Diabetes Care
American Diabetes Association

Interlaboratory Evaluation of Multiplex Autoantibody Assay Performance in the Islet Autoantibody Standardization Program 2024 Workshop

Ilaria Marzinotto, David L. Pittman, Peter Achenbach, Beena Akolkar, Clive H. Wasserfall, Vito Lampasona, and Anna E. Long

Diabetes Care 2026;49(00):1–9 | https://doi.org/10.2337/dc25-1326

Islet Autoantibody Standardization Program 2024: Interlaboratory Evaluation of Multiplex Islet Autoantibody Assays

- Promising performance for islet autoantibody detection
- Further optimization needed to increase specificity and concordance across laboratories

Advantages & Challenges

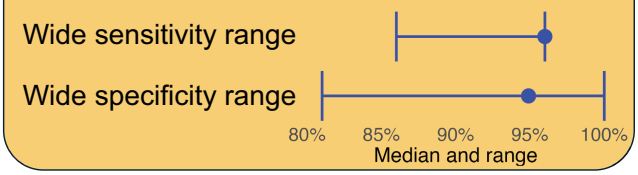
✓ One test, multiple autoAbs
✓ Nonradioactive
✓ Low sample volume
✓ High throughput

⚠ Performance variability across laboratories
⚠ High specificity needed

Study Design

- Blinded test:
  - 50 new-onset type 1 diabetes samples
  - 98 blood donor samples
- 11 single-readout assays (combined islet autoantibodies)
- Eight individual autoantibody readout assays

Results



Strategies to Improve Performance

- Standardize thresholds at 99% specificity
- Two-step testing strategies using alternative assay formats for retesting positive samples

ARTICLE HIGHLIGHTS

- **Why did we undertake this study?**
  Assays measuring multiple islet autoantibodies are increasingly being used in type 1 diabetes research and clinical practice. Interlaboratory evaluation is essential to understand their performance and support standardization efforts.

- **What is the specific question we wanted to answer?**
  We evaluated multiplex assays' sensitivity, specificity, and threshold standardization potential across diverse formats and laboratories.

- **What did we find?**
  Multiplex assays demonstrated high discriminatory capability (median receiver operating characteristic area under the curve analysis 0.98) with standardized thresholds achieving 93–97% sensitivity. Bridge-ELISA platforms showed promise for universal cutoff implementation, while substantial interlaboratory variation persisted in other formats.

- **What are the implications of our findings?**
  Multiplex assays showed promising performance for islet autoantibody detection. Our study suggests that continued standardization efforts and sequential testing strategies may be needed to optimize specificity for population-wide applications.

Interlaboratory Evaluation of Multiplex Autoantibody Assay Performance in the Islet Autoantibody Standardization Program 2024 Workshop

Ilaria Marzinotto,[1] David L. Pittman,[2] Peter Achenbach,[3] Beena Akolkar,[4] Clive H. Wasserfall,[2] Vito Lampasona,[1] and Anna E. Long[5]

https://doi.org/10.2337/dc25-1326

OBJECTIVE

The Islet Autoantibody Standardization Program (IASP) evaluates type 1 diabetes autoantibody assays through international interlaboratory comparison studies. This report describes 2024 IASP workshop results for multiplex assays that simultaneously measure multiple islet autoantibodies, increasingly adopted for population screening programs.

RESEARCH DESIGN AND METHODS

Participating laboratories analyzed coded sera from individuals with new-onset type 1 diabetes (*n* = 50), blood donors with no diabetes (*n* = 98), and replicate samples (*n* = 3). Performance was assessed using sensitivity, specificity, adjusted sensitivity at 95%, 99%, and 100% specificity, area under the receiver operating characteristic curve analysis (ROC-AUC), and partial ROC-AUC at 95% specificity (pAUC95).

RESULTS

Fifteen laboratories contributed results from 19 multiplex assays using five formats: antibody-dependent agglutination PCR (ADAP) (*n* = 6), bridge-ELISA (*n* = 8), electrochemiluminescence (*n* = 2), luciferase immunoprecipitation system (LIPS) (*n* = 1), and protein A luciferase/solid-phase capture LIPS (*n* = 2). Median ROC-AUC was 0.98 across all platforms, with a pAUC95 of 0.048 against a theoretical maximum of 0.05. Laboratory-reported sensitivity ranged from 86% to 96% with specificity 81% to 100%. Standardized adjusted sensitivity at 99% specificity thresholds eliminated most false positives without compromising sensitivity, achieving 93–97% adjusted sensitivity. Bridge-ELISA platforms demonstrated potential for universal cutoff standardization, while ADAP showed substantial interlaboratory threshold variation.

CONCLUSIONS

Multiplex assays demonstrated good overall performance for detecting islet autoantibodies and discriminating case subjects from control subjects. However, interlaboratory variability and platform-specific recognition patterns necessitate further standardization efforts. While not directly tested in this workshop, the results suggest that sequential testing strategies using complementary platforms may be necessary to achieve the stringent specificity required when these assays are used in a screening context.

[1]*San Raffaele Diabetes Research Institute, IRCCS Ospedale San Raffaele, Milan, Italy*
[2]*Department of Pathology, University of Florida, Gainesville, FL*
[3]*Institute of Diabetes Research, Helmholtz Munich, German Research Center for Environmental Health, Neuherberg, Germany*
[4]*Division of Diabetes, Endocrinology, and Metabolic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD*
[5]*Diabetes and Metabolism, Translational Health Sciences, University of Bristol, Bristol, U.K.*

Corresponding authors: Ilaria Marzinotto, marzinotto.ilaria@hsr.it, or Anna E. Long, anna.long@bristol.ac.uk

Received 5 June 2025 and accepted 7 November 2025

This article contains supplementary material online at https://doi.org/10.2337/figshare.30974740.

Autoantibodies that target pancreatic islet autoantigens, like glutamic acid decarboxylase 65 (GAD65, i.e., GADA), insulinoma-associated antigen-2 (IA-2A), zinc transporter 8 (ZnT8A), and insulin (IAA), serve as important biomarkers for identifying individuals at-risk of type 1 diabetes (T1D), long before clinical symptoms emerge (1,2).

Historically, islet autoantibodies have been measured using separate radiobinding assays (RBAs) for each specificity. While RBAs achieved high sensitivity and specificity, they present significant operational challenges for widespread implementation. These assays are laborious, require substantial sample volumes (typically 20–50 μL per autoantibody), cannot be multiplexed, and rely on radioactive isotopes that face increasing regulatory restrictions (3–5). Moreover, sequential testing of individual autoantibodies through separate assays introduces interassay variability and limits throughput capacity.

