

Electronic supplementary material: Validation of register-based diabetes classifiers in Danish data

S1: Codes used in register data

S1: Codes used in diabetes classifiers (including subcodes):			
Hospital diagnoses: ICD-10 codes	Prescription purchases: ATC codes	HbA1c samples: NPU codes	Health service codes: SPEC2 codes
Diabetes: E10-E14	Glucose-lowering drugs: A10	HbA1c (IFCC): NPU27300	Diabetes-specific podiatrist services: 54
Type 1 diabetes: E10	Insulins (including insulin analogues): A10A	HbA1c (DCCT): NPU03835 (samples also reported as NPU27300 were excluded)	
Type 2 diabetes: E11	Non-insulin glucose- lowering drugs: A10B		
Polycystic Ovary Syndrome: E282	Clomifene: G03GB02 Antiandrogens and estrogens: G03HB		
Gestational diabetes mellitus: O244			
Pregnancy endings: DO00 – DO06, DO80 - DO84, DZ33, DZ37 (pregnancies are defined as non- overlapping windows spanning 40 weeks before a registration to 12 weeks after)			
Maternal care: DO07, DO10 - DO49, DZ32, DZ34 – 36 (pregnancy windows are defined by maternal care codes only if no pregnancy ending is registered between 12 weeks before the maternal care record and 40 weeks after)			

S2: Detailed OSDC flow-charts

See the following two pages.

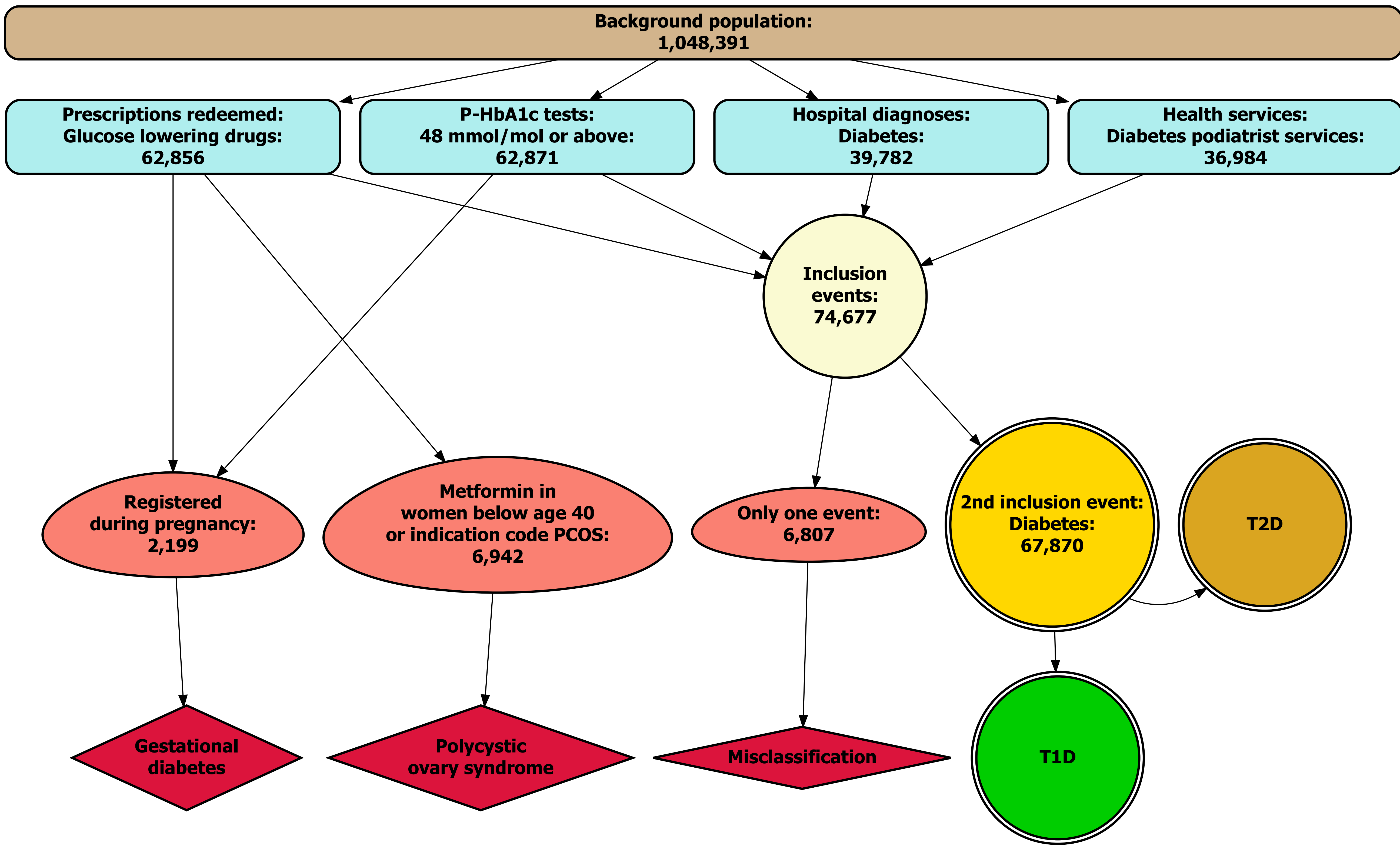
Flowchart of inclusion into study population:

