Diabetes Antibody Standardization Program: evaluation of assays for insulin autoantibodies

M. Schlosser · P. W. Mueller · C. Törn · E. Bonifacio ·
P. J. Bingley · Participating Laboratories

Received: 12 April 2010 / Accepted: 5 August 2010 / Published online: 25 September 2010
© Springer-Verlag 2010

Abstract

Aims/hypothesis Insulin autoantibodies (IAA) are important in type 1 diabetes risk assessment. However, their determination varies more between laboratories than other diabetes autoantibodies. The Diabetes Antibody Standardization Program (DASP) aims to improve and standardise mea-

Electronic supplementary material The online version of this article (doi:10.1007/s00125-010-1915-5) contains supplementary material, which is available to authorised users. This includes a list of Participating Laboratories.

M. Schlosser
Department of Medical Biochemistry and Molecular Biology,
Research Group of Predictive Diagnostics,
University of Greifswald,
Greifswald, Germany

M. Schlosser
Institute of Pathophysiology, Research Group of Predictive
Diagnostics, University of Greifswald,
Karlsburg, Germany

P. W. Mueller
Centers for Disease Control and Prevention,
Atlanta, GA, USA

C. Törn
Unit for Diabetes and Coeliac Disease, Institution of Clinical
Sciences, Clinical Research Centre, University Hospital MAS,
Malmö, Sweden

E. Bonifacio
Centre for Regenerative Therapies–Dresden,
Dresden University of Technology,
Dresden, Germany

P. J. Bingley (✉)
School of Clinical Sciences, University of Bristol,
Southmead Hospital,
Bristol BS10 5NB, UK
e-mail: polly.bingley@bristol.ac.uk

surement of autoantibodies associated with type 1 diabetes. We report the results of measurement of IAA from DASP workshops in 2002, 2003 and 2005.

Methods Up to 32 laboratories in 14 countries participated in each workshop. Aliquots of coded sera from 50 patients with newly diagnosed type 1 diabetes and 100 blood donor controls were circulated to participating laboratories. Reported results were analysed using receiver operator characteristic (ROC) curves. We compared concordance of antibody levels by ranking, IAA and insulin antibody (IA) indices and units derived from an IA standard curve.

Results In all three workshops IAA assay performance had improved compared with DASP 2000. The median area under the ROC curve was 0.73 in DASP 2002, 0.78 in 2003 and 0.80 in 2005 ($p=0.0012$), and median laboratory-assigned sensitivity was 26% in 2002, 36% in 2003 and 45% in 2005 ($p<0.0001$). There was, however, marked variation between assays. The range of AUC was 0.36–0.91 and that of laboratory-assigned sensitivity was 22–57%. Concordance of ranking of patient serum samples was related to AUC ($p<0.001$). Using an index related to common IAA and IA-positive or -negative control sera improved the concordance between assays ($p<0.0001$).

Conclusions/interpretation The overall performance of IAA assays has improved but there is still wide variation between laboratories. Concordance between assays would be improved by the use of a common reference reagent.

Keywords Adjusted sensitivity · AUC · Insulin autoantibodies · Insulin autoantibody index · Islet autoantibodies · Prediction · Reference serum · Sensitivity · Specificity

Abbreviations

AS95 Sensitivity adjusted to 95% specificity
CDC Centers for Disease Control and Prevention

DASP	Diabetes Antibody Standardization Program
GADA	GAD autoantibodies
IAA	Insulin autoantibodies
IA	Insulin antibodies
IA-2	Islet antigen-2
IA-2A	IA-2 autoantibodies
IDS	Immunology of Diabetes Society
ROC	Receiver operator characteristic

Introduction

Autoantibodies against islet cell antigens are important markers of type 1 diabetes-associated autoimmunity, and insulin autoantibodies (IAA) were the first beta cell specific autoantibody described [1]. IAA are important in the prediction of type 1 diabetes, especially in young children, as their level and prevalence at diagnosis correlate inversely with age [2, 3]. Measurement of IAA was initially limited by the high serum volume required for the early immunoprecipitation assays, which used polyethylene glycol to separate immune complexes [4]. The microassay format using protein A/G-Sepharose for precipitation allowed a major reduction of the amount of serum used and has improved assay specificity [5, 6]. IAA measurement using the microassay method is now performed in many laboratories and standardisation of assays is needed to compare the results of studies throughout the world.