As the field moves toward broader screening initiatives and comprehensive autoantibody profiling for risk stratification, these limitations have driven the development of alternative nonradioactive multiplex formats. These assays enable the simultaneous measurement of multiple islet autoantibodies (GADA, IA-2A, ZnT8A, and, in some formats, IAA) in a single analytical run (6–8). This integrated approach offers several distinct advantages over traditional single antibody testing methods: it reduces sample volume requirements, minimizes interassay variability, improves both qualitative and quantitative result consistency, and substantially enhances the efficiency and cost-effectiveness of comprehensive autoantibody profiling (2) and identification of individuals at highest risk for disease progression (i.e., positive for two or more autoantibodies).

Significant challenges persist in standardizing islet autoantibody measurements across different laboratories and platforms (9,10). To address these issues, the Islet Autoantibody Standardization Program (IASP) promotes standardization and improvement of islet autoantibody assays (11–15) through specialized workshops. IASP workshops consist of interlaboratory comparison studies where participating laboratories receive and test blinded serum samples from new-onset T1D cases and blood donors. Results are then centrally decoded and analyzed to evaluate assay performance across different laboratory settings.

This report presents the results of the IASP 2024 interlaboratory comparison study, with particular emphasis on multiplex assays. These results were preliminarily presented at the IASP 2024 workshop during the 20th Immunology of Diabetes Society (IDS) Congress.

Although the current IASP framework was not specifically designed to evaluate screening assays and their predictive capacity in presymptomatic individuals, the IASP 2024 interlaboratory comparison study nevertheless generated valuable data on contemporary multiplex islet autoantibody assays. These findings can inform both future assay optimization efforts and the implementation of more effective screening strategies.

RESEARCH DESIGN AND METHODS

Study Design

This study evaluated multiplex T1D autoantibody assay performance through an interlaboratory comparison using sera from patients with newly diagnosed T1D (collected within 14 days of initial insulin treatment from multiple international centers) and blood donors without a diabetes diagnosis collected in the United States. All samples were obtained with written informed consent and approval from local ethics committees, adhering to the ethical principles for medical research involving human subjects as in the Declaration of Helsinki (16).

The 2024 sample set included blinded specimens from 50 individuals with new-onset T1D, 98 blood donor controls, and three replicates of a selected T1D serum (IDS382) with GADA, IA-2A, ZnT8A, and IAA. Additional samples comprised six serial dilutions of the IgG1 anti-insulin mouse monoclonal antibody (mAb) HUI-018 (17) in normal human serum (156, 20, 5, 1.2, 0.6, and 0 ng/mL) and six National Institute of Diabetes and Digestive Kidney Diseases consortium standards, prepared from GADA-positive (493, 228, 92, 13, 4.8, and 0 NIDDK standard [DK] units per mL) and IA-2A–positive (235, 106, 42, 5.8, 2, and 0 DK units per mL) sera (3).

Participants' Demographics

Individuals with T1D had a median age of 13 years (range 4–43) and included 17 females and 25 males (demographic data missing for eight participants because of anonymization); 37 were White, 1 was Black, 2 were Hispanic, 1 was Asian, and 1 was of undisclosed ancestry. Blood donors without diabetes showed a median age of 34 years (range 20–40) and included 31 females and 67 males; 15 were White, 82 were Black, and 1 was of other ethnic backgrounds.

Data Analysis

Participating laboratories submitted assay protocol details, raw data, arbitrary units, and scores using standardized Excel reporting templates. All analyses were performed using the R statistical software (18).

Assay sensitivity was calculated as the percentage of T1D sera testing positive. Specificity was defined as the percentage of blood donor sera testing negative. For antibody-dependent agglutination PCR (ADAP) and electrochemiluminescence (ECL), which discriminate all four autoantibody reactivities, overall sensitivity and specificity were calculated by cumulating individual antibody scores (i.e., a sample was considered positive if any of the four measured autoantibodies was positive). To account for different diagnostic thresholds, we calculated each assay's adjusted sensitivity at 95% (AS95), 99% (AS99), and 100% (AS100) specificity by setting the positivity threshold at the 95th, 99th, and 100th percentiles, respectively, of values measured in blood donor samples.

Assay performance was evaluated using threshold-independent receiver operating characteristic (ROC) analysis. The area under the ROC curve (ROC-AUC) and the partial ROC-AUC at 95% specificity (pAUC95) were calculated using the pROC R package (19), while the median ROC curve was obtained using the cutpointr R package (20) (Supplementary Extended Methods).

For ADAP and ECL assays measuring individual antibodies, a cumulative ROC-AUC was calculated using logistic regression based on single autoantibody reactivities (Supplementary Extended Methods).

Fold over cutoff values were stratified into four logarithmic intensity tiers (<1, 1–10, 10–100, and >100) to accommodate the wide dynamic range of autoantibody signals across different platforms. This binning strategy enables classification of samples as negative (<1), weakly positive (1–10), moderately positive (10–100),

or strongly positive (>100), facilitating comparison of signal intensities across assays with differing absolute readout ranges.

Consensus scores were calculated based on agreement in >50% of assays measuring each autoantibody within each assay format.

RESULTS

Multiplex Assay Formats and Laboratory Participation

Fifteen laboratories contributed results from 19 multiplex assays, distributed across five detection platforms: five ADAP (Enable Biosciences Inc.); eight bridge-ELISA (RSR Limited); two ECL, one developed in-house and one from Meso Scale Diagnostics LLC); one luciferase immunoprecipitation system assay (LIPS) (in-house), and two assays based on solid-phase capture using plates coated with protein A (protein A luciferase [PAL] and solid-phase capture LIPS [scLIPS]) (in-house). All assays used nonradioactive tracers, with 21% (n = 4) being in-house (LIPS, PAL/scLIPS, and ECL) and 79% (n = 15) being commercial (ADAP, bridge-ELISA, ECL). The multiplex assays used different strategies for combining islet autoantibody measurements. Bridge-ELISA, LIPS, and PAL/scLIPS combined multiple antigens with a single readout, while ADAP and ECL measured individual autoantibodies separately within a multiplexed format. ADAP measures individual antibodies by running separate real-time PCR reactions after a common serum incubation and agglutination step. ECL measures distinct antibodies within a single well after capture of differently tagged antigens to specific spatial locations in each plate well, allowing simultaneous detection of different autoantibodies through spatially ECL signals. Major characteristics are reported in the Supplementary Material, Excel file.

Sensitivity and Specificity

Multiplex assays demonstrated consistently high sensitivity and specificity across different analyte combinations (Table 1 and Supplementary Fig. 1). Notably, one control sample (R587009) was identified as multiple antibody–positive by all assay formats and was subsequently excluded from specificity calculations.

Among assays combining antigens in one reaction and reporting a single readout (n = 11), the highest observed sensitivity and specificity were 96% and 100%, respectively. Bridge-ELISA assays multiplexing GADA and IA-2A (n = 2) showed 93% median sensitivity (range 92–94%) and 98.4% median specificity (range 97.9–99%). When bridge-ELISA assays included ZnT8A (GADA/IA-2A/ZnT8A, n = 6), sensitivity increased to 96% (range 94–96%) with a specificity of 98% (range 94.8–100%).

