

# Data-driven metabolic subtypes predict future adverse events in individuals with type 1 diabetes

Raija Lithovius<sup>1,2,3</sup> · Iiro Toppila<sup>1,2,3</sup> · Valma Harjutsalo<sup>1,2,3,4</sup> · Carol Forsblom<sup>1,2,3</sup> · Per-Henrik Groop<sup>1,2,3,5</sup> · Ville-Petteri Mäkinen<sup>6,7,8</sup> ·  
on behalf of the FinnDiane Study Group

Received: 16 December 2016 / Accepted: 14 March 2017 / Published online: 24 April 2017  
© Springer-Verlag Berlin Heidelberg 2017

## Abstract

**Aims/hypothesis** Previously, we proposed that data-driven metabolic subtypes predict mortality in type 1 diabetes. Here, we analysed new clinical endpoints and revisited the subtypes after 7 years of additional follow-up.

**Methods** Finnish individuals with type 1 diabetes (2059 men and 1924 women, insulin treatment before 35 years of age) were recruited by the national multicentre FinnDiane Study Group. The participants were assigned one of six metabolic subtypes according to a previously published self-organising map from 2008. Subtype-specific all-cause and cardiovascular mortality rates in the FinnDiane cohort were compared with registry data from the entire Finnish population. The rates of incident diabetic kidney disease and cardiovascular endpoints were estimated based on hospital records.

**Results** The advanced kidney disease subtype was associated with the highest incidence of kidney disease progression (67.5% per decade,  $p < 0.001$ ), ischaemic heart disease

(26.4% per decade,  $p < 0.001$ ) and all-cause mortality (41.5% per decade,  $p < 0.001$ ). Across all subtypes, mortality rates were lower in women compared with men, but standardised mortality ratios (SMRs) were higher in women. SMRs were indistinguishable between the original study period (1994–2007) and the new period (2008–2014). The metabolic syndrome subtype predicted cardiovascular deaths (SMR 11.0 for men, SMR 23.4 for women,  $p < 0.001$ ), and women with the high HDL-cholesterol subtype were also at high cardiovascular risk (SMR 16.3,  $p < 0.001$ ). Men with the low-cholesterol or good glycaemic control subtype showed no excess mortality. **Conclusions/interpretation** Data-driven multivariable metabolic subtypes predicted the divergence of complication burden across multiple clinical endpoints simultaneously. In particular, men with the metabolic syndrome and women with high HDL-cholesterol should be recognised as important subgroups in interventional studies and public health guidelines on type 1 diabetes.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00125-017-4273-8) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

✉ Per-Henrik Groop  
per-henrik.groop@helsinki.fi

✉ Ville-Petteri Mäkinen  
ville-petteri.makinen@sahmri.com

<sup>1</sup> Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, University of Helsinki, Haartmaninkatu 8, PO Box 63, 00014 Helsinki, Finland

<sup>2</sup> Abdominal Center Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

<sup>3</sup> Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland

<sup>4</sup> National Institute for Health and Welfare, Chronic Disease Prevention Unit, Helsinki, Finland

<sup>5</sup> The Baker IDI Heart and Diabetes Institute, Melbourne, VIC, Australia

<sup>6</sup> South Australian Health and Medical Research Institute, SAHMRI North Terrace, PO Box 11060, Adelaide, SA 5001, Australia

<sup>7</sup> School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia

<sup>8</sup> Computational Medicine, Faculty of Medicine, University of Oulu and Biocenter, Oulu, Finland

**Keywords** All-cause mortality · Cardiovascular mortality · Data-driven model · Diabetic kidney disease · Ischaemic heart disease · Metabolic subtypes · Self-organising map · Sex difference

### Abbreviations

ESRD	End-stage renal disease
FinnDiane	Finnish Diabetic Nephropathy Study
IQR	Interquartile range
SMR	Standardised mortality ratio
SOM	Self-organising map

### Introduction

The acute symptoms of type 1 diabetes can be managed by insulin therapy, but a significant number of individuals gradually develop chronic diabetic complications that reduce the quality of life [1], cause most of the public health burden [2] and shorten life expectancy [3]. Current prognostic tools for diabetic complications and cardiovascular disease rely on consensus guidelines that define broadly applicable univariate thresholds, or simplified algebraic formulas to combine a limited number of biomarkers [4, 5]. However, predicting the future is difficult, and consensus risk estimates may be too simplistic or too dependent on the type of cardiometabolic risk profiles in a given time period to reach sufficient precision [6]. To transcend the limitations of consensus guidelines, a new data-driven paradigm is now emerging that aims to collect massive repositories of clinical and molecular data (e.g. the US Precision Medicine Initiative and the UK Biobank), which are then integrated by advanced computational techniques to produce more precise and individualised diagnoses and risk estimates.

We postulate that the metabolic subtype, as observed via blood and urine biochemistry, carries important phenotypic information from the unique personal combination of genetics and environmental exposures that modulate long-term complication risk [7]. We also propose that diabetic complications comprise a complex spectrum of altered biomarker concentrations and comorbidities that may not be well captured by traditional statistical methods [8]. Therefore, a data-driven approach may provide important scientific insight into the associations between established clinical endpoints and prognostic markers. However, ‘big data’ in epidemiology remains a nascent area [9, 10], and significant challenges remain to ensure that the statistical models in data-driven precision medicine are accurate, interpretations are robust and unbiased, and that existing best practices are incorporated into new data-analysis frameworks.

In 2008, we described the connections between metabolic subtypes in type 1 diabetes and all-cause mortality [11]. From

a methodological point of view, the previous study provides us with an opportunity to evaluate the concept of molecular subtyping against new data that were not available at the time. In this study, we analysed new data on all-cause mortality after 2008 to find out what has happened to the subtypes since the original study, and if the mortality gap between type 1 diabetes and the general population has narrowed. The original study did not investigate the causes for the observed mortality, nor the incidence of micro- or macrovascular complications. Here, we report new findings on cardiovascular mortality, ischaemic heart disease and incident albuminuria that were not available in the earlier publication, and specifically test for sex differences to provide a clearer picture on subtype-specific risk in men and women. We are not aware of similar studies that have re-used a multivariable model after several years; this report is thus a novel example of how subtyping applications can be validated with new data, and demonstrates the usefulness of the self-organising map (SOM) approach for epidemiological research of diabetic complications.

