Simulating Screening for Risk of Childhood Diabetes: The Collaborative Open Outcomes tooL (COOL)

Mohamed Ghalwash, PhD[1], Eileen Koski, M.Phil[1], Riitta Veijola, MD, PhD[2], Jorma Toppari, MD, PhD[3], William Hagopian, MD, PhD[4], Marian Rewers, MD, PhD[5], Vibha Anand, PhD[6]

[1]Center for Computational Health, IBM Research, NY, USA; [2]Department of Pediatrics, PEDEGO Research Unit, University of Oulu and Oulu University Hospital, Oulu, Finland; [3]Department of Pediatrics, Turku University Hospital, Turku, Finland; [4]Pacific Northwest Research Institute, Seattle, WA, USA; [5]Barbara Davis Center for Diabetes, University of Colorado, Denver, CO, USA; [6]Center for Computational Health, IBM Research, Cambridge, MA;

Abstract *The Collaborative Open Outcomes tooL (COOL) is a novel, highly configurable application to simulate, evaluate and compare potential population-level screening schedules. Its first application is type 1 diabetes (T1D) screening, where known biomarkers for risk exist but clinical application lags behind. COOL was developed with the T1DI Study Group, in order to assess screening schedules for islet autoimmunity development based on existing datasets. This work shows clinical research utility, but the tool can be applied in other contexts. COOL helps the user define and evaluate a domain knowledge-driven screening schedule, which can be further refined with data-driven insights. COOL can also compare performance of alternative schedules using adjusted sensitivity, specificity, PPV and NPV metrics. Insights from COOL may support a variety of needs in disease screening and surveillance.*

1 Introduction

The Collaborative Open Outcomes tooL (COOL) was designed to help researchers and clinicians evaluate different potential screening schedules simulating their impact on optimal case identification rates prior to symptom onset using retrospective datasets. The initial application has been in type 1 diabetes (T1D), however, the tool can be used for a variety of conditions in which changes in biomarkers may signal progression to disease.

T1D is a complex, heterogeneous, autoimmune disorder in which insulin-producing pancreatic beta cells are mistakenly destroyed by the body's immune system. T1D has both genetic and familial components. Patients with T1D remain insulin dependent for life and are at high risk for serious long-term complications such as heart and kidney disease and diabetic retinopathy. For reasons that are not understood, T1D incidence rates have been rising dramatically. There is currently no cure or established prevention strategy for T1D and since newly diagnosed patients often present with diabetic ketoacidosis (DKA)[1], a life-threatening condition with potentially long-term consequences, research on prevention and early detection is increasingly critical.

There are known biomarkers for T1D, however, progression to diabetes is heterogeneous[2], with generally a 1 to 5 year horizon from birth, which has complicated efforts to establish a practical screening paradigm. This pre-symptomatic period presents an opportunity for both improved prediction and prevention. The NIH-funded TEDDY study[3] is currently investigating environmental determinants of T1D. Intervention trials of novel therapeutic agents focused on preventing or delaying (TrialNET and Innodia consortium[4]), increase the importance of early identification of at-risk patients for trial recruitment. These efforts, combined with the known clinical risk of DKA at onset, means that early population-based screening could make important contributions to both research and practice.

