Predicting Islet Cell Autoimmunity and Type 1 Diabetes: An 8-Year TEDDY Study Progress Report

*Jeffrey P. Krischer,[1] Xiang Liu,[1] Kendra Vehik,[1] Beena Akolkar,[2] William A. Hagopian,[3] Marian J. Rewers,[4] Jin-Xiong She,[5] Jorma Toppari,[6,7] Anette-G. Ziegler,[8] Åke Lernmark,[9] and the TEDDY Study Group**

Diabetes Care 2019;42:1051–1060 | https://doi.org/10.2337/dc18-2282

OBJECTIVE

Assessment of the predictive power of The Environmental Determinants of Diabetes in the Young (TEDDY)-identified risk factors for islet autoimmunity (IA), the type of autoantibody appearing first, and type 1 diabetes (T1D).

RESEARCH DESIGN AND METHODS

A total of 7,777 children were followed from birth to a median of 9.1 years of age for the development of islet autoantibodies and progression to T1D. Time-dependent sensitivity, specificity, and receiver operating characteristic (ROC) curves were calculated to provide estimates of their individual and collective ability to predict IA and T1D.

RESULTS

HLA genotype (DR3/4 vs. others) was the best predictor for IA (Youden's index $J = 0.117$) and single nucleotide polymorphism rs2476601, in PTPN22, was the best predictor for insulin autoantibodies (IAA) appearing first (IAA-first) ($J = 0.123$). For GAD autoantibodies (GADA)-first, weight at 1 year was the best predictor ($J = 0.114$). In a multivariate model, the area under the ROC curve (AUC) was 0.678 (95% CI 0.655, 0.701), 0.707 (95% CI 0.676, 0.739), and 0.686 (95% CI 0.651, 0.722) for IA, IAA-first, and GADA-first, respectively, at 6 years. The AUC of the prediction model for T1D at 3 years after the appearance of multiple autoantibodies reached 0.706 (95% CI 0.649, 0.762).

CONCLUSIONS

Prediction modeling statistics are valuable tools, when applied in a time-until-event setting, to evaluate the ability of risk factors to discriminate between those who will and those who will not get disease. Although significantly associated with IA and T1D, the TEDDY risk factors individually contribute little to prediction. However, in combination, these factors increased IA and T1D prediction substantially.

The Environmental Determinants of Diabetes in the Young (TEDDY) study is a large (N = 8,676) prospective cohort study designed to identify environmental factors influencing or protecting against development of islet autoimmunity (IA) and onset of type 1 diabetes (T1D) (1). This ongoing international study began in 2004. Published work to date has examined IA/T1D risk factors including HLA DR-DQ genotypes, sex, T1D-related single nucleotide polymorphisms (SNPs) (2,3), dietary factors (soluble fiber, probiotics, and infant formula type) (4–6), growth in early life (7), age at autoantibody seroconversion, and the first appearing autoantibody (3).

[1]*Health Informatics Institute, Morsani College of Medicine, University of South Florida, Tampa, FL*
[2]*National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD*
[3]*Pacific Northwest Research Institute, Seattle, WA*
[4]*Barbara Davis Center for Childhood Diabetes, University of Colorado, Aurora, CO*
[5]*Center for Biotechnology and Genomic Medicine, Medical College of Georgia, Augusta University, Augusta, GA*
[6]*Research Centre for Integrative Physiology and Pharmacology, Institute of Biomedicine, University of Turku, Turku, Finland*
[7]*Department of Pediatrics, Turku University Hospital, Turku, Finland*
[8]*Klinikum rechts der Isar, Technische Universität München, and Forschergruppe Diabetes e.V., Helmholtz Zentrum München, Neuherberg, Germany*
[9]*Department of Clinical Sciences, Clinical Research Centre, Lund University, Skane University Hospital, Malmö, Sweden*

Corresponding author: Jeffrey P. Krischer, jeffrey.krischer@epi.usf.edu

Received 1 November 2018 and accepted 12 March 2019

This article contains Supplementary Data online at http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc18-2282/-/DC1.

**A complete list of the TEDDY Study Group can be found in the Supplementary Data.*

TEDDY-identified risk factors are those factors with statistical significance in time-to-event analyses. However, statistically significant risk factors are not automatically good predictors of disease (8). For example, a risk factor may be associated with or cause a disease but be a poor predictor because it is only one of many causal mechanisms. Conversely, a risk factor may predict a disease and yet not be a component of its pathogenesis. Therefore, quantifying (9) the individual and cumulative/incremental predictive utility of the identified risk factors for IA and T1D is an important part of the investigative focus of the TEDDY study, as it should be in other similarly constructed prospective cohort studies.

The aim of this work is to evaluate the individual and combined ability of TEDDY-identified genetic and environmental risk factors to predict IA and progression to T1D. Specificity, sensitivity and receiver operating characteristic (ROC) curves are fundamental tools for diagnostic test evaluation. Since TEDDY is a prospective cohort study, the status of disease outcome (development of IA or onset of T1D) changes over time and the analysis of the cohort at any given point in time is subject to censoring. Hence, time dependency needs to be incorporated in an assessment of the predictive assessment of these risk factors, which may result in different values of sensitivity and specificity over the course of the study (10,11). In this work, time-dependent analysis of sensitivity, specificity, and ROC curve was constructed to evaluate the identified risk factor for predicting disease outcome at several time points of interest.