### Methods

In the original report from 2008, a total of 4197 individuals with type 1 diabetes were included from the nationwide multicentre Finnish Diabetic Nephropathy Study (FinnDiane) cohort [11]. In the current study, sufficient follow-up information was available for 3983 (95%) individuals. The initial data collection was cross-sectional between 1994 and 2007 (serum and urine samples from a single baseline visit), and longitudinal clinical records were obtained from hospitals and registries up to the end of year 2014 (see electronic supplementary material [ESM] Fig. 1). Type 1 diabetes was defined as age < 35 years at the onset of diabetes and transition to insulin treatment within 1 year of onset. The study protocol was approved by the Ethical Committee of Helsinki and Uusimaa Hospital District, and is in accordance with the Declaration of Helsinki. Written informed consent was received from each participant.

The FinnDiane individuals who participated in the first study and who had follow-up data on renal status, cardiovascular events and mortality were included. Diabetic kidney disease was categorised according to urinary AER in at least two out of three consecutive collections: normal AER < 30 mg/24 h, microalbuminuria 30 < AER < 300 mg/24 h and macroalbuminuria AER > 300 mg/24 h. Equivalent cut-offs for overnight urine samples were <20 µg/min for normal AER and >200 µg/min for macroalbuminuria. Patients on dialysis and recipients of kidney transplants were classified as having end-stage renal disease (ESRD). Retinopathy was defined as a history of laser treatment.

Cardiovascular disease was defined as ischaemic heart disease (ICD-10 codes I20–I25 [[www.who.int/classifications/icd/en/](http://www.who.int/classifications/icd/en/)]) or cerebrovascular disease (ICD-10 codes I60–I69). A

cardiovascular death was recorded if the death certificate listed any of the aforementioned codes as the underlying cause (narrow definition), or if any of the codes was listed as the underlying, immediate, intermediate or contributing cause (wide definition). Hospital data were used as another source of information: ischaemic heart disease was defined as a history of myocardial infarction or a coronary artery procedure such as bypass surgery or angioplasty, and stroke was defined as cerebral infarction or cerebral haemorrhage. The metabolic syndrome was defined according to a previously published version of the National Cholesterol Education Program Adult Treatment Panel (NCEP ATP) III guidelines [12].

Vitality status (until 31 December 2014) and death certificates (until 31 December 2013) for the study participants were obtained from the Finnish Causes of Death Register. Age-specific mortality rates and causes in the general population were provided by Statistics Finland from the national registries. Hospital-based information on ischaemic heart disease and stroke in the FinnDiane cohort (until 31 December 2013) was obtained from the national registry of admissions and discharges at the National Institute for Health and Welfare, and the latest available diabetic kidney disease status was derived from the individuals' healthcare records and multiple national registries (until 31 December 2013).

The SOM from 2008 was used to define six subgroups of individuals (ESM Fig. 2). Briefly, the SOM is a two-dimensional representation of multi-dimensional data where the distance on the map between two individuals corresponds to their similarity with respect to the biochemical profile from serum and urine. Therefore, individuals who are placed in the same area of the map share a similar metabolic subtype. By splitting the map into subregions, we assigned individuals into the corresponding subgroups, and then applied conventional statistics to evaluate the associations with clinical endpoints. All biochemical variables from the original SOM for the six metabolic subgroups are defined and listed in ESM Table 1.

**Statistical analyses** All analyses were performed for both sexes together and separately. Standardised mortality ratios (SMRs) were estimated by dividing the frequency of deaths in the FinnDiane cohort by the frequency of deaths in the general population. SMRs were initially estimated for each age group and observation year (an age–observation segment) separately. These raw SMRs were then weighted according to the follow-up duration and the number of FinnDiane participants in each segment. The weighted mean over all segments was reported as the final SMR. For all-cause SMRs, segments were defined by 1 year age intervals and annual observations. For vascular SMRs in the general population, the provided segments were defined according to 5 year age intervals and annual observations, and interpolated to 1 year age intervals before computing SMRs to ensure numerical stability for younger age groups with low event frequencies. The  $p$  values

and 95% CIs were estimated by bootstrap sampling of the FinnDiane participants.

The RR for clinical endpoints was estimated within the FinnDiane cohort. To eliminate the confounding effect of exposure time on complication burden, the individuals in a given subgroup ( $n$  cases) were paired with individuals with matching age and diabetes duration from the rest of the cohort (minimum  $n$  controls). RR was defined as the follow-up duration-weighted incidence in cases divided by the corresponding incidence in controls throughout all age–observation segments. Mortality and risk analyses were performed with R project statistical software, version R 3.0.2 [13] and Statistical Analysis System, version 9.4 (SAS Institute, Cary, NC, USA).

The summary statistics of continuous variables for different metabolic subtypes were expressed as medians and interquartile ranges (IQRs), and binary variables as percentages of prevalence. The statistical significance of the difference between two subtypes was estimated by the Kruskal–Wallis test or Pearson's  $\chi^2$  test where appropriate.

## Results

The clinical characteristics of the study participants at baseline are listed in Table 1, with results for men and women listed separately in ESM Table 2. Subtype A (good glycaemic control) was characterised by the lowest median HbA<sub>1c</sub>, 7.3% (56 mmol/mol), with IQR 6.7–8.0% (50–64 mmol/mol). Subtype B had the highest HDL-cholesterol (1.6 mmol/l, IQR 1.4–1.9 mmol/l). Subtype C (advanced kidney disease) showed the highest prevalence of ESRD (24%). Subtype D was characterised by the highest prevalence of the metabolic syndrome (82%). Subtype E was characterised by the lowest total cholesterol concentration (4.2 mmol/l, IQR 3.8–4.5 mmol/l). In addition, the region of the SOM that was not reported in the original paper (subtype U, high blood glucose) showed a high median HbA<sub>1c</sub> of 9.1% (76 mmol/mol) with IQR 8.3–10.1% (67–87 mmol/mol).

