EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

28 October 2021
EMA/580542/2021
Committee for Medicinal Products for Human Use (CHMP)

DRAFT Qualification Opinion of Islet Autoantibodies (AAs) as Enrichment Biomarkers for Type 1 Diabetes (T1D) Prevention Clinical Trials

Draft agreed by Scientific Advice Working Party (SAWP)	11 Feb 2021
Adopted by CHMP for release for consultation	25 March 2021[1]
Start of public consultation	03 November 2021 [2]
End of consultation (deadline for comments)	14 December 2021[3]

Comments should be provided using this template. The completed comments form should be sent to ScientificAdvice@ema.europa.eu

Keywords	Qualification of Novel Methodology, enrichment marker, patient selection, Type 1 Diabetes Mellitus, studies investigating prevention or delay in disease onset, islet autoantibodies, anti-IAA, anti-GAD65, anti-IA-2, anti-ZnT8, accelerated time-failure model

[1] Last day of relevant Committee meeting.
[2] Date of publication on the EMA public website.
[3] Last day of the month concerned.

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

1. Executive summary

The objective of this procedure was for the Critical Path Institute's Type 1 Diabetes Consortium (T1DC) to achieve a qualification opinion for a new drug development tool for Type 1 Diabetes (T1D) through EMA's qualification of novel methodologies for medicine drug development. The proposed context-of-use (COU) statement was that, in individuals at risk of developing T1D, the islet AAs can be used together with other patient features as enrichment biomarkers to optimize the selection of individuals for clinical trials of therapies intended to prevent or delay the clinical diagnosis of T1D. The islet AAs proposed include IAA, GAD65, IA-2, and ZnT8. Additional patient features include sex, baseline age, blood glucose measurements from the 120-minute timepoints of an oral glucose tolerance test (OGTT) and haemoglobin A1c (HbA1c) levels.

As of May 2020, the T1DC has obtained three datasets, The Environmental Determinants of Diabetes in the Young (TEDDY), the TrialNet Pathway to Prevention Study (TN01) and the Diabetes Autoimmunity Study in the Young (DAISY)[4]. The TEDDY and TN01 were aggregated to support the model-based qualification of islet AAs as enrichment biomarkers. This aggregated dataset was used to construct the statistical analysis plan presented in the T1DC's May 2019 submission for qualification advice. An accelerated time failure model provides the supporting evidence for the use of islet AAs anti-insulin AA (IAA), anti-glutamic acid decarboxylase 65 AA (GAD65), anti-insulinoma antigen-2 AA (IA-2), and zinc transporter 8 AA (ZnT8) as enrichment biomarkers in T1D prevention clinical trials. The developed model demonstrates that the islet AAs are statistically significant predictors of the time-varying probability of conversion to a diagnosis of T1D. Further when additional sources of variability, including, sex, baseline age, blood glucose measurements from the 120-minute timepoints of OGTT and HbA1c, are assessed with the islet AAs, it further improves the accuracy of predicting the time-varying probability of conversion to a T1D diagnosis. Since the may 2019 submission, the T1DC has acquired the data from DAISY, which was reserved to externally validate the model. In summary, analysis of TN01, TEDDY, and DAISY constitute data-driven evidence for using the presence of two or more islet AAs and other patient features as enrichment biomarkers for selection of subjects included in T1D prevention studies.

The presence of different numbers and combinations of islet AAs were analyzed in conjunction with other relevant sources of variability including, demographics, human leukocyte antigen (HLA) haplotype, first-degree relative (FDR), T1D status and blood glucose assessments. The specific sources of variability that were selected include sex, baseline age, blood glucose measurements from the 120-minute timepoints of an OGTT and HbA1c. The process by which these sources of variability were selected is outlined.

The developed models were shown to demonstrate that the baseline presence of various combinations of two or more islet AAs are statistically significant predictors of the time-varying probability of conversion to a diagnosis of T1D. Furthermore, glycemic measurements, sex, and baseline age within this multiple islet AA positive population were shown to further contribute as independent predictors, thereby increasing the accuracy of predicting the time-varying probability of conversion to a T1D diagnosis. The T1DC team considers that this model provides the supporting evidence for the application islet AAs as enrichment biomarkers as defined by the context of use statement.

2. Answers to applicant's questions

Based on the coordinators' reports the CHMP gave the following answers to the questions by the applicant:

Question 1:

Does EMA agree with the COU?

[4]The data from the TEDDY and TrialNet Study reported here were supplied by the NIDDK Central Repositories. This document/publication does not necessarily reflect the opinions or views of the TEDDY, TrialNet Study, the NIDDK Central Repositories, or the NIDDK.

T1DC's position: The proposed COU focuses on the application of islet AAs, together with other patient features, as enrichment biomarkers in individuals at risk of developing T1D to optimize the selection of individuals for clinical trials of therapies intended to prevent or delay the clinical diagnosis of T1D. The focus is on understanding the contribution of the positivity to these AAs as predictors of progressing towards a diagnosis of T1D. From a practical drug development standpoint, this proposed use is of added value because their intended application can help inform the definition of entry criteria, enrichment strategies, and stratification approaches in the field of T1D prevention.

CHMP answer

The qualification exercise included a modeling exercise that also identified the relevance of additional clinical parameters (sex, baseline age, blood glucose measurements from the 120-minute timepoints of oral glucose tolerance test (OGTT), and haemoglobin A1c (HbA1c) levels).
Individuals defined as 'At risk' were defined in this context as being a first degree relative (FDR) of a T1D patient or those having a specific human leukocyte antigen (HLA) subtype of risk (HLA-DR3/3, DR4/4, DR3/4, DR3/X [X≠ 3], DR4/X [X≠ 4]), excluding individuals with baseline fasting glucose ≥ 126mg/dL (7.0 mmol/L) or stimulated 2-hour glucose ≥ 200 mg/dL (11.1 mmol/L).
Positivity for two or more of the islet AAs, determined in this population, in addition to the relevant characteristics as described in the model, can be used for enrichment of clinical trials focusing on the delay or prevention of the clinical diagnosis of T1D.
The proposed COU is overall agreed. The clinical interest of identifying good biomarkers for Type 1 Diabetes (T1D) onset in an at-risk patient population is supported by the CHMP, and the unmet need for better means to optimize drug development in the field is acknowledged. There are no approved therapies to prevent or delay the onset of T1D and there is a lack of biomarkers to identify individuals and quantify risk of conversion to a diagnosis of T1D. In addition, there have been significant failures in late-stage development of therapies in new-onset T1D. These failures have been attributed in part to a high degree of heterogeneity in the patient population and a current inability to quantitatively describe the contributions of specific sources of variability to such heterogeneity. Second, intervening in new-onset T1D may be too late to significantly delay or halt disease progression and preserve endogenous β-cell function.
A practical problem foreseen is that in clinical trial recruitment, often the only parameter known is family history, which could limit the utility of this new screening/enriching tool unless mass screening efforts are taking place. During the discussion meeting (DM), the applicant clarified that there would also be the possibility to partner with pre-existing trial networks such as TrialNet and INNODIA that are carrying out screening efforts. This is supported.
The model-based approach proposed by the applicant is considered an acceptable method to address the question of interest, which is whether the combination of positivity to 2 or more of the 4 selected AAs can be considered acceptable predictors of a diagnosis of T1D, when combined with additional and well-defined patient characteristics.
It should however be noted that the modeling approach taken by the applicant is not a mechanistic disease model: a clear and fully quantitative description of the contribution of the different factors including positivity to these AAs as predictors of progressing towards a diagnosis of T1D is therefore not possible. The model allows confirming the existence of a significant statistical contribution of the different covariates and their relative relevance toward T1D diagnosis for patients at risk.
The analytical assays used to measure islet autoantibodies (AA) against glutamic acid decarboxylase 65 (GAD65), insulinoma antigen-2 (IA-2; also called ICA512), insulin (IAA) and zinc transporter 8 AA (ZnT8) are considered state of the art. It should be noted that the results and the conclusions of the modeling analysis as assessed during this qualification procedure are considered only applicable when

the islet autoantibodies are measured using these methods or methods proved to have at least equivalent analytical performances.

Target Population for Use of the Biomarkers: Individuals at risk of T1D, defined as being a FDR of a T1D patient, or having a specific HLA subtype of risk (HLA-DR3/3, DR4/4, DR3/4, DR3/X [X≠ 3], DR4/X [X≠ 4]), excluding individuals with baseline fasting glucose ≥ 126mg/dL (7.0 mmol/L) or stimulated 120-minute glucose ≥ 200 mg/dL (11.1 mmol/L). It is intended that positivity for two or more of the islet AAs be determined in this population, to be used as enrichment biomarkers for clinical trials focusing on the delay or prevention of the clinical diagnosis of T1D.

Stage of Drug Development for Use: All clinical efficacy evaluation stages of therapeutic interventions focused on the prevention or delay of T1D, including early signs of efficacy, proof-of-concept, dose-ranging, and registration studies.

Intended Application: To utilize the islet AAs as enrichment biomarkers for patient selection in clinical trials investigating therapies that are intended to prevent or delay the clinical diagnosis of T1D. These biomarkers, along with additional patient features, such as sex, baseline age, baseline HbA1c levels and the 120-minute time point from an OGTT, can be used as predictors to identify subpopulations at highest risk of a diagnosis of T1D during the course of T1D prevention clinical trials. The underlying time-to-event models that supports this qualification will be made available through the Critical Path Institute's website (https://www.c-path.org/).

Out-of-scope:

- The underlying evidence for the COU does not account for variability in the longitudinal seroconversion for the different islet AAs over the course of T1D prevention trials.
- The underlying time-to-event model that provides the evidence for the COU statement of the qualified biomarkers does not include the ability to generate virtual sub-populations for simulation purposes.

The COU is overall agreed. There is clearly an unmet need for biomarkers to aid development in T1DM prevention, a field with a long history of failed trials. This Qualification would only refer to the value of the positivity of two AAs in the risk assessment. The combination of AAs, numbers above two AAs and the reason for not assessing only one AA are well explained. With a validated method, this would clearly help with selection and stratification of subjects in clinical development. Having a model of the effect of two positive AAs cannot replace a placebo arm in a randomized trial setting.

Question 2:

Does EMA agree that the data sources are adequate to support the proposed COU?

T1DC's position: The available data sources, and their integration through data standardization and management, represents a unique opportunity to transform these data into valuable knowledge to provide the necessary evidence to support the qualification of islet AAs for the proposed context of use. The population captured in the data sources represents the population likely to be considered as candidates to participate in clinical trials of therapies intended to prevent or delay the clinical diagnosis of T1D.

CHMP answer

The data used for the model development and external validations to support the qualification of islet AAs as enrichment biomarkers originated from three datasets: the TN01, TEDDY, and DAISY registry studies. A summary of the three studies can be found in Table 1.TEDDY and TN01 were aggregated and used for model development and internal cross-validation. Data from the DAISY study was acquired and used to perform external validation on the final model.

Participants for TN01 were selected by the presence of a FDR with T1D, as this has been shown to be a risk factor for development of T1D. The criteria included (1) FDRs (age 1 –45 years) of T1D probands

or (2) second- and third-degree relatives (age 1 –20 years) of T1D probands (i.e., nieces, nephews, aunts, uncles, grandchildren, cousins, half-siblings). Based on these criteria, 211,230 subjects with positive FDRs were screened for the presence of islet AAs, as of November 2018. Between 2004-2009 subjects with the presence of 1 islet AA were considered eligible for follow-up. In 2009 the eligibility for follow-up changed to the presence of 2 islet AAs. As of December 2018, 4,524 subjects are being followed. Once subjects were selected for follow-up and opted in, they were monitored for 6 monthly visits using oral glucose tolerance test (OGTT), detection of islet AAs and measurement of HbA1c levels. TN01 is providing TrialNet with an active patient ready cohort and collaborative clinical trial network to evaluate novel therapies. TN01 is still enrolling new subjects and following current subjects. The TN01 data provided in this submission is locked as of December 2018.

TEDDY is longitudinally prospective study assessing a broad spectrum of environmental factors that may contribute to the stimulus or stimuli that are involved in the immune initiation of T1D. An assessment of these environmental factors that will not be part of this submission, include identification of infectious agents, dietary factors, or other environmental agents, including psychosocial factors. Children were screened and recruited during infancy based on high-risk HLA genotypes (361,518 initial screenings and 8,667 in initial prospective cohort), with separate inclusion criteria for GP children or FDR. Participants are monitored prospectively with study visits every 3 months for the first 4 years, and every 6 months thereafter to age 18. All children who are persistently positive for any islet AA are monitored every 3 months until the age of 15 years or diagnosis of T1D. As of November 2018, 9.1% of the participants had developed at least one islet AA; 3.8% had developed T1D and thus reached study endpoint. Of the original cohort who have not reached the study endpoint, 68% are still participating in follow-up. TEDDY data provided in this submission are locked as of June 2018. Of participants, 89% had no family history of T1D.

Diabetes Autoimmunity Study in the Young (DAISY) is a prospective cohort study of 2547 children who are at increased genetic risk for developing T1D. DAISY seeks to understand the environmental triggers for islet autoimmunity and progression to T1D. Children were screened and recruited in two groups (1) during infancy based on high-risk HLA genotypes or (2) during early childhood based on first-degree relative (FDR) status as described (Rewers et al. 1996a; Rewers et al. 1996b). Children in DAISY were monitored longitudinally for over 20 years, assessing a variety of environmental factors that may be involved in the development of islet autoimmunity. These included assessment of prenatal exposures, birth events, growth and puberty, dietary assessment, smoke exposure, daycare exposure, physical activity assessment, and biological samples for assessment of biomarkers and infectious agents (blood, urine, saliva, throat and rectal swabs). Participants were assessed at 9, 15 and 24 months of age and then annually thereafter. Those who developed islet autoimmunity were monitored every 6 months. Participants who were positive for more than one islet autoantibody were requested to follow up every 3 months until diagnosis of T1D. As of January 2020, 9.2% of the participants had developed at least one islet autoantibody and 4.2% had developed T1D. Of the original cohort, 42% were still engaged in follow-up. DAISY data provided in this submission are locked as of June 30, 2017.

In the TN01, TEDDY, and DAISY protocols, the diagnosis of T1D was a study endpoint. The diagnostic criteria pre-specified for each study differed slightly, but both were based on the American Diabetes Association (ADA) criteria. All studies are observational but certain features in their designs differ, including inclusion criteria and scheduled frequency of follow-up.

The data sources are judged largely relevant, consistent with the recommendation during the QA procedure. From a modeling perspective, this approach is endorsed, and the 3 data sources seem adequate. Potential covariate distribution and correlation were presented and discussed as requested during the qualification procedure.

The baseline data intended for modeling are relatively well defined, as well as the binary endpoint (T1D diagnosis).

Longitudinal assessments of islet AA positivity, OGTTs, C-peptide measurements, and HbA1c measurements are considered out of scope for the proposed analysis, and only baseline information were used for the modeling analysis.

The initial precise definition of baseline used for the analysis set is the first record (i.e., timepoint) for each individual in which the following criteria is satisfied:

- Presence of any two or more of the 4 islet AAs
- Complete, (i.e., non-missing) information for OGTT (0 and 120-minute time points), C-peptide measurements (0 and 120-minute time points), HbA1C measurements, age and sex.

204 ***Table 1. Overview TN01, TEDDY, and DAISY***

	TN01	**TEDDY**	**DAISY**
Type of study:	Observational	Observational	Observational
Years running:	2004-Present	2004-Present	1993-Present
Enrolment design:	Ongoing screening and active enrolment	Screening complete and fixed prospective cohort	Screening complete and fixed prospective cohort
Enrolment criteria:	Ages 1-45 must have FDR with T1D*, ages 1-20 must have extended family member** with T1D	Newborns (< 4 months old) with high-risk HLA*** or FDR	Newborns with high-risk HLA or FDR Sibling/offspring of individual with T1D, initial visit <7yo
Number of subjects:	209,394 initial screening 4,524 being followed (December 2018)	361,518 initial screening 8,667 in initial prospective cohort	31,881 initial newborn screening 2,547 in prospective cohort.
Primary Study Outcome:	T1D diagnosis	Appearance of one or more islet cell autoantibodies	T1D diagnosis
Secondary Study Outcome:	Metabolic and autoantibody measurements	T1D diagnosis	Detection of islet autoantibodies
Average age at entry:	19.1 years (<3 months to >49 years)	3 months	Average age at entry for newborn screened: 1.0 yr Average age at entry for sib/offspring cohort: 2.31 yr
Number of subjects who tested positive for 1 islet AA at or after screening:	13,058†	794	364
Number of subjects who tested positive for 2 islet AAs at or after screening:	4,550	535	136

205 * FDR is defined as a child, parent, or sibling.

