been validated by the company and have been used by our group and many others (for example PMID: 30572566, 28031368, 26628364).

Anti-VSV-m has been used extensively in immunofluorescence labeling and western blot analysis by Andrea Marzi’s lab and others (for example PMID: 30038228, 26091335, 12134006).

Eukaryotic cell lines

Policy information about cell lines

1. A549 - human lung epithelial - ATCC CCL-185
2. Artibeus jamaicensis - primary kidney - obtained from Anthony Schountz, Colorado State University (PMID: 26899616)
3. Artibeus jamaicensis - immortalized cells - AI-primary cells above were immortalized with SV40 T-antigen as described in our manuscript
4. BHK - hamster kidney - ATCC CCL-10
5. Caco-2 - human colon epithelial cells - ATCC HTB-37
<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huh-7.5</td>
<td>Human liver - obtained from Heinz Feldmann laboratory at Rocky Mountain Labs, Hamilton, MT, 59840</td>
</tr>
<tr>
<td>HypNi</td>
<td>Hypsignathus monstrosus kidney - obtained from Marcel Müller at Institute of Virology Charite - Universitätsmedizin Berlin (PMID: 25100832)</td>
</tr>
<tr>
<td>PK-15</td>
<td>Porcine kidney cells - ATCC CCL-33</td>
</tr>
<tr>
<td>RaKSM-2.5</td>
<td>Rousettus aegyptiacus kidney - generated in our lab and previously published (PMID: 31682727)</td>
</tr>
<tr>
<td>RhiLu</td>
<td>Rhinolophus aegyptiacus lung - obtained from Marcel Müller at Institute of Virology Charite - Universitätsmedizin Berlin (PMID: 23232719)</td>
</tr>
<tr>
<td>RhiNi</td>
<td>Rhinolophus landeri kidney - obtained from Marcel Müller at Institute of Virology Charite - Universitätsmedizin Berlin (<a href="https://web.expasy.org/cellosaurus/CVCL_RX64">https://web.expasy.org/cellosaurus/CVCL_RX64</a>)</td>
</tr>
<tr>
<td>Vero</td>
<td>African green monkey kidney - obtained from Heinz Feldmann laboratory at Rocky Mountain Labs, Hamilton, MT, 59840</td>
</tr>
</tbody>
</table>

### Authentication

- Cell species was confirmed by Cytochrome B sequencing

### Mycoplasma contamination

- All cell lines tested negative for mycoplasma by standard PCR prior to experiments

### Commonly misidentified lines

(See ICLAC register)

- Name any commonly misidentified cell lines used in the study and provide a rationale for their use.
Diabetes Mellitus, Fasting Glucose, and Risk of Cause-Specific Death

Sreenivasa Rao Kondapally Seshasai, M.D.,
University of Cambridge, Cambridge, United Kingdom

Stephen Kaptoge, Ph.D.,
University of Cambridge, Cambridge, United Kingdom

Alexander Thompson, Ph.D.#,
University of Cambridge, Cambridge, United Kingdom

Emanuele Di Angelantonio, M.D.#,
University of Cambridge, Cambridge, United Kingdom

Pei Gao, Ph.D.#,
University of Cambridge, Cambridge, United Kingdom

Nadeem Sarwar, Ph.D.#,
University of Cambridge, Cambridge, United Kingdom

Peter H. Whincup, F.R.C.P.,
St. George’s University of London, London

Kenneth J. Mukamal, M.D.,
Harvard University, Boston

Richard F. Gillum, M.D.,
Centers for Disease Control and Prevention, Atlanta

Ingar Holme, Ph.D.,
Ullevål University Hospital, Oslo

Inger Njølstad, M.D.,
University of Tromsø, Tromsø, Norway

Astrid Fletcher, Ph.D.,
London School of Hygiene and Tropical Medicine, London

Peter Nilsson, M.D.,
Lund University, Lund, Sweden
Abstract

BACKGROUND—The extent to which diabetes mellitus or hyperglycemia is related to risk of death from cancer or other nonvascular conditions is uncertain.

METHODS—We calculated hazard ratios for cause-specific death, according to baseline diabetes status or fasting glucose level, from individual-participant data on 123,205 deaths among 820,900 people in 97 prospective studies.