A total of 632 deaths occurred during a median of 14.0 (IQR 11.7–16.0) years of follow-up. To evaluate the accuracy of the subtype modelling, we calculated all-cause SMRs for the periods before (1994–2007) and after (2008–2014) the original analysis. Overall, the estimates were consistent between the two periods (Table 2, ESM Table 3). In particular, subtypes A and E, with favourable metabolic profiles, showed no excess mortality ( $p > 0.05$ ), whereas subtypes C and D were associated with significantly increased mortality risk in both periods (SMR between 4.6 and 6.8,  $p < 0.001$ ). The all-cause SMRs for all participants combined were 2.7 (95% CI 2.4, 3.0) for the earlier period and 2.6 (95% CI 2.3, 2.9) for the later period. For the entire follow-up (1994–2014), the highest all-cause mortality rate (41.5% per decade) and SMR (5.6) were observed for subgroup C (ESM Table 5).

**Table 1** Baseline characteristics of the study participants ( $n = 3983$ )

Characteristic	Good glycaemic control (A)	High HDL-cholesterol (B)	Advanced kidney disease (C)	Metabolic syndrome (D)	Low cholesterol (E)	High blood glucose (U)
Men/women ( $n$ )	420/433	294/267	202/190	317/286	577/511	249/237
Age (years)	37.2 (28.8–45.9)	42.2 (33.2–50.4)	42.3 (35.0–49.9)	36.3 (28.7–44.4)	32.4 (24.3–41.5)	35.9 (27.0–46.7)
Diabetes duration (years)	19.6 (9.4–29.5)	26.2 (18.0–35.1)	29.6 (23.1–36.8)	23.6 (15.1–31.0)	16.3 (9.1–26.8)	20.0 (12.3–30.1)
Microalbuminuria (%)	6.7	21.9	12.5	15.6	10.7	14.4
Macroalbuminuria (%)	2.5	18.0	54.6	32.3	2.7	4.9
ESRD (%)	2.0	12.1	24.0	13.3	0.9	3.5
Retinopathy (%)	18.5	47.6	80.3	54.8	17.3	27.5
Macrovascular disease (%) <sup>a</sup>	4.2	10.0	26.5	15.3	3.1	7.0
Metabolic syndrome (%)	11.2	24.1	53.1	81.6	27.1	33.3
HbA <sub>1c</sub> (%)	7.3 (6.7–8.0)	8.4 (7.7–9.1)	8.9 (8.1–9.8)	9.2 (8.2–10.2)	8.2 (7.4–9.0)	9.1 (8.3–10.1)
HbA <sub>1c</sub> (mmol/mol)	56 (50–64)	68 (61–76)	74 (65–84)	77 (67–88)	66 (57–75)	76 (67–87)
Urinary AER (mg/24 h)	6.7 (4.2–12.2)	23.3 (9.4–124.5)	251.1 (75.5–845.1)	43.6 (10.7–296.3)	7.7 (4.5–14.5)	9.5 (5.1–25.7)
Total cholesterol (mmol/l)	4.7 (4.3–5.2)	5.6 (5.2–6.1)	4.9 (4.3–5.4)	5.8 (5.2–6.4)	4.2 (3.8–4.5)	5.1 (4.7–5.5)
HDL-cholesterol (mmol/l)	1.5 (1.3–1.8)	1.6 (1.4–1.9)	1.2 (1.0–1.4)	1.0 (0.8–1.2)	1.2 (1.0–1.4)	1.2 (1.1–1.4)
Waist circumference (cm)	81.0 (75.0–88.0)	85.0 (77.0–93.0)	86.0 (79.0–94.0)	91.7 (83.0–101.0)	83.0 (77.0–90.0)	86.0 (78.0–93.0)

Data are median (IQR) or %

<sup>a</sup> Prevalent macrovascular disease was defined as cardiovascular disease, amputation or peripheral artery bypass

**Vascular diseases and mortality** Cardiovascular causes were identified for 37.2% of all deaths (222 out of 597) when using the narrow definition and for 57.8% of deaths (345 out of 597) when using the wide definition over a median of 13.1 (IQR 11.0–15.0) years of follow-up (Table 3). The proportion of cardiovascular deaths varied modestly between 33.9% and 41.7% across the subtypes when using the narrow definition. However, higher proportions of cardiovascular deaths were observed for subtypes C (61.6%) and D (62.5%) compared with subtypes A (43.5%) and E (46.3%) when using the wide definition.

There were no marked differences in cardiovascular SMRs between the first (1994–2007) and second (2008–2013) time periods (ESM Table 4). The results for the full period 1994–2013 are listed in Table 3. Subtype A with good glycaemic control and subtype E with low cholesterol showed

the lowest cardiovascular mortality rates (<2% per decade by either definition). Dramatically higher rates were observed for subtype C, with advanced kidney disease, (14.0% per decade, SMR 11.1,  $p < 0.001$ ) and subtype D, with the metabolic syndrome, (7.6% per decade, SMR 8.8,  $p < 0.001$ ) when the narrow definition was used. Similar patterns were observed with the wide definition of cardiovascular death. For the full cohort, cardiovascular mortality was significantly higher than in the general population according to the narrow (SMR 5.0,  $p < 0.001$ ) and wide definitions (SMR 7.8,  $p < 0.001$ ).