206 ** Extended family member is defined as a cousin, niece, nephew, aunt, uncle, grandparent, or half-sibling.

*** High risk HLA is defined as having an HLA genotype that is associated with higher incidences of HLA. In the TEDDY study these were HLA-DR3/3, DR4/4, DR3/4, DR3/X [X≠3], DR4/X [X≠4]

† Between 2004-2009 individuals with one islet AA were followed with six-monthly assessments. After 2009 this changed, and subjects required two or more islet AAs to be enrolled in the follow-up cohort

Question 3:

Does EMA agree the AFT survival model and its covariates represent adequate evidence for the qualification of islet AAs as enrichment biomarkers for T1D prevention trials?

T1DC's position: T1DC believes a survival model construct is adequate because the clinically relevant endpoint defined for the proposed model is a binary dependent variable and the need to understand the likelihood of conversion to a diagnosis of T1D over the course of a clinical trial for prevention or delay of T1D. The proposed survival model evaluating the contribution of subject's positivity to the different islet AAs taken in combination to understand the time-varying probability of conversion to a diagnosis of T1D also represents an adequate approach to provide the supporting evidence for this intended qualification procedure.

CHMP answer

The applicant developed a survival model to describe the time course of incidence of T1DM in patients included in the 2 datasets used for model building (TEDDY and TN01), given their baseline characteristics. The third dataset was used for model validation. The following hazard functions were tested and compared based on their Akaiké information criteria during the modeling process: Weibull, gamma, generalized gamma, generalized F, log logistic distributions. The patient baseline characteristics tested as covariates in the model, as well as their brief description are included in table 3 below. Table 4 and 5 provide their respective descriptive statistics.

Table 2. Covariates evaluated

Notation	Description of covariate at derived baseline	Type
$X_{\mathbf{GAD65\_IAA}}$	Positivity for GAD65, IAA	Binary
$X_{\mathbf{GAD65\_IA-2}}$	Positivity for GAD65, IA-2	Binary
$X_{\mathbf{GAD65\_ZnT8}}$	Positivity for GAD65, ZnT8	Binary
$X_{\mathbf{IA-2\_IAA}}$	Positivity for IA-2, IAA	Binary
$X_{\mathbf{IA-2\_ZnT8}}$	Positivity for IA-2, ZnT8	Binary
$X_{\mathbf{IAA\_ZnT8}}$	Positivity for IAA, ZnT8	Binary
$X_{\mathbf{GAD65\_IAA\_ZnT8}}$	Positivity for GAD65, IAA, ZnT8	Binary
$X_{\mathbf{GAD65\_IAA\_IA-2}}$	Positivity for GAD65, IAA, IA-2	Binary
$X_{\mathbf{GAD65\_IA-2\_ZnT8}}$	Positivity for GAD65, IA-2, ZnT8	Binary
$X_{\mathbf{IA-2\_IAA\_ZNT8}}$	Positivity for IA-2, IAA, ZnT8	Binary
$X_{\mathbf{GAD65\_IA-2\_IAA\_ZNT8}}$	Positivity for GAD65, IA-2, IAA, ZnT8	Binary
X_{STUDY}	Flag for being in TN01 or TEDDY	Binary
X_{HR_HLA}	Flag for high risk HLA subtype*	Binary
X_{FDR}	Flag for first-degree relative with T1D **	Binary
$X_{\mathbf{SEX}}$	Male or female	Binary
$X_{bAGE\ s}$	Age	Continuous
X_{BMI_s}	Body mass index	Continuous
X_{HbA1c_s}	HbA1c test result (%)	Continuous
$X_{Log\_GLU0\_s}$	Log transformed and standardized and 0-minute results from OGTT	Continuous
$X_{Log\_GLU120\_s}$	Log transformed and standardized and 120-minute results from OGTT	Continuous

* High-risk HLA is defined in Section 4.3.3.2

** In TN01, the actual FDR was listed, and required a derivation into a binary outcome for the FDR status.

Table 3. Data summary of covariates and diagnoses by study for analysis set

Study	TN01		TEDDY	
	Value	% Missingness	Value	% Missingness
Subjects	1669	-	353	-
Age at Derived Baseline (sd)	13.0 years (10.0)	0	5.7 years (2.5)	0
Sex (% Female)	45.5%	0	41.6%	0.06
Number of Islet AA measurements	1669	0	353	0
Has FDR %	1519	9%	65	0
Mean 0 Min OGTT in mg/dL (sd)	88.9 (9.7)	0	87.0 (8.9)	0
Mean 120 Min OGTT in mg/dL (sd)	120.3 (29.6)	0	108.1 (24.0)	0
HbA1C % (sd)	5.1 (0.3)	0	5.2 (0.2)	0
Number of HLA Measurements	1622	2.8	351	0.6
Mean BMI	21.2 (8.5)	67.6%	16.5 (2.4)	3.1%
Diagnoses	383	NA	138	NA

Table 4. T1D diagnoses in the analysis set by autoantibody combination

	TEDDY			TN01		
Islet AA combination	Subjects	Diagnoses	% Conversion	Subjects	Diagnoses	% Conversion
GAD65_IA-2	34	15	44%	150	35	23%
GAD65_IA-2_IAA	28	13	46%	64	16	25%
GAD65_IA-2_IAA_ZnT8	74	39	53%	280	83	30%
GAD65_IA-2_ZnT8	24	12	50%	315	85	27%
GAD65_IAA	74	15	20%	290	37	13%
GAD65_IAA_ZnT8	26	9	35%	164	28	17%
GAD65_ZnT8	41	3	7%	233	36	15%
IA-2_IAA	10	6	60%	16	4	25%
IA-2_IAA_ZnT8	24	18	75%	51	20	39%
IA-2_ZnT8	12	5	42%	71	32	45%
IAA_ZnT8	6	3	50%	35	7	20%

Given the empirical nature of the model, the results obtained by the applicant are also considered highly dependent on tested covariate distribution and correlation/collinearity.

The covariates remaining after the univariate analysis were analyzed for multicollinearity and associations prior to performing multivariate analysis. Pearson's correlation was used to test the correlation between continuous covariates, with a correlation value above 0.3 chosen as significant. The Wilcoxon test was used to test the association between continuous and categorical covariates, and the Chi-square test of independence was used to test the association between categorical covariates. In both cases, a p-value < 0.001 (multiplicity adjusted) was chosen as the threshold for significance.

The correlation between the continuous covariates (Figure 4) did not reveal any covariate pairs with high correlation, defined as correlations above 0.3. The Wilcoxon test (Table 11) and the chi-square test of independence (Table 12) showed that the baseline Age (bAGE_s) and SEX were highly associated with AA combinations. Association between islet AA combinations was not considered relevant as their presence is mutually exclusive (i.e., only one islet AA combination is possible for a given subject at a single measurement).

Figure 4. Pearson's correlation between continuous covariates

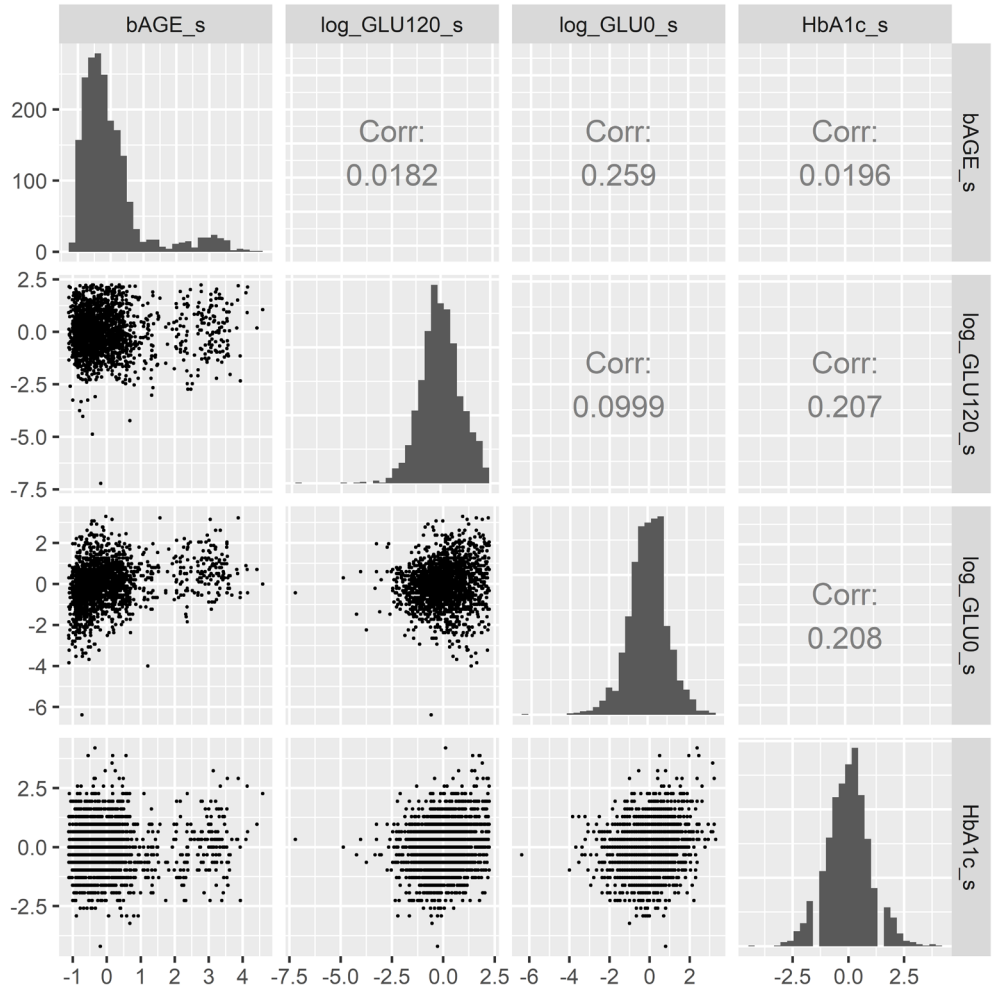


Table 5. Wilcoxon test between continuous and categorical covariates

Covariate	SEX	GAD65_ IAA	GAD65_ ZnT8	IA-2_ ZnT8	IA-2_ IAA_ZnT8	GAD65_IA-2 _IAA_ZnT8
bAGE_s	1.28E-02	3.31E-07	1.05E-16	3.51E-01	2.81E-10	1.14E-07
Log_GLU120_s	9.26E-02	7.38E-03	2.17E-03	3.76E-03	1.31E-03	5.45E-02
Log_GLU0_s	2.60E-04	6.85E-01	2.67E-01	2.29E-01	5.58E-01	4.10E-01
HbA1c_s	1.56E-01	4.37E-01	1.05E-01	2.30E-01	1.36E-01	7.22E-02

Table 6. Chi-square test of independence between categorical covariates

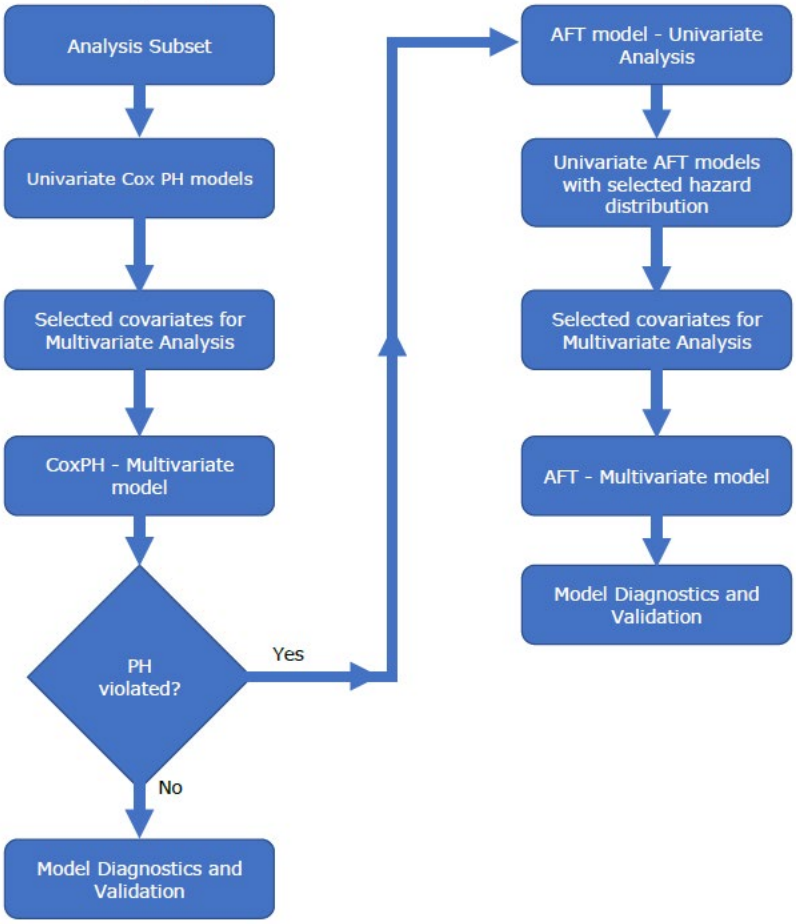
	GAD65_ IAA	GAD65_ ZnT8	IA-2_ ZnT8	IA-2_ IAA_ZnT8	GAD65_IA-2_ IAA_ZnT8
SEX	7.55E-01	4.07E-02	6.57E-05	4.13E-03	7.96E-01

Modeling Analysis Methodology

As per the original statistical analysis plan, the first approach was to analyze predictors of T1D diagnosis using a Cox proportional hazard (PH) model, (i.e., a semi-parametric approach), as this was the most parsimonious first step. Based on reviewer recommendations, a fully parametric approach was requested. With knowledge of prior quantitative analyses from the literature, consideration of the drug development context, and the available data, the full modeling analysis was executed. The flow chart (Figure 3) displays the progression of the modeling analysis, where subsequent steps were

executed based on best practices for model building and learnings from previous steps. All analysis was carried out in the R programming language. In completion, the model building process followed three main steps: (a) Analysis of Cox PH model using the TN01 and TEDDY datasets and testing the PH assumption; (b) Development of a parametric accelerated failure time model using the TN01 and TEDDY datasets; (c) Evaluation of model performance with k-fold cross-validation and external validation with DAISY as a separate independent dataset.

Figure 3. Modeling development workflow



Software

Model building, visualization, model assumptions, diagnostics and external validation was conducted in R (version 4.0.0; Vienna, Austria, R Core Team, 2018) using the packages "survival" (Therneau 2020), "flexsurv" (Jackson 2016), "survminer" (Kassambara and Kosinski, n.d.), "dplyr" (Wickham et al. 2020), "survAUC" (Potapov, Adler, and Schmid 2015), "rms" (Harrell 2019) and "riskRegression" (Ozenne et al. 2017).

Cox Proportional Hazard Model

The semiparametric Cox PH model relates the T1D diagnosis events with the covariates,

$$h_i(t) = h_0(t)\exp\left(\sum_{j\in I}\beta_j X_{ij}\right) \qquad \text{(E1)}$$

where $h_i(t)$ is hazard function for individual *i* determined by a set of *j* covariates $[X_{ij}]$ and corresponding (estimated) coefficients $[\beta_j]$, *t* is the survival time, and $h_0(t)$ is the baseline hazard. The use of a Cox PH model implies that the underlying baseline hazard function is not specified to have a parametric distribution and that the PH assumption holds, (i.e., the ratio of hazards between different individuals remains constant over time).