The LIPS assay showed 96% sensitivity and 94.9% specificity, while PAL and scLIPS assays demonstrated a median sensitivity of 92% (range 88–96%) and specificity of 97.9%.

For ADAP and ECL, which multiplex and discriminate all four autoantibody specificities, overall sensitivity and specificity were calculated by cumulating individual antibody scores. ADAP (n = 6) exhibited a median sensitivity of 96% (range 86–96%) and specificity of 92.3% (range 81.4–94.8%), while ECL (n = 2) showed a median sensitivity of 96% and specificity of 95.3% (range 92.8–97.9%).

Threshold-Independent ROC-AUC Analysis

To assess multiplex assay performance beyond laboratory-defined thresholds, we calculated ROC-AUC values and the pAUC95. The pAUC95 provides a more clinically relevant performance indicator than full ROC-AUC because it focuses on the high-specificity region of the ROC curve, where false-positive rates are reduced (Table 1 and Supplementary Fig. 2).

The median full ROC-AUC was 0.98 (range 0.96–0.99) across all submitted multiplex assays. No significant differences were found in ROC-AUC values across different assays and formats ($P > 0.05$ for all pairwise comparisons and comparisons against the median, after adjustment for multiple comparisons).

The median pAUC95 was 0.048 (range 0.044–0.049) against a theoretical maximum of 0.05. The only significant difference was observed for the 152-ADAP assay pAUC95, which was lower than the median and other ECL and ADAP assays ($P \leq 0.049$), although this difference was no longer significant after multiple comparison adjustment.

Autoantibody Level in Discordant Samples

Antibody levels, expressed as fold change over laboratory-specific cutoffs and stratified into four intensity tiers (<1, 1–10, 10–100, and >100 fold over cutoff), revealed distinct patterns across assay formats (Fig. 1 and Supplementary Fig. 3). After excluding samples with complete concordance (61 control subjects and 40 T1D case subjects) (Fig. 1), discordant T1D samples exhibited variable antibody levels across formats. The most notable pattern was the predominance of low-intensity IAA signals in ADAP and ECL assays for these samples. GADA, IA-2A, and ZnT8A levels tended to show more frequently higher signal intensities in these platforms. In contrast, bridge-ELISA and PAL/scLIPS formats typically produced intermediate to low binding signals for these samples. Control samples uniformly demonstrated signal intensities in the lowest tier across all assays and formats, with the notable exception of sample R587009.

DK and HUI-018 (IAA) standard samples exhibited expected concentration-dependent positivity and antibody level patterns, with ADAP and ECL formats revealing the appropriate detection of single autoantibodies corresponding to known reference material composition. The IDS382 replicate samples demonstrated complete interassay reproducibility.

Autoantibody Number in Discordant Samples

Among discordant T1D cases, individual samples tested positive in 80–100% of submitted assays, indicating that even samples lacking complete interassay agreement were detected by most platforms (Fig. 2). ADAP and ECL formats, which provide single autoantibody resolution, revealed notable differences in the number of autoantibodies detected by individual assays, ranging from none or single autoantibody reactivity to detection of all four measured autoantibodies in the same samples (Supplementary Fig. 4). Bridge-ELISA, LIPS, and PAL/scLIPS platforms showed comparable overall sensitivity for detecting autoantibodies in these T1D samples, although they provide binary positive/negative results without autoantibody enumeration.

Consensus Autoantibody Positivity Patterns Across Assay Formats

To evaluate autoantibody detection consistency across multiplex assay formats, we applied a four-tier consensus scoring system to both multiplex and single autoantibody assays. Samples were classified

Table 1—Performance of multiplex assays in the IASP 2024 workshop

Laboratory ID	Sensitivity (%)	Specificity (%)	AS95 (%)	AS99 (%)	AS100 (%)	ROC-AUC	pAUC95
ADAP: GADA/IA-2A/ZnT8A/IAA							
123	96.0	81.4	96.0	94.0	94.0	0.97	0.048
152	86.0	94.8	88.0	86.0	84.0	0.96	0.042
156	96.0	92.8	96.0	96.0	94.0	0.97	0.048
1603	96.0	93.8	98.0	96.0	92.0	0.98	0.048
1803	96.0	91.8	98.0	94.0	94.0	0.98	0.048
2404	94.0	91.8	96.0	94.0	94.0	0.97	0.047
Median	96.0	92.3	96.0	94.0	94.0	0.97	0.048
(range)	(86.0–96.0)	(81.4–94.8)	(88.0–98.0)	(86.0–96.0)	(84.0–94.0)	(0.96–0.98)	(0.042–0.048)
Bridge-ELISA: GADA/IA-2A							
132	94.0	99.0	94.0	94.0	92.0	0.97	0.047
132	92.0	97.9	94.0	92.0	92.0	0.97	0.047
Median	93.0	98.4	94.0 (94.0-94.0)	93.0	92.0	0.97	0.047
(range)	(92.0–94.0)	(97.9–99.0)		(92.0–94.0)	(92.0–92.0)	(0.97–0.97)	(0.047–0.047)
Bridge-ELISA: GADA/IA-2A/ZnT8A							
1801	96.0	94.8	96.0	96.0	96.0	0.98	0.048
2408	96.0	94.8	96.0	96.0	96.0	0.99	0.048
121	94.0	100.0	96.0	96.0	94.0	0.98	0.048
132	96.0	96.9	96.0	96.0	94.0	0.98	0.048
153	94.0	99.0	96.0	96.0	94.0	0.98	0.048
1602	96.0	99.0	96.0	96.0	94.0	0.99	0.048
Median	96.0	98.0	96.0	96.0	94.0	0.98	0.048
(range)	(94.0–96.0)	(94.8–100)	(96.0–96.0)	(96.0–96.0)	(94.0–96.0)	(0.98–0.99)	(0.048–0.048)
ECL: GADA/IA-2A/ZnT8A/IAA							
1306	96.0	92.8	98.0	98.0	96.0	0.98	0.049
133	96.0	97.9	98.0	96.0	94.0	0.98	0.048
Median	96.0	95.3	98.0	97.0	95.0	0.98	0.048
(range)	(96.0–96.0)	(92.8–97.9)	(98.0–98.0)	(96.0–98.0)	(94.0–96.0)	(0.98–0.98)	(0.048–0.049)
LIPS: GADA/IA-2A/ZnT8A							
116	96.0	94.8	96.0	84.0	84.0	0.97	0.045
PAL/scLIPS: GADA/IA-2A/ZnT8A							
153	96.0	97.9	96.0	94.0	88.0	0.97	0.047
153	88.0	97.9	96.0	88.0	88.0	0.96	0.046
Median	92.0	97.9	96.0 (96.0-96.0)	91.0	88.0	0.96	0.046
(range)	(88.0–96.0)	(97.9–97.9)		(88.0–94.0)	(88.0–88.0)	(0.96–0.97)	(0.046–0.047)

Performance metrics for individual assays organized by format and measured antibody combinations. A final number of 50 T1D and 97 control samples was used for all calculations (the control sample R587009, which tested positive across all assays, was excluded from specificity and ROC calculations). Sensitivity and specificity values are expressed as percentages. AS95, AS99, and AS100 indicate the adjusted sensitivity at thresholds corresponding to the 95th, 99th, and 100th percentile of controls. ROC-AUC and pAUC95 are also provided for each assay. For each format group, results are summarized as median values with ranges.

based on the proportion of assays detecting positivity: *1*) high consensus (positive in >50% of assays), *2*) moderate consensus (positive in ≤50% of assays), *3*) positive in only one assay, and *4*) negative in all assays. For ADAP and ECL assays measuring individual autoantibodies, cumulative consensus scores were calculated by combining individual autoantibody results (Supplementary Fig. 5).