The T1DI study group[5] is a collaboration among IBM Research, JDRF and five of its academic partners: in the US (DAISY, DEW-IT)[6,7], Sweden(DiPiS)[8], Finland(DIPP)[9] and Germany (BABYDIAB/BABYDIET)[10]. These studies have followed newborns and young children from the general population who are at genetic or familial risk for up to a period of 15 years for islet autoimmunity development or until diagnosis of clinical diabetes. The goal of the T1DI Study is to use advanced machine learning and statistical methods to derive novel insights into T1D disease processes and to develop risk stratification methods from the integrated T1DI cohort of over 24,000 subjects (with ¿2.5 million visits), approximately 3% of whom eventually received a T1D diagnosis. The T1DI cohort represents the largest data set for any natural history T1D study to date and the study aims to inform both research and clinical practice.
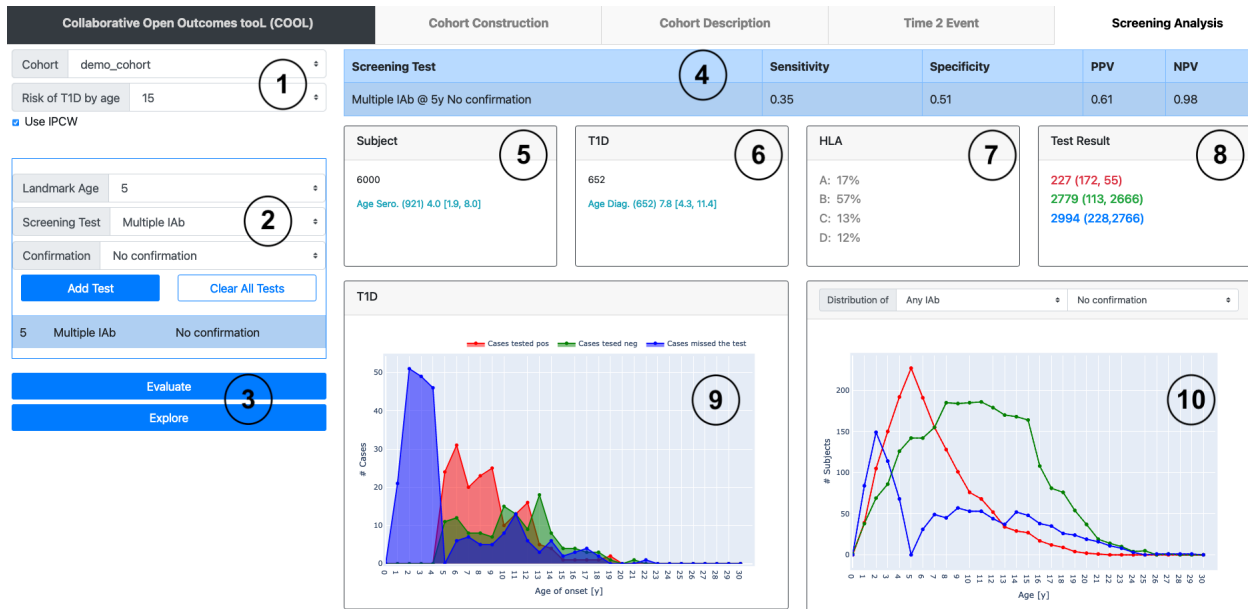
Figure 1: Screening for multiple biomarkers at age 5. COOL has 10 main panels - P1-P10. P1 defines the evaluation cohort. P2 defines the screening schedules. P3 allows the user to evaluate and explore the screening results. P4 shows the screening performance. P5-P7 show high level information about the cohort. P8 shows the number of subjects by category based on screening results. P9-P10 show further insights that could guide the user to refine the screening.

Cohorts, such as those included in this study, are expensive to maintain, and population-based screening also faces logistical obstacles in the case of a disease with heterogeneous onset patterns spanning many years. To complicate matters, geographical and socio-demographic factors also affect outcomes. The challenge for the practitioner and the researcher is to understand the nature of their own data enough to devise an effective screening paradigm.

Given the number and potential variety of variables involved, this poses a computational challenge for most clinicians and clinical researchers, and we knew of no tools available for this purpose. We developed COOL to help address this need, first applying it to our real-world case study of risk screening for T1D using the T1DI cohort. We have worked closely with our clinical collaborators to assure that the tool offers them the flexibility they need to iteratively ask "what-if" questions to determine how varying screening frequencies and ages can affect their ability to identify potential cases in a timely manner, and tuned to their specific population. This tool is currently being used actively in the T1DI Study in our efforts to define potential screening schedules for T1D.

2 Method

COOL is designed to allow users to simulate screening schedules for a given population to maximize identification of at risk patients based on known biomarkers. Four questions from clinical investigators in the T1DI collaboration will illustrate the utility of this highly configurable tool to gain new insights, and iteratively generate new hypotheses.

2.1 Research Question 1: When to screen subjects for biomarkers?

Clinical investigators start with a hypothesis such as: early detection of multiple biomarkers is likely to identify subjects with high risk of disease[11]. Once validated, they must determine when screening should occur. COOL provides an intuitive interface, shown in Figure 1, that helps users define, evaluate, and interactively explore the results of proposed screening schedules. Panels 1-3 allow the user to define a proposed schedule and panels 4-8 show the results for the available population. Panels 9-10 show additional insights that can guide the user to further refinements. These panels are described in more detail in the context of our results below.

COOL can also help evaluate and compare performance metrics for proposed screening schedules by providing functionality to address common challenges with retrospective cohorts. For example, a clinician may want to evaluate if screening for multiple biomarkers at the age of 5 years is sensitive enough to identify at-risk children. The question is domain knowledge driven, but the user may require special analytic techniques to utilize available data since not every subject may have been tested for known biomarkers at age 5. COOL can examine a subject's data for a biomarker sample within a 6-month window before and after the target age. If found, biomarker test results classify the subject as positive or negative. If a subject has no biomarker sample in the given window, the test is deemed missing. Subjects' biomarkers are thus categorized as: positive, negative, or no test. To evaluate screening performance given these categories, we compute cumulative sensitivity (sen), dynamic specificity (spc), positive predictive value (ppv), and negative predictive value (npv)[12]. These metrics are computed as:

$$ sen = \frac{tp}{tp + fn + np}, \qquad spc = \frac{tn}{tn + fp + nn}, \qquad ppv = \frac{tp}{tp + fp}, \qquad npv = \frac{tn}{tn + fn} $$

where tp is the number of diagnosed subjects diagnosed who tested positive for the biomarker. tn is the number of diagnosis-free subjects who tested negative. fp and fn are the numbers of false positives and false negatives, respectively. np is the number of diagnosed subjects with no test and nn is the number of diagnosis-free subjects with no test. Note that the denominator of sensitivity includes diagnosed subjects who had no test, i.e. np. This penalizes screening sensitivity if many subjects were diagnosed before the proposed screening age, but enables the metric to capture data on as many subjects as possible.