The Diabetes Antibody Standardization Program (DASP), a collaboration of the Immunology of Diabetes Society (IDS) and the US Centers for Disease Control and Prevention (CDC), was established as an extension of previous IDS antibody workshops to improve and standardise the measurement of the diabetes-associated autoantibodies [5–8]. The objective of the programme is to assist laboratories to improve their methods by providing technical support, training and information as well as assessing proficiency in order to harmonise antibody measurement between key laboratories worldwide. DASP proficiency evaluations since 2000 have shown that the majority of participating laboratories achieved a high concordance in measurement of antibodies to GAD and islet antigen-2 (IA-2) using the new WHO international reference reagent [7–9]. In contrast, the first DASP proficiency testing demonstrated wide variation among IAA microassays, with poor overall performance and low sensitivity [8]. Nonetheless, this first workshop also showed that some laboratories were able to achieve high sensitivity and specificity with good levels of concordance.

We now report the results for IAA assays of the follow-up proficiency-evaluation rounds performed in 2002, 2003 and 2005, and compare them with those of DASP 2000.

The format of the evaluations, which involved a relatively large number of coded serum samples from patients and controls, allowed a comprehensive comparison of different laboratories and assays. We used the same format for each round, allowing analysis of changes in assay performance over time. The main aims of these workshops were to allow participating laboratories to determine the sensitivity and specificity of their IAA assays, to assess concordance between laboratories and to examine whether the overall performance of assays had improved. We also evaluated the concordance between laboratories in quantifying antibody levels by ranking and using common reference standards and a potential IA standard curve based on dilutions of sera from two patients with long-standing insulin-treated type 1 diabetes.

Methods

Study design

Participating laboratories received uniquely coded sets of frozen 100 µl aliquots of sera from 50 patients with newly diagnosed type 1 diabetes and 100 healthy controls as described by Törn et al. [9]. Sets were distributed to 26–32 laboratories in up to 14 countries in each round (see Electronic supplementary material [ESM] for a list of participating laboratories). In addition to the proficiency samples, participating laboratories received the DASP IAA standard (sample 686) and the DASP IAA-negative serum. In 2005, serial dilutions of two potential standards of IA-positive type 1 diabetic patients' sera (IA standards IB4.4, IB4.6–IB4.10 and IC9.3, IC9.5–IC9.9) were also distributed. Laboratories were asked to test the sera using their usual method and to provide some details of the local assay. Results were reported as raw data, units calculated according to the local protocol and the classification of each serum sample as IAA negative or positive using the local cut-off. Informed consent has been obtained from all patients who have donated samples for DASP, and the investigations were carried out in accordance with the Declaration of Helsinki as revised in 2000.

Data analysis

Data were analysed as in previous DASP proficiency evaluations [8, 9]. Laboratory-assigned sensitivity and specificity were based on the local cut-off. Receiver operator characteristic (ROC) curves were used to evaluate the performance of each assay in discriminating disease from non-disease on the basis of the area under the curve. To facilitate comparison between laboratories, the coordi-

nates of the ROC curve were used to determine the level of sensitivity that corresponded to a specificity of 95%, defined as the adjusted sensitivity 95 (AS95). The combined ROC curve was compiled from the median values for each patient and control sample from all assay measurements in DASP 2005.

The reported relative levels of autoantibody in different assays in DASP 2005 were compared by ranking the patient sera for each assay. Concordance was assessed by linear regression of individual assay rank against the median rank for all sera from patients with type 1 diabetes.