Selection of Parametric Distribution

Multiple parametric distributions were tested for their ability to approximate the underlying hazard function including exponential, Weibull, gamma, generalized gamma, generalized F, log logistic, log normal and Gompertz. Resulting Akaike information criterion (AIC) values and graphical methods for survival and hazard function fits were compared to select an appropriate parametric form. The 'flexsurvreg' function in the 'flexsurv' R package was used for the selection of parametric distribution analysis.

Univariate Analysis

A univariate analysis was performed by estimating a Cox PH model for of the covariates in Table 3. The 'coxph' function in the 'survival' R package was used for Cox PH analysis (Therneau 2020). Covariates with no significant univariate association (p-value ≥ 0.1) with T1D diagnosis were not considered for the full model development. The p-value was computed using the Wald test, which evaluates whether the covariate coefficient is statistically different from zero. A multiplicity adjusted alpha value (Bonferroni correction) was used for univariate analysis.

Analysis of Correlation and Association between Covariates

The covariates remaining after the univariate analysis were analyzed for multicollinearity and associations prior to performing multivariate analysis. Pearson's correlation was used to test the correlation between continuous covariates, with a correlation value above 0.3 chosen as significant. The Wilcoxon test was used to test the association between continuous and categorical covariates, and the Chi-square test of independence was used to test the association between categorical covariates. In both cases, a p-value < 0.001 (multiplicity adjusted) was chosen as the threshold for significance.

Multivariate Analysis

The multivariate analysis was performed by testing all possible combinations of remaining covariates, as the number of covariates for multivariate analysis were reasonable. The comparison between possible models was conducted using Akaike's Information Criteria (AIC). A reduction in AIC value greater than or equal to 10 suggests a strong evidence in favor of the model with lower AIC (Burnham and Anderson 2016).

Model Diagnostics

To assess if the PH assumption was satisfied, Schoenfeld residuals were utilized. The expected value of these residuals can be used to quantify potential time-dependency on survival times. The Pearson product-moment correlation between the scaled Schoenfeld residuals and log(time) for each covariate was computed using the 'cox.zph' function in R. Values below a significance threshold indicated a violation of the PH assumption. Additional model diagnostics were not performed for the Cox PH model due to a violation of the PH assumption observed with the above-mentioned test.

Parametric Accelerated Failure Time Model

The AFT model was chosen as the modeling methodology after assessing the Cox PH model because it does not require satisfaction of the PH assumption. It assumes that the effect of a covariate is to adjust (accelerate or decelerate) the time course of the event of interest and is given by,

$$h_i(t) = h_0\left(t/\exp\left(\sum_{j\in I}\beta_j X_{ij}\right)\right)\exp\left(-\sum_{j\in I}\beta X_{ij}\right) \qquad \text{(E2)}$$

where $h_i(t)$ is hazard function for individual i determined by a set of j covariates $\{X_{ij}\}$ and corresponding (estimated) coefficients $\{\beta_j\}$, t is the survival time, and $h_0(t)$ is the baseline hazard defined by a parametric form with an underlying probability distribution such as Weibull, exponential, or gamma. The β-parameter value specifies the effect each covariate has on the survival time, where negative β values indicate that the survival time increases with positive-valued covariates, and positive β values indicate that the survival time decreases with positive-valued covariates.

Selection of Parametric Distribution

Multiple parametric distributions were tested for their ability to approximate the underlying hazard function including exponential, Weibull, gamma, generalized gamma, generalized F, log logistic, log normal and Gompertz. Resulting Akaike information criterion (AIC) values and graphical methods for survival and hazard function fits were compared to select an appropriate parametric form. The

'flexsurvreg' function in the 'flexsurv' R package was used for the selection of parametric distribution analysis.

Univariate Analysis

A univariate analysis was performed by estimating an AFT model using the parametric distribution selected from Section 4.3.6.1, for each of the covariates in Table 3. The 'flexsurvreg' function in the 'flexsurv' R package was used to perform parametric AFT model analysis. Individual covariates with no significant association (P-value ≥ 0.05) with T1D diagnosis were not considered for the full model development. The p-value was computed using the Wald test, as described. A multiplicity adjusted alpha value (Bonferroni correction) was used for univariate analysis. The remaining covariates were analyzed for multicollinearity and associations prior to performing multivariate analysis.

Analysis of Correlation and Association between Covariates

The analysis defined in Section 4.3.5.3 was repeated for the covariates remaining after the AFT univariate analysis.

Multivariate Analysis

The multivariate analysis was performed by testing all possible combinations of remaining covariates, as the number of covariates for multivariate analysis were reasonable. The comparison between possible models was conducted using Akaike's Information Criteria (AIC). A reduction in AIC value greater than or equal to 10 suggests a strong evidence in favor of the model with lower AIC (Burnham and Anderson 2016).

Model Diagnostics

Quantile-Quantile (Q-Q) plots were used to assess the validity of the AFT model assumption for two groups of survival data. In this case, such groups correspond to the presence or absence of an AA combination. Under the AFT model assumption, the presence of one islet AA combination has a multiplicative effect on survival time. Conceptually, a Q-Q plot examines various percentiles for which the survival times are computed for the two groups. A plot of the survival times for the chosen percentiles should give a straight line if the AFT model is appropriate, where the straight line is an estimate of the acceleration factor. Such plots were generated for each AA combination in the AFT model. To analyze continuous covariates, binary groups were formed using thresholds to allow for the generation of Q-Q plots.

Model Performance and internal Validation

Model Performance

To assess the model's predictive performance on the analysis set, time-dependent receiver operating characteristic (ROC) curves were generated (Heagerty and Zheng 2005). Conceptually, the methodology of this metric is that model predictions on all at-risk individuals up to a time t are derived, and true/false positive rates based on model predictions versus the observed data are computed. This is repeated across multiple timepoints to generate ROC curves. The area under the ROC curves (AUC) are computed, which are interpreted as the concordance between the model prediction and data. This methodology is an appropriate model performance metric as an individual's risk for developing T1D changes over time. Further, it provides metrics as to the model's predictive power for time frames over which a trial of reasonable duration would be conducted.

K-fold cross validation

Model validation was performed using the k-fold cross-validation technique (Breiman and Spector 1992). Data was split into k=5 subsets with roughly equal numbers of subjects. Four of the five subsets were used as a training set, and the remaining set was used as an individual test set. This process was repeated by assigning one of the five subsets as the new test set, while the remaining were used as the training set for all combinations. Goodness-of-fit plots were created by overlaying the model estimated survival on Kaplan-Meier curves for all five folds. The concordance index was computed for each of the five folds estimated by time increments of one year up to six years. Goodness-of-fit plots were created for visual assessments of models fits.

Cross-validation on Paediatric population

An internal validation was performed by analysing predictive performance on paediatric subpopulations in the data. A randomly selected portion (50%) of individuals aged less than an age threshold was extracted and used as a test data set. The remaining data constituted the training data used to fit the model. Goodness-of-fit plots were created by overlaying model estimated survival on Kaplan-Meier curves. The concordance index was computed for time increments of one year up to six years.

Model External Validation

External validation was performed using the DAISY dataset described. The definition of the derived baseline was applied to the data to arrive at a validation set. The AFT model within this subset. Goodness-of-fit plots were created by overlaying model estimated survival on Kaplan-Meier curves. The concordance index was computed for time increments of one year up to six years.

Modeling results

A parametric AFT model was chosen using a Weibull distribution. Model diagnostic, performance, and validation exercises were performed to assess the model's ability to quantify the time-varying effect of islet AAs and glycaemic markers on risk to T1D diagnosis with overall satisfactory results. Results of univariate and multivariate modeling are included in tables 17 and 19 below.

Table 7. Univariate analysis for each covariate using AFT model with Weibull distribution

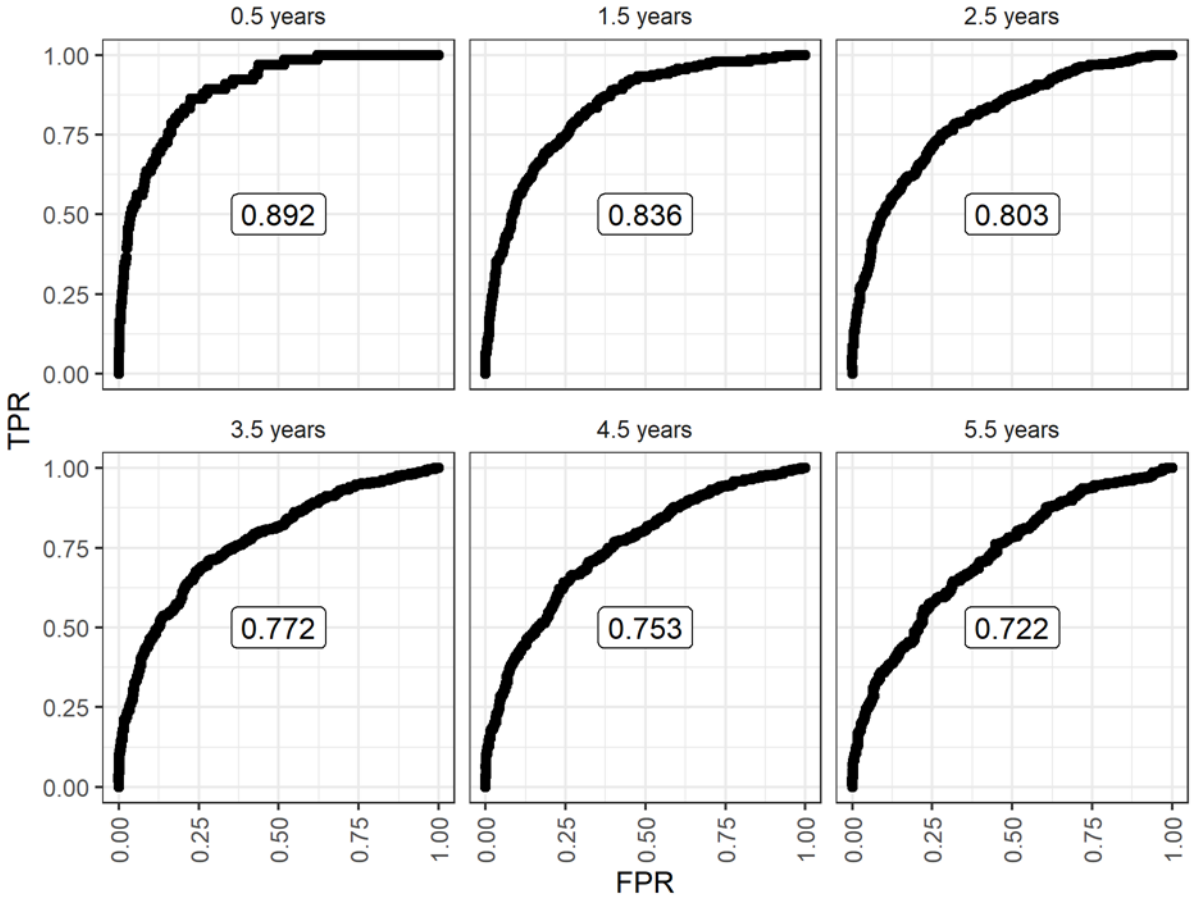
Covariate	beta	95% lower CI	95% upper CI	p-value	Significant
TEDDY_Trial	0.0109	-0.151	0.173	0.895	No
SEX	0.218	0.0755	0.361	0.00273	No
bAGE_s	0.217	0.129	0.306	1.56E-06	Yes
HR_HLA	-0.0684	-0.213	0.0765	0.355	No
FDR	-0.00096	-0.175	0.173	0.991	No
BMI	0.0212	0.000217	0.0421	0.0477	No
GAD65_IAA	0.587	0.348	0.826	1.50E-06	Yes
GAD65_ZnT8	0.663	0.392	0.935	1.66E-06	Yes
GAD65_IA-2	-0.0571	-0.298	0.184	0.643	No
IA-2_IAA	-0.329	-0.846	0.189	0.214	No
IA-2_ZnT8	-0.614	-0.892	-0.337	1.40E-05	Yes
IAA_ZnT8	0.0653	-0.452	0.583	0.805	No
GAD65_IA-2_IAA	-0.163	-0.473	0.147	0.303	No
GAD65_IAA_ZnT8	0.221	-0.056	0.498	0.118	No
GAD65_IA-2_ZnT8	-0.117	-0.299	0.0656	0.209	No
IA-2_IAA_ZnT8	-0.592	-0.868	-0.316	2.57E-05	Yes
GAD65_IA-2_IAA_ZnT8	-0.368	-0.536	-0.199	1.91E-05	Yes
Log_GLU120_s	-0.607	-0.687	-0.526	2.07E-49	Yes
Log_GLU0_s	-0.156	-0.232	-0.0789	7.01E-05	Yes
HbA1c_s	-0.449	-0.529	-0.369	5.08E-28	Yes

Table 8. Model 6 (orig_mod) parameter estimates

Covariates	Beta	95% lower CI	95% upper CI	p-value
Shape	1.350	1.260	1.440	NA
Scale	7.710	6.901	8.634	NA
GAD65_IAA	0.434	0.210	0.659	1.50E-04
GAD65_ZnT8	0.539	0.286	0.792	2.95E-05
IA-2_ZnT8	-0.303	-0.562	-0.043	2.21E-02
IA-2_IAA_ZnT8	-0.342	-0.597	-0.086	8.69E-03
GAD65_IA-2_IAA_ZnT8	-0.143	-0.306	0.021	8.78E-02
Log_GLU120_s	-0.518	-0.594	-0.441	5.64E-40
HbA1c_s	-0.309	-0.379	-0.239	3.42E-18

The time-dependent ROC curves and AUC values showed good prediction performance, especially for up to 2.5 years with AUC values greater than 0.8 (Figure 8).

Figure 8. Evaluation of model performance using time dependent Receiver operating characteristic (ROC) analysis on Final AFT model



Cross-Validation on Paediatric Population

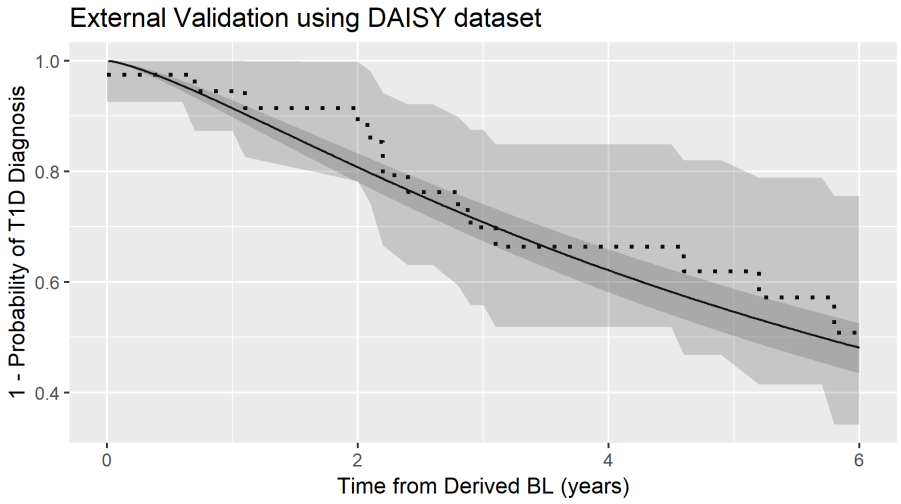
The paediatric population (age < 12) in the analysis dataset comprised of 1330 subjects, with 345 from TEDDY and 985 from TN01. Half of this population i.e. 665 were randomly selected as test set for this cross-validation analysis. A c-index of 0.8 or higher was obtained till 2 years and c-index of 0.75 or higher were obtained up to 6 years indicating good model performance (Table 20) The visual predictive check (VPC) performed on the survival plot for cross-validation on the paediatric population (age < 12) showed reasonable graphical fit (Figure 10). The dotted curve represents the Kaplan–Meier estimate, and the solid curve represent model prediction. The mean survival curve was within the 95% CI band of the estimated Kaplan-Meier curve.