The incidence of diabetic vascular complications was investigated within the FinnDiane cohort (Table 4). Similar to the earlier results, the favourable metabolic subtypes did well: no more than 6.3% of participants with subtypes A or E were affected by any specific new complication per decade. On the

**Table 2** Comparison between all-cause mortality in individuals with type 1 diabetes and the general population during the two time periods

Subtype	1994–2007			2008–2014		
	Follow-up (person-years)	Mortality (% died/decade)	SMR (95% CI)	Follow-up (person-years)	Mortality (% died/decade)	SMR (95% CI)
Good glycaemic control (A)	5989	3.2	0.92 (0.53, 1.37)	5723	5.8	1.18 (0.81, 1.58)
High HDL-cholesterol (B)	4113	16.3	3.25 (2.53, 4.13) <sup>***</sup>	3212	20.2	2.80 (2.20, 3.44) <sup>***</sup>
Advanced kidney disease (C)	2769	39.0	6.78 (5.66, 8.02) <sup>***</sup>	1715	45.5	6.52 (5.15, 8.02) <sup>***</sup>
Metabolic syndrome (D)	4569	19.9	4.89 (3.98, 5.93) <sup>***</sup>	3345	19.7	4.58 (3.62, 5.58) <sup>***</sup>
Low cholesterol (E)	7422	1.9	0.78 (0.39, 1.17)	7417	4.0	1.17 (0.78, 1.60)
High blood glucose (U)	3403	7.3	1.91 (1.17, 2.89) <sup>*</sup>	3106	11.6	2.22 (1.58, 2.92) <sup>***</sup>
All individuals	28,266	11.5	2.72 (2.44, 3.03) <sup>***</sup>	24,519	12.6	2.58 (2.31, 2.86) <sup>***</sup>

<sup>\*</sup>  $p < 0.05$ , <sup>\*\*\*</sup>  $p < 0.001$

**Table 3** Comparison between cardiovascular mortality in individuals with type 1 diabetes and the general population between 1994 and 2013 based on the narrow and wide definitions of cardiovascular deaths

Subtype	Follow-up (person-years)	Narrow definition			Wide definition		
		Mortality (% died/decade)	Proportion of all deaths (%)	SMR (95% CI)	Mortality (% died/decade)	Proportion of all deaths (%)	SMR (95% CI)
Good glycaemic control (A)	10,911	1.6	37.0	1.98 (1.09, 3.17)*	1.8	43.5	2.25 (1.29, 3.53)**
High HDL-cholesterol (B)	6895	7.7	41.7	6.40 (4.54, 9.05)***	10.9	59.1	8.79 (6.74, 11.95)***
Advanced kidney disease (C)	4276	14.0	33.9	11.14 (7.90, 14.74)***	25.5	61.6	19.65 (15.48, 26.27)***
Metabolic syndrome (D)	7468	7.6	37.5	8.83 (6.69, 13.08)***	12.7	62.5	14.66 (11.80, 19.38)***
Low cholesterol (E)	13,797	1.1	36.6	2.35 (1.26, 3.54)*	1.4	46.3	3.68 (2.00, 5.40)***
High blood glucose (U)	6083	3.3	37.0	3.94 (2.29, 6.27)***	4.4	50.0	5.30 (3.46, 7.89)***
All individuals	49,431	4.5	37.2	4.96 (4.33, 5.70)***	7.0	57.8	7.76 (6.99, 8.74)***

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ 

other hand, the majority of participants with the advanced kidney disease subtype (C) progressed in the categorical albuminuria scale (67.5% per decade), and many developed ESRD (59.4% per decade). Accordingly, the incidence values for ischaemic heart disease (26.4% per decade) and stroke (17.5% per decade) were also higher. ESRD and ischaemic heart disease had higher incidences in subtype C compared with subtype D (59.4% vs 19.9% per decade [threefold difference] and 26.4% vs 18.1% per decade [1.5-fold difference], respectively). When the gap in age and diabetes duration between the two subtypes was taken into account, the risk of ESRD was higher in the advanced kidney disease subtype (RR 5.0 [subtype C] vs RR 2.5 [subtype D]), as expected, but the RR for ischaemic heart disease was not (RR 1.7 for subtype C vs RR 2.0 for subtype D).

**Sex differences** The differences in mortality between men and women are summarised in Fig. 1, and the corresponding numerical values are reported in ESM Tables 5 and 6. The point estimate for 10 year all-cause mortality rate was lower in women across all subtypes (Fig. 1a) and significantly lower, comparing all women with all men. On a subgroup level, the rate was significantly lower in women only for subgroup C. However, all-cause SMR was higher in women compared with men (Fig. 1b) for subtypes B (SMR 2.2 in men vs 3.8 in women), C (SMR 4.0 vs 8.9) and D (SMR 3.2 vs 5.3). Of note, women with subtypes B, C or D were over 4 years older than men with the same subtypes ( $p < 0.001$ ), whereas women with the low-cholesterol subtype had 3.8 years' longer duration of diabetes than men ( $p < 0.001$ ).

The pattern for cardiovascular mortality was similar to that for all-cause mortality. Again, the overall 10 year cardiovascular mortality rate was significantly lower in women, comparing all women with all men (5.5% in women vs 8.4% in men) but the cardiovascular SMRs were substantially higher in women compared with men (Fig. 1d) for the subtypes B

(SMR 16.3 vs 6.0), C (SMR 46.0 vs 11.5) and D (SMR 23.4 vs 11.0) when using the wide definition. Surprisingly, the cardiovascular SMRs for subtypes C (SMR 11.5) and D (SMR 11.0) were almost identical within men, whereas a two-fold difference was observed within women.

The baseline metabolic subtype was predictive of diabetic vascular complications in both sexes (Fig. 2, ESM Table 7). The 10 year risk of kidney disease progression, incident ESRD and stroke was higher in men, comparing all men with all women. Men with subtype D and U had a significantly higher 10 year risk of kidney disease progression compared with women (D, 42.6% in men vs 24.2% in women; U, 21.0% in men vs 4.7% in women), and similarly men with subtype C had higher 10 year risk of incident ESRD (75.8% in men vs 46.5% in women) and stroke (25.2% in men vs 10.7% in women) (Fig. 2b). However, there was no difference in the 10 year incidence of ischaemic heart disease between all men and all women. On a subgroup level, the incidence was higher in women for subgroup E. Notably, age bias did not explain the equalisation of risk as women with subtype B (39.0 years) or C (40.2 years) were younger than the men with the same subtypes (B 43.8 years,  $p < 0.001$ ; C 44.6 years,  $p < 0.003$ ).