To evaluate the use of the IAA index and IA standard curve in antibody quantification, the reported Δcpm without and with unlabelled insulin was used in assays that included competition and cpm without unlabelled insulin in those that did not.

The IAA index for each serum in each laboratory compared with the DASP IAA standard and DASP negative serum was calculated as follows:

$$\text{IAA index} = 100 \times \frac{[\Delta\text{cpm unknown} - \Delta\text{cpm negative serum}]}{[\Delta\text{cpm positive standard} - \Delta\text{cpm negative serum}]}$$

The IA index for each serum in each laboratory was calculated from the IB4.4 standard dilution, which was reported positive by all laboratories, and the DASP negative serum as follows:

$$\text{IA index} = 100 \times \frac{[\Delta\text{cpm unknown} - \Delta\text{cpm negative serum}]}{[\Delta\text{cpm IB4.4 standard} - \Delta\text{cpm negative serum}]}$$

In addition IA units were derived from a logarithmic standard curve constructed from dilutions of the IB4 standard provided. The IA units assigned to the dilutions were: 125 IA units for IB4.4; 31.25 IA units for IB4.6; 15.6 IA units for IB4.7; 7.8 IA units for IB4.8; 3.9 IA units for IB4.9; and 1.95 IA units for IB4.10. Values outside the standard curve were calculated by extrapolation.

Differences in inter-laboratory concordance between ranking by laboratory-reported IAA levels, calculated IAA and IA indices and standard-curve-derived IA units were analysed by comparing the variance of the regression using the *F* test.

Non-parametric tests were used to compare antibody levels in patient and control samples, and for comparisons between laboratories, workshops and assay methods. Assays for which results for more than 10% of samples were missing were not included in inter-laboratory comparisons. These are indicated in Table 2. For all statistical analyses, two-tailed *p* values <0.05 were considered significant.

Results

IAA assay characteristics in DASP proficiency testing

The results of IAA determination in DASP 2000, 2002, 2003 and 2005 are summarised in Table 1. Twenty-three laboratories reported the results of 23 assays in 2000, 32 reported results of 35 assays in 2002, 28 reported results of 28 assays in 2003, and 26 reported results of 30 assays in 2005. Data-reporting errors resulted in poor performance for two laboratories in DASP 2003, and incomplete results (<90%) were reported by two laboratories in 2002, two in 2003 and one in 2005. In DASP 2005, 25 laboratories used competitive assays with displacement of IAA binding with unlabelled insulin and five laboratories used non-competitive assays. As shown in Table 1, the median AUC improved progressively from DASP 2000 to DASP 2005 for all participating laboratories ($p=0.001$; Fig. 1a), and in laboratories participating in three or four workshops ($p=0.011$; Table 1). There was no overall difference in AS95 between the workshop 2002, 2003 and 2005 ($p=0.268$; Fig. 1b). Laboratory-assigned sensitivity using local thresholds ($p<0.0001$) also improved between workshops. In particular, the median sensitivity was up to fourfold higher in 2005 compared with 2000 in laboratories that participated in three or four workshops (53% [IQR 33–58%] vs 14% [IQR 9–31%]; $p=0.0001$). In contrast, the median laboratory-assigned specificity decreased from 2000 to 2005 ($p<0.0001$), and this occurred also in the subset of laboratories that participated in three or four workshops ($p=0.0009$). Full results for individual laboratories are given in Table 2.

Of 22 laboratories with assay performance below the median AUC in 2002 and/or 2003, ten did not register for DASP 2005 (five participants in 2002, one in 2003 and four in 2002 and in 2003), and the performance of a further five laboratories remained below the median AUC in DASP 2005.