Numbers indicate individuals in the source data and the register-classified diabetes population. For each type of inclusion event, this corresponds to any individual with at least one event of the respective type of inclusion criteria. For censored events, this corresponds to any individual with at least one censored event due to the respective type of censoring.

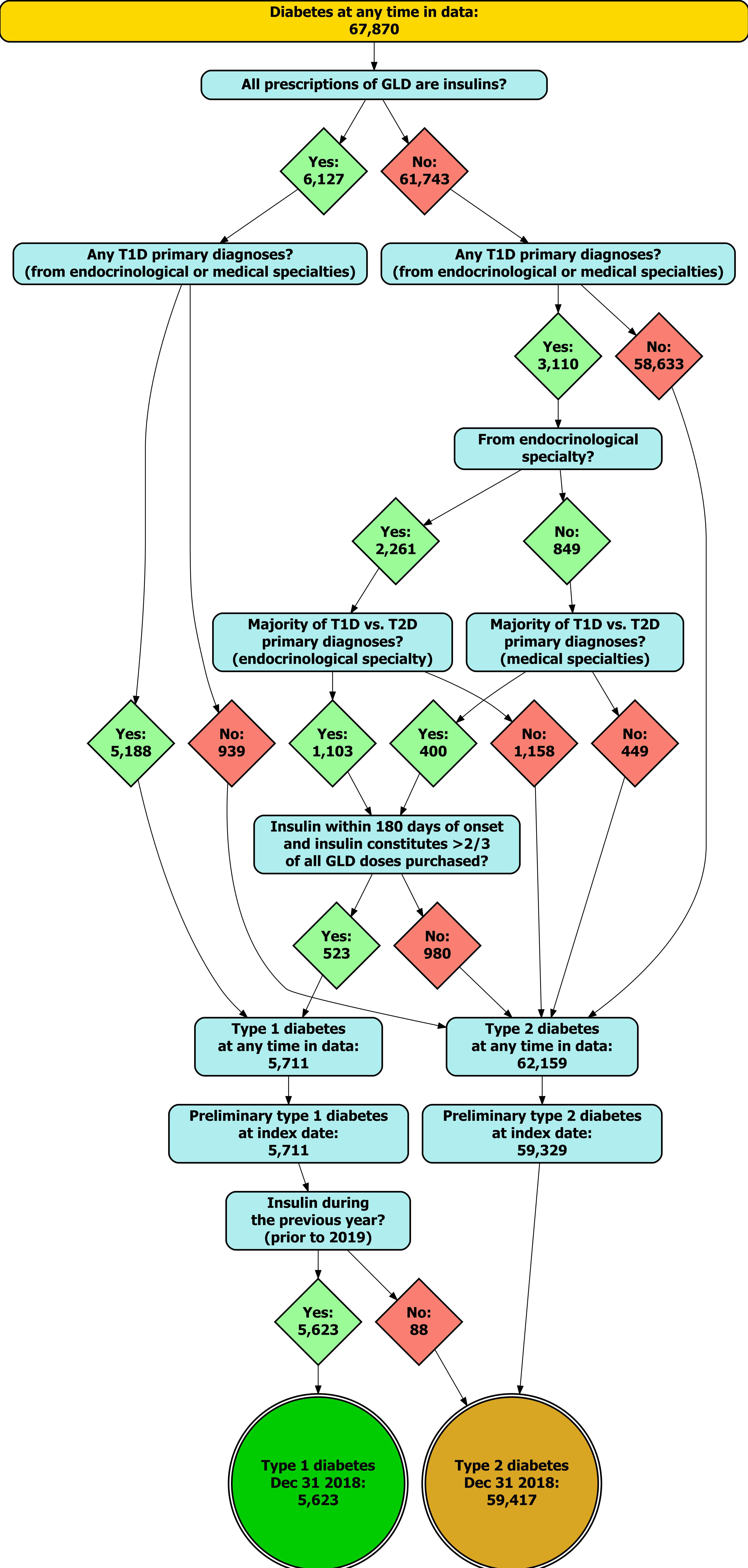
Flowchart of inclusion into study population: Diabetes type