RESEARCH DESIGN AND METHODS

Participants

TEDDY includes six clinical research centers: Colorado, Georgia/Florida, Washington, Finland, Germany, and Sweden. Written informed consent was obtained for all study participants from a parent or primary caretaker, separately, for genetic screening and participation in the prospective follow-up, beginning at birth. The high-risk genotypes selected for inclusion for participants screened from the general population (GP) were as follows: DRB1*04-DQA1*03-DQB1*03:02/DRB1*03-DQA1*05-DQB1*02:01 (DR3/4), DRB1*04-DQA1*03-DQB1*03:02/DRB1*04-DQA1*03-DQB1*03:02 (DR4/4), DRB1*04-DQA1*03-DQB1*03:02/DRB1*08-DQA1*04-DQB1*04:02 (DR4/8), and DRB1*03-DQA1*05-DQB1*02:01/DRB1*03-DQA1*05-DQB1*02:01 (DR3/3). Additional study inclusion genotypes for first-degree relatives (FDRs) of a subject with T1D were DRB1*04-DQA1*03-DQB1*03:02/DRB1*04-DQA1*03-DQB1*02:02 (DR4/4b), DRB1*04-DQA1*03-DQB1*03:02/DRB1*01-DQA1*01-DQB1*05:01 (DR4/1), DRB1*04-DQA1*03-DQB1*03:02/DRB1*13-DQA1*01-DQB1*06:04 (DR4/13), DRB1*04-DQA1*03-DQB1*03:02/DRB1*09-DQA1*03-DQB1*03:03 (DR4/9), and DRB1*03-DQA1*05-DQB1*02:01/DRB1*09-DQA1*03-DQB1*03:03 (DR3/9). The HLA DR-DQ genotype abbreviations shown in parentheses will be used throughout this article. Genotyping was confirmed by reverse blot hybridization at the central HLA Reference Laboratory at Roche Molecular Systems, Oakland, CA (12), along with the *INS*-23 Hph1 (rs689), *CTLA4* T17A (rs231775), and *PTPN22* R620 W (rs2476601) SNP primer pairs. Children enrolled are followed prospectively from 3 months to 15 years, with study visits every 3 months until 4 years of age and every 3 or 6 months thereafter depending on autoantibody positivity. All children who are persistent positive for any autoantibody are followed every 3 months until the age of 15 years or onset of T1D. T1D was defined according to the American Diabetes Association's criteria for diagnosis. Detailed study design and methods have previously been published (1,13). The study was approved by local institutional review or ethics boards and is monitored by an external evaluation committee formed by the National Institutes of Health.

SNP analysis was performed by the Center for Public Health Genomics at University of Virginia, using the Illumina Immunochip, which is a custom array for genotyping of SNPs selected from regions of the human genome firmly associated with autoimmune diseases (14). The final selection of SNPs containing ~186,000 SNPs in 186 regions, for 12 autoimmune diseases, was decided by the Immunochip Consortium. TEDDY previously examined whether any of 41 non-HLA SNPs previously shown to be associated with T1D conferred risk for IA (2).

Islet Autoantibodies and Identified Risk Factors

Autoantibodies to insulin (IAA), GAD (GADA), or insulinoma antigen-2 (IA-2A) were measured in two laboratories by radiobinding assays (1,13). In the U.S., all sera were assayed at the Barbara Davis Center for Childhood Diabetes at the University of Colorado Denver; in Europe, all sera were assayed at the University of Bristol, U.K. Both laboratories demonstrated high sensitivity and specificity as well as concordance (15). All positive islet autoantibodies and 5% of negative samples were retested in the other reference laboratory and deemed confirmed if concordant. Persistent IA was defined as confirmed positive autoantibodies to IAA, GADA, or IA-2A in at least two consecutive samples.

For the development of IA, the TEDDY study group has examined HLA DR-DQ genotypes, sex, family history of T1D, T1D-associated non-HLA SNPs presented by the Type 1 Diabetes Genetics Consortium (T1DGC) meta-analysis (2,14), complement genes (16), dietary factors (soluble fiber, probiotic, and infant formula type) (4–6), and growth in early life (6,7).

TEDDY has previously published that diabetes-related persistent confirmed islet autoantibodies first appear singly, and autoantibodies to IAA appear at an earlier age than those to GADA (17). The order of appearance was shown to be related to HLA DR-DQ genotypes. Therefore, TEDDY also examined all of the aforementioned IA-associated risk factors, along with early infection-related conditions, on the risks of developing IAA or GADA as the first appearing indication of autoimmunity (IAA appearing first [IAA-first] or GADA-first, respectively) using multivariate cause-specific proportional hazards modeling (18).

T1D and Identified Risk Factors

For T1D, the analysis included age at multiple autoantibodies, more than a single autoantibody as the first appearing indication of seroconversion, and female sex along with SNPs rs10517086, rs1534422, and rs2327832 in *TNFAIP3*, which were associated with an increased risk of progression to T1D, and SNP rs1004446 in *INS*, which was associated with decreased risk (3).

Statistical Methods

For each outcome (IA, IAA-first, GADA-first, or progression to T1D), we evaluated individual predictive ability of the TEDDY-identified risk factors by