## Discussion

We propose that the combination of non-optimal glycaemic control, unhealthy lifestyle and genetic susceptibility to diabetic end-organ damage manifests as systemic metabolic subtypes in individuals with type 1 diabetes [7]. Data-driven representative subtypes are an appealing approach to address the inherent complexity of biomedical data on diabetic complications, but rigorous testing is needed to establish the usefulness of the subtypes. For this study, we had the unique opportunity to revisit a subtype model from 2008 [11] and investigate what had happened in the subsequent 7 years.

**Table 4** Incidence of vascular complications in the FinnDiane cohort

Subtypes	Progression of diabetic kidney disease		ESRD		Ischaemic heart disease		Stroke	
	Incidence (%/decade)	RR (95% CI)	Incidence (%/decade)	RR (95% CI)	Incidence (%/decade)	RR (95% CI)	Incidence (%/decade)	RR (95% CI)
Good glycaemic control (A)	3.2	0.19 (0.12, 0.26)*	0.2	NC	3.6	0.33 (0.23, 0.44)*	1.9	0.29 (0.17, 0.44)*
High HDL-cholesterol (B)	15.9	0.90 (0.72, 1.11)	7.2	0.60 (0.43, 0.80)*	11.6	0.70 (0.55, 0.88)*	9.3	1.10 (0.84, 1.39)
Advanced kidney disease (C)	67.5	3.30 (2.93, 3.72)*	59.4	4.94 (4.31, 5.66)*	26.4	1.65 (1.32, 1.93)*	17.5	1.95 (1.54, 2.57)*
Metabolic syndrome (D)	33.0	2.24 (1.93, 2.60)*	19.9	2.46 (2.05, 2.94)*	18.1	1.99 (1.66, 2.46)*	8.3	1.25 (0.95, 1.62)*
Low cholesterol (E)	6.3	0.49 (0.38, 0.60)*	0.4	NC	4.0	0.78 (0.55, 1.04)	1.6	0.38 (0.23, 0.56)*
High blood glucose (U)	12.6	0.87 (0.67, 1.09)	3.4	0.51 (0.29, 0.77)*	7.2	0.78 (0.55, 1.04)	4.1	0.67 (0.41, 0.99)*
All individuals	14.8	NA	7.7	NA	9.1	NA	5.3	NA

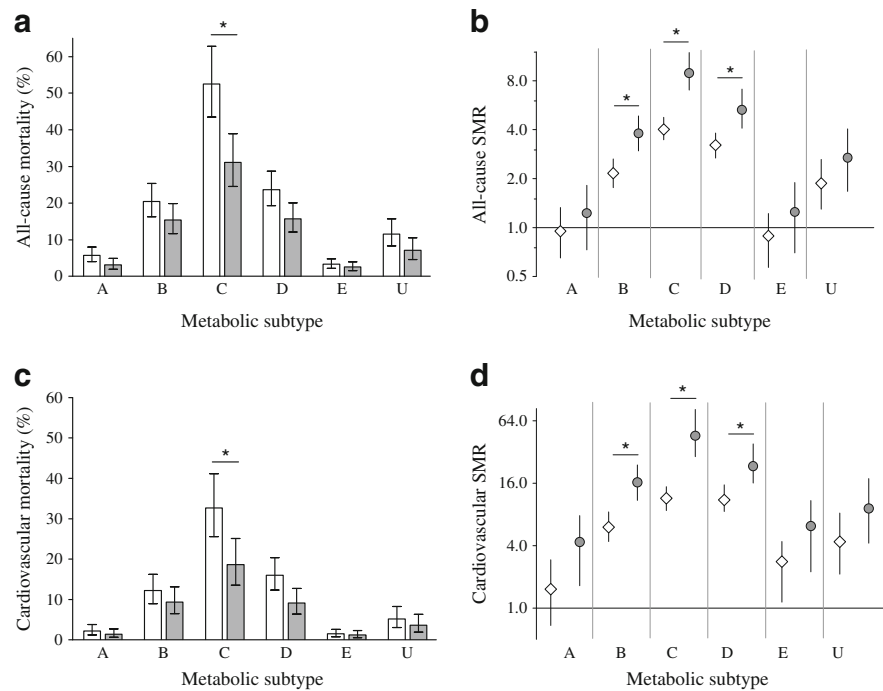
RR was not calculated if there were fewer than ten incident events during follow-up (see additional details in Methods)

By definition, the risk estimates were not available for the full cohort

\*  $p < 0.05$

NA, not available; NC, not calculated

**Fig. 1** Sex differences in all-cause (a, b) and cardiovascular (c, d) mortality for the metabolic subtypes A–E and U. Mortality rates are presented per decade. Error bars represent 95% CIs. White bars/squares, men; grey bars/circles, women. \* $p < 0.05$ ; all men vs all women: (a) in men 14.3% (12.9, 15.8), in women 9.5% (8.4, 10.8),  $p < 0.05$ ; (c) in men 8.4% (7.3, 9.6), in women 5.5% (4.6, 6.5),  $p < 0.05$