Figure 10. Survival plot for cross-validation on the paediatric population. (Dotted curve represents Kaplan–Meier estimate, and the solid curve represent model prediction)

External Validation

The external validation performed using DAISY data achieved a c-index 0.91 and 0.80 in years one and two, respectively, even with a limited number of subjects, 40, in the external dataset (Table 24). The c-index for subsequent years till six years was over 0.7. The VPC performed on the survival plot showed good graphical fit (Figure 11). These results provide strong evidence for good predictive power for time frames over which a trial of reasonable duration would be conducted.

Figure 11. Survival plot for cross-validation on DAISY external validation dataset (dotted curve represents Kaplan–Meier estimate and solid curve represents model prediction)



The survival modelling approach proposed by the applicant is overall consistent with previous recommendation and agreed upon in principle. The endpoint of interest (diagnosis of T1DM) is very well defined and usually non questionable from a clinical standpoint.

However, several methodological issues were identified in the initial modelling implementation approach as included in the initial proposal by the applicant, that were discussed during the DM, as summarized below:

- The applicant was invited to discuss the value of having a library of models included in the tool rather than a single model (as well as alternative approaches) to allow for flexibility in patient inclusion criteria in the studies.

- In the briefing package, the applicant described the parametric AFT model. However, statistical notation and the description of the model was incorrect.
- The applicant suggested that covariates that were introduced in the model influence the baseline hazard (h0), which was only partially correct.
- The statistical notation in the original briefing document submitted by the applicant needed to be adjusted to better reflect this modelling approach. This inaccuracy has no influence on the presented simulations but is of importance when interpreting the estimated coefficients in the model (β).
- The applicant found that baseline age (bAGE_s) and SEX were highly associated with AA combinations (4.4.1.2. Analysis of Correlation and Association between Covariates). Hence, it was decided to not include bAGE_s and SEX in the subsequent multivariate analysis. This deserved additional justification.
- While it is acknowledged that the introduction of correlated covariates in a model can be problematic, especially when trying to predict in another dataset where this correlation between the covariates might be absent, it seems that the correlations between baseline Age and SEX and the AA combinations are similar for the TrialNet and TEDDY dataset. It also seems that adding SEX and baseline Age to the final AFT model would further reduce the AIC in a statistically significant manner.
- The consistency of covariate correlation across datasets was therefore crucial and it was requested that the applicant provides these data.
- Results of comparison of predictive performance of the proposed model with that of alternative models with other combinations of covariates were also requested, including a model with baseline Age and SEX in addition to the covariates identified by the applicant as final AFT model.
- Moreover, the prediction interval for the survival curves were missing and should be displayed in the figures, along with the R-code used to generate the VPCs that needed to be provided.

As regards the statistical notation and the description of the model, the suggested modifications were implemented by the applicant. visual predictive check"-style figures and R code were provided as requested.

During the DM, in response to these issues, T1DC developed alternative models, including additional variables: baseline age and sex. The original model improved when age and sex were included, as indicated by the lower Akaike's Information Criteria (AIC) value. The time-dependent ROC curves and AUC values demonstrated good prediction performance (AUC > 0.75). Visual-predictive-check (VPC)-style plots showed good graphical fit for internal and external validation of this selected model which included age and sex.

This was acknowledged by the qualification team (QT). It is considered important the applicant provides documented instructions to ensure the model is used correctly.

Alternative models were tested with different combinations of covariates including baseline age and sex in addition to the covariates previously included in the model. A table (table 20) was provided showing the selected covariates for the alternative models. The predictive performance for these models was compared using the AIC. The AIC value of alternative model 3 (alt_mod3) was significantly lower (with a reduction > 10) compared to all other alternative models and the original model. Hence, alternative model 3 (alt_mod3) was chosen as the selected model. Table 21 shows the parameter estimates for the selected model (alt_mod3).

Table 20. Value of AIC for the original model and other alternative models

Model	Covariates	AIC
Original Model (orig_mod)	GAD65_IAA + GAD65_ZnT8 + IA-2_ZnT8 + IA-2_IAA_ZnT8 + GAD65_IA-2_IAA_ZnT8+ Log_GLU120_s + HbA1c_s	2982
Alternative Model 1 (alt_mod1)	GAD65_IAA + GAD65_ZnT8 + IA-2_ZnT8 + IA-2_IAA_ZnT8 + GAD65_IA-2_IAA_ZnT8+ Log_GLU120_s + HbA1c_s + SEX	2972
Alternative Model 2 (alt_mod2)	GAD65_IAA + GAD65_ZnT8 + IA-2_ZnT8 + IA-2_IAA_ZnT8 + GAD65_IA-2_IAA_ZnT8+ Log_GLU120_s + HbA1c_s + bAGE_s	2937
Alternative Model 3 (alt_mod3)	GAD65_IAA + GAD65_ZnT8 + IA-2_ZnT8 + IA-2_IAA_ZnT8 + GAD65_IA-2_IAA_ZnT8+ Log_GLU120_s + HbA1c_s + bAGE_s + SEX	2921

Table 9. Selected model (alt_mod3) parameter estimates

Covariates	Beta	95% lower CI	95% upper CI	p-value
Shape	1.370	1.280	1.470	4.31E-192
Scale	6.780	5.990	7.670	4.36E-56
log_GLU120_s	-0.546	-0.623	-0.469	1.54E-43
HbA1c_s	-0.322	-0.392	-0.252	1.33E-19
SEX	0.275	0.147	0.403	2.65E-05
bAGE_s	0.267	0.183	0.350	3.57E-10
GAD65_IAA	0.506	0.284	0.728	7.95E-06
GAD65_ZnT8	0.474	0.225	0.723	1.88E-04
IA-2_ZnT8	-0.346	-0.603	-0.087	8.42E-03
IA-2_IAA_ZnT8	-0.257	-0.512	-0.002	4.82E-02
GAD65_IA-2_IAA_ZnT8	-0.064	-0.226	0.099	4.40E-01

Model performance for the selected model (alt_mod3) was assessed using time dependent Receiver Operating Characteristic (ROC) curves and associated area under the curve (AUC) values (figure 12). The internal validation for the selected model (alt_mod3) was performed using visual predictive check (VPC)-style plots for a k-fold cross-validation and an internal validation with a paediatric population. An external validation was performed with the DAISY dataset (Figures 9-11) and c-index values over 6 years. The VPC-style plots overlaying observed data over model predictions showed good graphical fit. The "survParamSim" package was used to generate the VPC-style plots.

Figure 12. Evaluation of model performance using time dependent receiver operation characteristic (ROC) analysis

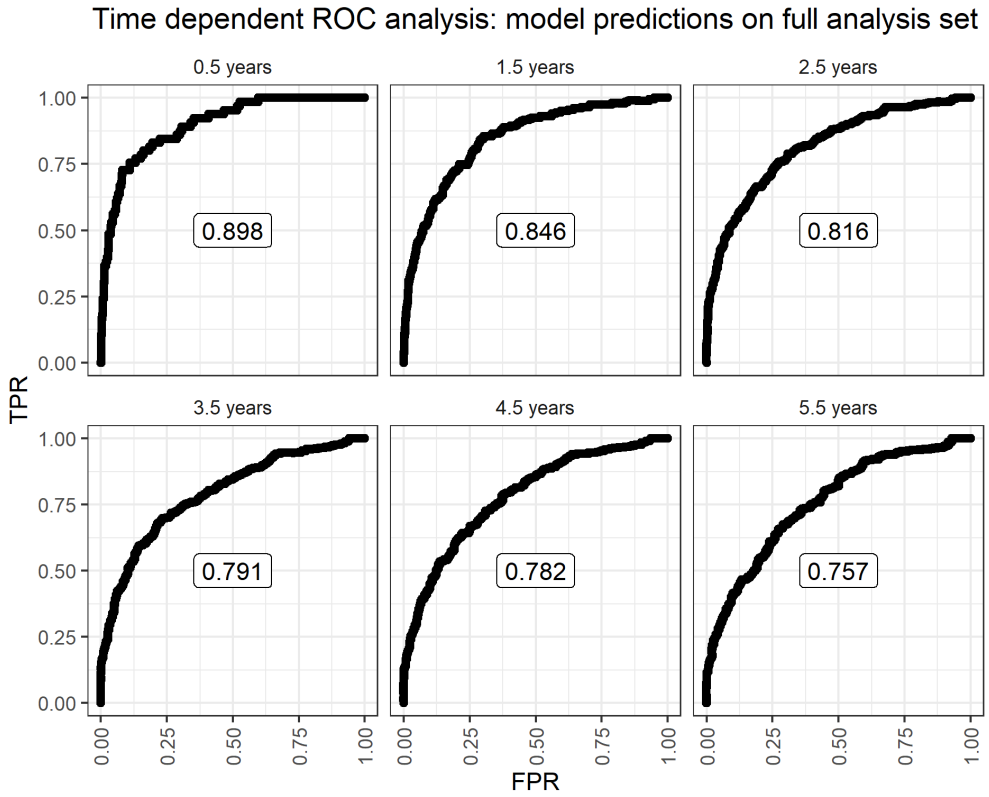


Figure 1. VPC-style plots for k-fold cross validation (red shaded region shows the 95% prediction interval and the black shaded region shows the 95% CI for the observed data)

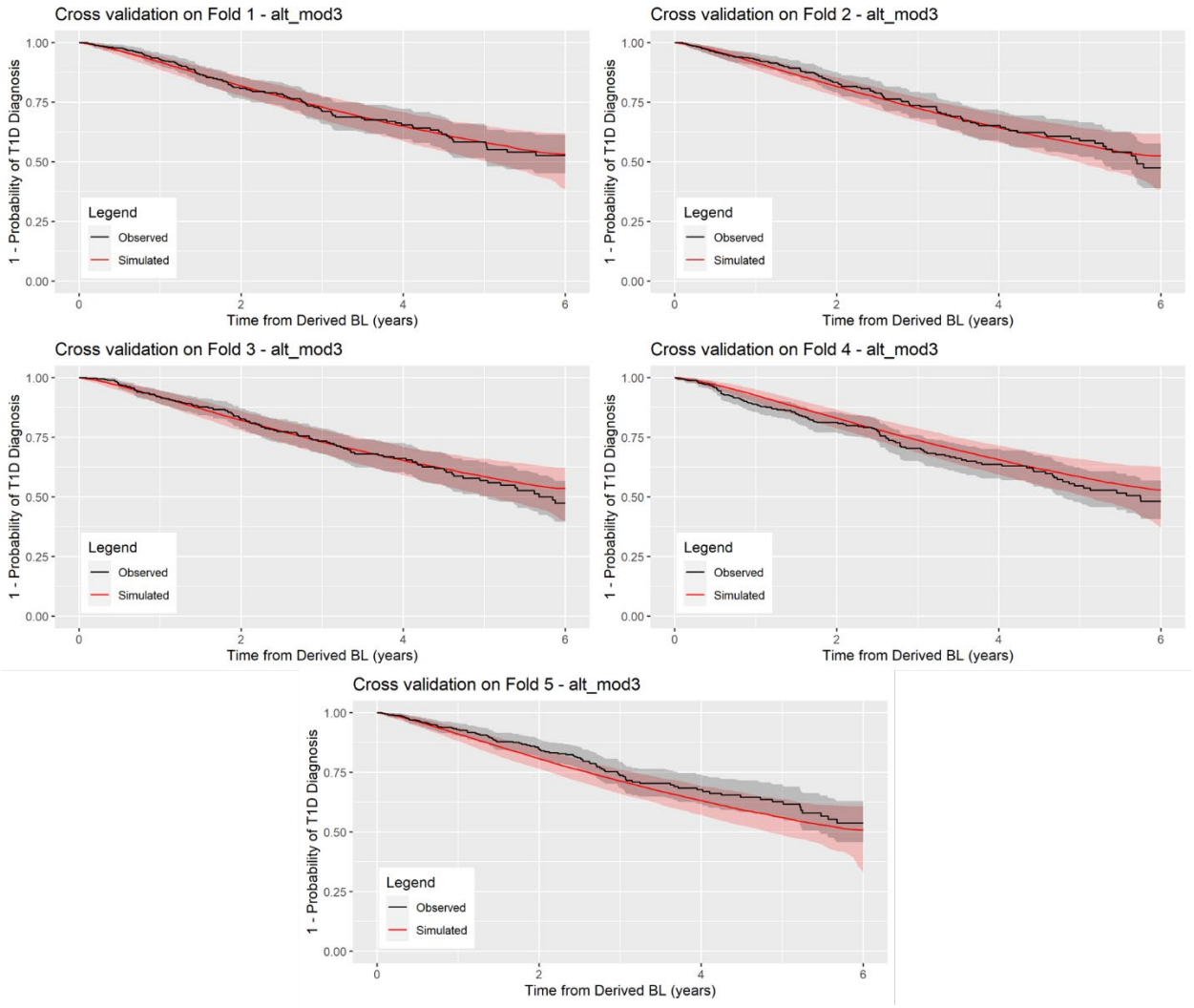


Figure 10. VPC-style plot for internal cross validation (CV) using pediatric population (red shaded region shows the 95% prediction interval and the black shaded region shows the 95% CI for the observed data)

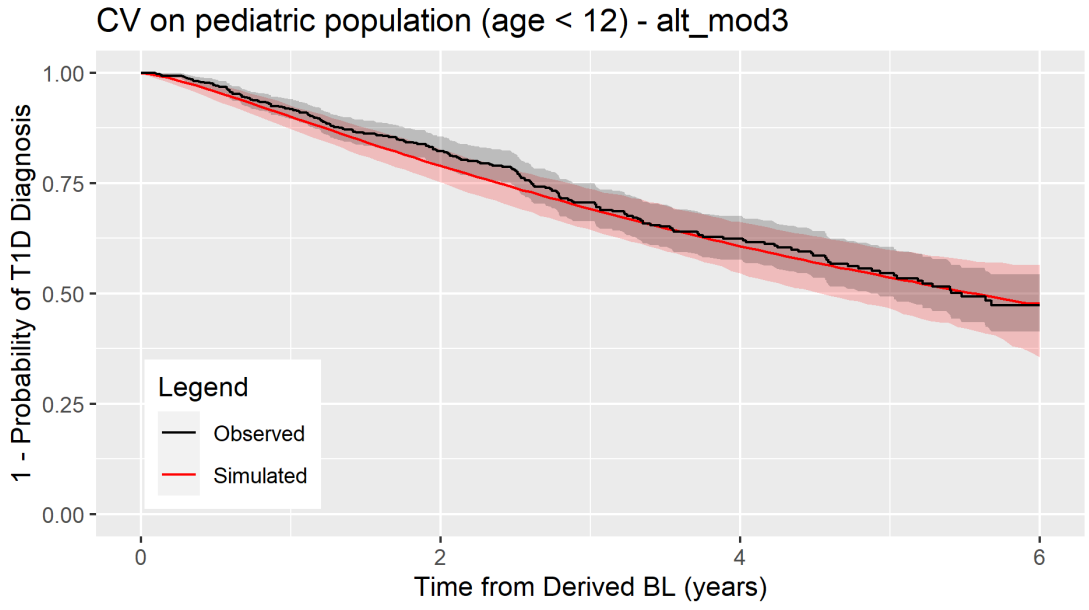
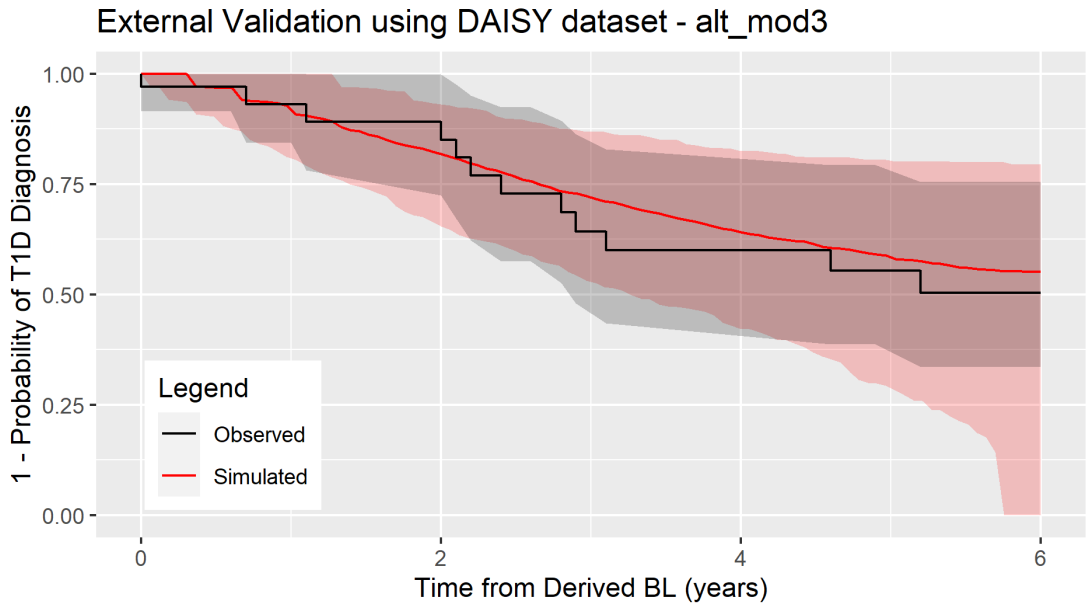


Figure 11. VPC-style plot for external validation using the DAISY analysis dataset (red shaded region shows the 95% prediction interval and the black shaded region shows the 95% CI for the observed data)



The time-dependent ROC curves and AUC values showed good prediction performance especially for up to 2.5 years with AUC values greater than 0.8. The AUC values for subsequent years for up to 5.5 years were greater than 0.75. These results provide evidence for good predictive power for time frames over which clinical trials of reasonable duration would be conducted. The c-index for the selected model (alt_mod3) for all five folds over six years was in most cases close to or higher than 0.8, suggesting good predictive performance.