RESULTS—After adjustment for age, sex, smoking status, and body-mass index, hazard ratios among persons with diabetes as compared with persons without diabetes were as follows: 1.80 (95% confidence interval [CI], 1.71 to 1.90) for death from any cause, 1.25 (95% CI, 1.19 to 1.31) for death from cancer, 2.32 (95% CI, 2.11 to 2.56) for death from vascular causes, and 1.73 (95% CI, 1.62 to 1.85) for death from other causes. Diabetes (vs. no diabetes) was moderately associated with death from cancers of the liver, pancreas, ovary, colorectum, lung, bladder, and breast. Aside from cancer and vascular disease, diabetes (vs. no diabetes) was also associated with death from renal disease, liver disease, pneumonia and other infectious diseases, mental disorders, nonhepatic digestive diseases, external causes, intentional self-harm, nervous-system disorders, and chronic obstructive pulmonary disease. Hazard ratios were appreciably reduced after further adjustment for glycemia measures, but not after adjustment for systolic blood pressure, lipid levels, inflammation or renal markers. Fasting glucose levels exceeding 100 mg per deciliter (5.6 mmol per liter), but not levels of 70 to 100 mg per deciliter (3.9 to 5.6 mmol per liter), were associated with death. A 50-year-old with diabetes died, on average, 6 years earlier than a counterpart...
without diabetes, with about 40% of the difference in survival attributable to excess nonvascular deaths.

**CONCLUSIONS**—In addition to vascular disease, diabetes is associated with substantial premature death from several cancers, infectious diseases, external causes, intentional self-harm, and degenerative disorders, independent of several major risk factors. (Funded by the British Heart Foundation and others.)

The presence of diabetes mellitus approximately doubles the risk of a wide range of vascular diseases.\(^1\) Evidence is also emerging that diabetes is associated with nonvascular conditions, including positive associations with certain cancers (e.g., liver cancer) and negative associations with other cancers (e.g., prostate cancer).\(^2\)\(^-\)\(^4\) However, a joint consensus statement of the American Diabetes Association and the American Cancer Society indicated that it is unclear whether such associations are direct (e.g., due to hyperglycemia) or indirect (e.g., due to diabetes as a marker of underlying biologic factors such as insulin resistance or hyperinsulinemia that alter the risk of cancer) or due to shared risk factors (e.g., obesity).\(^5\)\(^-\)\(^8\) Furthermore, many previous reports have considered diabetes in relation to only one or a few selected cancers or other nonvascular conditions. Since diabetes is a multisystem disorder, there is a need for adequately powered, standardized assessment of associations of diabetes with the risk of death from a broad range of causes.\(^9\)\(^,\)\(^10\)

We aimed to provide reliable estimates of any independent associations of baseline diabetes and fasting blood glucose level with the risk of cause-specific death by analyzing data from 820,900 people who were at risk for a total of 12.3 million person-years. We also estimated the effect of diabetes on life expectancy in adults.

**METHODS**

The study was designed and conducted by the independent academic coordinating center of the Emerging Risk Factors Collaboration (ERFC). Members of the coordinating center vouch for the accuracy and completeness of the data, the data analysis, and the results and made the decision to submit the article for publication. The sponsors had no role in the design, analysis, or interpretation of the study. The study was approved by the Cambridgeshire Ethics Review Committee.

Details of the ERFC have been published previously\(^11\) (also see the Supplementary Appendix, available with the full text of this article at NEJM.org). Specifically, we have already published reports on the associations of lipids, lipoproteins, and inflammatory markers with the risk of vascular disease and cause-specific death\(^12\)\(^-\)\(^14\) — the foregoing being risk factors that were the initial focus of the ERFC.\(^11\) In 2009, the ERFC agreed to extend analyses to diabetes and other metabolic markers in relation to the risk of incident fatal and nonfatal vascular disease outcomes\(^1\) and cause-specific death (see the Supplementary Appendix). The current analyses focus on individual-participant data from 97 prospective studies that had information about the diagnosis of diabetes or the fasting blood-glucose level at baseline, that did not select participants on the basis of having previous chronic disease (including vascular disease or diabetes), that included recording of cause-specific deaths classified according to clearly defined criteria, and that had accrued
more than 1 year of follow-up data. Study details are presented in Table 1 in the Supplementary Appendix (with references also listed). There were 820,900 participants who had no known preexisting vascular disease at baseline and for whom there was complete information about age, sex, smoking status (current smoker vs. any other status), body-mass index (BMI), history of diabetes or fasting glucose level (measured after ≥8 hours of fasting or overnight fasting), and subsequent cause-specific death recorded during follow-up. The contributing studies classified deaths according to the primary cause (or, in its absence, the underlying cause), on the basis of coding from the *International Classification of Diseases*, revisions 8 through 10, to at least three digits, or according to study-specific classification systems; ascertainment was based on death certificates. Attribution of death refers to the primary cause provided. A total of 67 of the 97 contributing studies also used medical records, findings on autopsy, and other supplementary sources to help classify deaths. We sought information on a range of risk factors and medication use at baseline (e.g., agents for lowering blood pressure, cholesterol, and glucose). Information on diabetes type (i.e., type 1 or 2) was generally not available, though the age of the participants suggests that the large majority with diabetes would have type 2.

Following the example of previous reports from the ERFC,12-15 we assessed whether baseline diabetes status (ascertained on the basis of self-report, medication use, fasting glucose level ≥126 mg per deciliter [7.0 mmol per liter], or a combination of these) and baseline fasting glucose level relate to death from any cause and its main components, including deaths from cancer, vascular disease, and nonvascular conditions not attributed to cancer, as well as to further subdivisions of these outcomes (e.g., site-specific cancers) (see definitions in Table 2 in the Supplementary Appendix).