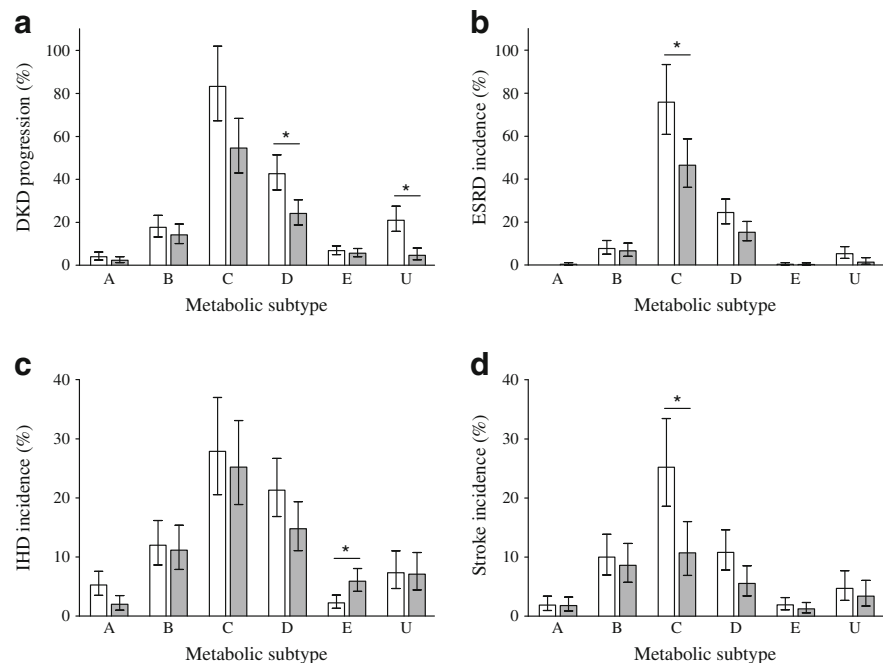


We combined the old models with new previously unseen clinical data to extend the original findings in coverage and over time. Importantly, we show compelling results on the mortality gap between type 1 diabetes and the general population, and describe the incidence of vascular diseases and sex divergence within the FinnDiane cohort across six pre-defined metabolic subtypes.

The relative gap in mortality between type 1 diabetes and the general population has remained stable over the past two

decades (Table 2), and the results are highly consistent across the subtypes. In general, consistent patterns of absolute mortality and SMRs have been observed across several Nordic studies that represent similar ethnic and socioeconomic circumstances [14–17]. In particular, the absolute mortality rate in the FinnDiane cohort (12.0% per decade) was similar to 13.9% in a Danish cohort [14] and 10.4% in a Swedish registry [15]. Previously published time trends in Finland showed a declining SMR time trend in the early-onset subset, but an

**Fig. 2** Sex differences in vascular disease incidence. (a) Diabetic kidney disease progression; (b) ESRD incidence; (c) ischaemic heart disease incidence; and (d) stroke incidence. Incidence is presented per decade. Error bars represent 95% CIs. White bars, men; grey bars, women. DKD, diabetic kidney disease; IHD, ischaemic heart disease. \* $p < 0.05$ ; all men vs all women: (a) in men 18.0% (16.2, 20.0), in women 11.6% (10.1, 13.1),  $p < 0.05$ ; (b) in men 9.0 (7.8, 10.3), in women 6.4 (5.4, 7.5),  $p < 0.05$ ; (c) in men 9.5 (8.3, 10.8), in women 8.8 (7.6, 10.1),  $p > 0.05$ ; (d) in men 6.4 (5.5, 7.5), in women 4.1 (3.3, 5.0),  $p < 0.05$



increasing SMR in the late-onset subset when stratified by the calendar year of diabetes onset [16]. Overall, a meta-analysis of type 1 diabetes mortality indicated that the relative mortality gap has stabilised since 1990 [18]. In absolute terms, life expectancy has increased in people with type 1 diabetes [19], but it seems that this has been largely offset by a concomitant increase in longevity across the whole population in Finland.

In this study, the at-risk population comprised mostly the same individuals between the two time periods (fewer than 10% died within the first period). As a consequence, the SMRs before and after 2008 apply to different age segments of the Finnish population, which may make them less accurate measures of age-controlled time trends. The consistency across all subtypes is, however, important: we postulate that the most effective way to reduce the overall SMR is by reducing the prevalence of adverse subtypes. For example, reducing the prevalence of the obesity-associated metabolic syndrome subtype (D) in men may lead to a relative increase in the prevalence of the nearby low-cholesterol subtype (E), which will eventually translate into a narrower gap in mortality with respect to non-diabetic peers.

Cardiovascular disease was the underlying cause in over 30% of deaths and a contributing cause in over 60% of deaths for the adverse kidney disease subtype (C) and the metabolic syndrome subtype (D). In an Australian study of 87,047 individuals followed for 15.2 years [20], approximately one-third of deaths were attributed to cardiovascular causes, which fits with the results from FinnDiane when using the narrow definition of cardiovascular disease. Chronic kidney disease on its own increases cardiovascular risk substantially in all age groups [21], and there is also a plausible link between the metabolic syndrome or its components and cardiovascular mortality in type 1 diabetes [22–24]. Against this backdrop, the patterns of cardiovascular mortality observed here (i.e. high-risk subtypes C and D) are in accordance with the literature. Importantly, the metabolically favourable subtype A showed no excess mortality in men, showing that type 1 diabetes per se, if well controlled, does not substantially increase cardiovascular risk even after 20 years of diabetes in metabolically robust individuals. It is logical to assume that individuals of subtype A have experienced 20 years of good metabolic health since type 1 diabetes onset. This may explain why subtype A is associated with a substantially better prognosis than the cohort as a whole, whereas in trials such as the DCCT/Epidemiology of Diabetes Interventions and Complications (EDIC) a shorter 6.5 years of intensive glucose control had a modest impact on all-cause mortality decades later [25]. We propose subtype A as an ideal, but achievable, treatment target for new multifactorial trials that combine tailored lifestyle and molecular interventions beyond more conventional approaches that target only one biomarker at a time.

There is a prevailing health gap between men and premenopausal women in the general population [26], whereas a narrower gap was observed in the FinnDiane cohort

(standardised risk estimates were higher in women despite lower absolute mortality or morbidity rates). The vulnerability of women to diabetic complications has been observed before [27, 28], but this is the first study to dissect the phenomenon into metabolic subtypes. In particular, a subset of women with high HDL-cholesterol and type 1 diabetes (subtype B) were not protected against ischaemic heart disease, and even lost their advantage over men, which resulted in dramatically increased standardised mortality and relative cardiovascular risk. Although men with subtype A were doing as well as their non-diabetic peers, subtype A in women was associated with a fourfold increased cardiovascular SMR. It is thus plausible that type 1 diabetes itself may disrupt hormonal cycles and metabolic homeostasis in women [29], beyond the corresponding effects in men. From a practical point of view, metabolic interventions that are specifically targeted for diabetic women at risk of microvascular complications may represent a cost-effective means to mitigate the adverse sequelae of type 1 diabetes.