Strong concordance was observed across multiplex assay formats for T1D samples, with 48 of 50 (96%) showing high-consensus positivity across all tested formats (i.e., positivity in >50% multiplex assays). Among control samples, different formats showed varying false-positive patterns. ADAP assays demonstrated high-consensus positivity (>50% of assays) for 5 controls, moderate consensus in 5 additional controls (≤50% of assays), and single-assay positivity in 15 controls. ECL platforms identified nine controls as positive across assays, all in only single assays. Bridge-ELISA showed high consensus in two controls and moderate consensus in three additional controls. LIPS identified five controls as positive, while PAL/scLIPS showed high consensus in two controls.

Adjusted Sensitivity at Different Specificity Thresholds

Laboratory-defined positivity thresholds vary between assays, making direct performance comparisons difficult. Lower thresholds increase sensitivity but reduce specificity, while higher thresholds have the opposite effect. Importantly, even assays demonstrating excellent threshold-independent discriminatory performance (as evidenced by high ROC-AUC values) may still misclassify control samples when using laboratory-defined cutoffs. To enable fair comparison, we calculated adjusted sensitivity using standardized cutoffs corresponding to the 95th, 99th, and 100th percentiles of control values (AS95, AS99, and AS100, respectively) (Table 1 and Supplementary Fig. 6). This approach allows evaluation of how each assay would perform under identical specificity constraints.

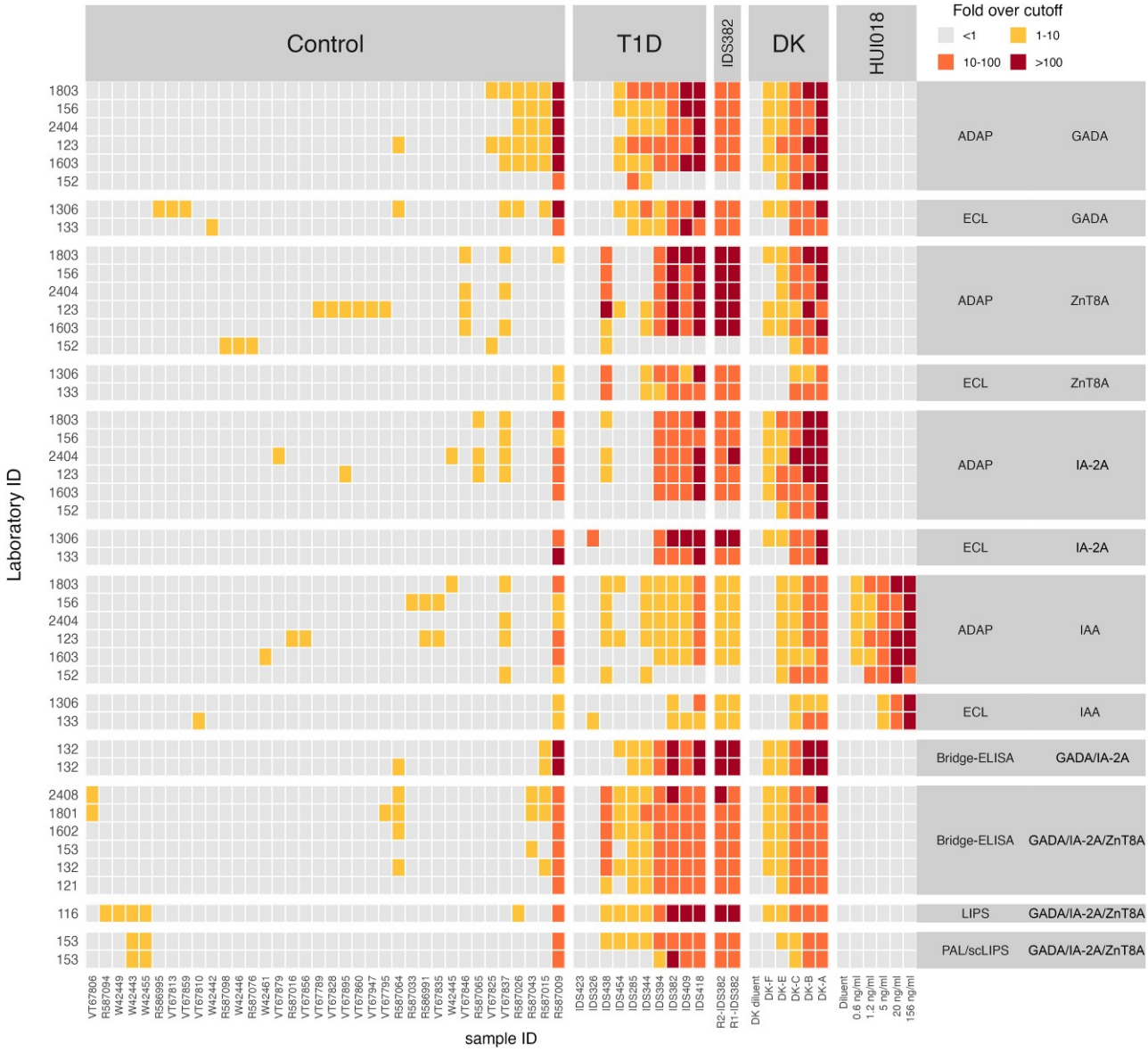


Figure 1—Heat map of signal over cutoff values in discordant samples by multiple assays submitted to IASP 2024. This graph displays the fold over cutoff values assigned by laboratories to selected samples across multiple autoantibody assays (62 control samples called negative by all assays and 40 T1D samples called positive by all assays were excluded). Samples are clustered horizontally to highlight format-specific and antibody combination patterns, and are grouped by specimen type (controls, new-onset T1D, DK standards, IDS382 replicates, and HUI-018 mAb dilutions). DK standards are shown in order of decreasing concentration from right to left (DK-A 493, DK-B 228, DK-C 92, DK-E 13, DK-F 4.8, and DK diluent 0 DK units/mL for GADA; DK-A 235, DK-B 106, DK-C 42, DK-E 5.8, DK-F 2, and DK diluent 0 DK units/mL for IA-2A). HUI-018 mAb dilutions are shown from highest (right) to lowest (left) concentration (156, 20, 5, 1.2, 0.6, and 0 ng/mL). The *y*-axis displays results from different laboratory assays, organized by analyte (separate antigens for ADAP and ECL, antibody combinations for all the other formats) and by assay format (ADAP, ECL, bridge-ELISA, LIPS, RBA, and CLIA), and ordered by their AUC95 values. Samples with fold over cutoff values <1 (below threshold) are shown with light gray tiles, values between 1 and 10 fold are indicated by yellow tiles, values between 10 and 100 fold are indicated by orange tiles, and values >100 fold over cutoff are indicated by red tiles.