The alternative models developed by the applicant represent underlying evidence for the qualification of islet AAs as enrichment biomarkers and include clinically relevant glycaemic assessments (i.e., OGTT and HbA1c) as well as demographics (i.e., sex and baseline age) to allow for flexibility in patient inclusion criteria for T1D prevention studies. T1DC indicated that language will be drafted to guide sponsors to discuss with the regulatory agencies the use of this model to inform their drug development strategies.

Patient-level data from DAISY for the derived baseline showed similar distribution and correlation of covariates (including age, sex and AAs) compared to TEDDY and TN01 for the derived baseline. The selected model showed adequate predictive performance across the three datasets for the selected covariates. The addition of age and sex improved model performance. T1DC indicated that they are open to continuing to test covariate correlation and updating the model as more data becomes available.

Conclusion

After the interactions with the SAWP, the applicant has provided a library of models, resulting in acceptable predictive performances for T1DM onset over a 6 years period.

It should be noted that additional covariates were also included in each of the proposed models beside positivity to at least 2 Islet AAs. These additional predictors include HbA1c, blood glucose measurements from the 120-minute timepoints of an OGTT, baseline age and sex of patients. The magnitude of the covariate effects for each of these predictors as well as their combination (OGTT, HbA1c, age and sex) was found to be higher than that of the Islet AAs alone. As a consequence, the impact of the added-value of the positivity will for example be much less important for the patients with already impaired OGTT (120-minute value between 7.8 and 11.1 mmol/L) and pre-diabetes (fasting b-glucose 5.6 to 6.9 mmol/L).

The use of the Islet AAs as a biomarker to optimize the design of clinical trials for the prevention of T1DM should therefore always be done also considering these additional patient characteristics.

Question 4:
EMA agree that the validation is adequate?

T1DC's position: The k-fold cross-validation approach is an adequate method to assess model performance, given all observations are used for training and validation and each observation is used for validation exactly once. This approach has been successfully used in prior qualification procedures with EMA for different novel methodologies in drug development, including biomarkers and quantitative drug development tools. While additional validation using published meta-data was not deemed feasible, an additional external independent patient-level dataset, (i.e., DAISY), was acquired by the T1DC and used to perform patient-level external validation. This approach provided further evidence of robust model performance.

CHMP answer

VPC-style plots overlaying Kaplan-Meier curves over the selected model predictions showed good graphical fit for folds 1, 2, 3 and 4 while fold 5 only performed well within the first year. For the internal cross validation using a paediatric population (age < 12), a c-index of 0.8 or higher was obtained until 3 years and a c-index of 0.75 or higher was obtained up to 6 years for the selected model (alt_mod3) indicating good model performance. The visual predictive check (VPC) performed on the survival plot for cross-validation on the paediatric population (age < 12) showed reasonable graphical fit. For external validation with DAISY dataset, the selected model (alt_mod3) achieved a c-index 0.91 and 0.82 in years one and two, respectively, even with a limited number of subjects (n=34). However, the c-index values beyond three years were relatively lower than up to 2 years, likely attributable to the sparsity of T1D diagnoses during the later years in the DAISY analysis set. The VPC performed on the survival plot showed good graphical fit (Figure 4).
It is agreed that these results provide strong enough evidence for good predictive power for time frames over which a trial of reasonable duration would be conducted.
External validation was considered lacking in the qualification advice procedure. The applicant claims difficulty using published studies. The DAISY dataset was obtained for this purpose. In many ways it is similar to the prior two datasets but, limited to one clinical centre over a long time period. The numbers reaching the T1DM endpoint are low (N=19) compared to the other datasets. The clinical presentation (Table 2) differs significantly, with none of the patients developing DKA in DAISY. This could be due to the small numbers but could also indicate other differences.

Question 5:
Does EMA agree the presented results represent adequate supporting evidence for a qualification opinion?

T1D Consortium position: The presented results demonstrate that the combinations of islet AA for which subjects are seropositive at a sensible baseline for clinical trials independent and statistically significant time-varying predictors of T1D. The presented analyses also show that the use of positivity for combinations of at least 2 islet AAs together with patient characteristics (sex, baseline age) and measures of glycaemic control (blood glucose measurements from the 120-minute timepoints of oral glucose tolerance test (OGTT)), and haemoglobin A1c (HbA1c) levels can help inform the definition of entry criteria, enrichment strategies, and stratification approaches for T1D prevention clinical trials.

CHMP answer

The consortium has done much appreciated work in validating the model, in line with the feedback from the regulators during the qualification advice and opinion procedures. A qualification is therefore recommended for the positivity to at least 2 of the following islet AAs (IAA, GAD65, IA-2, and ZnT8), as measured using the analytical methods described in appendices A and B, and Addendum 1 as a biomarker of incidence of TD1M, when combined with the following additional baseline patients characteristics of OGTT, HbA1c, age and sex.

3. Qualification opinion statement

Positivity to at least 2 of the following islet AAs; IAA, GAD65, IA-2, and ZnT8 is qualified for use as enrichment biomarker, in combination with clinical parameters (sex, baseline age, blood glucose measurements from the 120-minute timepoints of oral glucose tolerance test (OGTT), and hemoglobin A1c (HbA1c) levels) in T1D prevention trials targeting individuals at risk of developing T1D. A survival modelling approach was used to describe how the islet antibodies can be combined to the other patient baseline characteristics for predicting timing to a T1D diagnosis.

The modeling exercise identified the relevance of additional clinical parameters (sex, baseline age, blood glucose measurements from the 120-minute timepoints of an oral glucose tolerance test (OGTT), and hemoglobin A1c (HbA1c) levels).

"At risk" was defined in this context as being a first degree relative (FDR) of a T1D patient, or having a specific human leukocyte antigen (HLA) subtype of risk (HLA-DR3/3, DR4/4, DR3/4, DR3/X [X≠3], DR4/X [X≠4]), excluding individuals with baseline fasting glucose ≥ 126mg/dL (7.0 mmol/L) or stimulated 2-hour glucose ≥ 200 mg/dL (11.1 mmol/L).

The present qualification opinion was requested for a new tool dedicated to enriching Type 1 Diabetes (T1D) prevention clinical trials. The proposed focus is on confirming the existence of a statistically significant contribution of the positivity of of two or more islet autoantibodies (AAs) as predictors of progressing towards a diagnosis of T1D, when combined with additional patient characteristics such as OGTT, HbA1c, age and sex, as described in a validated survival model.

The applicant used an empirical/data driven modeling approach. In the absence of a mechanistic disease model, a clear and fully quantitative description of the contribution of the different factors including positivity to these AAs as predictors of progressing towards a diagnosis of T1D is therefore not possible. The models, as proposed, only allow confirming the existence of a statistically significant contribution of the different (combinations of) covariates and their relative relevance toward theT1D diagnosis for patient at risk.

From a practical drug development standpoint, this proposed use is considered of added value because the intended application can help inform the definition of entry criteria, enrichment strategies, and stratification approaches in the field of T1D prevention. The clinical interest of identifying a good biomarker for Type 1 Diabetes (T1D) onset in an at-risk patient population is supported by the CHMP and the unmet need for better means to optimize drug development in the field is acknowledged.

The model-based approach proposed by the applicant is considered an acceptable method to address the question of interest which is whether the combination of positivity to 2 or more of the 4 selected islet AAs can be considered acceptable predictors of a diagnosis of T1D, when combined to additional and well-defined patient characteristics.

The analytical assays used to measure islet autoantibodies (AA) against glutamic acid decarboxylase 65 (GAD65), insulinoma antigen-2 (IA-2; also called ICA512), insulin (IAA), and zinc transporter 8 (ZnT8) in the three clinical studies contained in the modeling analysis are described in the 'key additional elements' section below. They are considered state of the art. It should be noted that the results and the conclusions of the modeling analysis as assessed during this qualification procedure are considered only applicable when the islet autoantibodies are measured using these methods or methods proved to have at least equivalent analytical performances.

It should importantly be noted that this Qualification only refers to the value of the positivity of at least two islet AAs in the risk assessment, when measured using the described analytical methods ('key additional elements' section below), or methods with comparable accuracy, sensitivity and specificity.

The data used for the model development and external validations to support the qualification of islet AAs as enrichment biomarkers originated from three independent datasets: The Environmental Determinants of Diabetes in the Young (TEDDY), the TrialNet Pathway to Prevention Study (TN01) and the Diabetes Autoimmunity Study in the Young (DAISY) the TN01, TEDDY, and DAISY registry studies. Details are provided in the answer to Question 2 by the applicant.

The data sources are judged largely relevant, consistent with the recommendation during the QA procedure. From a modeling perspective, this approach is endorsed, and the 3 data sources seem adequate. Potential covariate distribution and correlation were presented and discussed as requested during the qualification procedure.

The baseline data intended for modeling are relatively well defined, as well as the binary endpoint (T1D diagnosis). Longitudinal assessments of islet AA positivity, OGTTs, C-peptide measurements, and HbA1c measurements are considered out of scope for the proposed analysis, and only baseline information were used for the modeling analysis.

The precise definition of baseline used for the analysis set is the first record, (i.e., time point) for each individual in which the following criteria is satisfied:

- Presence of any two or more of the 4 islet AAs
- Complete, (i.e., non-missing) information for OGTT (0 and 120-minute time points), HbA1C measurements, age and sex.

The applicant developed a survival model to describe the time course of incidence of T1DM in patients included in the 2 datasets used for model building (TEDDY and TN01), given their baseline characteristics. The third dataset (DAISY) was used for model validation.

The details and different steps of modeling methodology, model development, internal and external validation are described as initially provided by the applicant in answer to Questions 3 and 4. Following the DM with SAWP, several components of the proposed modelling plan were updated according to SAWP feedback. The updated modelling analysis plan was executed, submitted to SAWP, and discussed at a subsequent DM.

Briefly, the applicant has provided a library of models, including a preferred selected model, resulting in acceptable predictive performances for T1DM onset over a 6-year period. It should be noted that additional covariates were also included in each of the proposed models beside the positivity to at least 2 islet AAs. These additional predictors are HbA1c, blood glucose measurements from the 120-minute timepoints of OGTT, baseline age and sex. The magnitude of the covariate effects for each of these predictors as well as their combination (OGTT, HbA1c, age and sex) was found to be higher than that of the IAAs. As a consequence, the impact of the added-value of the positivity will for example be much less important for the patients with already impaired OGTT (120-minute value between 7.8 and 11.1 mmol/L) and pre-diabetes (fasting b-glucose 5.6 to 6.9 mmol/L).

The models that provided the evidence for this qualification opinion are available in the 'key additional elements' section below as implemented in R software (The R code used to implement the model with the best predictive performances is provided below).

In conclusion, the use of the islet AAs as biomarkers to optimize the design of clinical trials for the prevention of T1DM should therefore always be done also considering these additional patient characteristics, as described in the models.

4. Key additional elements

4.1. Islet autoantibody analytical assays

General background on Islet autoantibody assays

Multiple assays were used to measure islet autoantibodies (AA) against glutamic acid decarboxylase 65 (GAD65), insulinoma antigen-2 (IA-2; also called ICA512), and insulin (IAA) in the two clinical studies contained in our modeling analysis: TrialNet Pathway to Prevention (TN01TN01, formally called TrialNet natural history study), and The Environmental Determinants of Diabetes in the Young (TEDDY). Data were collected over the period of 2004-2018 for TN01 and 2004-2016 for TEDDY. These dates were generated by reference laboratories in Denver, CO (Barbara Davis Center, University of Colorado; Dr. Liping Yu is Principal Investigator) for both TN01 and TEDDY studies and in Bristol, UK (University of Bristol; Mr. Alistair Williams is Principal Investigator) for only TEDDY study. Both laboratories have extensive experience in immunoassay development and validation with a strong record of publication in peer-reviewed journals. The laboratory in Denver was CLIA certified in 2000. Islet AAs were measured in serum using standardized radio-binding assays (RBAs) whose methodological details have

been published. A sample is determined as "positive" or "negative" for a particular islet AA according to pre-specified thresholds determined with reference samples (i.e., sera from patients with recently diagnosed with T1D diabetes as positives, and sera from normal patients as negatives). In addition, robust procedures were used by both laboratories to ensure accuracy of positive calls and the consistency of responses over time. Emphasis for this EMA submission document is placed on a binary determination of seropositivity or seronegativity of islet AAs, rather than on quantitation of continuous values.

Prior to 2010, data were generated using "local" assays developed and published by the Denver and Bristol laboratories. However, starting in 2010, both laboratories implemented NIDDK sponsored "harmonized" assays for autoantibodies to GAD65 and IA-2 (but not for ZnT8 or IAA autoantibody assays) which were developed under the direction of the Islet Autoantibody Harmonization Committee, which included the use of common reference standards (for generating standard curves and common units of autoantibody levels in serum) from the US National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). This project was also supported by the Islet Autoantibody Standardization Program (IASP), formerly known as the Diabetes Autoantibody Standardization Program (DASP), which is an international effort to improve and harmonize measurement of islet AAs associated with T1D through proficiency testing, and by providing advice, training, and information. The Centers for Disease Control and Prevention (CDC) have participated in this National Institutes of Health (NIH) sponsored standardization effort. Every 18 months IASP carries out a voluntary or opt-in assessment program for labs around the world that perform islet AA assays. In this assessment, IASP provides between 50-150 blinded seropositive and seronegative sera samples sets from T1D patients and control subjects as well as reference standard reagents to participating laboratories, and the results released to laboratories to continually compare and improve assay performance. Data from the DASP/IASP assessments for the Bristol and Denver labs are described later in this document under the discussion of concordance.

The qualitative, binary determination of seropositivity or seronegativity for each islet autoantibody is a key feature in the modeling plan outlined in Section 4.3.1 of the Briefing Document. Calling a particular sample positive for a given autoantibody is defined as when the measured value exceeds a cutoff that was set at an antibody prevalence in reference populations of healthy individuals and those with T1D. Ideally, the reference populations should have similar characteristics to the at-risk population and be large enough to achieve tight confidence intervals. For the determination of positivity cutoffs, positive controls are serum samples from patients newly diagnosed (within two weeks) with T1D, and negative controls are serum samples from healthy individuals. The cutoff is commonly set at the 99th percentile of the reference population, i.e. a level exceeded by only 1% of these healthy individuals. For the GAD65 and IA-2 harmonized assays (i.e., from 2010 onwards) from Denver and Bristol, NIDDK standards were provided to establish a six-point standard curve for the calculation of standardized Digestive and Kidney (DK) units that were then compared to pre-specified cutoffs for determination of seropositivity or negativity. These NIDDK standards were run in each assay and were provided as part of the harmonization program. For all IAA assays run in Denver, and for GAD65 and IA-2 assays prior to 2010 (termed "local" assays), positive control sera from newly diagnosed T1D patients and negative control sera from healthy subjects were used by the Denver lab to generate an index that enabled the determination of seropositivity or negativity. The index is a ratio of the signal in the test serum to the signal in a positive control; if that ratio exceeds the pre-specified cutoff, then the sample is called seropositive. In the GAD65 and IA-2 assays run before 2010 in Bristol, locally prepared standards were used to generate standard curves for the calculation of World Health Organization (WHO) units that were then compared to pre-specified cutoffs for determination of seropositivity or negativity. In addition, a detailed discussion of how seropositivity was confirmed can be found in Section 4.3 of the Briefing Package.