Hazard ratios were calculated with the use of Cox proportional-hazards regression models stratified according to study, sex, and when appropriate, trial group. To minimize potential bias, log_{e} hazard ratios were calculated separately within each study and then pooled across studies by means of a random-effects meta-analysis.16 For the analyses of death from specific causes (cancer and noncancer, nonvascular), a one-step fixed-effect meta-analysis was used. We used methods described previously to characterize shapes of associations with fasting glucose level and to investigate heterogeneity.1,16 Unless otherwise specified, hazard ratios were adjusted for baseline age, sex, smoking status, and BMI. In subsidiary analyses, hazard ratios were also adjusted for other characteristics.

We corrected for regression dilution in subsidiary analyses, using serial measurements in 331,515 participants (mean interval, 2.9 years).17 Participants were included in analyses of death irrespective of the previous occurrence of non-fatal events. For each specific cause of death, participants’ data were censored if the participant was lost to follow-up, died from other causes, or reached the end of the follow-up period. Estimates of cumulative survival from 35 years of age and older among those with and those without diabetes at baseline were calculated by applying hazard ratios (specific to age at risk and sex) for cause-specific mortality associated with diabetes to cause-specific rates of death at 35 years of age and older for residents of the European Union.18,19 The Supplementary Appendix provides further statistical details. Analyses were carried out with Stata software (release 11).
RESULTS

Among the 820,900 participants included in analyses of diabetes status or fasting glucose level, the mean (±SD) age at baseline was 55±9 years; 48% were women. The large majority of participants were enrolled in Europe (58%) or North America (36%). Of the 715,061 participants included in analyses of diabetes status, 40,116 (6%) had diabetes at the time of enrollment (Table 1). The baseline characteristics of participants included in the analyses of fasting glucose are listed in Table 3 in the Supplementary Appendix. During the 12.3 million person-years at risk (median time to death, 13.6 years), 123,205 deaths were recorded: 41,320 from cancer, 44,407 from vascular disease, 27,661 from other causes, and 9817 of unknown or ill-defined cause (Table 4 in the Supplementary Appendix).

DIABETES AND MORTALITY

The crude overall rates of death were higher among participants with diabetes than among those without diabetes: 29 per 1000 person-years versus 12 per 1000 person-years among men, respectively, and 23 per 1000 person-years versus 7 per 1000 person-years among women, respectively. The corresponding cause-specific rates of death were as follows: for cancer deaths, 7 versus 4 per 1000 person-years among men and 4 versus 3 per 1000 person-years among women; for vascular deaths, 13 versus 5 per 1000 person-years among men and 11 versus 2 per 1000 person-years among women; and for noncancer, nonvascular deaths, 6 versus 3 per 1000 person-years among men and 6 versus 2 per 1000 person-years among women.

Hazard ratios for death among participants with diabetes, as compared to those without diabetes, after adjustment for baseline age, sex, smoking status, and BMI, were as follows: 1.80 (95% confidence interval [CI], 1.71 to 1.90) for death from any cause, 1.25 (95% CI, 1.19 to 1.31) for death from cancer, 2.32 (95% CI, 2.11 to 2.56) for death from vascular causes, 1.73 (95% CI, 1.62 to 1.85) for death from nonvascular causes not attributed to cancer, and 1.88 (95% CI, 1.62 to 2.18) for deaths of unknown or ill-defined cause (Fig. 1 in the Supplementary Appendix). These hazard ratios were not appreciably reduced after additional adjustment for systolic blood pressure, lipid levels, C-reactive protein levels, fibrinogen levels, alcohol use, estimated glomerular filtration rate, or indicators of socioeconomic status, when such information was available. However, the hazard ratios were reduced considerably after adjustment for fasting glucose or glycated hemoglobin levels (Table 2).

Hazard ratios for death among participants with diabetes, as compared to those without diabetes, were significantly higher at younger ages and among women, except for death from any cancer (Fig. 2 in the Supplementary Appendix). Hazard ratios appeared to decline somewhat with increasing calendar decade of study enrollment (Fig. 3 in the Supplementary Appendix). Diabetes was moderately associated with deaths from cancers of the liver, pancreas, ovary, colorectum, lung, bladder, and breast (Fig. 1A, and Fig. 4 in the Supplementary Appendix). Aside from cancer and vascular disease, diabetes was also associated with deaths from renal disease, liver disease, pneumonia, other infectious diseases, mental disorders, nonhepatic digestive diseases, external causes, intentional self-harm, nervous system disorders, and chronic obstructive pulmonary disease (Fig. 1B).
FASTING GLUCOSE AND MORTALITY