Table 1—Characteristics of the TEDDY study population

	Category	IA negative (*n* = 7,041)	Any IA (*n* = 736)	IAA-first (*n* = 281)	GADA-first (*n* = 316)	Multiple IAs without progression to T1D (*n* = 215)	Multiple IAs with progression to T1D (*n* = 219)
Country	U.S.	2,972 (42.2)	246 (33.4)	83 (29.5)	123 (38.9)	80 (37.2)	65 (29.7)
	Finland	1,525 (21.7)	187 (25.4)	92 (32.7)	62 (19.6)	47 (21.9)	67 (30.6)
	Germany	475 (6.7)	53 (7.2)	20 (7.1)	14 (4.4)	16 (7.4)	21 (9.6)
	Sweden	2,069 (29.4)	250 (34.0)	86 (30.6)	117 (37.0)	72 (33.5)	66 (30.1)
Family history	FDR: mother	284 (4.0)	35 (4.8)	11 (3.9)	14 (4.4)	13 (6.0)	12 (5.5)
	FDR: father	346 (4.9)	74 (10.1)	29 (10.3)	32 (10.1)	26 (12.1)	30 (13.7)
	FDR: sibling	103 (1.5)	33 (4.5)	20 (7.1)	7 (2.2)	10 (4.7)	14 (6.4)
	GP	6,308 (89.6)	594 (80.7)	221 (78.6)	263 (83.2)	166 (77.2)	163 (74.4)
Sex	Male	3,562 (50.6)	404 (54.9)	160 (56.9)	167 (52.8)	136 (63.3)	109 (49.8)
	Female	3,479 (49.4)	332 (45.1)	121 (43.1)	149 (47.2)	79 (36.7)	110 (50.2)
HLA genotype	DR3/4	2,668 (37.9)	362 (49.2)	137 (48.8)	155 (49.1)	116 (54.0)	123 (56.2)
	DR4/4	1,396 (19.8)	139 (18.9)	51 (18.1)	52 (16.5)	49 (22.8)	37 (16.9)
	DR4/8	1,226 (17.4)	113 (15.4)	59 (21.0)	38 (12.0)	27 (12.6)	30 (13.7)
	DR3/3	1,523 (21.6)	96 (13.0)	21 (7.5)	68 (21.5)	15 (7.0)	17 (7.8)
	FDR specific*	228 (3.2)	26 (3.5)	13 (4.6)	3 (0.9)	8 (3.7)	12 (5.5)
Probiotics introduction age	≥28 days	6,522 (92.6)	694 (94.3)	267 (95.0)	300 (94.9)		
	<28 days	519 (7.4)	42 (5.7)	14 (5.0)	16 (5.1)		
Weight *z* score at 12 months, mean (SD)		−0.132 (1.022)	−0.008 (1.033)	−0.033 (1.055)	0.045 (1.038)		
First infant formula type during the first 7 days of life	No formula, no cow's milk	4,442 (63.1)	473 (64.3)	177 (63.0)	205 (64.9)		
	No formula, only cow's milk	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)		
	Nonhydrolyzed formula	2,237 (31.8)	211 (28.7)	79 (28.1)	94 (29.7)		
	Partially hydrolyzed formula	89 (1.3)	10 (1.4)	3 (1.1)	4 (1.3)		
	Extensively hydrolyzed formula	169 (2.4)	32 (4.3)	16 (5.7)	11 (3.5)		
	Other formula	103 (1.5)	10 (1.4)	6 (2.1)	2 (0.6)		
First infant formula type during the first 3 months of life	No formula, no cow's milk	1,880 (26.7)	213 (28.9)	87 (31.0)	88 (27.8)		
	No formula, only cow's milk	17 (0.2)	5 (0.7)	0 (0.0)	3 (0.9)		
	Nonhydrolyzed formula	4,523 (64.2)	450 (61.1)	163 (58.0)	202 (63.9)		
	Partially hydrolyzed formula	220 (3.1)	18 (2.4)	6 (2.1)	7 (2.2)		
	Extensively hydrolyzed formula	218 (3.1)	36 (4.9)	17 (6.0)	12 (3.8)		
	Other formula	183 (2.6)	14 (1.9)	8 (2.8)	4 (1.3)		
Age at multiple persistent confirmed autoantibodies (months), mean (SD)						60.4 (31.5)	30.2 (21.3)
Type of first autoantibody	GADA only					87 (40.5)	53 (24.2)
	IAA only					80 (37.2)	85 (38.8)
	IA-2A only					3 (1.4)	4 (1.8)
	Two or more autoantibodies					45 (20.9)	77 (35.2)
rs1004446 (*INS*)	GG	2,771 (39.4)	332 (45.1)	135 (48.0)	130 (41.1)	100 (46.5)	110 (50.2)
	AG	3,287 (46.7)	331 (45.0)	118 (42.0)	152 (48.1)	95 (44.2)	88 (40.2)

Continued on p. 1054

Table 1—Continued

	Category	IA negative (*n* = 7,041)	Any IA (*n* = 736)	IAA-first (*n* = 281)	GADA-first (*n* = 316)	Multiple IAs without progression to T1D (*n* = 215)	Multiple IAs with progression to T1D (*n* = 219)
	AA	983 (14.0)	73 (9.9)	28 (10.0)	34 (10.8)	20 (9.3)	21 (9.6)
rs10517086	GG	3,612 (51.3)	363 (49.3)	137 (48.8)	164 (51.9)	101 (47.0)	97 (44.3)
	AG	2,878 (40.9)	308 (41.8)	114 (40.6)	128 (40.5)	96 (44.7)	99 (45.2)
	AA	551 (7.8)	65 (8.8)	30 (10.7)	24 (7.6)	18 (8.4)	23 (10.5)
rs1534422	AA					76 (35.3)	50 (22.8)
	AG					106 (49.3)	115 (52.5)
	GG					33 (15.3)	54 (24.7)
rs2327832 (*TNFAIP3*)	AA					144 (67.0)	134 (61.2)
	AG					66 (30.7)	70 (32.0)
	GG					5 (2.3)	15 (6.8)
rs1143678 (*ITGAM*)	CC	5,162 (73.3)	545 (74.0)	206 (73.3)	234 (74.1)		
	CT	1,708 (24.3)	174 (23.6)	69 (24.6)	72 (22.8)		
	TT	171 (2.4)	17 (2.3)	6 (2.1)	10 (3.2)		
rs12708716 (*CLEC16A*)	AA	3,086 (43.9)	352 (47.9)	125 (44.5)	157 (49.8)		
	AG	3,100 (44.1)	322 (43.8)	130 (46.3)	136 (43.2)		
	GG	837 (11.9)	61 (8.3)	26 (9.3)	22 (7.0)		
rs2292239 (*ERBB3*)	GG	3,266 (46.4)	291 (39.5)	112 (39.9)	125 (39.6)		
	TG	3,070 (43.6)	359 (48.8)	141 (50.2)	152 (48.1)		
	TT	704 (10.0)	86 (11.7)	28 (10.0)	39 (12.3)		
rs2476601 (*PTPN22*)	GG	5,665 (80.5)	518 (70.4)	194 (69.0)	225 (71.2)		
	AG	1,292 (18.3)	205 (27.9)	79 (28.1)	87 (27.5)		
	AA	84 (1.2)	13 (1.8)	8 (2.8)	4 (1.3)		
rs2816316 (*RGS1*)	AA	4,693 (66.7)	485 (65.9)	188 (66.9)	210 (66.5)		
	AC	2,111 (30.0)	223 (30.3)	81 (28.8)	95 (30.1)		
	CC	237 (3.4)	28 (3.8)	12 (4.3)	11 (3.5)		
rs3184504 (*SH2B3*)	CC	2,224 (31.6)	175 (23.8)	71 (25.3)	70 (22.2)		
	CT	3,446 (48.9)	365 (49.6)	133 (47.3)	158 (50.0)		
	TT	1,371 (19.5)	196 (26.6)	77 (27.4)	88 (27.8)		
rs4597342 (*ITGAM*)	CC	3,135 (44.5)	311 (42.3)	132 (47.0)	128 (40.5)		
	CT	3,091 (43.9)	341 (46.3)	116 (41.3)	153 (48.4)		
	TT	812 (11.5)	84 (11.4)	33 (11.7)	35 (11.1)		
rs4948088 (*COBL*)	CC	6,401 (90.9)	689 (93.6)	265 (94.3)	293 (92.7)		
	AC	619 (8.8)	45 (6.1)	16 (5.7)	22 (7.0)		
	AA	21 (0.3)	2 (0.3)	0 (0.0)	1 (0.3)		