From a mechanistic perspective, it is plausible that the secondary effects of kidney disease cause the dramatic rise in vascular mortality in individuals with subtype C. Insulin resistance [22], lipoprotein trafficking abnormalities [30] and low-grade chronic inflammation [31] are the likely explanations for the increased risk in the metabolic syndrome subgroup who had equivalent or even higher insulin doses, abdominal obesity, triacylglycerol-cholesterol imbalance and elevated C-reactive protein compared with the subtypes A and E. The lack of protection from high HDL-cholesterol in women of subtype B is supported by Mendelian randomisation studies that have failed to show causality for HDL-cholesterol despite an inverse correlation with cardiovascular disease [32, 33]. The underlying metabolic subtype may be a crucial contextual confounder: the isolated high HDL-cholesterol in subtype B may indicate dysfunction, whereas the high HDL-cholesterol in subtype A fits with the overall favourable metabolic profile. Indeed, high HDL-cholesterol in women with type 1 diabetes has previously been linked to cardiovascular risk [34].

The use of pre-defined data-driven baseline subtypes with new prospective data is the unique strength of this study. In a traditional epidemiological study, emergent subgroups beyond established clinical criteria are rarely addressed, whereas purely data-driven studies such as this one have the potential to reveal unmet needs such as cardiovascular medication in women with high HDL-cholesterol. Furthermore, this is the first study of type 1 diabetes, to our knowledge, that has successfully replicated an epidemiological multivariable model across two time periods. On the other hand, the lack of replication in an independent cohort is a limitation, and caution is warranted when generalising the results from the FinnDiane cohort to other ethnic groups or socioeconomic settings. Another limitation is the study design (a single biochemical sample at baseline and long baseline period) that



does not directly address the time-dependent trajectories of individuals or disease incidence. Despite the limitations, we maintain that our statistical subtyping framework is a feasible and robust discovery tool for precision medicine initiatives that aim to integrate diverse clinical and epidemiological datasets. In particular, the original unsupervised SOM model was constructed without using any of the follow-up data (thus preventing over-fitting), and we believe this is why the SMRs replicated so accurately across the two time periods.

Big data in medicine remains an emerging area [9, 10], with significant technical and political challenges to tackle before new paradigms such as precision medicine can be successfully applied in the clinics. This study focuses on the epidemiological aspect of data-driven subtyping, but caution is warranted before these approaches can be translated into clinical practice. The prediction of vascular endpoints at an individual level from a single blood sample is difficult to achieve, regardless of statistical methods, because of standardisation issues between study centres and countries, stochasticity of physiological events, genetic heterogeneity and the diversity of personal circumstances within human populations. That said, there is great value in applying the SOM in large population cohorts to reveal subgroups that may not be covered by traditional approaches or accurately addressed by current treatment guidelines.

Baseline metabolic subtypes, derived from an array of quantitative biomarkers, contain a wealth of diagnostic and prognostic information on type 1 diabetes. Here, we have demonstrated how this information can be used to identify specific subgroups towards new designs of more effective public health interventions. For instance, we found that men with type 1 diabetes and the metabolic syndrome, and a subset of women with high HDL-cholesterol, may represent justifiable and cost-effective points of intervention to reduce the burden of late vascular complications. We also defined the ideal sex-specific metabolic profile in type 1 diabetes that can be adapted into a multivariable treatment target for precision medicine. Finally, the updated statistics from this study highlight the persisting health gap between those with type 1 diabetes and the Finnish background population, and the urgent need to develop effective targeted treatments to prevent the progression of diabetic kidney disease.

**Acknowledgements** The authors acknowledge the contributions by physicians and nurses at each of the FinnDiane study centres (see ESM). Some of the data were presented as an abstract at the 52nd EASD Annual Meeting, Munich 2016, at the ADS-ADEA (Australian Diabetes Society and Australian Diabetes Educators Association) Annual Scientific Meeting, Cold Coast, 2016, at the 76th ADA Scientific Sessions, New Orleans, 2016 and at the 29th EDNSG (European Diabetic Nephropathy Study Group) Annual Meeting, Pisa, 2016.

**Data availability** No data are available. The ethical statement and the informed consent do not allow for free data availability.

**Funding** This research was supported by grants from the Folkhälsan Research Foundation, the Wilhelm and Else Stockmann Foundation, the Päivikki and Sakari Sohlberg Foundation, the Novo Nordisk Foundation (NNF14SA0003) and the Academy of Finland (134379).

**Duality of interest** P-HG has received lecture honorary honoraria from Boehringer Ingelheim, Genzyme, Novartis, Novo Nordisk, MSD, Eli Lilly and Medscape. P-HG is an advisory board member for Boehringer Ingelheim, Eli Lilly, Novartis, Abbott and AbbVie. P-HG has received investigator-initiated study grants from Eli Lilly and Roche. The funding sources were not involved in the design or conduct of the study. The authors declare that there is no duality of interest associated with this manuscript.

**Author contributions** RL, IT and V-PM designed and carried out the data analyses, interpreted the results and drafted the manuscript. CF, VH and P-HG contributed to the acquisition of data and revised the manuscript critically for important intellectual content. V-PM and P-HG are the guarantors of this work and, as such, had full access to all the data in the study and take full responsibility for the integrity of the data and the accuracy of the data analysis. All authors gave their final approval of this version of the manuscript.