Fasting glucose level was nonlinearily related to the risk of death (Fig. 2). Fasting glucose levels exceeding 100 mg per deciliter (5.6 mmol per liter) but not levels of 70 to 100 mg per deciliter (3.9 to 5.6 mmol per liter) were associated with excess risk (Fig. 5 and 6 in the Supplementary Appendix). At fasting glucose levels above 100 mg per deciliter, hazard ratios for higher levels of glucose, assessed in increments of 18 mg per deciliter (1 mmol per liter), were 1.05 (95% CI, 1.03 to 1.06) for cancer deaths, 1.13 (95% CI, 1.11 to 1.15) for vascular deaths, 1.10 (95% CI, 1.07 to 1.12) for non-cancer, nonvascular deaths, and 1.10 (95% CI, 1.09 to 1.11) for death from any cause, assuming the existence of log-linear relationships above the threshold of 100 mg per deciliter (although associations with cancer deaths tended to plateau at higher levels).

There were generally too few deaths to characterize the shapes of associations of fasting glucose level with more specific causes of death. To avoid potential bias, we examined hazard ratios for various fasting glucose categories after excluding participants with a known history of diabetes at enrollment. As compared with the reference group (with a fasting glucose level of 70 to 100 mg per deciliter), hazard ratios for those with a fasting glucose level of 126 mg per deciliter (7.0 mmol per liter) or more were 1.39 (95% CI, 1.22 to 1.59) for cancer deaths, 1.89 (95% CI, 1.69 to 2.10) for vascular deaths, and 1.54 (95% CI, 1.31 to 1.81) for noncancer, nonvascular deaths. As compared with the same reference group, hazard ratios for participants with impaired fasting glucose levels (100 to <126 mg per deciliter [5.6 to <7.0 mmol per liter]) were 1.13 (95% CI, 1.06 to 1.20) for cancer deaths, 1.17 (95% CI, 1.08 to 1.26) for vascular deaths, and 1.12 (95% CI, 1.07 to 1.18) for noncancer, nonvascular deaths; whereas hazard ratios for participants with fasting glucose levels less than 70 mg per deciliter were 1.01 (95% CI, 0.93 to 1.10) for cancer deaths, 1.32 (95% CI, 1.12 to 1.56) for vascular deaths, and 1.05 (95% CI, 0.89 to 1.24) for noncancer, nonvascular deaths (Fig. 5 in the Supplementary Appendix). Among people with a history of diabetes at baseline, hazard ratios for death were higher among those who had a fasting glucose level of 126 mg per deciliter or more as compared with those who had a level below 126 mg per deciliter (hazard ratio, 2.16 vs. 1.51) (Fig. 5 in the Supplementary Appendix).

Qualitatively similar findings to those reported here were observed in a range of subsidiary analyses, such as those that included 75,195 participants with a history of cardiovascular disease at baseline (Fig. 7 in the Supplementary Appendix), excluded the initial 5 years of follow-up, excluded current smokers (Table 5 in the Supplementary Appendix), corrected for regression dilution in the fasting glucose level and in potential confounders (Fig. 8 and Table 6 in the Supplementary Appendix), assessed interactions with sex and age, excluded 56,766 participants known to be taking lipid-lowering or blood pressure–lowering medication at baseline (Table 5 in the Supplementary Appendix), standardized glucose values in studies that did not involve plasma measurements, assessed fasting glucose levels while ignoring a history of diabetes at baseline (Fig. 9 in the Supplementary Appendix), focused solely on the 60 cohort studies that recruited participants from population registries or general-practice lists (Fig. 10 in the Supplementary Appendix), and explored the effect of plausible degrees of potential misclassification of baseline diabetes status, new-onset...
diabetes, or attribution of cause of death (Table 7 and Fig. 11 in the Supplementary Appendix).

On average, middle-aged adults with diabetes but without known vascular disease at the time of enrollment died about 6 years younger than people without diabetes (Fig. 3). At 40, 50, and 60 years of age, men with diabetes but without a history of vascular disease would incur about 6.3, 5.8, and 4.5 years of life lost, respectively. The corresponding years of life lost for women with diabetes in middle age were 6.8, 6.4, and 5.4 years, respectively. About 58%, 9%, and 30% of this survival difference at 50 years of age can be attributed to excess vascular, cancer, and non-cancer, nonvascular deaths, respectively.

**DISCUSSION**

In addition to the excess risk of vascular disease, our data show that diabetes is associated with substantial premature mortality from several cancers, infectious diseases, external causes, intentional self-harm, and degenerative disorders, independent of several major risk factors. Our results suggest that, on average, a 50-year-old with diabetes but with no history of vascular disease is about 6 years younger at the time of death than a counterpart without diabetes; for comparison, the reduction in life expectancy from long-term cigarette smoking is about 10 years. About 40% of the years of life lost from diabetes can be attributed to nonvascular conditions, including about 10% attributable to death from cancer.