Data are presented as number (percentage) unless otherwise indicated. *FDR-specific HLA DR-DQ genotypes are DR4/4b, DR4/1, DR4/13, DR4/9, and DR3/9.

calculating time-dependent sensitivity (*Se*) and specificity (*Sp*) at 6 years of age (see Supplementary Data for 2 and 10 years follow-up). Family history of T1D was categorized as FDR versus GP, HLA genotypes were categorized as DR3/4 versus others, type of formula was categorized as extensively hydrolyzed cow's milk–based formula versus others, each of the SNPs was categorized as major (no copy of minor allele) versus minor/heterogeneous (one or two copies of minor alleles), and first appearing indication of seroconversion was categorized as more than a single autoantibody versus others. Continuous risk factors, e.g., the weight *z* score at 1 year and age at multiple autoantibodies onset, were evaluated using median value as the cutoff.

Since we were also interested in the incremental predictive ability of each risk factor, we performed a forward selection procedure, with an entry level ($P < 0.1$) to order and include risk factors from all of the identified risk factors by Cox proportional hazards regression. Then, we fitted Cox models by adding one risk factor at a time according to the order determined by the forward selection. For each of the fitted Cox models, a risk score was defined as the estimated linear predictor in the log hazard function. A time-dependent ROC curve at each of the specified time points was calculated for each of the risk scores (prediction models). The Youden's index (*J*), $Se + Sp - 1$ (9), and area under the curve of the time-dependent ROC (AUC) were calculated. For the analyses presented herein, *J* is defined at its maximum. The AUC quantifies the probability that the risk scores from a randomly selected pair of diseased and nondiseased individuals are correctly ordered.

The forward selection procedure was performed using the SAS software (version 9.4; SAS Institute, Cary, NC). The

package "survival" (19) was used to fit Cox models and the package "timeROC" (20) was used to calculate time-dependent sensitivity, specificity, and ROC in R 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

As of 31 July 2017, 7,777 of 8,676 (89.6%) enrolled children who were HLA DR-DQ genotype eligible and had Immunochip data available were included in the analyses. The median age of follow-up for the 7,777 children was 9.1 years (interquartile range 7.9–10.4). As summarized in Table 1, a total of 736 (9.5%) children developed one or more persistent autoantibodies (GADA, IAA, or IA-2A), and among them 281 (38%) children had IAA only as the first appearing autoantibody (IAA-first), 316 (43%) children had GADA only as the first appearing autoantibody (GADA-first), 17 (2%) children had IA-2A only as the first appearing autoantibody (IA-2A-first), and 104 (14%) children had two autoantibodies and 18 (2%) children had three autoantibodies appearing simultaneously. Altogether, 434 subjects developed two or more autoantibodies, and of these, 219 (50.5%) children developed T1D.

Predicting Development of IA

The sensitivity (*Se*), specificity (*Sp*), and Youden's index (*J*) of each risk factor for predicting IA, IAA-first, and GADA-first at 6 years are presented in Table 2. Among all the identified risk factors, HLA genotype (DR3/4 vs. others) is the best predictor for IA ($J = 0.117$) and SNP rs2476601 (PTPN22) is the best predictor for IAA-first ($J = 0.123$), whereas family history of T1D ($J = 0.100$ and 0.111) for both, SNP rs2476601 for IA ($J = 0.112$), and HLA genotype for IAA-first ($J = 0.104$) achieved at least 0.10 in Youden's index. For GADA-first, weight at 1 year is the best predictor ($J = 0.114$), followed by HLA genotype ($J = 0.103$) and SNP rs2292239 in ERBB3 ($J = 0.112$). SNP rs3184504 in SH2B3 also achieved >0.10 in Youden's index ($J = 0.105$). Similar results have been observed in predicting IA and IAA-first, while SNP rs2476601 is the best predictor for GADA-first ($J = 0.191$) at 2 years (Supplementary Table 1). For prediction of IA, IAA-first, and GADA-first at 10 years (Supplementary Table 2), HLA genotype is the best predictor, with Youden's index of $J = 0.105$, 0.113, and 0.09, respectively. No other risk factor achieved >0.10 in Youden's index. From the plots of (Se, $1 - Sp$) in Fig. 1, it is easy to see that these risk factors individually lie close to the diagonal line, which is interpreted as providing little predictive value for IA (IA, IAA-first, or GADA-first) or progression to T1D.

Table 2—*Se*, *Sp*, and *J* for each risk factor to predict any IA, IAA-first, or GADA-first at age of 6 years or progression to T1D at 3 years from the appearance of multiple autoantibodies

	At age of 6 years									Progression to T1D at year 3		
	Any IA			IAA-first			GADA-first					
Risk factor	*Se*	*Sp*	*J*	*Se*	*Sp*	*J*	*Se*	*Sp*	*J*	*Se*	*Sp*	*J*
Family history (FDR vs. GP)	0.214	0.886	0.100	0.226	0.886	0.112	0.193	0.886	0.079			
Sex (female vs. male)	0.547	0.491	0.038	0.564	0.491	0.055	0.532	0.491	0.023	0.517	0.598	0.115
HLA genotype (DR3/4 vs. others)	0.498	0.619	0.117	0.485	0.619	0.104	0.484	0.619	0.103			
Probiotics introduction age ($<$28 days vs. $\geq$28 days)	0.943	0.078	0.020	0.945	0.078	0.023	0.959	0.078	0.036			
Weight *z* score at 12 months ($\geq$median vs. $<$median)	0.550	0.509	0.059	0.511	0.509	0.020	0.605	0.509	0.114			
First infant formula type during the first 7 days of life (extensively hydrolyzed vs. other)	0.045	0.972	0.017	0.056	0.972	0.028	0.037	0.972	0.009			
First infant formula type during the first 3 months of life (extensively hydrolyzed vs. other)	0.050	0.966	0.015	0.060	0.966	0.025	0.041	0.966	0.006			
Age at multiple persistent confirmed autoantibodies (months) ($\geq$median vs. $<$median)										0.732	0.496	0.228
Type of first autoantibody ($\geq$2 autoantibodies vs. 1 autoantibody)										0.370	0.756	0.127
rs1004446 (*INS*)	0.463	0.606	0.069	0.488	0.606	0.093	0.416	0.606	0.022	0.538	0.551	0.089
rs10517086	0.511	0.514	0.025	0.511	0.514	0.025	0.508	0.486	−0.006	0.545	0.457	0.002
rs1534422										0.796	0.295	0.091
rs2327832 (*TNFAIP3*)										0.391	0.658	0.049
rs1143678 (*ITGAM*)	0.746	0.268	0.014	0.726	0.268	−0.006	0.762	0.268	0.030			
rs12708716 (*CLEC16A*)	0.475	0.561	0.036	0.442	0.561	0.003	0.480	0.561	0.042			
rs2292239 (*ERBB3*)	0.624	0.472	0.096	0.621	0.472	0.093	0.640	0.472	0.112			
rs2476601 (*PTPN22*)	0.309	0.802	0.111	0.321	0.802	0.123	0.294	0.802	0.095			
rs2816316 (*RGS1*)	0.344	0.673	0.016	0.328	0.673	0.000	0.337	0.673	0.010			
rs3184504 (*SH2B3*)	0.755	0.315	0.069	0.725	0.315	0.040	0.791	0.315	0.105			
rs4597342 (*ITGAM*)	0.582	0.436	0.018	0.456	0.564	0.020	0.595	0.436	0.031			
rs4948088 (*COBL*)	0.936	0.086	0.022	0.951	0.086	0.037	0.921	0.086	0.008			