Right-censored data from subjects who are lost to follow up pose challenges in evaluating screening schedules because their outcomes are often unknown. However, excluding these subjects introduces bias in risk estimation for disease onset. To account for censoring, we use inverse probability weighting or IPCW mechanism[13], which assigns weights to subject data based on probability of censoring. It assigns larger weights to subjects with known diagnosis to offset those for whom it is unknown by the study end. For example, if the probability of censoring after 10 years of followup since birth is 0.2, it means that for any subject diagnosed by age 10, there are on average 4 subjects censored before age 10 and one subject followed through age 10. IPCW assigns a weight of 5 to the followed subject to offset the 4 censored subjects. Thus when computing tp, tn, fp, fn, np, and nn metrics, each subject i is assigned weight w_i. This built-in feature enables users to get unbiased results despite incomplete outcome data.

2.2 Research Question 2: How to improve screening performance?

The objective of screening is to improve early identification of at risk subjects, but which screening strategy will be the most effective: a specific biomarker? any biomarker? multiple biomarkers? Can we modify the testing age? or add an additional testing age to obtain more sensitive (and specific) screening?

COOL allows the user to configure screening strategies by age and compare across different ages. In addition, the tool is flexible enough that a user can configure age-specific screening tests and chain them, such as: screening for *any biomarkers* at age 2 and for *a specific biomarker* at age 5. Figure 2 shows how COOL computes the results of a chained screening strategy. The result is positive if the subject tests positive at the first or second age or both. Subjects with no tests are categorized as missing. All other subjects (both negative or one negative and one missing) are categorized as negative. The advantage of this approach is that it can be applied recursively to chain any number of different screening tests at different ages.

2.3 Research Question 3: Should we confirm positive screening tests?

It is common practice to confirm positive (biomarker) tests[3,6] to reduce the impact of false positives although the confirmation strategy may vary by test. Subjects who test positive typically require subsequent tests.[14] For example, when a subject tests positive for multiple biomarkers in the screening sample, should the confirmatory test focus on the initial positive biomarker(s) or all biomarkers? This question was raised to consider different confirmatory strategies to use in practice. COOL provides 5 strategies for confirmation, shown in Figure 3a. These policies are illustrated in Figure 3b for the screen for *any biomarker* strategy. The table shows sample status of four biomarkers in screening (blue columns) and confirmatory (green columns) samples. Subject $S1$ initially tested positive for two biomarkers,
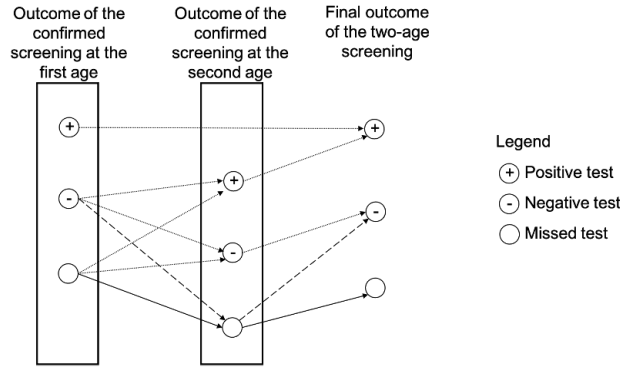
Outcome of the confirmed screening at the first age

Outcome of the confirmed screening at the second age

Final outcome of the two-age screening

Legend
⊕ Positive test
⊖ Negative test
◯ Missed test

Figure 2: Chain screening tests.

Subject	Biomarkers in the screen sample				Biomarkers in the confirmatory sample				Screen Results for any biomarker after Confirmation				
	B1	B2	B3	B4	B1	B2	B3	B4	No Confirmation	Confirm for any biomarker	Confirm for at least one initial +ve biomarker	Confirm for at least two initial +ve biomarkers	Confirm for all initial +ve biomarkers
S1	1	0	0	1	1	0	1	1	+	+	+	+	+
S2	1	0	0	1	1	0	0	0	+	+	+	-	-
S3	0	0	1	0	0	0	1	0	+	+	+	-	+
S4	0	1	0	0	1	0	0	0	+	+	-	-	-
S5	0	1	0	0	0	0	0	0	+	-	-	-	-
S6	0	0	0	0					-	NA	NA	NA	NA