We have also found that there are generally continuous associations between fasting glucose levels greater than 100 mg per deciliter and risk of death, supporting the view that hyperglycemia (or some factor closely related to it) may be directly relevant. This possibility is also consistent with substantial attenuation observed in hazard ratios for death from diabetes after adjustment for markers of glycemia (though interpretation of such analyses is complicated because glucose levels are used to define diabetes status). In contrast, we did not observe appreciable alteration in the associations between diabetes and mortality after adjustment for several other risk factors (e.g., systolic blood pressure, measures of adiposity, inflammation biomarkers, insulin, or renal function), even after using serial measurements to adjust for their long-term average levels. These findings reduce the likelihood that such risk factors are major mediators of the excess risk of death associated with diabetes in our study.

Our study indicates that diabetes is moderately associated with death from cancers of the liver, pancreas, ovary, colorectum, lung, bladder, and breast (although, because multiple disease outcomes were studied, any marginally significant findings should be evaluated further). Although causality has not been established for these associations, systemic and local factors have been proposed to explain them. For pancreatic cancer, however, diabetes can be a consequence of the cancer, rather than a cause, although the exclusion of initial follow-up data did not weaken the hazard ratios in this study. In contrast with some previous studies of the incidence of prostate cancer, however, we did not find a significant inverse association of diabetes with the incidence of death from prostate cancer, although our confidence intervals were compatible with the risk estimates reported previously.
Aside from cancers, we observed strong positive associations of diabetes with deaths from renal and digestive diseases and infectious diseases. These results may reflect associated nephropathy, fatty liver disease, and suppression of cellular immunity, respectively.\textsuperscript{26-28} Furthermore, diabetes is associated with death from injuries, which could be related to end-organ complications such as neuropathy and eye disease or to episodes of hypoglycemia. Associations observed between diabetes and deaths due to mental and neurologic diseases should be studied further, including disaggregation into more specific conditions and investigation of the possible link between diabetes and the onset of depression,\textsuperscript{29} particularly since we found that among people with diabetes there was a substantial excess of deaths due to intentional self-harm. Collectively, therefore, our findings broaden and intensify the need for efforts to prevent and understand diabetes and encourage detailed study of a broader range of disease outcomes than has been customary in randomized trials of diabetes prevention and treatment.\textsuperscript{30} These results also reinforce the need for people with diabetes to consider cancer screening appropriate for their age and sex.

Our study has several strengths. These include the large sample size (over 123,000 deaths recorded during more than 12 million person-years at risk), standardized approaches to adjustment for several potential confounding factors, serial assessment of risk factors in 331,515 participants, an extended follow-up period, and information about cause-specific deaths from a variety of conditions. Furthermore, we studied several factors proposed to mediate associations of diabetes and cancer.\textsuperscript{31}

The generalizability of our findings to populations in economically developed Western countries is supported by broadly consistent results across 97 prospective cohorts in 25 countries. Even though the 6\% overall prevalence of diabetes in our study was somewhat lower than the prevalence currently reported for some Western populations, this difference would not have influenced hazard ratios. To enhance the validity of our findings, we conducted a range of sensitivity analyses, including a focus on population-based cohorts, adjustment for several major risk factors, and exclusion of the initial years of the follow-up period. We investigated mortality rather than the incidence of nonfatal disease, which should have reduced the likelihood of finding artificial associations due to preferential diagnosis of certain conditions in people with diabetes. Indeed, we observed qualitatively similar results among people with no history of diabetes but a fasting glucose level of 126 mg per deciliter or more and among people with a history of diabetes at baseline. However, our study cannot determine whether diabetes might increase the incidence of such diseases or reduce survival (or both) among persons with such illnesses.

Despite the strengths of our study, residual bias could persist owing to unmeasured or imprecisely measured potential confounding factors (e.g., dietary intake and physical activity, respectively). Certain glucose-lowering agents have been reported to increase\textsuperscript{32} or decrease\textsuperscript{33} the risk of cancer, but our observational study cannot reliably address these questions. Because our study did not include comprehensive recording of use of aspirin (an agent that is known to prevent vascular disease and may also prevent colorectal cancer\textsuperscript{34,35}), we may have underestimated the hazard ratios for death from colorectal cancer. Although our data suggest that hazard ratios for death in patients with diabetes may have declined somewhat in recent decades,\textsuperscript{36} this finding may be subject to confounding, because cohort
studies initiated in earlier decades may differ in several ways from more recent cohort
studies. Were the decline to be confirmed, however, then it would imply that our study has
slightly overestimated the contemporary effect of diabetes on death. Future studies are
warranted to investigate additional (and potentially more specific) risk factors that may link
diabetes and chronic diseases, to study non-Western populations, and to explain
associations observed between very low glucose levels and vascular death in people without
diabetes.