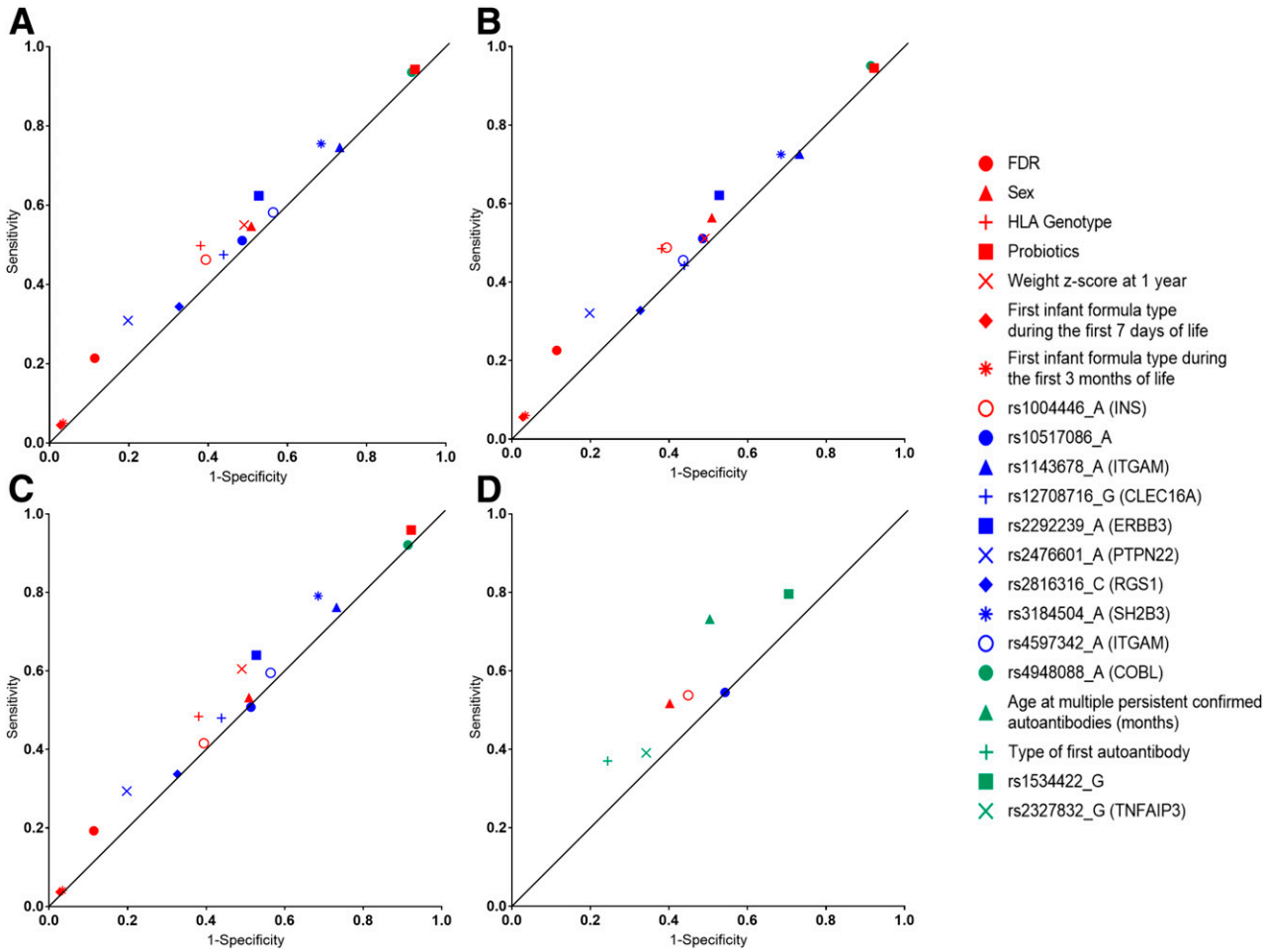

Figure 1—*Se* vs. 1 − *Sp* of each TEDDY-identified risk factor to predict IA, IAA-first, and GADA-first (*A*–*C*) at age 6 years and to predict progression to T1D at 3 years from the appearance of multiple autoantibodies (*D*).