✓ No confirmation
Confirm for any IAb
Confirm for at least one positive IAb
Confirm for at least two positive IAb
Confirm for all positive IAb

(a) Confirmation Options

(b) Various Confirmation Policies

Figure 3: (a) COOL provides five confirmation options. *No confirmation*: no confirmation required for positive screens. *Confirm for any IAb*: positive screen confirmed if any confirmatory sample biomarker is positive. *Confirm for at least one initial positive IAb*: positive screen confirmed only if at least one positive screening biomarker is positive in confirmatory sample. *Confirm for at least two positive IAb*: positive screen is confirmed only if at least two positive screening biomarkers are positive in confirmatory sample. *Confirm for all IAb*: the positive screen is confirmed only if all positive screening biomarkers are positive in confirmatory sample. (b) Blue columns show the status of four biomarkers, B1-B4, for subjects S1-S6. Green columns show biomarker status in the confirmatory sample. 1=positive(+), 0=negative(-). Screening results (+/-) are shown for each subject and confirmation strategy.

B1 and B4. If *no confirmation* strategy is used, then screening results would be positive for S1. The *Confirm for any biomarker* strategy confirms positive screen if any biomarker is positive in the confirmatory sample regardless of the biomarker status in the initial sample. Under this policy, S4 confirmed positive although only B2 was positive in the first sample and only B1 was positive in the second (confirmatory) sample. The *Confirm for at least one initial positive biomarker* strategy confirms positive only if at least one of the positive biomarkers in the first sample is confirmed in the confirmatory sample. For example, S2 would screen positive under this strategy because it had two positive biomarkers B1 and B4 in the first sample, and one of them, B1, was confirmed positive. Similarly, the *Confirm for at least two initial positive biomarkers* strategy confirms positive screen only if at least two biomarkers are confirmed positive. Finally, the *Confirm for all initial positive biomarkers* policy confirms positive screen only if all positive biomarkers in the first sample are confirmed. The user can create more confirmation policies that suit the problem domain (it is beyond the scope of this paper to explain how the user can configure the tool).

2.4 Research Question 4: Is the screening universal or population-specific?

A fundamental attribute of screening guidelines is whether or not they can be applied universally. For example, Finland has the highest incidence of T1D[15], while The United States has the second highest incidence in North America[16]. Is

the screening population-specific or universal? To answer this question, the same screening schedule can be applied to different cohorts with different characteristics, comparing performance in different populations. COOL provides a built-in cohort construction panel. The user can choose the evaluation cohorts in panel (1) in Figure 1 and compare screening performance on each chosen cohort. Whenever a screening schedule is evaluated, its results are appended to the table in panel (4) so users can compare performance across cohorts.

3 Results on Type 1 Diabetes

The following results are based on the T1DI cohort where islet autoantibodie were measured during 15-year follow-up or until T1D diagnosis.

3.1 Answer to Research Question 1: When to screen subjects for *multiple* biomarkers?

We use panel (1) to select the desired cohort and age, starting with age 5 years (Figure 4a), and to evaluate screening on the demo cohort for T1D risk by age 15. In panel (2) (Figure 4a) the user defines 3 screening parameters: target age, test and confirmation. Various screening strategies are implemented in COOL, such as multiple biomarkers, any biomarker, or specific combinations of biomarkers. These screening tests are shown in Figure 4b. For example, screening for *any GADA* tests the presence of Glutamic Acid Decarboxylase Autoantibodies in the blood sample. Screening for *Only IAA+IA2* tests for the presence of Insulin Autoantibodies (IAA) and Insulinoma Antigen-2 Autoantibody (IA-2A) in the absence of other antibodies (GADA). The screening tests are not hard coded but are configurable by the tool making it easy to add additional tests. In the current use case, we screen for multiple biomarkers. Let us start with the hypothesis of no confirmation to reduce the number of tests and based on the results we can refine these parameters.
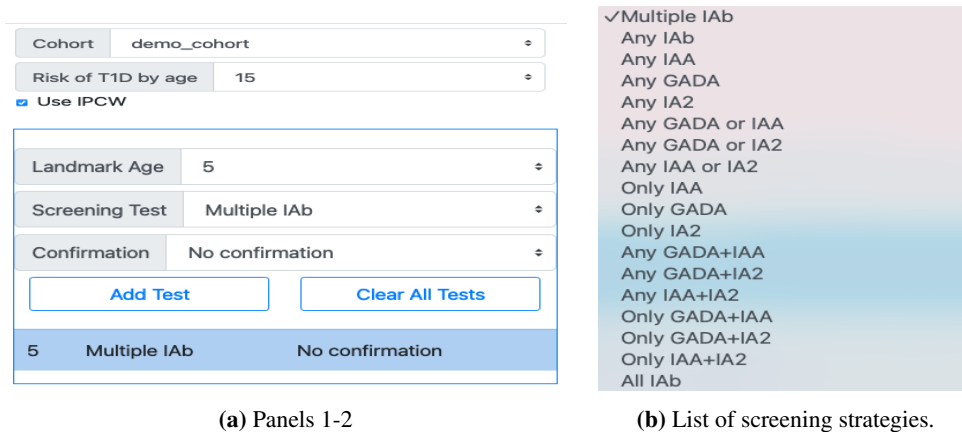

(a) Panels 1-2   (b) List of screening strategies.