In conclusion, in addition to vascular disease, diabetes is associated with substantial
premature death from several cancers, infectious diseases, external causes, intentional self-
harm, and degenerative disorders, independent of major risk factors. These findings
highlight the need to better understand and prevent the multisystem consequences of
diabetes.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

Supported by grants from the British Heart Foundation (RG/08/014), the U.K. Medical Research Council, and
Pfizer (to the ERFC Coordinating Centre), as well as the Gates Cambridge Trust Scholarship, an Overseas Research
Studentship Award, and an Addenbrooke's Charitable Trust Clinical Research Fellowship (to Dr. Kondapally
Seshasai). Various sources have supported recruitment, follow-up, and laboratory measurements in the cohorts
contributing to the ERFC. Investigators of several of these studies have contributed to a list (http://
ceu.phpc.cam.ac.uk/research/erfc/studies) naming relevant funding sources.

We thank Jill Boreham for information on mortality rates in the European Union and Paul Pharoah for helpful
comments on a draft of the manuscript.

**References**

1. The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration,
and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;
2. El-Serag HB, Hampel H, Javadi F. The association between diabetes and hepato-cellular carcinoma:
[PubMed: 16527702]
[PubMed: 15886996]
8. Atchison EA, Gridley G, Carreon JD, Leitzmann MF, McGlynn KA. Risk of cancer in a large


Figure 1. Hazard Ratios for Death from Cancer and from Noncancer, Nonvascular Causes among Participants with Diabetes as Compared with Those without Diabetes at Baseline

Panel A shows hazard ratios for deaths from cancer, and Panel B shows hazard ratios for deaths from noncancer, nonvascular causes. With the exception of the classifications “site unspecified or other” in Panel A and “other noncancer, nonvascular deaths” in Panel B, causes of death are presented in descending order according to their estimated hazard ratios. All analyses were stratified on the basis of study, sex, and trial group (where applicable) and adjusted for baseline age, smoking status (current smoker vs. any other status), and body-mass index. There was evidence of heterogeneity in hazard ratios among cancer sites and among the noncancer, nonvascular causes of death (P<0.001 for both comparisons).

Participants with known preexisting cardiovascular disease at baseline were excluded from all analyses. The sizes of the data markers are proportional to the inverse of the variance of the log hazard ratios. In Panel A, risk estimates for cancer of the colorectum were broadly similar to those for cancer at subsites (i.e., colon cancer vs. cancer of the rectosigmoid and anus). In Panel B, death from endocrine disorders does not include death coded as being due to...
to diabetes. Other noncancer, nonvascular deaths are those that could not be attributed to a major organ or system. COPD denotes chronic obstructive pulmonary disease.
Figure 2. Hazard Ratios for Major Causes of Death, According to Baseline Levels of Fasting Glucose