## References

- Jacobson AM, Braffett BH, Cleary PA, Gubitosi-Klug RA, Larkin ME, DCCT/EDIC Research Group (2013) The long-term effects of type 1 diabetes treatment and complications on health-related quality of life: a 23-year follow-up of the diabetes control and complications/epidemiology of diabetes interventions and complications cohort. *Diabetes Care* 36:3131–3138
- Ray JA, Valentine WJ, Secnik K et al (2005) Review of the cost of diabetes complications in Australia, Canada, France, Germany, Italy and Spain. *Curr Med Res Opin* 21:1617–1629
- Groop PH, Thomas MC, Moran JL et al (2009) The presence and severity of chronic kidney disease predicts all-cause mortality in type 1 diabetes. *Diabetes* 58:1651–1658
- Gross JL, de Azevedo MJ, Silveiro SP, Canani LH, Caramori ML, Zelmanovitz T (2005) Diabetic nephropathy: diagnosis, prevention, and treatment. *Diabetes Care* 28:164–176
- D'agostino RBS, Grundy S, Sullivan LM, Wilson P, CHD Risk Prediction Group (2001) Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA* 286:180–187
- Zgibor JC, Piatt GA, Ruppert K, Orchard TJ, Roberts MS (2006) Deficiencies of cardiovascular risk prediction models for type 1 diabetes. *Diabetes Care* 29:1860–1865
- Mäkinen VP, Kangas AJ, Soininen P, Wurtz P, Groop PH, Ala-Korpela M (2013) Metabolic phenotyping of diabetic nephropathy. *Clin Pharmacol Ther* 94:566–569
- Mäkinen VP, Soininen P, Forsblom C et al (2008) 1H NMR metabolomics approach to the disease continuum of diabetic complications and premature death. *Mol Syst Biol* 4:167
- Richardson A, Signor BM, Lidbury BA, Badrick T (2016) Clinical chemistry in higher dimensions: machine-learning and enhanced prediction from routine clinical chemistry data. *Clin Biochem* 49:1213–1220
- Alyass A, Turcotte M, Meyre D (2015) From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med Genet* 8:33
- Mäkinen VP, Forsblom C, Thorn LM et al (2008) Metabolic phenotypes, vascular complications, and premature deaths in a population of 4,197 patients with type 1 diabetes. *Diabetes* 57:2480–2487

12. Thorn LM, Forsblom C, Fagerudd J et al (2005) Metabolic syndrome in type 1 diabetes: association with diabetic nephropathy and glycaemic control (the FinnDiane study). *Diabetes Care* 28:2019–2024
13. R Development Core Team (2011) R: a language and environment for statistical computing. Vienna, Austria: the R foundation for statistical computing. ISBN 3-900051-07-0. Available online at <http://www.R-project.org/>
14. Jørgensen ME, Almdal TP, Carstensen B (2013) Time trends in mortality rates in type 1 diabetes from 2002 to 2011. *Diabetologia* 56:2401–2404
15. Petrie D, Lung TW, Rawshani A et al (2016) Recent trends in life expectancy for people with type 1 diabetes in Sweden. *Diabetologia* 59:1167–1176
16. Harjutsalo V, Forsblom C, Groop PH (2011) Time trends in mortality in patients with type 1 diabetes: nationwide population based cohort study. *BMJ* 343:d5364
17. Gagnum V, Stene LC, Sandvik L et al (2015) All-cause mortality in a nationwide cohort of childhood-onset diabetes in Norway 1973–2013. *Diabetologia* 58:1779–1786
18. Lung TW, Hayes AJ, Herman WH, Si L, Palmer AJ, Clarke PM (2014) A meta-analysis of the relative risk of mortality for type 1 diabetes patients compared to the general population: exploring temporal changes in relative mortality. *PLoS One* 9:e113635
19. Miller RG, Secrest AM, Sharma RK, Songer TJ, Orchard TJ (2012) Improvements in the life expectancy of type 1 diabetes. The Pittsburgh epidemiology of diabetes complications cohort study. *Diabetes* 61:2987–2992
20. Harding JL, Shaw JE, Peeters A, Davidson S, Magliano DJ (2016) Age-specific trends from 2000–2011 in all-cause and cause-specific mortality in type 1 and type 2 diabetes: a cohort study of more than one million people. *Diabetes Care* 39:1018–1026
21. Van der Zee S, Baber U, Elmariah S, Winston J, Fuster V (2009) Cardiovascular risk factors in patients with chronic kidney disease. *Nat Rev Cardiol* 6:580–589
22. Thorn LM, Forsblom C, Waden J et al (2009) Metabolic syndrome as a risk factor for cardiovascular disease, mortality, and progression of diabetic nephropathy in type 1 diabetes. *Diabetes Care* 32:950–952
23. Pambianco G, Costacou T, Orchard TJ (2007) The prediction of major outcomes of type 1 diabetes: a 12-year prospective evaluation of three separate definitions of the metabolic syndrome and their components and estimated glucose disposal rate: the Pittsburgh epidemiology of diabetes complications study experience. *Diabetes Care* 30:1248–1254
24. Kilpatrick ES, Rigby AS, Atkin SL (2007) Insulin resistance, the metabolic syndrome, and complication risk in type 1 diabetes: "double diabetes" in the diabetes control and complications trial. *Diabetes Care* 30:707–712
25. Orchard TJ, Nathan DM, Zinman B et al (2015) Association between seven years of intensive treatment of type 1 diabetes and long term mortality. *JAMA* 313:45–53
26. Mendelsohn ME, Karas RH (2005) Molecular and cellular basis of cardiovascular gender differences. *Science* 308:1583–1587
27. Huxley RR, Peters SA, Mishra GD, Woodward M (2015) Risk of all-cause mortality and vascular events in women versus men with type 1 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol* 3:198–206
28. Lloyd CE, Kuller LH, Ellis D, Becker DJ, Wing RR, Orchard TJ (1996) Coronary artery disease in IDDM. Gender difference in risk factors but not risk. *Atheroscler Thromb Vasc Biol* 16:720–726
29. Brown SA, Jiang B, McElwee-Malloy M, Wakeman C, Breton MD (2015) Fluctuations of hyperglycemia and insulin sensitivity are linked to menstrual cycle phases in women with T1D. *J Diabetes Sci Technol* 9:1192–1199
30. Tolonen N, Forsblom C, Thorn L et al (2009) Lipid abnormalities predict progression of renal disease in patients with type 1 diabetes. *Diabetologia* 52:2522–2530
31. Saraheimo M, Teppo AM, Forsblom C, Fagerudd J, Groop PH (2003) Diabetic nephropathy is associated with low-grade inflammation in type 1 diabetic patients. *Diabetologia* 6:1402–1407
32. Holmes MV, Asselbergs FW, Palmer TM et al (2015) Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* 36:539–550
33. Voight B, Peloso G, Orho-Melander M et al (2012) Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* 380:572–580
34. Costacou T, Evans RW, Orchard TJ (2011) High density lipoprotein cholesterol in diabetes: is high always better? *J Clin Lipidol* 5:387–394