Figure 4: (a) Panels 1-2: parameters to define a screening schedule. (b) Different screening strategies based on three biomarkers IAA, GADA, and IA2.

After the user defines the screening schedule they can evaluate and explore the results. The tool immediately computes the performance of the screening schedule to assess how likely subjects are to develop disease by the designated age. The table in panel (4) shows the results of the chosen screening as in Figure 5a. Sensitivity, specificity, PPV, and NPV metrics are computed for evaluation of the schedule.

Showing the accuracy of the screening test is valuable to estimate how likely the screened-positive subjects are to develop the disease. If the results are not satisfactory the user can alter the proposed schedule to obtain improved results. The tool displays helpful insights about the results to guide the user about possible changes to the screening schedule or strategy. Panel (8) (Figure 5b) shows the number of subjects screened positive (red group, n=227), screened negative (green group, n=2779), and not tested (blue group, n=2994). The blue group did not have biomarker samples at the screening age. In each subgroup, the panel shows the number of cases and controls between parentheses. This information highlights the fact that there are many subjects with missing tests. So, if the user wants to capture these subjects they may want to screen at a different age, but what age needs to be determined?

Screening Test	Sensitivity	Specificity	PPV	NPV
Multiple IAb @ 5y No confirmation	0.35	0.51	0.61	0.98

**Test Result**

227 (172, 55)
2779 (113, 2666)
2994 (228,2766)

**(a)** Panel 4: Screening Results   **(b)** Panel 8: #Subjects per group

Figure 5: (a) Screening results: sensitivity, specificity, ppv and npv. (b) Number of subjects who test positive (red), negative (green), missed the test (blue). The number of cases and controls are between parentheses subgroup.

Figure 6a shows the distribution of diagnosis by subgroup. Looking at the blue group reveals that many subjects developed type 1 diabetes earlier than 2-4 years and were missed by screening at age 5. Earlier screening might have captured these subjects. Panel (10) (Figure 6b) shows the distribution of multiple biomarkers at different ages by subgroup. Many subjects in the blue group developed multiple biomarkers by age 2. These two figures (6a and 6b) together indicate that screening at age 2 would capture many missed cases, which leads to the next analysis on how to improve the performance of screening by adding another test or changing the test.
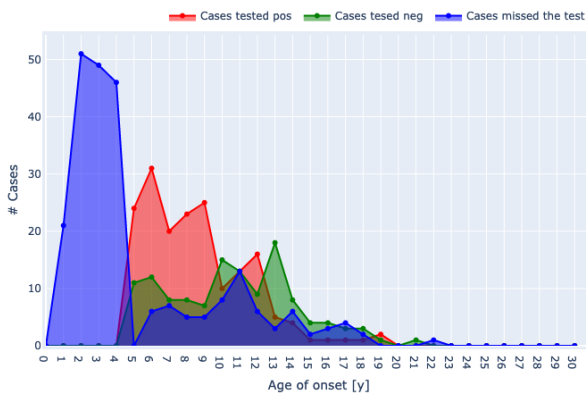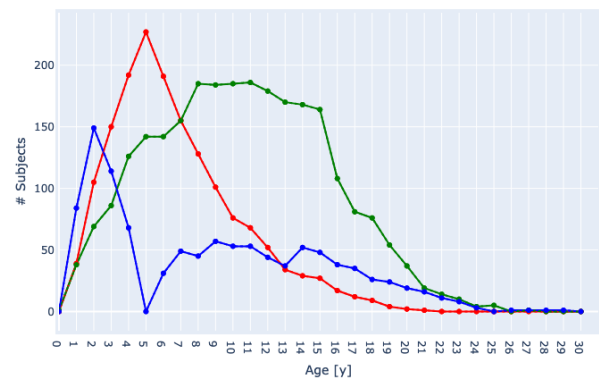
(a) Distribution of T1D onset age   **(b)** Distribution of IAb development age

Figure 6: Insights from evaluated screening schedule by sub-groups - tested positive (red), negative (green), and missed the test (blue). (a) Distribution of T1D onset age (b) Distribution of age for multiple IAb

3.2 Answer to Research Question 2: How to improve screening performance?

The previous section insights suggest that adding another test at an early age may improve screening performance. The user refines the screening schedule to two tests, at age 2 and 5 in panel (1), as shown in Figure 7a. The table in Figure 7b appends the results of the current screening strategy to the previous one for easy user comparison. From here, they can see that screening twice may be helpful - at ages 2 and 5 increases sensitivity from 0.35 to 0.51 and specificity from 0.51 to 0.56. The user can repeatedly chain different screening tests until improved results are obtained.