In the forward selection for IA, a total of 12 risk factors were added into the prediction models in the order of family history of T1D, HLA DR-DQ genotype, rs2476601 (PTPN22), rs3184504 (SH2B3), rs1004446 (INS), weight *z* score at 1 year of age, rs2292239 (ERBB3), rs1270876 (CLEC16A), sex, rs4948088 (COBL), probiotics exposure before 28 days, and type of first formula introduced before 3 months of age. The AUC and the Youden's index (*J*) of each prediction model for predicting IA at 6 years are given in Table 3. The AUC was 0.552 (95% CI 0.534, 0.569) with the inclusion of family history of T1D and increased to 0.678 (95% CI 0.655, 0.701) with the full model at 6 years. Both AUC and Youden's index increased when additional risk factors were added into the prediction models. The incremental AUC (ΔAUC) and Youden's index (Δ*J*) introduced by each newly added risk factor are also given in Table 3. The first four risk factors (family history of T1D, HLA genotype, SNP rs2476601 in *PTPN22*, and SNP rs3184504 in *SH2B3*) contributed the most. Similar results have been observed in predicting IA at 2 and 10 years (Supplementary Tables 3 and 4 and Fig. 1*A*). The ROC curves to predict IA at 2, 6, and 10 years using the prediction model with all of the 12 risk factors are plotted in Fig. 2*A*. As we can see, the ability to predict IA gets worse as the time frame gets lengthened.

For IAA-first, the AUCs at years 2, 6, and 10 reached 0.704 (95% CI 0.665, 0.744), 0.707 (95% CI 0.676, 0.739), and 0.683 (95% CI 0.647, 0.719), respectively, when all eight identified risk factors were combined into the prediction model (Table 3, Supplementary Tables 3 and 4, and Fig. 2*B*). The first five added risk factors (family history of T1D, HLA DR-DQ genotype, SNP rs2476601 in *PTPN22*, SNP rs1004446 in INS, and SNP rs3184504 in *SH2B3*) contributed the most in terms of the AUC. For GADA-first, the AUCs at years 2, 6, and 10 reached 0.683 (95% CI 0.618, 0.747), 0.686 (95% CI 0.651, 0.722), and 0.639 (95% CI 0.603, 0.674), respectively, when all risk factors were combined into the prediction model (Table 3, Fig. 2*C*, and Supplementary Tables 3 and 4).

Predicting Progression to T1D From Multiple Autoantibodies

The sensitivity and specificity of each risk factor for predicting progression to T1D at 3 years after onset of multiple autoantibodies are presented in Table 2 and are plotted in Fig. 1*D*. Among all identified risk factors, age at onset of multiple autoantibodies is the best predictor for progression to T1D at 3 years in terms of Youden's index ($J = 0.228$). First appearing indication of seroconversion ranked second in terms of Youden's index ($J = 0.127$). Similar results have been observed in predicting T1D at 1 and 5 years

Table 3—AUC and *J* of each prediction model, together with ΔAUC and Δ*J* introduced by the newly added risk factor in each prediction model factor to predict any IA, IAA-first, or GADA-first at age of 6 years or progression to T1D at 3 years from the appearance of multiple autoantibodies

# of risk factors in the model	At age of 6 years															Progression to T1D at year 3				
	Any IA					IAA-first					GADA-first									
	Newly added risk factor	AUC	ΔAUC	*J*	Δ*J*	Newly added risk factor	AUC	ΔAUC	*J*	Δ*J*	Newly added risk factor	AUC	ΔAUC	*J*	Δ*J*	Newly added risk factor	AUC	ΔAUC	*J*	Δ*J*
1	Family history	0.552		0.100		Family history	0.560		0.115		rs3184504 (SH2B3)	0.587		0.129		Age at multiple IAs	0.668		0.321	
2	HLA genotype	0.605	0.054	0.161	0.062	HLA genotype	0.639	0.080	0.191	0.076	rs2476601 (PTPN22)	0.606	0.019	0.183	0.054	Type of first autoantibody	0.672	0.004	0.310	−0.011
3	rs2476601 (PTPN22)	0.627	0.022	0.189	0.028	rs2476601 (PTPN22)	0.668	0.029	0.231	0.040	HLA genotype	0.631	0.025	0.183	−0.001	rs1534422	0.690	0.018	0.314	0.004
4	rs3184504 (SH2B3)	0.646	0.018	0.207	0.018	rs1004446 (INS)	0.679	0.011	0.276	0.045	Family history	0.655	0.024	0.236	0.053	Sex	0.704	0.014	0.364	0.050
5	rs1004446 (INS)	0.656	0.010	0.231	0.024	rs3184504 (SH2B3)	0.694	0.015	0.316	0.040	Weight *z* score at 12 months	0.674	0.018	0.253	0.017	rs2327832 (TNFAIP3)	0.702	−0.002	0.346	−0.018
6	Weight *z* score at 12 months	0.665	0.009	0.250	0.018	Sex	0.698	0.004	0.313	−0.002	rs12708716 (CLEC16A)	0.677	0.003	0.280	0.027	rs1004446 (INS)	0.707	0.005	0.355	0.009
7	rs2292239 (ERBB3)	0.671	0.006	0.250	0.001	Weight *z* score at 12 months	0.702	0.004	0.322	0.009	rs2292239 (ERBB3)	0.686	0.009	0.287	0.007	rs10517086	0.706	−0.002	0.352	−0.003
8	rs12708716 (CLEC16A)	0.673	0.002	0.252	0.002	rs4948088 (COBL)	0.706	0.004	0.326	0.004										
9	Sex	0.675	0.002	0.250	−0.002	Probiotics introduction age	0.707	0.002	0.316	−0.010										
10	rs4948088 (COBL)	0.675	0.000	0.248	−0.002															
11	Probiotics introduction age	0.676	0.002	0.251	0.003															
12	First infant formula type during the first 3 months of life	0.678	0.002	0.256	0.006															