The assays for GAD65 and IA-2 AAs that generated data for this submission are not quantitative and are only being used in this submission to determine the presence or absence of an individual AA. Some of the features of these islet AA assays that prevent them from being used quantitively are:

- The amount of radio-labeled antigen generated in the *in vitro* transcription/translation reaction is not quantitated.
- The radio-labeled antigen does not fully saturate binding sites of the serum AAs.
- There is no step to compete off non-specific binding using excess cold antigen.

For these reasons, the absolute lower limit of quantitation (LLOQ) and upper limit of quantitation (ULOQ) are not determined for these assays. In addition to the points stated above, because the autoantibodies being detected are a composition of polyclonal antibodies that differ in affinity and concentration, parallelism studies and linearity assessments have not been performed. Although these factors prevent the use of the continuous measure from these islet AA assays, robust positive and negative controls enable the binary adjudication of seropositivity or negativity. Similarly, the Islet AA assay as performed by the Denver and Bristol labs is not quantitative, despite the ability to quantify the antigen and the inclusion of unlabeled insulin to reduce non-specific binding. Although quantitative IAA assays could be developed, those used in this submission were not run in a quantitative format and only the binary output is being utilized.

Although FDA 501k-cleared assays are available to measure some of the AAs, samples for the two studies were analyzed in two different laboratories using either local or harmonized radiobinding assays (RBAs) that were published by the participating laboratories as summarized in Table 1 of this document. This assay format is commonly used for measurement of AAs because it is high throughput, relatively inexpensive, uses small serum volumes, is easily adapted for detection of different AAs (by changing the radiolabeled antigen) and performed better than other immunoassays such as ELISA because of the RBAs solution phase format that facilitates antigen-antibody binding. Should sponsors want to measure islet AAs in future clinical studies, they may choose to use different assays, including those that do not require radiolabels. To verify that these future assays are indeed fit for purpose, a proficiency test consisting of a panel of samples comprising different levels of islet AAs should be performed. This proficiency test would evaluate the same panel of 7 samples in both the RBAs described here and these future alternative assays. This proposed proficiency test is not discussed any further as it is not the focus of this submission. Users of any proposed future islet AAs assay will be required to provide detailed information on precision and relative accuracy.

As with the assays for GAD65, IA-2, and insulin AAs, the ZnT8 AA assay that generated data for this submission is not quantitative and is only being used in this submission to determine the presence or absence of an individual AA. Some of the features of these islet AA assays that prevent them from being used quantitively are:

- The amount of radio-labeled antigen generated in the *in vitro* transcription/translation reaction is not quantitated.
- The radio-labeled antigen does not fully saturate binding sites of the serum AAs.
- There is no step to compete off non-specific binding using excess cold antigen.

For these reasons, the absolute lower limit of quantitation (LLOQ) and upper limit of quantitation (ULOQ) are not determined for these assays. In addition to the points stated above, because the AA being detected are a composition of polyclonal antibodies that differ in affinity and concentration, parallelism studies and linearity assessments have not been performed. Although these factors prevent the use of the continuous measure from these islet AA assays, robust positive and negative controls enable the binary adjudication of seropositivity or negativity.

Samples were analyzed using a local radiobinding assay (RBA) assay format that is commonly used for measurement of AAs because it is high throughput, relatively inexpensive, uses small serum volumes, and is easily adapted for detection of different AAs (by changing the radiolabeled antigen). In addition, the assay performed better than other immunoassays such as ELISA because of the RBA's solution-phase format that facilitates antigen-antibody binding. Should sponsors want to measure islet AAs in future clinical studies, they may choose to use different assays, including those that do not require radiolabels. To verify that these future assays are indeed fit for purpose, a proficiency test consisting of a panel of samples comprising different levels of islet AAs should be performed. This proficiency test

would evaluate the same panel of samples in both the RBAs described here and these future alternative assays. This proposed proficiency test is not discussed any further as it is not the focus of this submission. Users of any proposed future islet AA assays will be required to provide detailed information on precision and relative accuracy.

Table 1. Autoantibody assay summary

Autoantibody	Trial/study name	Site Measured*	RBA Assay Type**	Assay Documentation
GAD65	TN01	UC	Local, Harmonized	2019 Briefing Pkg
GAD65	TEDDY	UC/Bristol	Local, Harmonized	2019 Briefing Pkg
GAD65	DAISY	UC	Local, Harmonized	2019 Briefing Pkg
IA-2	TN01	UC	Local, Harmonized	2019 Briefing Pkg
IA-2	TEDDY	UC/Bristol	Local, Harmonized	2019 Briefing Pkg
IA-2	DAISY	UC	Local, Harmonized	2019 Briefing Pkg
IAA	TN01	UC	Local	2019 Briefing Pkg
IAA	TEDDY	UC/Bristol	Local	2019 Briefing Pkg
IAA	DAISY	UC	Local	2019 Briefing Pkg
ZnT8	TN01	UC	Local	2020 Update
ZnT8	TEDDY	UC	Local	2020 Update
ZnT8	DAISY	UC	Local	2020 Update

* UC = UC Core Facility

** For GAD65 and IA-2, local assays were used for samples analyzed before 2010 and harmonized assays were used for samples analyzed starting in 2010.

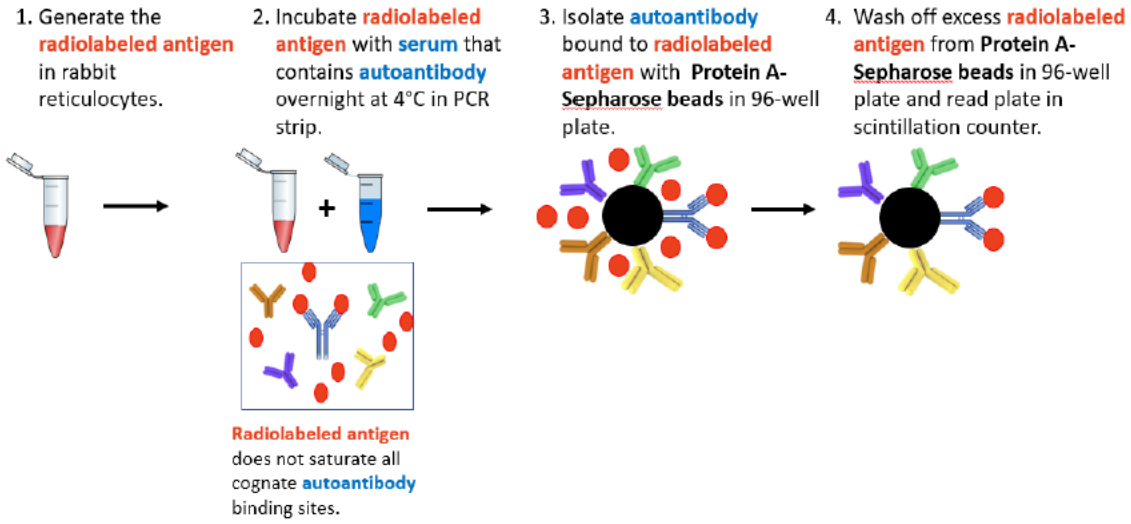
In summary, the assays used to generate the islet AA data were performed in central laboratories that have been participating in multi-center diabetes studies and international islet AA assay harmonization workshops for more than 20 years and the methodologies for all assays have been published in peer-reviewed journals. Importantly, robust procedures, including the use of QC controls that have shown strong concordance between labs and minimal variability over time, were used by both laboratories to ensure accuracy of positive calls or seropositivity or seronegativity and the consistency of responses over time.

Summary of GAD65 and IA-2 AA assays

Overview

Radiobinding assays are used to qualitatively determine the presence or absence, as seropositivity or seronegativity, of the AAs to GAD65 and IA-2 (the local Denver IA-2 was originally called islet cell antigen 512 [ICA512]) in serum samples from patients. For most of these RBAs, one autoantibody is assessed per well (i.e., using one radiolabeled antigen), except for the local GAD65 and IA-2 assays in Denver that are multiplexed using different radiolabels for each antigen. In this assay format, *in vitro* transcription and translation (IVTT) is used to generate a specific radiolabeled human antigen (either GAD65 or IA-2) using a radiolabeled amino acid in rabbit reticulocyte lysates. Once prepared, the radio-labeled antigen is incubated with patient serum overnight. A non-specific immunoglobulin precipitation is then carried out with Protein A-Sepharose beads to isolate radiolabeled antigen-islet autoantibody complex to enable removal of unincorporated radiolabeled amino acids. The washed, isolated beads are then assessed via scintillation counting to evaluate the levels of radiolabeled antigen that have been isolated. These levels are then compared to positive controls for the determination of seropositivity.

Figure 1: Schematic of Radiobinding Assay format used for GAD65 and IA-2 AA assessments



The harmonized GAD65 and IA-2 autoantibody assays both use calibrators/standards developed by the NIDDK that were prepared from a set of positive and negative calibrators prepared from pooled sera (Bonifacio et al. 2010). For the positive calibrator, 25–50 ml serum was collected from each of 21 patients with T1D aged 18 to 30 years with a median time since diagnosis of 1.1 year (range, 0.2 to 2.2 year). The presence of moderate/high levels of antibodies to GAD and IA-2 in individual sera was confirmed in the Denver, Bristol, and Munich (a laboratory from Helmholtz Zentrum München participated in harmonization efforts but did not assay samples for TEDDY and TN01) laboratories before the sera were pooled. For the negative serum diluents and calibrator, 12 frozen serum donations (median volume for each sample, 228 ml) were obtained from the blood bank of the Städt Klinikum München GmbH, Munich and pooled.

While there are general similarities in how the different RBAs are performed for all autoantibodies, there are also differences when comparing the local and harmonized assays from the same site, as well as the assays from Denver and Bristol. For example, each site prepares its own local QC standards that are used to assess performance over time and to ensure the assay is functioning properly. Other differences specific for GAD65 and IA-2 are highlighted in the text and tables below.

GAD65 and IA-2 Assay Characteristics

GAD65: In comparing the local assays from Denver and Bristol, there are several differences. The local Denver assay measures GAD65 in a multiplexed format with IA-2 (called ICA512 in the SOP) in which GAD65 is labeled with 3H-leucine and IA-2 is labeled with 35S-methionine in separate IVTT reactions and then the two labeled antigens are mixed with the serum in the assay. Also, the Denver assay uses an index for determining seropositivity whereas the Bristol assay uses WHO units, which were established at the first DASP workshop in 2000 (Mire-Sluis et al. 2000). All versions of the GAD65 assay used expression plasmids encoding the full-length protein. In comparing the harmonized assays, the methods are highly similar, but as mentioned, different local QC controls are used. Table 2 compares the local and harmonized Denver and Bristol GAD65 assays. In addition, only the Bristol lab uses a confirmatory threshold (20 DK units, which is set below the positivity threshold of 33 DK units to avoid introducing a negative bias); samples that exceed the threshold are repeated in a separate assay and reported as the mean of the two results. Finally, the positivity cutoff for the harmonized assay run in Denver is 20 DK units/ml, whereas it is 33 DK units/ml for the Bristol assay.

Table 2. Comparison of local and harmonized GAD65 assays from Denver and Bristol

	GAD65			
Local or Harmonized	Local	Local	Harmonized	Harmonized
Site	Denver	Bristol	Denver	Bristol
Antigen source	IVTT	IVTT	IVTT	IVTT
Plasmid clone	pEx9-GAD65	pGEM3-GAD65	pTH-GAD65	pTH-GAD65
Source of plasmid	A. Lernmark	E. Bonifacio	A. Lernmark	A. Lernmark
Plasmid Reference	Grubin 1994	Bonifacio 1995	Hansson 2010	Hansson 2010
Amino acids expressed	Full length	Full length	Full length	Full length
Local QC controls	High pos, low pos, neg	High pos, med pos, low pos, neg	High pos, low pos, neg	High pos, med pos, low pos, neg
Calibrator/Standards	Same as QC controls	Locally prepared	7 from NIDDK	7 from NIDDK
Radiolabel	^{3}H-Leucine	^{35}S-Methionine	^{35}S-Methionine	^{35}S-Methionine
Sepharose beads	Protein A	Protein A	Protein A	Protein A
Multiplexed	Yes (with IA-2)	No	No	No
Assay Units	Index	WHO units/ml	DK units/ml	DK units/ml

IA-2: In comparing the local assays from Denver and Bristol, there are several differences. The local Denver assay measures IA-2 (called ICA512 in SOP) in a multiplexed format in which the IA-2 is labeled with 35S-methionine and GAD65 is labeled with 3H-leucine in separate IVTT reactions and then the two labeled antigens are mixed with the serum in the assay. Also, as with GAD65, the local Denver assay uses an index for determining seropositivity whereas the Bristol assay uses WHO units, which were established at the first DASP workshop in 2000 (Mire-Sluis et al. 2000). Finally, the antigen expressed in the local Denver IA-2 assay (pCRII-ICA512bdc, amino acids 256-979) is different from local Bristol assay (pSP64 IA-2ic, 605 to 979) and the antigen in the harmonized assay (pSP64-PolyA-IA-2ic, amino acids 606 to 979). Table 3 compares the local and harmonized Denver and Bristol IA-2 assays.

To demonstrate alignment between the local Denver IA-2 assay, which utilized a long-form construct (256-979), and the Bristol local and harmonized assays, which used a shorter construct (606 to 979), a comparison carried out between both labs was performed using 2,172 TN01 samples. These 2,172 samples included: 1,089 samples positive for any AA with the "local" TrialNet assays (GAD65, IAA and IA-2) and 1,074 randomly selected antibody negative samples. The local Denver IA-2 and harmonized assays from Denver were 95% concordant for positives or negatives with r2= 0.72 for IA-2 AAs. In comparing the harmonized assays from Bristol and Denver, the methods are highly similar, but as mentioned, different local QC controls are used. In addition, only the Bristol lab uses a confirmatory threshold (1.4 DK units, which is set below the positivity threshold of 5 DK units to avoid introducing a negative bias); samples that exceed the threshold are repeated in a separate assay and reported as the mean of the two results.

Table 3. Comparison of local and harmonized IA-2 assays from Bristol and Denver.