The results show that screening for multiple biomarkers twice at ages 2 and 5 is likely to identify subjects who will develop type 1 diabetes by age 15, but, can we improve performance further without an additional screening age, for example by exploring different testing strategies? COOL allows the user to explore different strategies as shown in Figure 4b. Figure 8 compares two screening policies at same ages but with different tests, screening for *any* biomarker versus screening for *multiple* biomarkers at ages 2 and 5. Screening for *any biomarker* has 13 points higher sensitivity than screening for *multiple biomarkers*, yet retaining similar positive predictive value. A user in a policy making capacity can perhaps use these results to begin to optimize the most effective screening schedule and tests for the intended population. In addition, the tool is flexible enough to allow the user to experiment. For example, screen for *any biomarkers* at age 2 and then screen for *a specific biomarker* at age 5 (Figure 9).

Landmark Age	5	⇕
Screening Test	Multiple IAb	⇕
Confirmation	No confirmation	⇕
Add Test		Clear All Tests
2	Multiple IAb	No confirmation
5	Multiple IAb	No confirmation

Screening Test	Sensitivity	Specificity	PPV	NPV
Multiple IAb @ 5y No confirmation	0.35	0.51	0.61	0.98
Multiple IAb @ 5y No confirmation Multiple IAb @ 2y No confirmation	0.51	0.56	0.65	0.98

(a) Chaining screen tests   **(b)** Results of a chain of screenings

Figure 7: (a) Chain screen tests. (b) Compare screening schedules.

Screening Test	Sensitivity	Specificity	PPV	NPV
Multiple IAb @ 2y No confirmation Multiple IAb @ 5y No confirmation	0.51	0.56	0.65	0.98
Any IAb @ 2y No confirmation Any IAb @ 5y No confirmation	0.64	0.54	0.43	0.99

Figure 8: Comparing different screen tests.

3.3 Answer to Research Question 3: Should we confirm positive screen?

As shown in Figure 10, various confirmation strategies are evaluated for screening for *any IAb*. The results indicate that confirming for *any IAb* has comparable sensitivity to *no confirmation* but a slightly higher PPV.

3.4 Answer to Research Question 4: Is the screening universal or population-specific?

The domain-knowledge driven screening schedule can be applied to each cohort, and results for each can be compared to assess whether the chosen screening schedule is applicable only to the selected cohort or universally, i.e. to all sub-cohorts in the available dataset. Furthermore, the user can evaluate them based on various domain attributes, e.g. male vs. female, high vs.low-risk HLA etc.

4 Discussion

We developed a pragmatic tool, (COOL), to evaluate efficacy of simulated screening strategies for chronic conditions. We applied COOL to screening for development of autoimmunity in progression to T1D onset. In this scenario, past research[17] had provided important insights from discovery of immune biomarkers in the pre-symptomatic phase of the disease[2]. These insights have yet to be translated into practice, specifically for screening children at risk for diabetes. However, the lack of screening in practice can be attributed to multiple factors, including unknowns of cost effectiveness and psychological readiness of those being screened, but the consequences are dire if not addressed expediently[18,19]. This report also stresses that the natural history of the disease is still not fully understood, particularly varying ages and rates of progression, and the influence of ethnicity and environment. A tool such as COOL can help users explore the underlying data from large natural history studies to unlock potential explanations.

Using the example of T1D in 15 years of follow up and by leveraging existing data on autoimmunity development from the T1DI study group[5], we have shown how COOL can simulate population-level screening. We have shown four practical applications of this tool for evaluating screening strategies. A user can craft one or more competing "knowledge-driven" screening strategies. Then the user can refine these strategies in a data-driven way by exploring the impact of underlying population distributions. These explorations can lead to further insights to improve risk screening performance, e.g. to maximize "at-risk catchment". Screening strategies can be further improved by "chaining" different screening tests by screening age, testing sequence, or from what may be otherwise known in the domain.

Screening Test	Sensitivity	Specificity	PPV	NPV
Any IAb @ 2y No confirmation Multiple IAb @ 5y No confirmation	0.57	0.55	0.54	0.98

Figure 9: Various screen tests at different ages.

Different strategies can be compared and a host of other alternative strategies devised for further exploration. Based on previously mentioned comparisons, a user can fine tune strategies for intended risk screening, e.g. universal or population specific. Within the framework of strategy development, a user can refine performance for varying time horizons (e.g T1D onset in 5, 10, or 15 years) as the tool accounts for right censored data in an unbiased way. Lastly, the built-in workbench capabilities of the tool retains alternatively evaluated strategies, which a user can recall later and further refine, for example when more data become available. The tool can also do a brute force search (evaluate all possible combinations configured) to find the optimal strategy.