Assay format

In house radioimmunoassays vs commercial kits In every DASP workshop, the highest laboratory-assigned sensitivity, specificity, AUC and AS95 for IAA were achieved by laboratories using in-house radioimmunoassays. In DASP 2002, two commercial RIA kits, one time-resolved immunofluorometric assay and one ELISA kit were tested in five different laboratories, but achieved lower sensitivity, specificity, AUC and AS95 (Table 2). In DASP 2003, three commercial RIA kits and one time-resolved immunofluorometric assay were tested in four laboratories, and in DASP 2005, six laboratories tested commercial RIA kits. The results obtained with the six commercial kits are shown together with those of the 26 in-house RIAs in Fig. 2.

Table 1 Results of IAA determinations in DASP proficiency evaluations

Data set	Assays (<i>n</i>)	Labs returning incomplete results	Lab-assigned sensitivity, median (IQR) (%)	Range of sensitivity (%)	Lab-assigned specificity, median (IQR) (%)	Range of specificity (%)	AUC, median (IQR)	Range of AUC	AS95, median (IQR) (%) ^a	Range of AS95 (%)
All labs										
DASP 2000	23	0	14 (8–30)	4–42	100 (98–100)	90–100	0.67 (0.59–0.75)	0.32–0.80		
DASP 2002	35	2	26 (21–45)	2–62	98 (96–99)	44–99	0.73 (0.65–0.81)	0.45–0.88	42 (24–58)	6–70
DASP 2003	29	2	36 (17–49)	6–78	97 (92–98)	52–100	0.78 (0.61–0.81)	0.45–0.92	42 (16–55)	4–82
DASP 2005	30	0	45 (22–57)	8–74	98 (95–99)	24–100	0.80 (0.71–0.83)	0.36–0.91	43 (32–63)	6–74
Labs participating in three or more workshops										
DASP 2000	16	0	14 (9–31)	4–42	100 (98–100)	90–100	0.71 (0.64–0.76)	0.55–0.80		
DASP 2002	22	1	42 (24–51)	8–62	98 (96–99)	44–99	0.78 (0.67–0.82)	0.45–0.88	50 (30–60)	6–70
DASP 2003	21	1	38 (27–53)	12–78	97 (95–98)	52–100	0.79 (0.76–0.83)	0.45–0.92	48 (30–60)	12–82
DASP 2005	20	0	53 (33–58)	14–74	98 (96–99)	24–100	0.80 (0.71–0.84)	0.36–0.88	50 (33–64)	6–68

^a In DASP 2000, sensitivity was adjusted to 90% specificity of 50 control samples

Fig. 1 The changes in area under the ROC curve ($p=0.001$) (a) and AS95 (b) for IAA in DASP 2000–5. In DASP 2000, only 50 control samples were circulated and the AS95 was therefore not calculated

Variation between commercial kits The performance of the kits was variable. In DASP 2005, the median laboratory-assigned sensitivity for assays using kits was 33% (IQR 18–49%) vs 52% (IQR 25–58%) for in-house RIA ($p=0.147$), median specificity was 96% (IQR 58.5–99%) vs 98% (IQR 96–99%; $p=0.35$), median AUC was 0.78 (IQR 0.48–0.86) vs 0.81 (IQR 0.72–0.83; $p=0.539$) and median AS95 was 37% (IQR 14–65%) vs 47% (IQR 33–63%; $p=0.351$). In DASP 2002–2005, only one commercial RIA kit (laboratory 132) achieved sensitivity, specificity, AUC and/or AS95 above the median values of all participating laboratories. One RIA kit (laboratory 209) achieved the highest AUC and AS95 of all assays in DASP 2005, but the laboratory-assigned sensitivity was only 22%. Of note, the two RIA kits with lowest AUC and AS95 values used the non-competitive assay format without displacement of IAA binding with unlabelled insulin (Fig. 2a, b; white circles). In DASP 2005, four assays (laboratories 121, 150, 153 and 209) reported values for both AUC and AS95 in the upper quartile.