Bridge-ELISA assays testing GADA and IA-2A (n = 2) showed a median adjusted sensitivity of 93% for AS99 (range 92–94%). When combined with ZnT8A (n = 6), the median AS99 increased to 96%. The LIPS assay showed an AS99 value of 84%, while PAL and scLIPS assays (n = 2) demonstrated a median AS99 value of 91% (range 88–94%). ADAP and ECL assays, which calculated overall adjusted sensitivity by combining individual antibody scores, achieved median AS99 sensitivities of 94% (range 86–96%) and 97% (range 96–98%), respectively (Table 1 and Supplementary Fig. 6).

Overall, when the AS99 threshold was imposed, most multiplex assays eliminated several false-positive controls without affecting sensitivity for T1D cases (Fig. 3).

We evaluated whether universal cutoffs could achieve equivalent AS99 performance across laboratories that use standardized assay platforms (Supplementary Fig. 7). Bridge-ELISA assays, which share identical protocols, reagents, and standards, achieved equivalent sensitivity with a universal 40 AU cutoff compared with laboratory-specific AS99 thresholds. This universal threshold was often more stringent than local AS99 values yet maintained equivalent sensitivity, suggesting that universal cutoffs could improve specificity without compromising sensitivity.

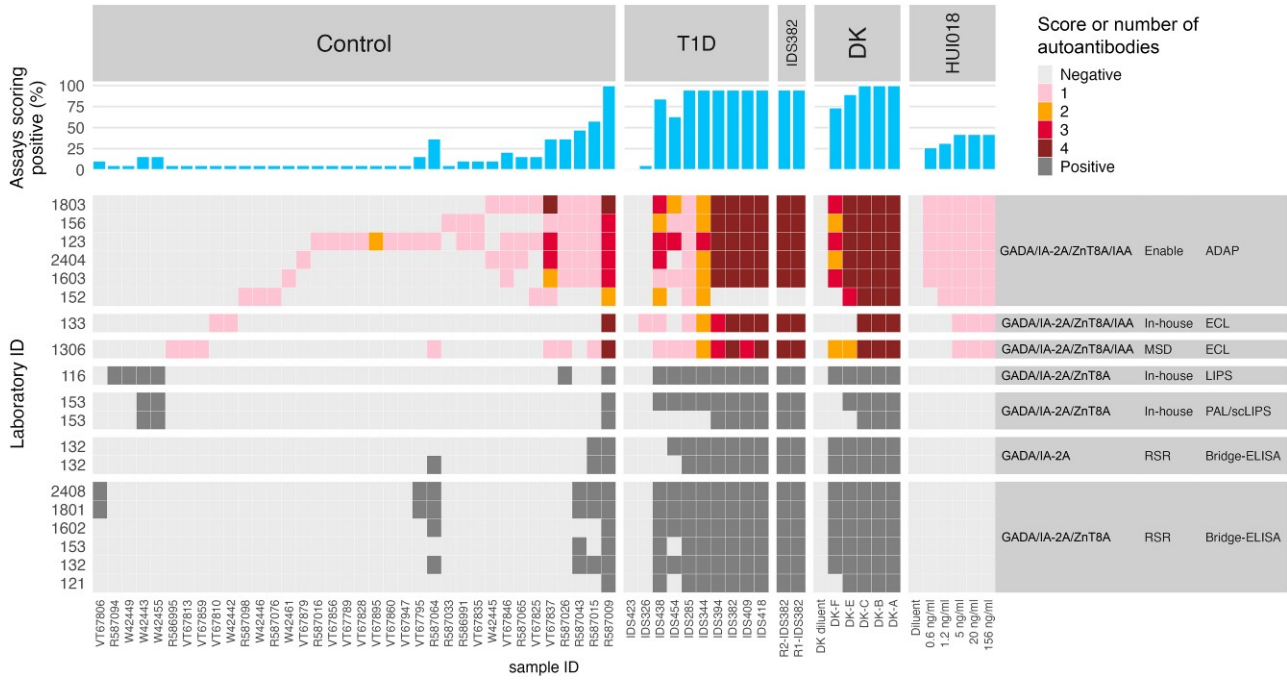


Figure 2—Tile map of scores assigned to discordant samples by multiplex assays submitted to IASP 2024. This graph displays the positivity scores of multiplex assays assigned by laboratories to individual selected samples. Control samples called negative by all assays (n = 62) and T1D samples called positive by all assays (n = 40) were excluded. The upper panel shows the percentage of assays scoring positive for each sample, while the lower panel presents a detailed tile map of individual assay results across different sample types. Samples are clustered horizontally to highlight format-specific and antibody combination patterns, and are grouped by specimen type (controls, new-onset T1D, DK standards, IDS382 replicates, and HUI-018 mAb dilutions). The y-axis displays results from different laboratory assays, organized by assay format (ADAP, ECL, LIPS, PAL/scLIPS and bridge-ELISA), analyzed antibody combinations and commercial vs in-house status (manufacturer shown for commercial assays: RSR [RSR Limited], MSD [Meso Scale Diagnostics LLC], Enable [Enable Biosciences Inc.]), and ordered by their AUC95 values. Samples scored negative are shown with light gray tiles, while samples scored positive are indicated by dark gray tiles for assays not distinguishing across single autoantibody reactivities. For ADAP and ECL formats, the positive samples are colored by the number of autoantibodies detected (light pink for one autoantibody, orange for two autoantibodies, red for three autoantibodies, and dark brown for four autoantibodies).

In contrast, ADAP assays showed substantial interlaboratory variation in AS99 thresholds despite standardized protocols: GADA 12.5–85 AU, IA-2A 10–26 AU (one outlier at 90 AU), IAA 2–4 AU (one outlier at 26 AU), and ZnT8A 3–10 AU (one outlier at 41 AU). No single universal threshold maintained equivalent sensitivity across all laboratories. This variability may reflect the use, in ADAP assays, of ΔCt values relative to blank controls rather than calibrated arbitrary units, which may contribute to interlaboratory threshold variability. No universal threshold maintained equivalent sensitivity across all sites, suggesting that the use of calibrators to standardize signal conversion might facilitate universal cutoffs for ADAP platform.