History of diabetes at baseline was defined according to a self-reported history of diabetes or treatment for diabetes. Glucose levels for participants without a known history of diabetes at baseline were classified as less than 4.0, 4.0 to less than 4.5, 4.5 to less than 5.0, 5.0 to less than 5.5, 5.5 to less than 6.0, 6.0 to less than 6.5, 6.5 to less than 7.0, 7.0 to less than 7.5, and 7.5 mmol per liter or higher. Hazard ratios were plotted against the mean fasting glucose level in each group (reference category, 5.0 to <5.5 mmol per liter). The sizes of the data markers are proportional to the inverse of the variance of the log hazard ratios. All analyses were stratified or adjusted for sex and adjusted for baseline age, smoking status (current smoker vs. any other status), and body-mass index. Participants with known preexisting cardiovascular disease at baseline were excluded from all analyses. To convert values for fasting glucose to milligrams per deciliter, divide by 0.05551.
Panel A shows estimated survival curves that were plotted by applying hazard ratios for death from any cause (specific for sex and age at risk) from the present analyses to mortality data for the European Union in 2000. Panel B shows the estimated numbers of years of life lost owing to diabetes. Participants with known preexisting cardiovascular disease at baseline were excluded from both analyses.
Table 1
Baseline Data Used in Analyses of Diabetes and Cause-Specific Death, According to Participants’ Diabetes Status.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Diabetes (N = 40,116)</th>
<th>No Diabetes (N = 674,945)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age — yr</td>
<td>58±8</td>
<td>55±8</td>
</tr>
<tr>
<td>Sex — no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>21,213 (53)</td>
<td>337,178 (50)</td>
</tr>
<tr>
<td>Female</td>
<td>18,903 (47)</td>
<td>337,767 (50)</td>
</tr>
<tr>
<td>Geographic region or country — no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>23,844 (59)</td>
<td>257,818 (38)</td>
</tr>
<tr>
<td>Europe</td>
<td>11,883 (30)</td>
<td>371,717 (55)</td>
</tr>
<tr>
<td>Japan</td>
<td>2,835 (7)</td>
<td>20,170 (3)</td>
</tr>
<tr>
<td>Other</td>
<td>1,554 (4)</td>
<td>25,240 (4)</td>
</tr>
<tr>
<td><strong>Lifestyle factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking status — no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>8,894 (22)</td>
<td>225,331 (33)</td>
</tr>
<tr>
<td>Other</td>
<td>31,222 (78)</td>
<td>449,614 (67)</td>
</tr>
<tr>
<td>Alcohol use — no./total no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current drinker</td>
<td>10,177/20,447 (50)</td>
<td>232,476/365,375 (64)</td>
</tr>
<tr>
<td>Other</td>
<td>10,270/20,447 (50)</td>
<td>132,899/365,375 (36)</td>
</tr>
<tr>
<td>Physical activity — no./total no. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not active</td>
<td>4117/8723 (47)</td>
<td>132,795/258,071 (51)</td>
</tr>
<tr>
<td>Active</td>
<td>4606/8723 (53)</td>
<td>125,276/258,071 (49)</td>
</tr>
<tr>
<td><strong>Anthropometric markers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body-mass index‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>40,116</td>
<td>674,945</td>
</tr>
<tr>
<td>Mean</td>
<td>28±5</td>
<td>26±4</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>36,434</td>
<td>504,516</td>
</tr>
<tr>
<td>Mean — mm Hg</td>
<td>144±19</td>
<td>135±18</td>
</tr>
<tr>
<td>Waist circumference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>10,324</td>
<td>127,636</td>
</tr>
<tr>
<td>Mean — cm</td>
<td>97±13</td>
<td>89±12</td>
</tr>
<tr>
<td>Waist-to-hip ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>10,288</td>
<td>127,167</td>
</tr>
<tr>
<td>Mean</td>
<td>0.93±0.07</td>
<td>0.89±0.08</td>
</tr>
<tr>
<td><strong>Lipids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cholesterol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>35,016</td>
<td>494,678</td>
</tr>
<tr>
<td>Mean — mmol/liter</td>
<td>5.9±1.2</td>
<td>5.9±1.1</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Diabetes (N = 40,116)</td>
<td>No Diabetes (N = 674,945)</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-----------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Non-HDL cholesterol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>25,906</td>
<td>321,267</td>
</tr>
<tr>
<td>Mean — mmol/liter</td>
<td>4.6±1.1</td>
<td>4.5±1.1</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>25,928</td>
<td>321,537</td>
</tr>
<tr>
<td>Mean — mmol/liter</td>
<td>1.2±0.4</td>
<td>1.4±0.4</td>
</tr>
<tr>
<td>Log&lt;sub&gt;e&lt;/sub&gt; triglycerides</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>27,515</td>
<td>397,593</td>
</tr>
<tr>
<td>Mean — mmol/liter</td>
<td>0.5±0.6</td>
<td>0.3±0.5</td>
</tr>
<tr>
<td>Inflammatory markers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrinogen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>12,598</td>
<td>162,820</td>
</tr>
<tr>
<td>Mean — μmol/liter</td>
<td>9.8±2.3</td>
<td>9.3±2.1</td>
</tr>
<tr>
<td>Log&lt;sub&gt;e&lt;/sub&gt; C-reactive protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>6,979</td>
<td>104,642</td>
</tr>
<tr>
<td>Mean — mg/liter</td>
<td>0.9±1.1</td>
<td>0.6±1.1</td>
</tr>
<tr>
<td>Metabolic and renal markers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fasting glucose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>22,015</td>
<td>157,023</td>
</tr>
<tr>
<td>Mean — mmol/liter</td>
<td>8.6±3.6</td>
<td>5.2±0.6</td>
</tr>
<tr>
<td>Log&lt;sub&gt;e&lt;/sub&gt; insulin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>6,911</td>
<td>53,711</td>
</tr>
<tr>
<td>Mean — pmol/liter</td>
<td>4.6±0.7</td>
<td>4.0±0.6</td>
</tr>
<tr>
<td>Glycated hemoglobin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>4,971</td>
<td>48,468</td>
</tr>
<tr>
<td>Mean — %</td>
<td>7.2±2.0</td>
<td>5.3±0.5</td>
</tr>
<tr>
<td>Log&lt;sub&gt;e&lt;/sub&gt; estimated glomerular filtration rate&lt;sup&gt;‡&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. with data</td>
<td>22,122</td>
<td>216,585</td>
</tr>
<tr>
<td>Mean — ml/min/1.73 m&lt;sup&gt;2&lt;/sup&gt;</td>
<td>4.3±0.3</td>
<td>4.4±0.2</td>
</tr>
</tbody>
</table>

* Plus–minus values are means ±SD. Data are shown for the 715,061 of the 820,900 participants without known preexisting cardiovascular disease at baseline and for whom there was complete information about age, sex, smoking status (current smoker vs. any other status) and body-mass index. Information about the other characteristics in the table was not available for all participants. Diabetes was ascertained on the basis of self-report, medication use, fasting glucose level of 126 mg per deciliter (7.0 mmol per liter) or more, or a combination of these. To convert values for cholesterol to milligrams per deciliter, divide by 0.02586. To convert values for triglycerides to milligrams per deciliter, divide by 0.01129. To convert values for fibrinogen to grams per deciliter, divide by 29.41. To convert values for insulin to microunits per milliliter, divide by 6.945. To convert values for fasting glucose to milligrams per deciliter, divide by 0.05551. HDL denotes high-density lipoprotein.