Concordance of laboratory-reported measurements In DASP 2005, serum samples from nine patients and one healthy control were reported positive in $\geq 75\%$ of assays. An additional 12 patient samples, but none of the control samples, were reported positive in $\geq 50\%$ of assays, and an additional nine patient samples and another two control samples were positive in $\geq 25\%$ of assays. There was agreement on positive/negative status in $\geq 75\%$ of assays for 108 samples (nine patient samples and 99 control samples; ESM Fig. 1a, b). In three of four laboratories with assay performances for both AUC and AS95 in the upper quartile, there was agreement for either positivity or negativity in 127 samples (27 patients samples and 100 control samples; data not shown).

The concordance of ranking of the IAA level in the patient samples between all laboratories by linear regression analysis was highly significant ($r^2=0.642$, variance=73.7, $p<0.0001$; Fig. 3). As expected, concordance in ranking of patient samples was lower between the assays with both AUC and

AS95 below the 25th centile ($n=5$ assays, $r^2=0.392$, variance=126) than between the assays with AUC and AS95 between the 25th and 75th centile ($n=21$ assays, $r^2=0.669$, variance=67.9; $p<0.0001$) and between assays with AUC and AS95 above the 75th centile ($n=4$ assays, $r^2=0.861$, variance=29.3; $p<0.0001$ vs lower 25th centile, and $p<0.0001$ vs 25th–75th centiles using the F test).

Concordance of laboratory-reported IAA levels, common IAA index and common IA index IAA and IA indices were calculated in 27 of the 30 assays in DASP 2005. Three laboratories failed to include the standards in their measurements. The DASP IAA standard was reported positive in all assays. The median IAA index of the IB4.4 IA standard was 56.1 (IQR 42.5–82.1) and the median IAA index of the IC9.3-IA standard was 43.3 (IQR 22.4–53.1; $p=0.003$); IB4.4 was reported positive in all 27 assays and IC9.3 was reported positive in 26 assays. Further analyses were therefore based on units derived from the IB4.4 standard curve.

The ranking of patient samples by laboratory-reported IAA levels varied greatly between the 27 assays ($r^2=0.088$, variance 128,000; $p<0.0001$) and also between the four assays with AUC and AS95 performances in the upper quartile ($r^2=0.467$, variance 468,000; $p<0.0001$). The overall concordance of ranking was markedly improved by expressing results as an index in relation to either the IAA or IA common standard ($r^2=0.779$, variance 385, $p<0.0001$, and $r^2=0.747$, variance 1,100, $p<0.0001$, respectively; F test IAA index and IA index vs laboratory-reported IAA level, $p<0.0001$; Fig. 4a, c). This was particularly apparent in the four laboratories with AUC and AS95 performances above the 75th centile (IAA index: $r^2=0.904$, variance 173, $p<0.0001$; IA index: $r^2=0.918$, variance 356, $p<0.0001$, respectively; F test, IAA index and IA index vs laboratory-reported IAA level, $p<0.0001$; Fig. 4b, d). In all assays, the variance of ranking was lower using the IAA index compared with the IA index (F test, $p<0.0001$). The ranking by units derived from the complete IB4 standard curve did not improve the inter-laboratory