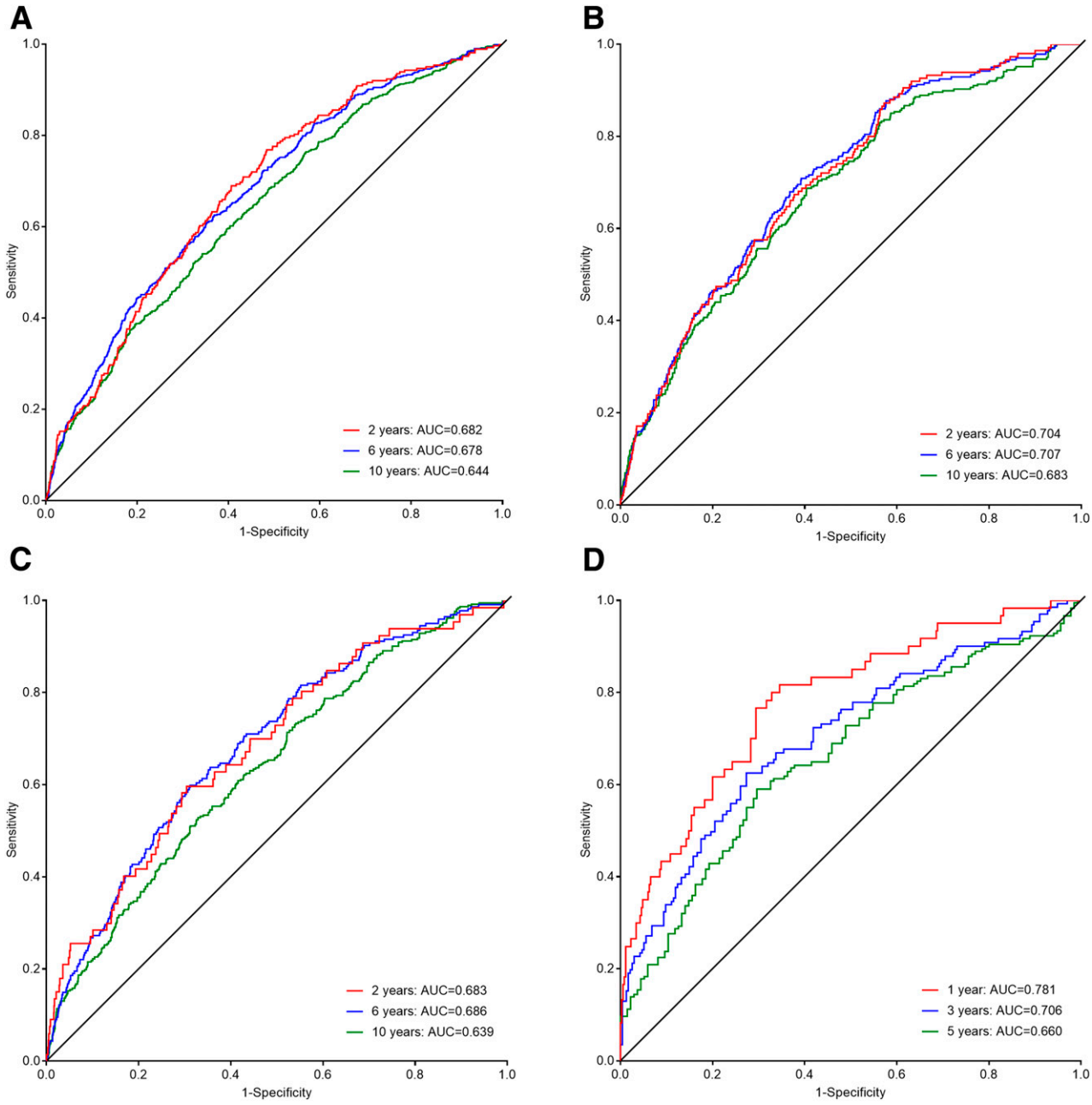

Figure 2—ROC curves with associated AUC of prediction models with TEDDY-identified risk factors selected by forward selection to predict IA, IAA-first, and GADA-first (*A*–*C*) at ages 2, 6, and 10 years and predict T1D at 1, 3, and 5 years from the appearance of multiple autoantibodies (*D*).

(Supplementary Tables 1 and 2 and Supplementary Figs. 1*D* and 2*D*).

In the forward selection, all seven risk factors were added into the predictive model in the order of age at onset of multiple antibodies, first appearing indication of seroconversion, SNP rs1534422, sex, and SNPs rs2327832, rs1004446, and rs10517086. The AUC and Youden's index of each prediction model for predicting progression to T1D at 3 years are given in Table 3 and at 1 and 5 years are given in Supplementary Tables 1 and 2. Age at onset of multiple antibodies is a strong predictor for progression to T1D, and other risk factors contributed only a little in additional predictive value. The AUCs of the prediction model combining all identified risk factors reached 0.781 (95% CI 0.716, 0.845), 0.706 (95% CI 0.649, 0.762), and 0.660 (95% CI 0.600, 0.720), respectively, to predict progression to T1D at years 1, 3, and 5 (Fig. 2*D*). As to the ability to predict IA, the prediction models gave a better prediction of progression to T1D at an earlier time point compared with a later time point.

CONCLUSIONS

While TEDDY has identified genetic and environmental markers significantly associated with the risk of IA and T1D, this work has evaluated their individual and combined contribution to predicting IA and progression to T1D based on time-dependent specificity, sensitivity, and the corresponding ROC curve. Despite the strong statistical associations, the results described herein show that these associations are relatively weak predictors. The Youden indices are generally

<0.4 and the AUC <0.8, which put the predictors no better than fair. Interestingly, the ability to predict IA or T1D decreases with an increasing time horizon. In addition, the ability to predict IAA-first and GADA-first is no better than IA, suggesting the strong influence of age (most IAA-first occurs at an early age compared with GADA-first, which is consistent with our previous results [17,18]).

With respect to prediction of IAA-first, HLA genotype contributed the most toward prediction after family history of T1D. The contribution of three other genetic factors (rs2476601 in PTPN22, rs1004446 in INS, and rs3184504 in SH2B3) contributed equally, raising the J index from 0.115 to 0.316 with the others contributing little improvement. With respect to predicting GADA-first, rs3184504 in SH2B3_T contributed the most followed by rs2476601 in PTPN22 and family history of T1D contributing equally. After consideration of these risk factors, additional contribution to prediction from the others was really nominal. Age at onset of multiple autoantibodies was the most significant predictor of T1D, with the other risk factors also making a very nominal contribution. Of note, the TEDDY cohort is young, and these results may be affected by the early seroconversion to IAA-first autoantibody positivity, which diminishes as age increases. Additional follow-up may increase the ability to predict T1D following GADA-first seroconversion.

The ability to estimate sensitivity and specificity, in a time-until-event setting, provides a means by which to evaluate the contribution of statistically significant risk factors identified by using Cox proportional hazards models. Those models describe contributions to risk estimates (hazard rates) that are only the positive predictive value. The methods included herein provide an ability to assess the extent to which risk factors can discriminate between those who will and will not get disease.

The incorporation of significant risk factors identified from TEDDY's nested case-control study, such as the observation of lower IA odds ratio associated with higher average 25(OH)D concentration over all time points using conditional logistic regression analysis (21), into the prediction modeling requires analytic methods that extend the results to the entire cohort. The matched concordance index (C-index) (22) of 0.56 (95% CI 0.53–0.59) suggests that it adds little to the data presented herein.