Single Autoantibody Assay Formats

In addition to multiplex formats, the workshop included results from 132 single autoantibody assays measuring the major islet autoantibodies independently. Individual autoantibody results for GADA, IA-2A, ZnT8A, IAA, and TSPAN7A submitted to the IASP 2024 workshop are detailed in Supplementary Figs. 8–16. Traditional RBAs remained one of the most common formats for single autoantibody detection (n = 36, 27.3%) alongside bridge-ELISA (n = 39, 29.5%), followed by ADAP (n = 24, 18.2%), LIPS (n = 16, 12.1%), ECL (n = 8, 6.1%), CLIA (n = 7, 5.3%), and PAL/scLIPS (n = 2, 1.5%). Performance metrics and manufacturer for each assay are detailed in Supplementary Tables 1–5.

CONCLUSIONS

The development of nonradioactive multiplex platforms for islet autoantibody detection addresses a critical need as regulatory restrictions on radioactive assays continue to tighten globally (3–5). Our IASP 2024 interlaboratory evaluation provides the first comprehensive assessment of how well these alternative formats perform in a standardized comparison setting, demonstrating that several multiplex technologies can achieve the discriminatory performance historically associated with RBAs. While our case-control design was not intended for screening validation, the data generated offer important technical insights into assay capabilities, performance variability, and standardization opportunities that are essential prerequisites for any future adaptation of these platforms to broader screening applications.

Traditional RBAs were once the standard method for detecting T1D autoantibodies but have significant limitations in addition to the use of radioactive materials: they require larger sample volumes, are labor-intensive, and detect only one antibody at a time. Modern multiplex platforms have addressed these drawbacks by detecting multiple autoantibodies simultaneously in a single test. These newer methods require less sample, avoid radioactive tracers, and are more cost-effective (6–8). Additionally, they can identify people with multiple positive autoantibody markers, a particularly important capability, since individuals with multiple markers have up to a 70% risk of developing symptomatic T1D within 10 years (21).

Figure 3—Tile map of lab-assigned vs. AS99 scores in multiplex assays submitted to the IASP 2024. The upper panel shows the percentage of assays yielding positive results for each sample across different specimen types: controls, new-onset T1D, DK standards, IDS382 replicates, and HUI-018 mAb dilutions. Yellow bars represent the percentage of assays that were positive using the original laboratory-assigned thresholds, while deep magenta bars represent the percentage of assays that were positive after imposing stringent thresholds at 99th percentile of control values (AS99 scores). The overlapping visualization of bars allows comparison between the two scoring methods. Control samples called negative by all assays (before and after the imposition of AS99 thresholds, $n = 60$) were removed, while control sample R587009, which tested positive across all assays, was excluded for specificity calculation but included in the graph. The lower panel provides a tile map of local and AS99 positivity scores from multiplex assays. Samples positive according to the original laboratory-assigned threshold but negative after imposing the AS99 threshold (lost positives) are shown in yellow, while samples remaining positive using the AS99 threshold are indicated in deep magenta. Negative scores are shown in light gray. Laboratory assays are organized vertically by format, antigen combinations, and commercial vs in-house status (manufacturer shown for commercial assays: RSR [RSR Limited], MSD [Meso Scale Diagnostics LLC], Enable [Enable Biosciences Inc.]). For ADAP and ECL formats, displayed scores represent combinations of individual antibody AS99 scores, allowing for multiple positive results in controls.

The IASP 2024 interlaboratory evaluation demonstrated that multiplex platforms have excellent performance in distinguishing new-onset T1D case subjects from control subjects, with a median ROC-AUC of 0.98. The pAUC95, with a median of 0.048 out of a maximum of 0.05, confirmed that these assays remain highly discriminatory at high specificity levels in blood donors and comparable, in this workshop, to RBAs that were originally used for screening. While we observed no statistically significant differences between formats' ROC curves, this finding should be interpreted cautiously, as it may reflect limited statistical power rather than true equivalence between platforms.

Despite encouraging overall performance, substantial interassay variability emerged as a critical consideration, as already observed in previous studies (9). Although median sensitivities exceeded 90% across all tested formats, individual assay performance varied considerably, with sensitivity and specificity as low as 86% and 81%, respectively. Even bridge-ELISA assays, which achieved the highest median specificity, showed specificity range from 95% to 100%. This variability indicates that, although the technology for high-accuracy performance is available, substantial optimization is still needed to ensure consistent results across laboratories.

Our adjusted sensitivity analysis revealed that differences in threshold selection often underlie apparent performance discrepancies between assays. Applying standardized cutoffs, we demonstrated that AS99 thresholds typically eliminated false positives without sacrificing sensitivity for T1D detection. Bridge-ELISA platforms showed particular promise for standardization, as a harmonized cutoff of 40 AU achieved equivalent performance across laboratories following identical protocols. In contrast, ADAP assays exhibited significant performance variation even after threshold adjustment, with AS99 sensitivity ranging from 86% to 96%, and no universal cutoff maintained equivalent sensitivity across all sites. For multiplex assays measuring individual autoantibodies independently (ADAP and ECL) (7,8), lower specificity may also result from a "cumulative false positivity effect," where each additional marker introduces its own false-positive risk.

Our study was not designed to assess screening performance in the general population, where very high specificity is needed because of the low disease prevalence (22), but our findings once again highlight the crucial relationship between specificity and positive predictive value (PPV) in different clinical contexts. Assuming the sensitivity and specificity observed in our case-control cohort hold true in a general pediatric population, the implications for PPV would be significant. For instance, an assay with 99% specificity and 90% sensitivity would have a PPV of 98% in a diagnostic setting with 50% disease prevalence, but only 27% in a population screening setting with a prevalence of 0.4%, resulting in more false positives than true positives. This mathematical relationship between prevalence and PPV means that different screening contexts might require different performance thresholds. Population screening programs targeting asymptomatic individuals need assays with specificity >99% to achieve acceptable PPVs

and minimize false positives. In contrast, high-sensitivity screening approaches with confirmatory testing can tolerate lower specificity (95–97%) in exchange for higher sensitivity (22).

Sequential testing strategies can help balance sensitivity and specificity (23). This involves an initial high-sensitivity multiplex assay, followed by confirmatory testing using a different assay format. The confirmatory test verifies the presence of autoantibodies, helping to distinguish true positives from false positives.

Our data show that most false-positive results in controls were not shared across assay formats, suggesting that, if similar patterns held in screening populations, such two-step sequential testing strategies would be well-supported (24,25). The high sensitivity of multiplex assays (exceeding 90%) in this workshop suggests they could potentially capture at-risk individuals during initial screening, while standardized thresholds might help minimize false-positive burden on confirmatory testing resources, although this would need confirmation in actual screening populations.