Numbers indicate individuals classified as type 1 diabetes or type 2 diabetes across each step of the algorithm. Note that due to the source data on prescription purchases and hospital diagnoses being restricted to December 31 2018 and earlier, only individuals with register-classified type 2 diabetes were excluded due to onset after the index date.

Flowchart of inclusion into study population



Flowchart of inclusion into study population: Diabetes type



S3: Bootstrapped main analyses

Computed means and 95% confidence intervals for each diabetes type, classifiers and validation metric in 1000 bootstrap samples

	Sensitivity	Specificity	PPV	NPV
Type 1 diabetes				
OSDC	0.774 (0.739; 0.809)	0.999 (0.999; 1.000)	0.951 (0.929; 0.972)	0.997 (0.996; 0.997)
RSCD	0.702 (0.661; 0.740)	0.999 (0.999; 1.000)	0.952 (0.930; 0.973)	0.996 (0.995; 0.996)
Type 2 diabetes				
OSDC	0.943 (0.940; 0.946)	0.989 (0.988; 0.990)	0.878 (0.866; 0.889)	0.995 (0.995; 0.996)
RSCD	0.899 (0.891; 0.908)	0.992 (0.991; 0.993)	0.899 (0.888; 0.910)	0.992 (0.991; 0.992)

S4: Health in Central Denmark respondent characteristics without downsampling:

Supplementary 4: Characteristics of diabetes cohorts by diabetes status without downsampling		T1D	T2D	No diabetes
Self-reported diabetes cohort				
Diabetes type (self-reported)	N (%)	3,241 (6.8)	16,024 (33.8)	28,085 (59.3)
Age at onset (self-reported)	Mean (SD)	26.6 (16.2)	51.4 (11.5)	
Sex: Female	N (%)	1,449 (44.7)	6,328 (39.5)	12,463 (44.4)
Age	Mean (SD)	50.2 (14.9)	61.9 (9.3)	59.9 (11.4)
Migrant origin	N (%)	151 (4.7)	1,135 (7.1)	1,324 (4.7)
OSDC-classified diabetes cohort				
Diabetes type (OSDC-classified)	2,674 (5.6)	18,018 (38.1)	26,658 (56.3)	2,674 (5.6)
Age at onset (OSDC-classified)	30.4 (13.1)	53.4 (10.2)		30.4 (13.1)
Sex: Female	1,218 (45.5)	7,395 (41.0)	11,627 (43.6)	1,218 (45.5)
Age	48.6 (15.1)	61.5 (9.6)	60.0 (11.4)	48.6 (15.1)
Migrant origin	83 (3.1)	1,338 (7.4)	1,189 (4.5)	83 (3.1)
RSCD-classified diabetes cohort				
Diabetes type (RSCD-classified)	N (%)	2,419 (5.1)	16,647 (35.2)	28,284 (59.7)
Age at onset (RSCD-classified)	Mean (SD)	28.5 (13.2)	52.5 (10.7)	
Sex: Female	N (%)	1,076 (44.5)	6,688 (40.2)	12,476 (44.1)
Age	Mean (SD)	48.0 (15.3)	61.4 (9.6)	60.1 (11.4)
Migrant origin	N (%)	80 (3.3)	1,222 (7.3)	1,308 (4.6)