In many ways the T1DI cohort is a unique dataset for understanding screening requirements, and for development of enabling tools. Its large sample size, variations in follow-up schedules among constituent studies and recruitment of newborns from both the general population and children and adolescents with family history[6–10] have helped with experimentation on the domain problem as well as the tool design. The data lends itself to understanding variations in population-level screening policies to yield early results. This may be important because large scale screening studies in the US and EU are currently ongoing[20,21]. The TEDDY study[3], which has rigorously followed at-risk children from young ages for development of autoimmunity and T1D is currently ongoing. There is no doubt TEDDY will shed more light on currently unknown disease processes and consequently inform screening policies[22] but it will not reach the 15 year follow up mark until 2024. In the interim, as we have shown, evaluation of competing screening policies with existing T1D cohorts may help in the T1D domain. All stakeholders such as clinicians, policymakers, patient families may be able to pose pertinent questions, generate new hypothesis to test in-silico with COOL and get rapid answers. These advances will ultimately pave the way for faster adoption of screening in practice.

COOL differs from other tools in many respects. It is a general purpose and highly configurable evaluation tool, with built-in cohort selection and exploration capabilities that gives the user guidance based on underlying data and questions. Similar tools have generally focused on one question at a time, for example, risk prediction in a given time horizon[23], or evaluating drug design metrics[24] or a web-based application for health-screening[25,26] and have rarely sought to add functionality such as "chaining" alternate or competing strategies for risk evaluation and, to the best of our knowledge, none has sought to retain and compare all user-defined strategies in one tool.

However, as with all work, this tool also has some limitations. We do not address evaluation of repeated screening (or monitoring) strategies for a single patient (personalized), though we aim to investigate it in the future using state of

Screening Test	Sensitivity	Specificity	PPV	NPV
Any IAb @ 2y No confirmation Any IAb @ 5y No confirmation	0.64	0.54	0.43	0.99
Any IAb @ 2y Confirm for any IAb Any IAb @ 5y Confirm for any IAb	0.64	0.55	0.48	0.98
Any IAb @ 2y Confirm for at least one positive IAb Any IAb @ 5y Confirm for at least one positive IAb	0.42	0.55	0.48	0.96
Any IAb @ 2y Confirm for at least two positive IAb Any IAb @ 5y Confirm for at least two positive IAb	0.49	0.55	0.48	0.97
Any IAb @ 2y Confirm for all positive IAb Any IAb @ 5y Confirm for all positive IAb	0.49	0.55	0.48	0.97

Figure 10: Evaluation of different confirmation strategies.

the art machine-learning (e.g. reinforcement learning) approaches and incorporate it into the tool. In this work, we also specifically aimed at risk screening policies in the pre-symptomatic period of a disease. As currently there is no readily available data on cost or yield, the cost effectiveness of screening (to prevent disease or cost of testing) was not evaluated. However, we believe that when these data[20,21] become available, COOL can be enhanced to attach cost to the results of a screening test for evaluating cost-effectiveness

5   Conclusion

In conclusion, the present analysis shows that Collaborative Open Outcomes tooL (COOL) can unlock invaluable datasets, particularly from natural history studies of chronic conditions, to evaluate the efficacy and propel adoption of screening programs and potentially even inform current clinical practices. One example may be in a similar autoimmune condition - inflammatory bowel disease[27], where there are known bio-markers but no general population screening in practice. The tool's capabilities can be applied to many forms of real world evidence such as data from electronic health records from hospitals, clinics and physician practices, as well as disease registries. A tool such as COOL can greatly facilitate exploration and evaluation of potential screening paradigms by clinical staff, using available data, which can have important implications for both research and clinical practice.

6   Acknowledgements

References

1. Duca LM, Reboussin BA, Pihoker C, Imperatore G, Saydah S, Mayer-Davis E, et al. Diabetic ketoacidosis at diagnosis of type 1 diabetes and glycemic control over time: The SEARCH for diabetes in youth study;20(2):172–179.

2. Insel RA, Dunne JL, Atkinson MA, Chiang JL, Dabelea D, Gottlieb PA, et al. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. Diabetes care. 2015;38(10):1964–1974.

3. Hagopian WA, Erlich H, Lernmark Å, Rewers M, Ziegler AG, Simell O, et al. The Environmental Determinants of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. Pediatric diabetes. 2011;12(8):733–743.

4. Herold KC, Bundy BN, Long SA, Bluestone JA, DiMeglio LA, Dufort MJ, et al. An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. New England Journal of Medicine. 2019;381(7):603–613.