The factors identified and the analysis presented in this article are not without their limitations. The TEDDY cohort is relatively young, with a selected high-risk HLA from selected geographic settings. The results reported herein may not remain the same when the cohort ages or if applied to other cohorts. But this point speaks to generalizability of each of the risk factors, not to the methods described here, which provide a means to assess their relative contribution to prediction in the same cohort from which they were identified. It is important to have a validation data set, and we accept this limitation in TEDDY overall. Finally, the analytic methodology that we used to combine the risk factors has its own limitations, yet follows accepted practice for such an analysis (23).

As additional risk factors are identified and prediction improves, the results can provide insights into causal mechanisms. TEDDY is uniquely poised to make this contribution as it adds environmental exposures, pathogens, and extensive genomics to the risk factors identified thus far. Going forward, the presented evaluation framework will serve as a summary tool to evaluate these new risk factors as they are identified in TEDDY.

Acknowledgments. The authors thank Sarah Austin-Gonzalez with the Health Informatics Institute at the University of South Florida for assistance with preparing the figures.

Funding. The TEDDY Study is funded by the National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Allergy and Infectious Diseases, National Institute of Child Health and Human Development, National Institute of Environmental Health Sciences, JDRF, and the Centers for Disease Control and Prevention (grants U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63836, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63836, UC4 DK95300, UC4 DK100238, UC4 DK106955, UC4 DK112243, and UC4 DK117483, and contract no. HHSN267200700014C). This work was supported in part by National Institutes of Health/National Center for Advancing Translational Sciences Clinical and Translational Science Awards to the University of Florida (UL1 TR000064) and the University of Colorado (UL1 TR001082).

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. J.P.K., X.L., K.V., B.A., W.A.H., M.J.R., J.-X.S., J.T., A.-G.Z., and Å.L. attest to meeting ICMJE (International Committee of Medical Journal Editors) uniform requirements for authorship by making substantial contributions to conception and design of this manuscript; acquisition, analysis, and interpretation of the data; drafting or revising the manuscript for intellectual content; and giving final approval of the published version. J.P.K. designed the study, proposed the analysis, interpreted the findings, and wrote the manuscript. X.L. performed the analysis and drafted/revised the manuscript. K.V. provided input on the analytical plan and interpretation of the results and drafted/revised the manuscript. B.A., W.A.H., M.J.R., J.-X.S., J.T., A.-G.Z., and Å.L. designed the study and reviewed/edited the manuscript. J.P.K. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented in abstract form at the 78th Scientific Sessions of the American Diabetes Association, Orlando, FL, 22–26 June 2018.

References

1. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. Ann N Y Acad Sci 2008;1150:1–13
2. Törn C, Hadley D, Lee HS, et al.; TEDDY Study Group. Role of type 1 diabetes-associated SNPs on risk of autoantibody positivity in the TEDDY study. Diabetes 2015;64:1818–1829
3. Krischer JP, Liu X, Lernmark Å, et al.; TEDDY Study Group. The influence of type 1 diabetes genetic susceptibility regions, age, sex, and family history on the progression from multiple autoantibodies to type 1 diabetes: a TEDDY study report. Diabetes 2017;66:3122–3129
4. Hummel S, Beyerlein A, Tamura R, et al.; TEDDY Study Group. First infant formula type and risk of islet autoimmunity in The Environmental Determinants of Diabetes in the Young (TEDDY) study. Diabetes Care 2017;40:398–404
5. Uusitalo U, Liu X, Yang J, et al.; TEDDY Study Group. Association of early exposure of probiotics and islet autoimmunity in the TEDDY study. JAMA Pediatr 2016;170:20–28
6. Beyerlein A, Liu X, Uusitalo UM, et al.; TEDDY study group. Dietary intake of soluble fiber and risk of islet autoimmunity by 5 y of age: results from the TEDDY study. Am J Clin Nutr 2015;102:345–352
7. Elding Larsson H, Vehik K, Haller MJ, et al.; TEDDY Study Group. Growth and risk for islet autoimmunity and progression to type 1 diabetes in early childhood: The Environmental Determinants of Diabetes in the Young study. Diabetes 2016;65:1988–1995
8. Lo A, Chernoff H, Zheng T, Lo SH. Why significant variables aren't automatically good predictors. Proc Natl Acad Sci U S A 2015;112:13892–13897
9. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–35
10. Krischer J. The biostatistics of prediction. Autoimmunity 2004;37:261–263
11. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000;56:337–344
12. Hagopian WA, Erlich H, Lernmark A, et al.; TEDDY Study Group. The Environmental Determinants

of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. Pediatr Diabetes 2011;12:733–743
13. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. Pediatr Diabetes 2007;8:286–298
14. Barrett JC, Clayton DG, Concannon P, et al.; Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41:703–707
15. Bonifacio E, Yu L, Williams AK, et al. Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for National Institute of Diabetes and Digestive and Kidney Diseases consortia. J Clin Endocrinol Metab 2010;95:3360–3367
16. Törn C, Liu X, Hagopian W, et al.; TEDDY Study Group. Complement gene variants in relation to autoantibodies to beta cell specific antigens and type 1 diabetes in the TEDDY Study. Sci Rep 2016;6:27887
17. Krischer JP, Lynch KF, Schatz DA, et al.; TEDDY Study Group. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. Diabetologia 2015;58:980–987
18. Krischer JP, Lynch KF, Lernmark Å, et al.; TEDDY Study Group. Genetic and environmental interactions modify the risk of diabetes-related autoimmunity by 6 years of age: the TEDDY study. Diabetes Care 2017;40:1194–1202
19. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York, Springer, 2000
20. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med 2013;32:5381–5397
21. Norris JM, Lee HS, Frederiksen B, et al.; TEDDY Study Group. Plasma 25-hydroxyvitamin D concentration and risk of islet autoimmunity. Diabetes 2018;67:146–154
22. Brentnall AR, Cuzick J, Field J, Duffy SW. A concordance index for matched case-control studies with applications in cancer risk. Stat Med 2015;34:396–405
23. Yan Q, Bantis LE, Stanford JL, Feng Z. Combining multiple biomarkers linearly to maximize the partial area under the ROC curve. Stat Med 2018;37:627–642