S5: Validation analyses of T1D stratified by sex

Supplementary 5: Analyses of T1D stratified by sex				
Men				
OSDC	Survey: +T1D	Survey: -T1D	Total N	
OSDC: +T1D	163	11	174	PPV: 0.937 (0.890, 0.968)
OSDC: -T1D	65	16,386	16,451	NPV: 0.996 (0.995, 0.997)
Total N	228	16,397	16,625	
	Sens: 0.715 (0.652, 0.773)	Spec: 0.999 (0.999, 1.000)		
Women				
OSDC	Survey: +T1D	Survey: -T1D	Total N	
OSDC: +T1D	154	8	162	PPV: 0.951 (0.905, 0.978)
OSDC: -T1D	28	12,576	12,604	NPV: 0.998 (0.997, 0.999)
Total N	182	12,584	12,766	
	Sens: 0.846 (0.785, 0.895)	Spec: 0.999 (0.999, 1.000)		
Men				
RSCD	Survey: +T1D	Survey: -T1D	Total N	
RSCD: +T1D	155	10	165	PPV: 0.939 (0.891, 0.971)
RSCD: -T1D	73	16,387	16,460	NPV: 0.996 (0.994, 0.997)
Total N	228	16,397	16,625	
	Sens: 0.680 (0.615, 0.740)	Spec: 0.999 (0.999, 1.000)		
Women				
RSCD	Survey: +T1D	Survey: -T1D	Total N	
RSCD: +T1D	132	7	139	PPV: 0.950 (0.899, 0.980)
RSCD: -T1D	50	12,577	12,627	NPV: 0.996 (0.995, 0.997)
Total N	182	12,584	12,766	
	Sens: 0.725 (0.654, 0.789)	Spec: 0.999 (0.999, 1.000)		