Table 2 DASP results for insulin autoantibody assays

Lab no.	DASP 2002, median ROC AUC: 0.73				DASP 2003, median ROC AUC: 0.78				DASP 2005, median ROC AUC: 0.80				Assay type	Label
	Lab-assigned sensitivity (%)	Lab-assigned specificity (%)	ROC AUC (%)	AS95 (%)	Lab-assigned sensitivity (%)	Lab-assigned specificity (%)	ROC AUC (%)	AS95 (%)	Lab-assigned sensitivity (%)	Lab-assigned specificity (%)	ROC AUC (%)	AS95 (%)		
105 ^{a,j}	2	99											RIA	125 ⁱ
109 ^f	42	87	0.65	14	48	91	0.77	36	56	100	0.83 (0.74–0.91)	<0.0001	68	125 ⁱ
109									56	95	0.77 (0.68–0.86)	<0.0001	26	125 ⁱ
110	22	98	0.64	30	7	99	0.73	38.6					RIA	125 ⁱ
111	24	96	0.60	24									RIA	125 ⁱ
113 ^f	16	98	0.60	16.3					16	99	0.60 (0.493–0.705)	0.048	22	125 ⁱ
114	44	98	0.83	58	40	98	0.85	62	58	98	0.81 (0.73–0.90)	<0.0001	64	125 ⁱ
115					6 ⁱ	98 ⁱ							RIA	125 ⁱ
116	52	97	0.88	70	34	98	0.87	62	56	97	0.87 (0.80–0.93)	<0.0001	62	125 ⁱ
117	30	93	0.67	24									RIA	125 ⁱ
118	8	99	0.55		14	94	0.60	12	20	98	0.69 (0.59–0.79)	<0.0001	34	125 ⁱ
120	36	98	0.83	58	38	98	0.78	52	38	98	0.71 (0.61–0.81)	<0.0001	44	125 ⁱ
121	50	98	0.81	58	64	99	0.85	74	70	99	0.87 (0.80–0.94)	<0.0001	64	125 ⁱ
123	24	94	0.70	20	36	90	0.74	22					RIA	125 ⁱ
126 ^f	44	97	0.73	44	78	96	0.92	82	74	93	0.87 (0.81–0.93)	<0.0001	40	125 ⁱ
132 ^a	34	99	0.78	44	46	97	0.81	54	46	95	0.80 (0.71–0.89)	<0.0001	56	125 ⁱ
133	62	98	0.77	62	74	90	0.81	34	58	99	0.83 (0.74–0.91)	<0.0001	64	123 ⁱ
133 ^g									54	99	0.80 (0.73–0.88)	<0.0001	54	123 ⁱ
135	22	97	0.64	24	12	97	0.59	16	14	99	0.70 (0.60–0.79)	<0.0001	32	125 ⁱ
137	16	98	0.72	24									RIA	125 ⁱ
138	24	98	0.66	28									RIA	125 ⁱ
140	28	98	0.65	28									RIA	125 ⁱ
140	18	99	0.68	32									RIA	125 ⁱ
140	22	99	0.74	42									RIA	125 ⁱ
140	46	96	0.80	54	56	98	0.89	74	54	93	0.81 (0.73–0.89)	<0.0001	42	125 ⁱ
140 ^h					36	72	0.56	12	36	88	0.60 (0.49–0.71)	0.052	27	125 ⁱ
141													RIA	125 ⁱ
148	62	94	0.81	62	48	98	0.71	48	58	98	0.81 (0.72–0.89)	<0.0001	60	125 ⁱ
149	20	99	0.74	36	20	98	0.78	48	40	92	0.74 (0.65–0.83)	<0.0001	22	125 ⁱ
149 ^b									44	93	0.77 (0.69–0.85)	<0.0001	38	125 ⁱ
150	26	99	0.85	60	28	98	0.76	42	58	97	0.85 (0.77–0.93)	<0.0001	64	125 ⁱ
152	24	98	0.77	50	26	95	0.79	26	22	99	0.76 (0.68–0.85)	<0.0001	32	125 ⁱ
153	56	95	0.82	60	58	96	0.79	58	64	98	0.88 (0.82–0.94)	<0.0001	66	125 ⁱ

^aRSR Ltd, Cardiff, UK (RIA)
^bRSR Kronus, Boise, ID, USA (RIA)
^cMerckodia AB, Uppsala, Sweden (ELISA)
^dPerkin Elmer Life Sciences Wallac Oy, Turku, Finland. Time-resolved immunofluorometric assay (TR-IFMA)
^eYamasa Corporation, Tokyo, Japan
^fNon-competitive RIA
^g1 h incubation assay
^hPolyethylene glycol precipitation assay
ⁱData-tracking errors
^jIncomplete results ('zero' for samples with missing lab values)
^kValues refer to the sensitivity/specificity characteristics of an individual assay in DASP 2005, based on ROC analyses