† Physical activity was defined as any form of aerobic or anaerobic exercise.

‡ The body-mass index is the weight in kilograms divided by the square of the height in meters.

§ Estimated glomerular filtration rate was calculated with the use of the Modification of Diet in Renal Disease equation.
Table 2

Hazard Ratios for Death among Participants with Diabetes as Compared with Those without Diabetes at Baseline, after Adjustment for Potential Risk Factors and Mediators, According to Cause of Death.

<table>
<thead>
<tr>
<th>Model Variables</th>
<th>Cancer Death</th>
<th>Vascular Death</th>
<th>Noncancer, Nonvascular Death</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total No. of Participants</td>
<td>No. of Deaths</td>
<td>Total No. of Participants</td>
</tr>
<tr>
<td>Progressive adjustment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age and sex</td>
<td>486,807</td>
<td>22,399</td>
<td>513,951</td>
</tr>
<tr>
<td>Plus smoking status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plus body-mass index</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plus systolic blood pressure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plus total cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional adjustment*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic model</td>
<td>242,531</td>
<td>8,804</td>
<td>260,975</td>
</tr>
<tr>
<td>Plus non-HDL cholesterol, HDL cholesterol, and loge triglycerides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflammatory markers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic model</td>
<td>150,758</td>
<td>5,803</td>
<td>168,997</td>
</tr>
<tr>
<td>Plus fibrinogen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic model</td>
<td>84,172</td>
<td>4,402</td>
<td>101,681</td>
</tr>
<tr>
<td>Plus loge CRP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifestyle factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic model</td>
<td>338,476</td>
<td>12,704</td>
<td>353,614</td>
</tr>
<tr>
<td>Plus alcohol use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic model</td>
<td>247,831</td>
<td>12,291</td>
<td>255,535</td>
</tr>
<tr>
<td>Plus physical activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic model</td>
<td>242,977</td>
<td>10,699</td>
<td>254,215</td>
</tr>
<tr>
<td>Metabolic markers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic model</td>
<td>204,609</td>
<td>11,130</td>
<td>222,376</td>
</tr>
</tbody>
</table>

N Engl J Med. Author manuscript; available in PMC 2014 July 24.
<table>
<thead>
<tr>
<th>Model Variables</th>
<th>Total No. of Participants</th>
<th>No. of Deaths</th>
<th>Cancer Death</th>
<th>Vascular Death</th>
<th>Noncancer, Nonvascular Death</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hazard Ratio (95% CI)</td>
<td>Total No. of Participants</td>
<td>No. of Deaths</td>
</tr>
<tr>
<td>Plus log_e estimated GFR ‡</td>
<td>155,049</td>
<td>9,943</td>
<td>1.25 (1.16–1.34)</td>
<td>176,288</td>
<td>13,078</td>
</tr>
<tr>
<td>Basic model</td>
<td></td>
<td></td>
<td>1.25 (1.16–1.35)</td>
<td>176,288</td>
<td>13,078</td>
</tr>
<tr>
<td>Plus fasting glucose</td>
<td></td>
<td></td>
<td>1.08 (0.98–1.19)</td>
<td>176,288</td>
<td>13,078</td>
</tr>
<tr>
<td>Basic model</td>
<td>47,456</td>
<td>1,464</td>
<td>1.27 (1.09–1.48)</td>
<td>48,295</td>
<td>1,480</td>
</tr>
<tr>
<td>Plus glycated hemoglobin</td>
<td></td>
<td></td>
<td>1.10 (0.91–1.33)</td>
<td>48,295</td>
<td>1,480</td>
</tr>
<tr>
<td>Basic model</td>
<td>48,361</td>
<td>3,363</td>
<td>1.21 (1.08–1.37)</td>
<td>59,350</td>
<td>4,483</td>
</tr>
<tr>
<td>Plus log_e insulin</td>
<td></td>
<td></td>
<td>1.19 (1.05–1.35)</td>
<td>59,350</td>
<td>4,483</td>
</tr>
</tbody>
</table>

* All basic models were adjusted for age, sex, smoking status (current smoker vs. any other status), body-mass index, systolic blood pressure, and total cholesterol. Total cholesterol was not included in the analysis that further adjusted for high-density lipoprotein (HDL) cholesterol, non-HDL cholesterol, and triglycerides. Participants with known preexisting cardiovascular disease at baseline were excluded from all analyses. CRP denotes C-reactive protein.

‡ Educational level was categorized as no education, completion of primary school, completion of secondary school, or completion of vocational school or university. Hazard ratios adjusted for occupation were similar (data not shown).

The estimated glomerular filtration rate (GFR) was calculated with the use of the Modification of Diet in Renal Disease equation.