S6: Validation analyses of T1D stratified by sex

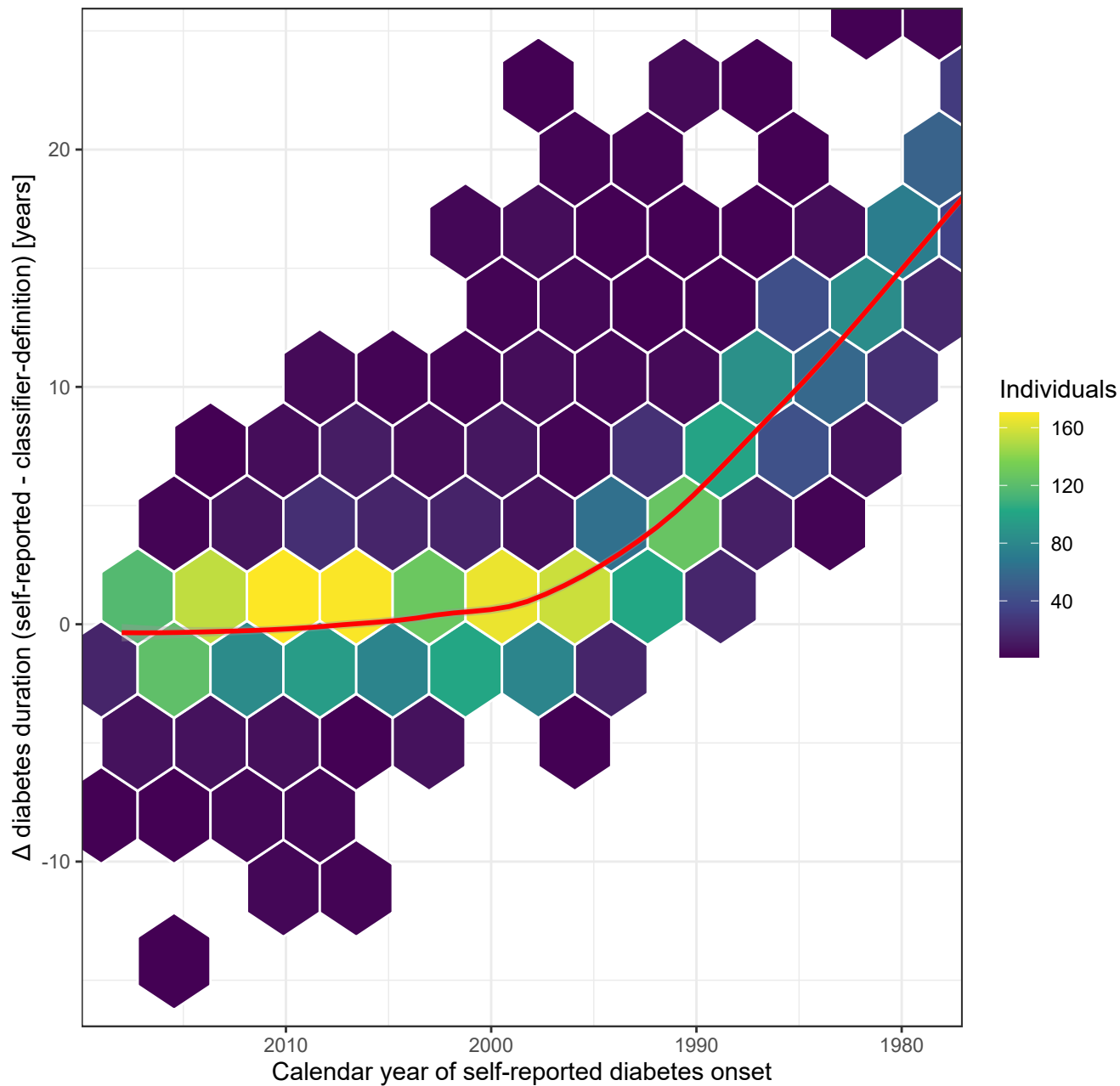
Supplementary 6: Analyses of T2D stratified by sex				
Men				
OSDC	Survey: +T2D	Survey: -T2D	Total N	
OSDC: +T2D	1,264	156	1,420	PPV: 0.890 (0.873, 0.906)
OSDC: -T2D	74	15,131	15,205	NPV: 0.995 (0.994, 0.996)
Total N	1,338	15,287	16,625	
	Sens: 0.945 (0.931, 0.956)	Spec: 0.990 (0.988, 0.991)		
RSCD				
RSCD	Survey: +T2D	Survey: -T2D	Total N	
RSCD: +T2D	1,223	108	1,331	PPV: 0.919 (0.903, 0.933)
RSCD: -T2D	115	15,179	15,294	NPV: 0.992 (0.991, 0.994)
Total N	1,338	15,287	16,625	
	Sens: 0.914 (0.898, 0.929)	Spec: 0.993 (0.991, 0.994)		
Women				
OSDC	Survey: +T2D	Survey: -T2D	Total N	
OSDC: +T2D	834	143	977	PPV: 0.854 (0.830, 0.875)
OSDC: -T2D	51	11,738	11,789	NPV: 0.996 (0.994, 0.997)
Total N	885	11,881	12,766	
	Sens: 0.942 (0.925, 0.957)	Spec: 0.988 (0.986, 0.990)		
RSCD				
RSCD	Survey: +T2D	Survey: -T2D	Total N	
RSCD: +T2D	788	121	909	PPV: 0.867 (0.843, 0.888)
RSCD: -T2D	97	11,760	11,857	NPV: 0.992 (0.990, 0.993)
Total N	885	11,881	12,766	
	Sens: 0.890 (0.868, 0.910)	Spec: 0.990 (0.988, 0.992)		

Supplementary 7: Bias between self-reported and register-classified diabetes duration across calendar year of diabetes onset:

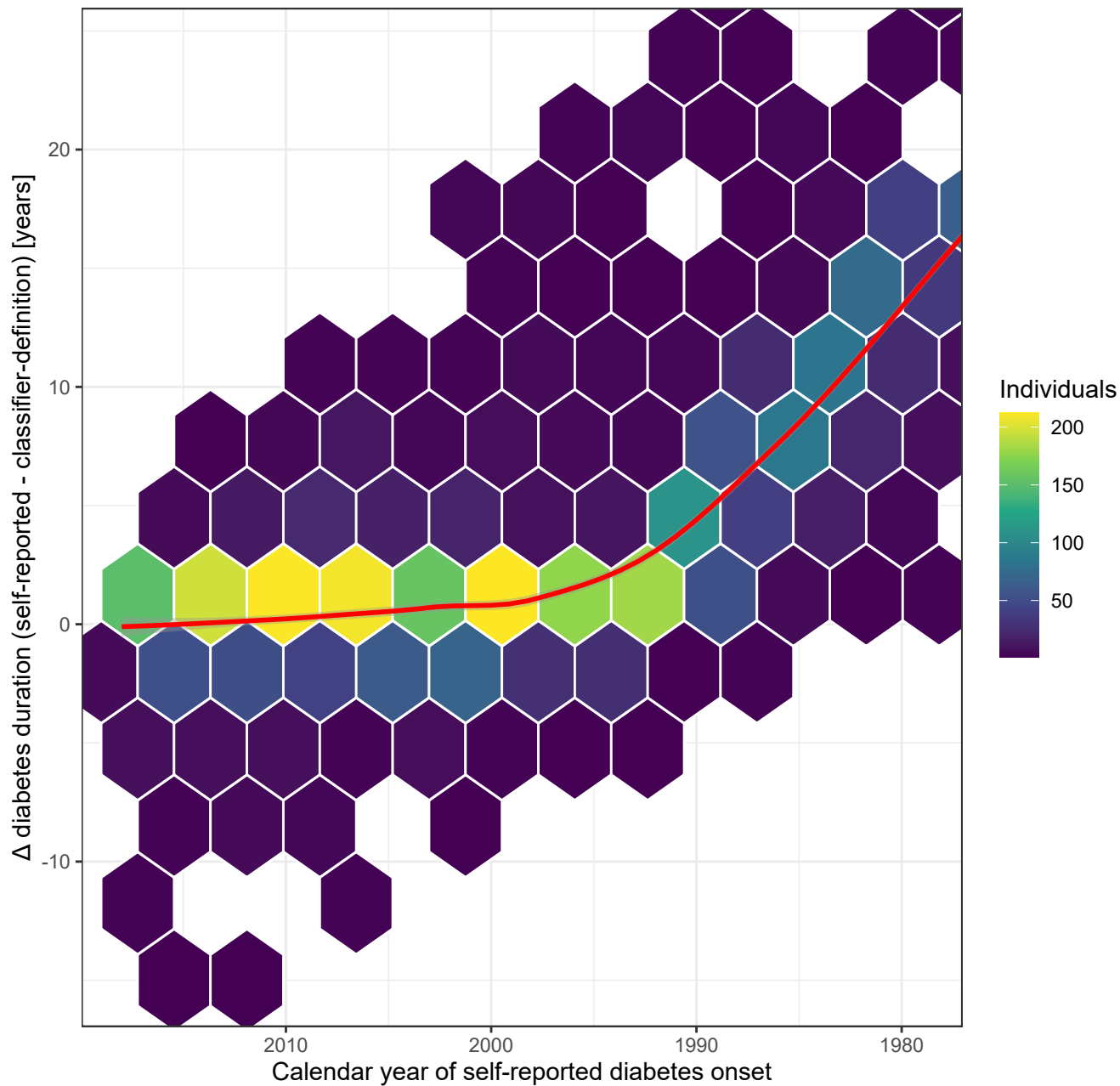
See the following four pages of plots for each diabetes type and register-based classifier.

Number of individuals within each hexagon is denoted by its color. A smoothed loess regression line with 95% confidence intervals as shades is plotted in red. To comply with statistical disclosure requirements to protect individual-level data privacy, analyses were performed in the survey respondent population without subsampling.