Several limitations of our study must be acknowledged. Its design, with a 1:2 case-to-control ratio, differs from real-world populations where disease prevalence would be much lower (∼0.4%), resulting in more than 200 control subjects per case subject. The relatively small control sample size (∼100) is insufficient to precisely estimate specificity at the levels (>99%) that would be required for population screening, and small changes in specificity that fall within our CIs could still have significant clinical and economic implications. Accurately assessing specificity for population screening would require considerably larger control cohorts representative of the general population.

The demographic and clinical composition of our cohort further limits generalizability. The case cohort (median age 13 years) was enriched for adolescent and adult-onset cases, particularly affecting IAA assessment, which peaks in very young children and declines with age. More importantly, the vast majority of T1D case subjects presented with three or four autoantibodies, providing limited representation of single antibody–positive individuals or other autoantibody combinations that might be encountered during earlier disease stages in screening programs (26). This high-multiplicity bias may overestimate assay performance for detecting individuals with fewer autoantibodies, who represent important targets for early intervention strategies.

Technical considerations also limit the generalizability of our findings to field applications. While all samples in our study consisted of high-quality venous sera without hemolysis, many large-scale screening programs rely on minimally invasive collection methods, particularly in pediatric populations (24). Capillary blood and fingerstick collection methods frequently exhibit hemolysis and matrix effects that could impact assay performance in ways not captured by our evaluation with optimal sample quality. Additionally, our focus on qualitative concordance rather than quantitative harmonization represents a limitation, as precise quantification across assays could provide valuable insights for risk stratification and disease monitoring in screening contexts.

Nevertheless, demonstrating that several nonradioactive multiplex platforms achieve strong discriminatory performance in this case-control setting provides valuable technical information. As regulatory restrictions on radioactive assays continue to increase globally, these alternatives offer viable options for screening applications, with the diversity of platforms providing flexibility to optimize for specific operational, financial, and regulatory constraints across health care settings.

The standardization framework presented here provides a foundation for harmonizing assay performance across laboratories and health care systems. Further development of calibration approaches, standardized reference materials, and analytical protocols is needed to achieve consistency. As the field advances toward broader screening initiatives driven by disease-modifying therapies like teplizumab (27) and the paradigm shift toward detecting presymptomatic T1D stages (28–33), these multiplex assays demonstrate the technical capabilities required to support such efforts, consistent with earlier validation studies in selected at-risk populations (9,24). Ongoing standardization efforts through programs like IASP will be essential to ensuring these platforms can realize their potential for supporting T1D prevention strategies and ultimately reducing disease burden on individuals, families, and health care systems worldwide.

APPENDIX

This appendix lists the participating investigators and their institutional affiliations: C. Wasserfall, Wasserfall Laboratory, University of Florida, Gainesville, FL; M.R. Batstra, Reinier Haga MDC, Department of Immunology, Delft, the Netherlands; R. Veijola, Diabetes Research Laboratory, Department of Pediatrics, University of Oulu, Oulu, Finland; A.E. Long, Diabetes & Metabolism, Learning & Research, University of Bristol, Bristol, U.K.; K. Verhaeghen and B. Blomme, Uzbrussel Vub, Clinical Biology of Diabetes-Diabetes Research Center, Brussel, Belgium; P. Achenbach, Institute of Diabetes Research, Helmholtz Zentrum München, Neuherberg, Germany; L. Castaño, Endocrinology and Diabetes Research Group, Cruces University Hospital, Bizkaia, Spain; W.A. Hagopian, Pacific Northwest Diabetes Research Institute, Seattle, WA; S. Chen, FIRS Laboratories, RSR Ltd., Cardiff, U.K.; L. Yu, Barbara Davis Center, Aurora, CO; S. Berndt, MEDIPAN GMBH, Blankenfelde-Mahlow OT Dahlewitz, Germany; J. Wentworth, Endocrine Laboratory, Royal Melbourne Hospital, Victoria, Australia; I. Marzinotto and V. Lampasona, Diabetes Research Institute, IRCCS Ospedale San Raffaele, Milano, Italy; A. Ramelius, Diabetes and Celiac Disease Research Unit, Lund University, Malmö, Sweden; M. Knip and T. Härkönen, Children's Hospital, Scientific Laboratory, University of Helsinki, Helsinki, Finland; D. Pittman, Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL; M. Groesvik, The Hormone Laboratory, Haukeland University Hospital, Bergen, Norway; K. Wyka, Immunopathology and Genetic Laboratory, Medical University of Lodz, Łódź, Poland; J. Wood, Royal Devon & Exeter Blood Sciences, Exeter, U.K.; N. Fabien, Immunology-Autoimmunity, Hôpital Lyon Sud, Pierre-Bénite, France; A. Mathew, Meso Scale Diagnostics, LLC, Rockville, MD; D. Mueller, Preclinical Approaches to Stem Cell Therapy/Diabetes, Dresden, Germany; B. Brandao and R. Mallone, Mallone and You Laboratory, INSERM U1016 Cochin Institute, Paris, France; C. Mann, Quest Diagnostics Nichols Institute, San Juan Capistrano, CA; E. Kawasaki, Diabetes Center, Shin-Koga Hospital, Kurume, Japan; J. Tsai, Enable Biosciences, Inc., San Francisco,

CA; K. Martin, Randox Teoranta, Donegal, Ireland; V. Hirschberg Jensen, Steno Diabetes Center Copenhagen, Department for Diabetes Treatment, Herlev, Denmark; L. Babini, SOD Medicina di Laboratorio, Azienda Ospedaliero-Universitaria, Ancona, Italy; O. Cinek, University Hospital Motol, Prague, Czechoslavakia.

Acknowledgments. During the course of preparing this work, the authors used Claude 4 Sonnet for the purpose of drafting, language editing, statistical analysis code generation, and editing. Following the use of this tool, the authors formally reviewed the content for its accuracy and edited it as necessary. The authors take full responsibility for all the content of this publication.

The authors thank all the people and the clinicians who contributed blood samples, without whose generosity and effort the IASP workshops and collaborative studies could not take place. We also remember the precious contribution to both studies' conception and data analysis of our friend and IASP committee member Alistair J.K. Williams, who passed away in 2020.

Funding. The authors acknowledge the support received by the IASP study through a The Environmental Determinants of Diabetes in the Young (TEDDY) grant administered at the University of South Florida (Tampa, FL) and by the Type 1 Diabetes TrialNet/IASP Ancillary studies arm.

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

Author Contributions. I.M., D.L.P., and V.L. contributed to the collection, preliminary analysis, and interpretation of data. I.M., V.L., and A.E.L. drafted the initial manuscript. D.L.P., P.A., B.A., and C.H.W. critically revised the data analysis and the manuscript. All authors approved the publication. I.M. and A.E.L. are the guarantors of the paper and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented at the Islet Autoantibody Standardization Program 2024 Workshop at the 20th Immunology of Diabetes Society Congress, Bruges, Belgium, 4–8 November 2024.