5. Anand V, Liu B, Ghalwash M, Koski E, Ng K, Dunne JL, et al.. 1345-P: Presymptomatic and Clinical T1D in The US, Sweden, and Finland: Joint Analysis of Four Birth Cohort Studies. Am Diabetes Assoc; 2019.

6. Rewers M, Bugawan T, Norris J, Blair A, Beaty B, Hoffman M, et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). Diabetologia. 1996;39(7):807–812.

7. Wion E, Brantley M, Stevens J, Gallinger S, Peng H, Glass M, et al. Population-wide Infant Screening for HLA-Based Type 1 Diabetes Risk via Dried Blood Spots from the Public Health Infrastructure. Annals of the New York Academy of Sciences. 2003;1005(1):400–403.

8. Elding Larsson H. A Swedish approach to the prevention of type 1 diabetes. Pediatric diabetes. 2016;17:73–77.

9. Kupila A, Muona P, Simell T, Arvilommi P, Savolainen H, Hämäläinen AM, et al. Feasibility of genetic and immunological prediction of type I diabetes in a population-based birth cohort. Diabetologia. 2001;44(3):290–297.

10. Ziegler AG, Hummel M, Schenker M, Bonifacio E. Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes: the 2-year analysis of the German BABYDIAB Study. Diabetes. 1999;48(3):460–468.

11. Ziegler AG, Rewers M, Simell O, Simell T, Lempainen J, Steck A, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. Jama. 2013;309(23):2473–2479.

12. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. BMC medical research methodology. 2017;17(1):1–19.

13. Vock DM, Wolfson J, Bandyopadhyay S, Adomavicius G, Johnson PE, Vazquez-Benitez G, et al. Adapting machine learning techniques to censored time-to-event health record data: A general-purpose approach using inverse probability of censoring weighting. Journal of biomedical informatics. 2016;61:119–131.

14. Maxim LD, Niebo R, Utell MJ. Screening tests: a review with examples. Inhalation toxicology. 2014;26(13):811–828.

15. Mobasseri M, Shirmohammadi M, Amiri T, Vahed N, Fard HH, Ghojazadeh M. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. Health promotion perspectives. 2020;10(2):98.

16. Patterson CC, Karuranga S, Salpea P, Saeedi P, Dahlquist G, Soltesz G, et al. Worldwide estimates of incidence, prevalence and mortality of type 1 diabetes in children and adolescents: Results from the International Diabetes Federation Diabetes Atlas. Diabetes research and clinical practice. 2019;157:107842.

17. Bingley PJ, Bonifacio E, Ziegler AG, Schatz DA, Atkinson MA, Eisenbarth GS. Proposed guidelines on screening for risk of type 1 diabetes. Diabetes Care. 2001;24(2):398–398.

18. Becker D, Insel R. Screening, staging, and naming of presymptomatic type 1 diabetes. Pediatric diabetes. 2017;19(1):7–10.

19. Narendran P. Screening for type 1 diabetes: are we nearly there yet? Diabetologia. 2019;62(1):24–27.

20. McQueen RB, Rasmussen CG, Waugh K, Frohnert BI, Steck AK, Yu L, et al. Cost and cost-effectiveness of large-scale screening for type 1 diabetes in Colorado. Diabetes care. 2020;43(7):1496–1503.

21. Ziegler AG, Kick K, Bonifacio E, Haupt F, Hippich M, Dunstheimer D, et al. Yield of a public health screening of children for islet autoantibodies in Bavaria, Germany. Jama. 2020;323(4):339–351.

22. Bonifacio E, Weiß A, Winkler C, Hippich M, Rewers MJ, Toppari J, et al. An Age-Related Exponential Decline in the Risk of Multiple Islet Autoantibody Seroconversion During Childhood. Diabetes Care. 2021.

23. Senders JT, Staples P, Mehrtash A, Cote DJ, Taphoorn MJ, Reardon DA, et al. An online calculator for the prediction of survival in glioblastoma patients using classical statistics and machine learning. Neurosurgery. 2020;86(2):E184–E192.

24. Empereur-Mot C, Zagury JF, Montes M. Screening explorer–An interactive tool for the analysis of screening results. Journal of chemical information and modeling. 2016;56(12):2281–2286.

25. Ooi CY, Ng CJ, Sales AE, Lim HM. Implementation Strategies for Web-Based Apps for Screening: Scoping Review. Journal of medical Internet research. 2020;22(7):e15591.

26. Partanen VM, Anttila A, Heinävaara S, Pankakoski M, Sarkeala T, Bzhalava Z, et al. NordScreen–an interactive tool for presenting cervical cancer screening indicators in the Nordic countries. Acta Oncologica. 2019;58(9):1199–1204.

27. Rosen MJ, Dhawan A, Saeed SA. Inflammatory bowel disease in children and adolescents. JAMA pediatrics. 2015;169(11):1053–1060.