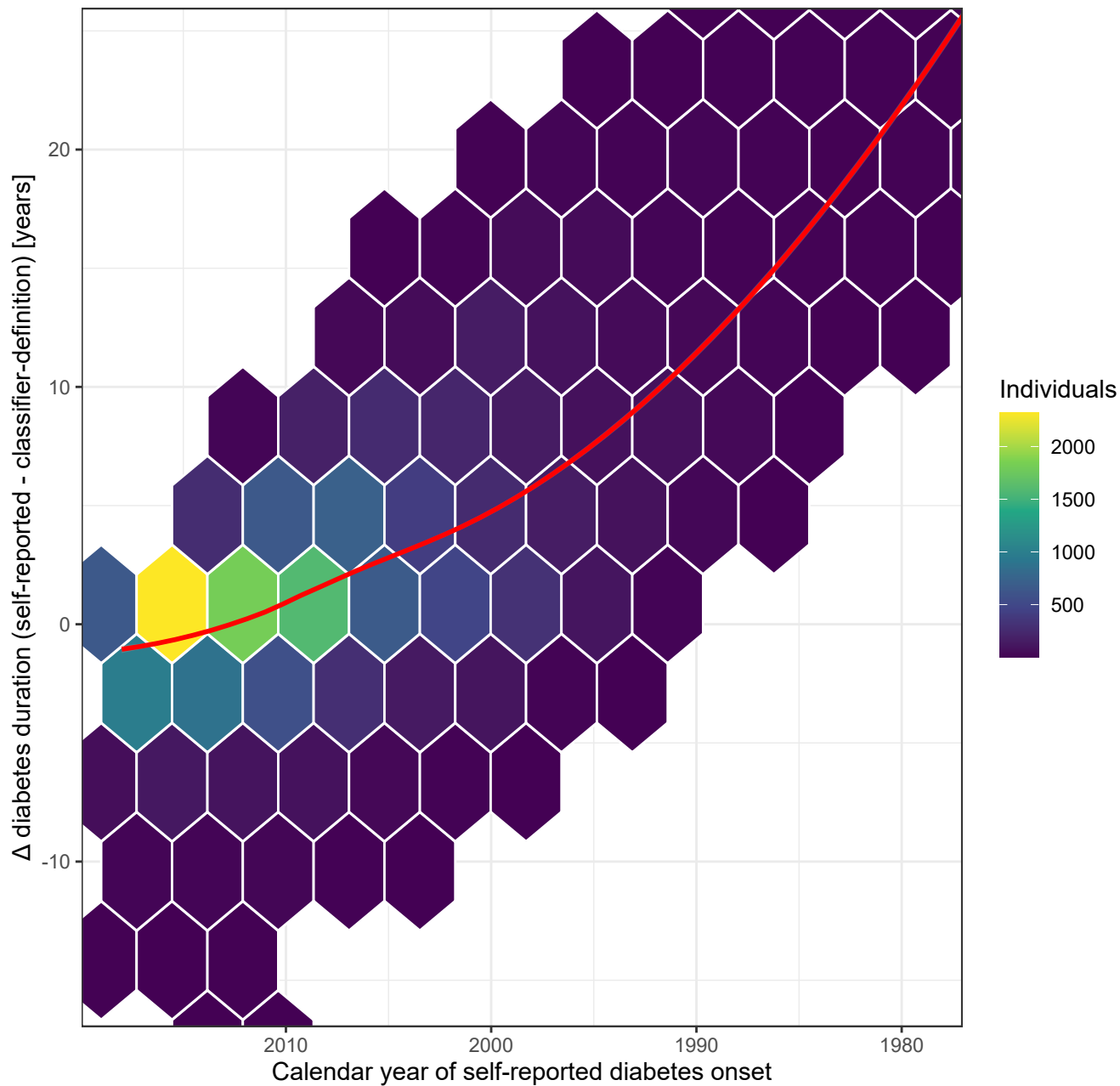
Diabetes duration: T1D in OSDC



Diabetes duration: T1D in RSCD



Diabetes duration: T2D in OSDC



Diabetes duration: T2D in RSCD