Handling Editors. The journal editors responsible for overseeing the review of the manuscript were John B. Buse and Michael J. Haller.

References

1. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. Lancet 2014;383:69–82
2. Jia X, Yu L. Understanding islet autoantibodies in prediction of type 1 diabetes. J Endocr Soc 2023;8:bvad160
3. Bonifacio E, Yu L, Williams AK, et al. Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for national institute of diabetes and digestive and kidney diseases consortia. J Clin Endocrinol Metab 2010;95:3360–3367
4. Williams AJ, Bingley PJ, Bonifacio E, Palmer JP, Gale EA. A novel micro-assay for insulin autoantibodies. J Autoimmun 1997;10:473–478
5. Wenzlau JM, Liu Y, Yu L, et al. A common nonsynonymous single nucleotide polymorphism in the SLC30A8 gene determines ZnT8 autoantibody specificity in type 1 diabetes. Diabetes 2008;57:2693–2697
6. Amoroso M, Achenbach P, Powell M, et al. 3 Screen islet cell autoantibody ELISA: a sensitive and specific ELISA for the combined measurement of autoantibodies to GAD65, to IA-2 and to ZnT8. Clin Chim Acta 2016;462:60–64
7. He L, Jia X, Rasmussen CG, et al. High-throughput multiplex electrochemiluminescence assay applicable to general population screening for type 1 diabetes and celiac disease. Diabetes Technol Ther 2022;24:502–509
8. Cortez FdJ, Gebhart D, Tandel D, et al. Automation of a multiplex agglutination-PCR (ADAP) type 1 diabetes (T1D) assay for the rapid analysis of islet autoantibodies. SLAS Technol 2022;27:26–31
9. Krischer JP, Muller S, You L, et al.; Type 1 Diabetes TrialNet Study Group. Comparative analysis of the sensitivity, specificity, concordance, and 5-year predictive power of diabetes-related autoantibody assays. Diabetes 2025;74:1535–1546
10. Williams AJK, Lampasona V, Schlosser M, et al.; Participating Laboratories. Detection of antibodies directed to the N-terminal region of GAD is dependent on assay format and contributes to differences in the specificity of GAD autoantibody assays for type 1 diabetes. Diabetes 2015;64:3239–3246
11. Bingley PJ, Bonifacio E, Mueller PW. Diabetes Antibody Standardization Program: first assay proficiency evaluation. Diabetes 2003;52:1128–1136
12. Schlosser M, Mueller PW, Achenbach P, Lampasona V, Bingley PJ, Participating Laboratories. Diabetes Antibody Standardization Program: first evaluation of assays for autoantibodies to IA-2β. Diabetes Care 2011;34:2410–2412
13. Lampasona V, Schlosser M, Mueller PW, et al. Diabetes Antibody Standardization Program: first proficiency evaluation of assays for autoantibodies to zinc transporter 8. Clin Chem 2011;57:1693–1702
14. Lampasona V, Pittman DL, Williams AJ, et al.; Participating Laboratories. Islet Autoantibody Standardization Program 2018 Workshop: interlaboratory comparison of glutamic acid decarboxylase autoantibody assay performance. Clin Chem 2019;65:1141–1152
15. Marzinotto I, Pittman DL, Williams AJK, et al.; participating laboratories. Islet Autoantibody Standardization Program: interlaboratory comparison of insulin autoantibody assay performance in 2018 and 2020 workshops. Diabetologia 2023;66:897–912
16. World Medical Association. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. Bull World Health Organ 2001;79:373–374
17. Johansson E, Wu X, Yu B, et al. Insulin binding to the analytical antibody sandwich pair OXI-005 and HUI-018: epitope mapping and binding properties. Protein Sci 2021;30:485–496
18. R Core Team. R: a language and environment for statistical computing. Version 4.5.0. Accessed 11 April 2025. Available from https://www.R-project.org/
19. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77
20. Thiele C, Hirschfeld G. cutpointr: improved estimation and validation of optimal cutpoints in R. J Stat Soft 2021;98:1–27
21. Ziegler AG, Rewers M, Simell O, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. JAMA 2013;309:2473–2479
22. Bonifacio E, Ziegler A-G. Type 1 diabetes risk factors, risk prediction and presymptomatic detection: evidence and guidance for screening. Diabetes Obes Metab 2025;27(Suppl. 6):28–39
23. Phillip M, Achenbach P, Addala A, et al. Consensus guidance for monitoring individuals with islet autoantibody-positive pre-stage 3 type 1 diabetes. Diabetologia 2024;67:1731–1759
24. Raab J, Haupt F, Scholz M, et al.; Fr1da Study Group. Capillary blood islet autoantibody screening for identifying pre-type 1 diabetes in the general population: design and initial results of the Fr1da study. BMJ Open 2016;6:e011144
25. Lind A, Freyhult E, de Jesus Cortez F, et al.; Better Diabetes Diagnosis (BDD) study group. Childhood screening for type 1 diabetes comparing automated multiplex Antibody Detection by Agglutination-PCR (ADAP) with single plex islet autoantibody radiobinding assays. EBioMedicine 2024;104:105144
26. Insel RA, Dunne JL, Atkinson MA, et al. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. Diabetes Care 2015;38:1964–1974
27. Herold KC, Bundy BN, Long SA, et al.; Type 1 Diabetes TrialNet Study Group. An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. N Engl J Med 2019;381:603–613
28. Ziegler A-G, Kick K, Bonifacio E, et al.; Fr1da Study Group. Yield of a public health screening of children for islet autoantibodies in Bavaria, Germany. JAMA 2020;323:339–351
29. Quinn LM, Dias RP, Greenfield SM, et al. Protocol for a feasibility and acceptability study for UK general population paediatric type 1 diabetes screening—the EarLy Surveillance for Autoimmune diabetes (ELSA) study. Diabet Med 2025;42:e15490
30. Cherubini V, Mozzillo E, Iafusco D, et al. Follow-up and monitoring programme in children identified in early-stage type 1 diabetes during screening in the general population of Italy. Diabetes Obes Metab 2024;26:4197–4202
31. Hoffmann L, Kohls M, Arnolds S, et al.; EDENT1FI consortium. EDENT1FI Master Protocol for screening of presymptomatic early-stage type 1 diabetes in children and adolescents. BMJ Open 2025;15:e088522
32. McQueen RB, Geno Rasmussen C, Waugh K, et al. Cost and cost-effectiveness of large-scale screening for type 1 diabetes in Colorado. Diabetes Care 2020;43:1496–1503
33. Wentworth JM, Sing ABE, Naselli G, et al. Islet autoantibody screening throughout Australia using in-home blood spot sampling: 2-year outcomes of Type1Screen. Diabetes Care 2025;48:556–563