Fig. 2 The effects of IAA assay format on AUC (a) and AS95 (b) in DASP 2005. A wide variation was seen in the results for both commercial kits and in-house assays. Commercial RIA kits using a competitive assay format (black circles) achieved assay performance comparable with that of the in-house RIA, but those using the non-competitive assay format (white circles) had a low assay performance

concordance of all assays, or of the four assays with AUC and AS95 in the upper quartile ($r^2=0.147$, variance 204, $p<0.0001$, and $r^2=0.786$, variance 711, $p<0.0001$; F test IAA index and IA index vs IB4-IA units, $p<0.0001$; Fig. 4e, f).

Combined ROC curve The median IAA index values for each patient and control sample compiled from 27 assay measurements including the standards provided in DASP 2005 were used to construct a combined ROC curve with AUC 0.89 (CI 95% 0.824–0.957, $p<0.0001$; Fig. 5). Using this combined curve, the AS95 was defined at 70%. The cut-off IAA index value of 1.5 corresponded to a specificity of 98% and a sensitivity of 54%. In comparison, for autoantibodies to GAD (GADA), the AUC was 0.95 (95% CI 0.91–1.0), and at specificity 98% sensitivity was 88%. For IA-2A, the AUC was 0.86 (95% CI 0.78–0.94), and at specificity 98% sensitivity it was 74% [9].

Discussion

In the first DASP proficiency testing performed in 2000 we reported that, in contrast to GADA and IA-2A assays, IAA

Fig. 3 IAA in samples from 50 patients with newly diagnosed type 1 diabetes in the DASP 2005 proficiency evaluation. The rank of individual samples in each assay is plotted against the median rank obtained for all 30 participating assays

Fig. 4 IAA index (a, b), IA index (c, d) and standard-curve-derived IA units (e, f) in samples from 50 patients with newly diagnosed type 1 diabetes in the DASP 2005 proficiency evaluation. Samples are arranged in order of ascending median rank of 27 assays (a, c, e) and of the four assays with AUC and AS95 performances in the upper quartile (b, d, f). Boxes represent the median and interquartile range

microassays generally achieved low sensitivity, with poor inter-laboratory concordance [8]. The overall performance of the IAA microassays, however, improved in the three subsequent proficiency-testing rounds. In particular, there was a stepwise increase in the laboratory-assigned sensitivity, which was fourfold higher in DASP 2005 than in DASP 2000. Although there was some reduction in specificity, the improved sensitivity was associated with improvement of the overall ability to discriminate diabetes and control sera as demonstrated by ROC curve analysis.

Although there was still wide variation between assays in the ranking of IAA levels in different serum samples, there was overall concordance. As expected, this varied with assay sensitivity; assays with the highest AUC and adjusted sensitivity were the most concordant and, as the overall performance of IAA microassays has improved, concordance between laboratories in reporting patients as positive has also been enhanced. Whereas in DASP 2000

Fig. 5 Generalised ROC curve for IAA (solid line), GADA (dotted line), and IA-2A (dashed line) in DASP 2005. These were compiled from 27 IAA assays, 50 GADA assays and 50 IA-2A assays of all proficiency samples. For IAA, the AUC was 0.89 (95% CI 0.82–0.96), and a cut-off IAA index value of 1.5 corresponded to a specificity of 98% and a sensitivity of 54%. For GADA, the AUC was 0.95 (95% CI 0.91–1.0), and at a specificity of 98% the sensitivity was 88%. For IA-2A, the AUC was 0.86 (95% CI 0.78–0.94), and at a specificity of 98% the sensitivity was 74%

only three patients were reported positive in $\geq 75\%$ of assays and six patients were reported positive in $\geq 50\%$ of assays [8], in DASP 2005, nine patients were reported positive by $\geq 75\%$ and 21 by $\geq 50\%$. Three out of four laboratories with AUC and AS95 values in the upper quartile concordantly reported 27 patients as positive.