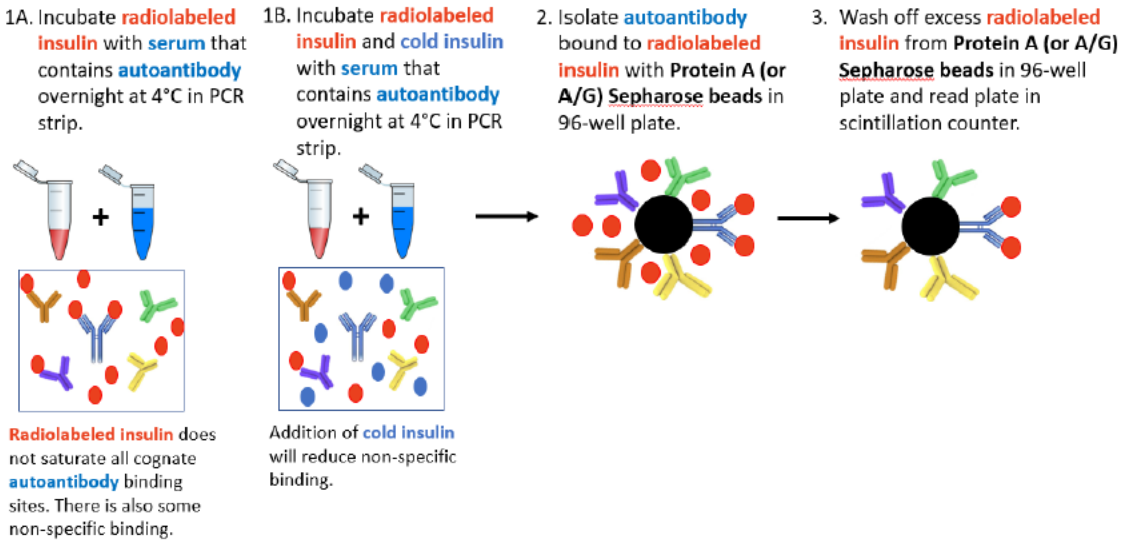
	IA-2			
Local or Harmonized	Local	Local	Harmonized	Harmonized
Site	Denver	Bristol	Denver	Bristol
Antigen source	IVTT	IVTT	IVTT	IVTT
Plasmid clone	pCRII-ICA512bdc	pSP64 IA-2ic	pSP64-PolyA-IA-2ic	pSP64-PolyA-IA-2ic
Source of plasmid	Barbara Davis Center	M. Christie	Ezio Bonifacio	V. Lampasona
Plasmid Reference	Gianani 1995	Hatfield 1997	Bonifacio 2010	Bonifacio 2010
Amino acids expressed	256-979	605-979	606-979	606-979
Local QC controls	High pos, low pos, neg	High pos, med pos, low pos, neg	High pos, low pos, neg	High pos, med pos, low pos, neg
Calibrator/Standards	Same as QC controls	Locally prepared	7 from NIDDK	7 from NIDDK
Radiolabel	^{35}S-Methionine	^{35}S-Methionine	^{35}S-Methionine	^{35}S-Methionine
Sepharose beads	Protein A	Protein A	Protein A	Protein A
Multiplexed	Yes (with GAD65)	No	No	No
Assay Units	Index	WHO units/ml	DK units/ml	DK units/ml

3 Summary of the Insulin AA Assay

Overview

For detection of insulin AAs, 125I-insulin is used as the antigen rather than *in vitro* transcription and translation (Figure 2). There is also an additional step that includes competition with unlabeled ("cold") insulin (to reduce non-specific binding), and immunoglobulin-binding Sepharose beads are used to isolate the radiolabeled antigen-islet AA complex to enable removal of unincorporated radiolabeled amino acids. In parallel, samples are incubated with either 125I-insulin alone, or with a combination of 125I-insulin and cold insulin, and the results are calculated based on the difference in radioactivity between the two for each sample. In all assays, an islet AA is called positive, if the measurement in the assay exceeds a predefined positivity threshold/cutoff.

Figure 2: Schematic of Insulin AA Radiobinding Assay Format



Insulin AA Assay Characteristics

In Bristol, the assay is run in two stages: first, a screening assay (IAA) in which samples are tested for insulin binding using 125I-insulin alone (hot label) is run; if above the screening threshold then a competition assay (CIAA) is run in which specificity of insulin binding is confirmed by displacement of

binding to 125I-insulin label by addition of excess recombinant insulin (cold label). The CIAA results are calculated using the mean delta cpm (cpm with hot label – cpm with cold label) for each sample. The Denver assay is run similarly, except that, starting in 2015, if the signal of the duplicate with cold insulin is ≥ 20 cpm, then one of two next steps is taken: If the count is less than 1,000 cpm and is confirmed by a 2nd run, then the sample will be considered not reportable (due to non-specific binding). However, if the counts are greater than or equal to 1,000 cpm and confirmed by 2nd run, then the assay will be re-run with 10x more cold insulin. The reason for this two-step process is that in rare instances, some samples gave false positive signals in the original IAA assay. It was determined that these false positives were due to the presence of 125I-labeled peptides other than insulin, which were related to insulin and present as a contaminant of the purchased 125I-insulin reagent. This was demonstrated because the false positive signal could not be competed with cold insulin and was hence likely due to the presence of non-insulin 125I-labeled peptides being present in the purchased 125I-insulin reagent.

The Denver assays uses a combination of Protein A-Sepharose and Protein G-Sepharose, whereas the Bristol lab only uses Protein A-Sepharose. The determination of positivity in the Denver assay is based on an index using local QC controls, whereas a standard curve of arbitrary units is used for the Bristol assay. Table 5 compares the local Denver and Bristol IAA assays.

Table 5. Comparison of local IAA assays from Denver and Bristol

Local or Harmonized	Local	Local
Site	Denver	Bristol
Antigen source	Amersham	Amersham
Plasmid clone	NA	NA
Source of plasmid	Perkin Elmer	Perkin Elmer
Plasmid Reference	NA	NA
Amino acids expressed	Full length	Full length
Local QC controls	High pos, low pos, ultra-low pos, neg	High pos, med pos, low pos, neg
Calibrator/Standards	Same as QC controls	Locally prepared
Radiolabel	^{125}I-Insulin	^{125}I-Insulin
Sepharose beads	Protein A and Protein G	Protein A
Multiplexed	No	No
Assay Units	Index	Arbitrary units

ZnT8 Assays

Data from the ZnT8 assay were generated by the Autoantibody/HLA Core Facility at the University of Colorado (UC), Aurora, CO, USA; (referred to as the UC Core Facility throughout the rest of this document and referred to as the "Denver lab"). Islet AAs were measured in serum using standardized radio-binding assays (RBAs) whose methodological details have been published [1, 2]. A sample is determined as "positive" or "negative" for a particular islet AA according to pre-specified thresholds determined with reference samples (i.e., sera from patients with recently diagnosed with T1D diabetes as positives, and sera from normal patients as negatives).

1. Lampasona V, Schlosser M, Mueller PW, et al (2011) Diabetes Antibody Standardization Program: First Proficiency Evaluation of Assays for Autoantibodies to Zinc Transporter 8. Clinical Chemistry 57(12):1693–1702. https://doi.org/10.1373/clinchem.2011.170662
2. Yu L, Herold K, Krause-Steinrauf H, et al (2011) Rituximab selectively suppresses specific islet antibodies. Diabetes 60(10):2560–2565. https://doi.org/10.2337/db11-0674

R code for the final model (i.e. with the best predictive performances)

```
---
R markdown file number: "4"
title: "Model validation - Islet AA for EMA qualification"
author: "T1DC modeling team at C-Path"
last updated: 12 May 2020
---
This R markdown file contains code for model validation including K-fold and external validation with
DAISY dataset. The result from running a code block can be viewed under the code block. Additionally,
the figures and tables generated from these code blocks will be saved in "figures" and "tables" folder
under "deliv" folder. The associated file names for the figures and tables describe the data being
visualized or tabulated.
```{r Check if relevant libraries are installed on local machine, install otherwise}
#Function to check whether a package is installed
is.installed <- function(mypkg) {
            is.element(mypkg, installed.packages()[, 1])
}
#A tool for fast aggregation of large data
if (is.installed("data.table") == FALSE) {
      install.packages("data.table" , dependencies = TRUE)
}
#A library for computing survival analyses
if (is.installed("survival") == FALSE) {
       install.packages("survival" , dependencies = TRUE)
}
#A library for visualizing survival analysis results
if (is.installed("survminer") == FALSE) {
       install.packages("survminer" , dependencies = TRUE)
}
#A library of r packages to perform data science tasks
if (is.installed("tidyverse") == FALSE) {
      install.packages("tidyverse" , dependencies = TRUE)
}
#A package to generate correlation plots
if (is.installed("corrplot") == FALSE) {
       install.packages("corrplot" , dependencies = TRUE)
}
#A package to perform survival analysis
if (is.installed("flexsurv") == FALSE) {
      install.packages("flexsurv" , dependencies = TRUE)
}
#A package to compute time-dependent ROC curve from censored survival data
if (is.installed("survivalROC") == FALSE) {
      install.packages("survivalROC" , dependencies = TRUE)
}
#A toolbox for assessing and comparing performance of risk predictions
if (is.installed("riskRegression") == FALSE) {
      install.packages("riskRegression" , dependencies = TRUE)
}
#A package for estimation of prediction accuracy for time-to-event data
if (is.installed("survAUC") == FALSE) {
```
```

```
        install.packages("survAUC" , dependencies = TRUE)
}
```
```{r load libraries}
library(data.table) #A tool for fast aggregation of large data
library(survival) #A library for computing survival analyses
library(survminer) #A library for visualizing survival analysis results
library(tidyverse) #A library for r packages for perform data science tasks
library(corrplot)#A package to generate correlation plots
library(flexsurv)#A package to perform survival analysis
library(survivalROC)#A package to compute time-dependent ROC curve from censored survival data
library(riskRegression) #A toolbox for assessing and comparing performance of risk predictions
library(survAUC) #A package for estimation of prediction accuracy for time-to-event data
#library(rms)
```
```{r Clear environment}
rm(list=ls())
```
```{r Load modeling analysis datasets generated from R markdown file 1 from the "data" folder}
#Model analysis dataset from TN01 and TEDDY
data <- readRDS("../data/final_EMA_islet_AA_datamart.rds")
#External validation dataset from DAISY
data_daisy <- readRDS("../data/final_EMA_daisy_datamart.rds")
```
```{r Recode subject IDs to be consecutive integers}
data$IDp <- data$IDp_new
```
```{r K-fold cross-validation analysis as discussed in section 4.3.7.2 - generating random k-folds}
#Set a seed value for random split
set.seed(1)
#set number of folds to 5
n <- 5
#Generate 5 random data splits
cv <- getSplitMethod(paste0("cv",n), B=1, N=2022)
folds <- cv[[3]]
folds <- as.factor(folds)
splits <- split(data, folds)
```
```{r K-fold cross-validation analysis as discussed in section 4.3.7.2}
#Set a seed value
set.seed(1)
#Assign maximum year for c-index calculation
yrs_for_cindex <- 6
#Create a matrix to store c-index values
cindex_k_fold <- matrix(NA,nrow = n, ncol = yrs_for_cindex)
#Apply for loop to rotate folds for cross-validation
for(i in 1:n){

  train <- data.frame()
  train_inds <- c(1:n)
  train_inds <- train_inds[-i]
```

```
  test_ind <- i
  for(j in 1: (n-1)) {train <- rbind(train,splits[[train_inds[j]]])}
  test <- splits[[test_ind]]


    #Fit model using 'flexsurvreg" function with final multivariate AFT model described in section
4.4.2.4
    surv_obj_train <- Surv(train$T_event, train$status)
    fit_train <- do.call(flexsurvreg, list(formula = surv_obj_train ~ GAD65_IAA + GAD65_ZNT8 +
IA2A_ZNT8 + IA2A_IAA_ZNT8 + GAD65_IA2A_IAA_ZNT8 + HbA1c_s +         log_GLU120_s, data =
train, dist = "Weibull"))

    #Use "survreg" to compute c-index
    fit_train_concordance<- survreg(Surv(T_event, status) ~ GAD65_IAA + GAD65_ZNT8 +
IA2A_ZNT8 + IA2A_IAA_ZNT8 + GAD65_IA2A_IAA_ZNT8 +  HbA1c_s +log_GLU120_s , data =
train ,dist = "weibull" )
    #Check model fit with test fold
    fit_test <- survfit(Surv(T_event, status) ~ 1, data = test)
    surv1 <- summary(fit_train, newdata = test, type = "survival", B=1, tidy = TRUE)
    varnames <- c("time", "surv", "lower", "upper")
    fit_test_data <- cbind(fit_test$time, fit_test$surv, fit_test$lower, fit_test$upper)
    fit_test_data <- as.data.frame(fit_test_data)
    names(fit_test_data) <- varnames

    surv_avg <- surv1 %>%
      group_by(time) %>%
      summarise(mean_est = mean(est, na.rm=TRUE),
                mean_lcl = mean(lcl, na.rm=TRUE),
                mean_ucl = mean(ucl, na.rm=TRUE),
                )
    #Generate plot to check goodness-of-fit
    p <-ggplot() +
    ggtitle(paste("Cross validation on Fold ",i, sep = "")) +
    geom_line(data = surv_avg, aes(x = time, y = mean_est)) +
    geom_step(data = fit_test_data, aes(x = time, y = surv), linetype = 3, size = 1) +
    geom_ribbon(data = fit_test_data, aes(x = time, ymin = lower, ymax = upper ), linetype = 0,
alpha = .2, show.legend = FALSE) +
    geom_ribbon(data = surv_avg, aes(x = time, ymin = mean_lcl, ymax = mean_ucl), linetype = 0,
alpha = .2, show.legend = FALSE) +
    xlab("Time from Derived BL (years)") +
    ylab("1 - Probability of T1D Diagnosis")

    #View goodness-of-fit plot
     p

    #Export cross-validation plots
    ggsave(paste("../deliv/figures/",i," fold_validation",".png", sep = ""), p, width = 16, height = 9,
units = "cm")
   #Compute c-index for model prediction on kth fold
    for(q in 1:yrs_for_cindex){
```

```
      c_index_tmp <- concordance(object = fit_train_concordance, newdata = test, ymin = 0,ymax =
q)
      cindex_k_fold[i,q] <- c_index_tmp$concordance
    }
}
#Store c-index value in a data frame
cindex_k_fold <- as.data.frame(cindex_k_fold)
#Assign column and row names for c-index table
colnames(cindex_k_fold)<-c("year 1","year 2", "year 3", "year 4", "year 5", "year 6")
rownames(cindex_k_fold)<-c("fold 1","fold 2", "fold 3", "fold 4", "fold 5")
#export results
write.csv(cindex_k_fold, "../deliv/tables/cindex_k_fold.csv", row.names = TRUE)
```
```{r K-fold cross-validation analysis stratified by each of the islet AA combinations and continuous
covariates using binary groups as discussed in Appendix H Figure 39-73}
#Set a seed value
set.seed(1)
#Apply for-loop to rotate folds for cross-validation
for(i in 1:n){

  train <- data.frame()
  train_inds <- c(1:n)
  train_inds <- train_inds[-i]
  test_ind <- i
  for(j in 1: (n-1)) {train <- rbind(train,splits[[train_inds[j]]])}
  test <- splits[[test_ind]]

  #Create a covariate list for stratification
  strat_vars <-
c("GAD65_IAA","GAD65_ZNT8","IA2A_ZNT8","IA2A_IAA_ZNT8","GAD65_IA2A_IAA_ZNT8",
"A1c_binary", "GLU120_binary")
  #Create a list for populating the plot titles
  strat_vars_title <- c("GAD65_IAA", "GAD65_ZnT8", "IA-2_ZnT8", "IA-2_IAA_ZnT8", "GAD65_IA-
2_IAA_ZnT8", "HbA1c_binary", "GLU120_binary")

  #Create a variable with threshold value for continuous covariates
  binary_cutoffs <- c("5.25 %","100 mg/dl")

  #Store the number of covariates being used for stratification
  n_vars <- length(strat_vars)

  #Apply for loop to rotate folds for cross-validation
  for(k in 1:n_vars) {

    m <- ifelse(k >= 6,k,0)

    #For the test fold, split the covariate being used for stratification into presence or absence
    test_1 <- test %>% filter(.data[[strat_vars[[k]]]] == 1)
    test_2 <- test %>% filter(.data[[strat_vars[[k]]]] == 0)

    #Create "surv" object
```

```
    surv_obj_train <- Surv(train$T_event, train$status)

    #Fit model using 'flexsurvreg" function with final multivariate AFT model described in section
4.4.2.4
    fit_train <- do.call(flexsurvreg, list(formula = surv_obj_train ~ GAD65_IAA + GAD65_ZNT8 +
IA2A_ZNT8 + IA2A_IAA_ZNT8 + GAD65_IA2A_IAA_ZNT8 + HbA1c_s    +log_GLU120_s, data = train,
dist = "Weibull"))

    #Check model fit with test fold
    fit_test_1 <- survfit(Surv(T_event, status) ~ 1, data = test_1)
    fit_test_2 <- survfit(Surv(T_event, status) ~ 1, data = test_2)

    surv1 <- summary(fit_train, newdata = test_1, type = "survival", B=50, tidy = TRUE)
    surv2 <- summary(fit_train, newdata = test_2, type = "survival", B=50, tidy = TRUE)

    varnames <- c("time", "surv", "lower", "upper")

    fit_test_1_data <- cbind(fit_test_1$time, fit_test_1$surv, fit_test_1$lower, fit_test_1$upper)
    fit_test_1_data <- as.data.frame(fit_test_1_data)
    names(fit_test_1_data) <- varnames
    fit_test_1_data$var <- as.factor(paste(strat_vars[k], ": 1"))

    fit_test_2_data <- cbind(fit_test_2$time, fit_test_2$surv, fit_test_2$lower, fit_test_2$upper)
    fit_test_2_data <- as.data.frame(fit_test_2_data)
    names(fit_test_2_data) <- varnames
    fit_test_2_data$var <- as.factor(paste(strat_vars[k], ": 0"))

    surv_1_avg <- surv1 %>%
      group_by(time) %>%
      summarise(mean_est = mean(est, na.rm=TRUE),
                mean_lcl = mean(lcl, na.rm=TRUE),
                mean_ucl = mean(ucl, na.rm=TRUE),
                var = as.factor(paste(strat_vars[k], ": 1")))

    surv_2_avg <- surv2 %>%
      group_by(time) %>%
      summarise(mean_est = mean(est, na.rm=TRUE),
                mean_lcl = mean(lcl, na.rm=TRUE),
                mean_ucl = mean(ucl, na.rm=TRUE),
                 var = as.factor(paste(strat_vars[k], ": 0")))
 #Generate plots to check goodness-of-fit
 if(m != k){
    p <-ggplot() +
      ggtitle(paste("Fold ",i, " Stratified by ", strat_vars_title[k], sep = "")) +
      geom_line(data = surv_1_avg, aes(x = time, y = mean_est, colour = var)) +
      geom_line(data = surv_2_avg, aes(x = time, y = mean_est, colour = var)) +

      geom_step(data = fit_test_1_data, aes(x = time, y = surv, colour = var), linetype = 3, size = 1)
+
      geom_step(data = fit_test_2_data, aes(x = time, y = surv, colour = var), linetype = 3, size = 1)
+
```

```
    geom_ribbon(data = fit_test_1_data, aes(x = time, ymin = lower, ymax = upper, colour = var, fill
= var), linetype = 0, alpha = .2, show.legend = FALSE) +
    geom_ribbon(data = fit_test_2_data, aes(x = time, ymin = lower, ymax = upper, colour = var, fill
= var), linetype = 0, alpha = .2, show.legend = FALSE) +

    geom_ribbon(data = surv_1_avg, aes(x = time, ymin = mean_lcl, ymax = mean_ucl, colour =
var, fill = var), linetype = 0, alpha = .2, show.legend = FALSE) +
    geom_ribbon(data = surv_2_avg, aes(x = time, ymin = mean_lcl, ymax = mean_ucl, colour =
var, fill = var), linetype = 0, alpha = .2, show.legend = FALSE) +

    xlab("Time from Derived BL (years)") +
    ylab("1 - Probability of T1D Diagnosis")

    #View goodness-of-fit plots
    p

  #Export cross-validation plots
  ggsave(paste("../deliv/figures/",i,strat_vars[k],".png", sep = ""), p, width = 16, height = 9, units =
"cm")
  }

  #Generate plot to check goodness-of-fit
   if(m == k){
    p <-ggplot() +
    ggtitle(paste("Fold ",i, " Stratified by ", strat_vars_title[k]," threshold of ",binary_cutoffs[m-5], sep
= "")) +
    geom_line(data = surv_1_avg, aes(x = time, y = mean_est, colour = var)) +
    geom_line(data = surv_2_avg, aes(x = time, y = mean_est, colour = var)) +

    geom_step(data = fit_test_1_data, aes(x = time, y = surv, colour = var), linetype = 3, size = 1)
+
    geom_step(data = fit_test_2_data, aes(x = time, y = surv, colour = var), linetype = 3, size = 1)
+

    geom_ribbon(data = fit_test_1_data, aes(x = time, ymin = lower, ymax = upper, colour = var, fill
= var), linetype = 0, alpha = .2, show.legend = FALSE) +
    geom_ribbon(data = fit_test_2_data, aes(x = time, ymin = lower, ymax = upper, colour = var, fill
= var), linetype = 0, alpha = .2, show.legend = FALSE) +

    geom_ribbon(data = surv_1_avg, aes(x = time, ymin = mean_lcl, ymax = mean_ucl, colour =
var, fill = var), linetype = 0, alpha = .2, show.legend = FALSE) +
    geom_ribbon(data = surv_2_avg, aes(x = time, ymin = mean_lcl, ymax = mean_ucl, colour =
var, fill = var), linetype = 0, alpha = .2, show.legend = FALSE)  +

    xlab("Time from Derived BL (years)") +
    ylab("1 - Probability of T1D Diagnosis")

    #View goodness-of-fit plots
    p
```

```
    #Export cross-validation plots
    ggsave(paste("../deliv/figures/",i,strat_vars[k],".png", sep = ""), p, width = 16, height = 9, units =
"cm")


    }
  }
}
```
```{r Cross-validation analysis on pediatric population (age < 12) as discussed in section 4.3.7.3}
#Set a seed value
set.seed(1)
#Assign age threshold of 12
age_thres <- 12
#Extract 50% of the pediatric population (age < 12) from the data as test set
ped_inds <- data$IDp[which(data$bAGE < age_thres)]
ped_inds_test <- sample(ped_inds,round(length(ped_inds)/2), replace = FALSE)
#Extract remaining data for model training
ped_inds_train <- setdiff(data$IDp,ped_inds_test)
#Prepare train and test data for cross-validation analysis
train <- data[ped_inds_train,]
test <- data[ped_inds_test,]
#Create "surv" object
surv_obj_train <- Surv(train$T_event, train$status)
#Fit model using 'flexsurvreg" function - final multivariate AFT model described in section 4.4.2.4
fit_train <- do.call(flexsurvreg, list(formula = surv_obj_train ~ GAD65_IAA + GAD65_ZNT8 +
IA2A_ZNT8 + IA2A_IAA_ZNT8 + GAD65_IA2A_IAA_ZNT8 + HbA1c_s +log_GLU120_s, data = train,
dist = "Weibull"))

#Test model fit with test data
fit_test <- survfit(Surv(T_event, status) ~ 1, data = test)
surv <- summary(fit_train, newdata = test, type = "survival", B=50, tidy = TRUE)

varnames <- c("time", "surv", "lower", "upper")

fit_test_data <- cbind(fit_test$time, fit_test$surv, fit_test$lower, fit_test$upper)
fit_test_data <- as.data.frame(fit_test_data)
names(fit_test_data) <- varnames
surv_avg <- surv %>%
    group_by(time) %>%
    summarise(mean_est = mean(est, na.rm=TRUE),
              mean_lcl = mean(lcl, na.rm=TRUE),
              mean_ucl = mean(ucl, na.rm=TRUE),
              )
#Generate goodness-of-fit plot
p <-ggplot() +
ggtitle(paste("Cross validation on pediatric population: Age < ",age_thres, sep = "")) +
geom_line(data = surv_avg, aes(x = time, y = mean_est)) +
geom_step(data = fit_test_data, aes(x = time, y = surv), linetype = 3, size = 1) +
geom_ribbon(data = fit_test_data, aes(x = time, ymin = lower, ymax = upper ), linetype = 0, alpha
= .2, show.legend = FALSE) +
```

```
geom_ribbon(data = surv_avg, aes(x = time, ymin = mean_lcl, ymax = mean_ucl), linetype = 0,
alpha = .2, show.legend = FALSE) +
#Add x and y labels
xlab("Time from Derived Baseline (years)") +
ylab("1 - Probability of T1D Diagnosis")
#view plot
p
#Export plot to "Figures" folder
ggsave(paste("../deliv/figures/ped_validation_",age_thres,"c.png", sep = ""), p, width = 16, height =
9, units = "cm")

```
```{r Cross-validation analysis on pediatric population (age < 12) as discussed in seciton 4.3.7.3 - C-
index table}
#Assign maximum year for c-index calculation
yrs_for_cindex <- 6
#Create a matrix to store c-index values
cindex_peds <- matrix(NA,nrow = 1, ncol = yrs_for_cindex)
#Use "survreg" with the final multivariate AFT model described in section 4.4.2.4 to compute c-index
fit_train_concordance<- survreg(Surv(T_event, status) ~ GAD65_IAA + GAD65_ZNT8 + IA2A_ZNT8 +
IA2A_IAA_ZNT8 + GAD65_IA2A_IAA_ZNT8 +  HbA1c_s +log_GLU120_s , data = train ,dist =
"weibull" )
#Compute c-index till six years with one-year increments
 for(q in 1:yrs_for_cindex){
     c_index_tmp <- concordance(object = fit_train_concordance, newdata = test, ymin = 0,ymax =
q)
     cindex_peds[1,q] <- c_index_tmp$concordance
   }
#Store the c-index values in a data frame
cindex_peds <- as.data.frame(cindex_peds)
#Create columns and rows names for c-index table
colnames(cindex_peds)<-c("year 1","year 2", "year 3", "year 4", "year 5", "year 6")
rownames(cindex_peds)<-c("Peds c-index")
#Export the c-index table
write.csv(cindex_peds, "../deliv/tables/cindex_peds.csv", row.names = TRUE)
```
```{r Model performance using time dependent ROC as discussed in section 4.3.7.1}
#Select data for time dependent ROC analysis and convert status to 0 and 1 to use predict function
data_for_ROC<-data %>%
  select(IDp,T_event,status,GAD65_IAA,GAD65_ZNT8 , IA2A_ZNT8 , IA2A_IAA_ZNT8 ,
GAD65_IA2A_IAA_ZNT8 ,  log_GLU0_s ,HbA1c_s ,log_GLU120_s ) %>%
  mutate(status=status-1)
#Identify missing covariate value
aa=which(complete.cases(data_for_ROC)==F)
#Fit the model using the "survreg" function
fit_weib<- survreg(Surv(T_event, status) ~ GAD65_IAA + GAD65_ZNT8 + IA2A_ZNT8 +
IA2A_IAA_ZNT8 + GAD65_IA2A_IAA_ZNT8 +  HbA1c_s +log_GLU120_s , data = data_for_ROC,dist =
"weibull" )
#Extract the linear predictor
data_for_ROC$lp <- predict(fit_weib, type = "lp")
#Define a helper function to evaluate at various time points
```

```
survivalROC_helper <- function(t) {
  survivalROC(Stime        = data_for_ROC$T_event,
              status       = data_for_ROC$status,
              marker       = data_for_ROC$lp,
              predict.time = t,
              method       = "KM")#,span = 0.25 * nrow(data_for_ROC)^(-0.20))
}
#Evaluate every 0.5 years
survivalROC_data <- tibble(t =seq(0.5,5.5,by=1)) %>%
  mutate(survivalROC = map(t, survivalROC_helper),
         ## Extract scalar AUC
         auc = map_dbl(survivalROC, magrittr::extract2, "AUC"),
         ## Put cut off dependent values in a data_frame
         df_survivalROC = map(survivalROC, function(obj) {
           as_data_frame(obj[c("cut.values","TP","FP")])
         })) %>%
  dplyr::select(-survivalROC) %>%
  unnest() %>%
  arrange(t, FP, TP) %>%
  mutate(FP=1-FP,TP=1-TP,auc=1-auc)
#Generate ROC curves
p_ROC <-ggplot(data = survivalROC_data,mapping = aes(x = FP, y = TP)) +
    ggtitle("Time dependent ROC analysis: model predictions on full analysis set")+
    geom_point() +
    geom_line() +
    geom_label(data = survivalROC_data %>% dplyr::select(t,auc) %>% unique,
               mapping = aes(label = sprintf("%.3f", auc)), x = 0.5, y = 0.5) +
    facet_wrap( ~ t, labeller = labeller(t = c("0.5" = "0.5 years", "1.5" = "1.5 years", "2.5" = "2.5
years", "3.5" = "3.5 years", "4.5" = "4.5 years", "5.5" = "5.5 years"))) +
    xlab("FPR")+
    ylab("TPR") +
    theme_bw() +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5),
          legend.key = element_blank(),
          plot.title = element_text(hjust = 0.5),
          strip.background = element_blank())
#View ROC curves
p_ROC
#Export plot
ggsave(paste("../deliv/figures/survival_ROC.png", sep = ""),p_ROC , width = 16, height = 13, units =
"cm")
```
```
```{r External validation using DAISY dataset as discussion in section 4.3.7.4}
#Create a "surv" object
surv_obj_train <- Surv(data$T_event, data$status)
#Train model - final multivariate AFT model described in section 4.4.2.4
fit_train <- do.call(flexsurvreg, list(formula = surv_obj_train ~ GAD65_IAA + GAD65_ZNT8 +
IA2A_ZNT8 + IA2A_IAA_ZNT8 + GAD65_IA2A_IAA_ZNT8  +HbA1c_s +log_GLU120_s, data = data,
dist = "Weibull"))
#Test the model with external data from DAISY study
test <- data_daisy
```
```

```
fit_test <- survfit(Surv(T_event, status) ~ 1, data = test)
surv <- summary(fit_train, newdata = test, type = "survival", B=50, tidy = TRUE)
varnames <- c("time", "surv", "lower", "upper")
fit_test_data <- cbind(fit_test$time, fit_test$surv, fit_test$lower, fit_test$upper)
fit_test_data <- as.data.frame(fit_test_data)
names(fit_test_data) <- varnames
surv_avg <- surv %>%
  group_by(time) %>%
  summarise(mean_est = mean(est, na.rm=TRUE),
            mean_lcl = mean(lcl, na.rm=TRUE),
            mean_ucl = mean(ucl, na.rm=TRUE))
#Generate plot to check goodness-of-fit
 p <-ggplot() +
    ggtitle("External Validation using DAISY dataset") +
    geom_line(data = surv_avg, aes(x = time, y = mean_est)) +
    geom_step(data = fit_test_data, aes(x = time, y = surv), linetype = 3, size = 1) +
    geom_ribbon(data = fit_test_data, aes(x = time, ymin = lower, ymax = upper ), linetype = 0,
alpha = .2, show.legend = FALSE) +
    geom_ribbon(data = surv_avg, aes(x = time, ymin = mean_lcl, ymax = mean_ucl), linetype = 0,
alpha = .2, show.legend = FALSE) +
    xlab("Time from Derived BL (years)") +
    ylab("1 - Probability of T1D Diagnosis")
    #View goodness-of-fit
    p

  #Export cross-validation plot
  ggsave(paste("../deliv/figures/Daisy_External_Validation.png", sep = ""), p, width = 16, height = 9,
units = "cm")
#Assign maximum year for c-index calculation
yrs_for_cindex <- 6
#Create a matrix to store c-index values
cindex_daisy <- matrix(NA,nrow = 1, ncol = yrs_for_cindex)

#Use "survreg" with the final multivariate AFT model described in section 4.4.2.4 to compute c-index
fit_train_concordance<- survreg(Surv(T_event, status) ~ GAD65_IAA + GAD65_ZNT8 + IA2A_ZNT8 +
IA2A_IAA_ZNT8 + GAD65_IA2A_IAA_ZNT8 +  HbA1c_s +log_GLU120_s , data = data ,dist =
"weibull" )
 #Compute c-index till six years with one-year increments
 for(q in 1:yrs_for_cindex){
     c_index_tmp <- concordance(object = fit_train_concordance, newdata = data_daisy, ymin =
0,ymax = q)
     cindex_daisy[1,q] <- c_index_tmp$concordance
   }
#Store the c-index values in a data frame
cindex_daisy <- as.data.frame(cindex_daisy)
#Create columns and rows names for c-index table
colnames(cindex_daisy)<-c("year 1","year 2", "year 3", "year 4", "year 5", "year 6")
rownames(cindex_daisy)<-c("Daisy c-index")
#Export the c-index table
write.csv(cindex_daisy, "../deliv/tables/cindex_daisy.csv", row.names = TRUE)
```