Some caveats are needed in analysing changes between workshops. First, there were some differences in the laboratories taking part. The improvement in sensitivity and AUC were, however, also seen in the subset of laboratories that took part in three or four workshops. Second, only a minority of serum samples were included in more than one workshop and improvements in sensitivity could potentially be an artefact of differences between the serum samples included.

In all three workshops the highest laboratory-assigned sensitivity, specificity, AUC and AS95 values were achieved in laboratories using in-house radioimmunoassays. This is in accordance with the outcomes from the 4th International Workshop on the Standardization of Insulin Autoantibody Measurements and the first DASP workshop [8, 10]. Time-resolved immunofluorometric

assays, an ELISA kit and the majority of commercial RIA kits performed less well, although one RIA kit did achieve sensitivity and specificity comparable with that of the reliable in-house RIAs. One further commercial RIA kit achieved the highest AUC and AS95 of all DASP 2005 IAA assays, but the laboratory-assigned sensitivity was low (22%), suggesting that the threshold for positivity was inappropriately high. Some kits did not include competition with unlabelled insulin and these performed relatively poorly. Displacement of IAA binding with unlabelled insulin is strongly recommended for laboratories using RIA kits.

The format of the workshops also allowed us to evaluate the potential benefits of introducing an IAA reference standard preparation, as has been done for GADA and IA-2A [11]. The DASP IAA standard and negative sera were tested in all DASP workshops. Using this common standard to quantify levels as an IAA index clearly improved concordance between assays, particularly among those with the highest sensitivity. Sources of this standard are, however, very limited and, because IAA are usually found in young children, there will always be problems with obtaining a sample of large volume from a single patient, particularly at the time of diagnosis. We therefore included serial dilutions of sera obtained from two insulin-treated patients in DASP 2005 to see whether the same improvement could result from using an index derived from an IA-positive serum and to evaluate whether units derived from an IA standard curve by regression analysis could further improve concordance between laboratories. Such serum samples are more readily available, can be diluted, and can therefore be used in several workshops. The IB4.4 dilution was selected because the antibody level was closest to the IAA standard and was designated positive by all participating laboratories. Using the IA index based on this standard and the negative serum, the concordance between assays was clearly improved, though to a lesser extent than was achieved when using an index based on the IAA standard. This may be explained by the lower antibody levels in the selected standard dilution resulting in a higher variance of the calculated indices. The lower IAA levels in the alternative standard (IC9.3) meant that it was reported negative by one laboratory and was thus unsuitable for use as a common standard. Although in a previous workshop, we demonstrated that concordance between laboratories was improved by reporting IAA results in units derived from a standard curve based on the DASP IAA standard (P. Bingley, unpublished data), reporting results in units derived from the standard curve based on dilutions of the IA-positive standard, IB4, did not improve concordance between laboratories. This difference may be due to the lower antibody levels in the top standard of the IA curve, with high variability in the lower binding range—particularly

in less sensitive assays—leading to lower concordance. In addition, extrapolation to values outside the range of the standard curve will have exaggerated differences between laboratories.

The combined ROC curves for IAA, GADA and IA-2A in the 50 patients and 100 control samples in DASP 2005 show that IAA were less sensitive than GADA or IA-2A at the level giving 98% specificity, perhaps reflecting the age distribution of the patients in DASP, which generally includes few young children. As demonstrated in the DASP Insulin Autoantibody Affinity Workshop, the combined measurement of affinity and titre has the potential to markedly improve the sensitivity, specificity and concordance of IAA measurement [12]. At present, the techniques and analyses to determine affinity are cumbersome, but this issue will be addressed again in future DASP workshops.

In summary, the reported DASP proficiency evaluations, involving blinded testing of large numbers of serum samples in a wide range of laboratories, have shown that insulin autoantibody measurement by microassay methods has markedly improved since the first DASP workshop in 2000. Comprehensive comparison of assay performance has shown that both in-house RIAs and commercial RIA kits can achieve high levels of sensitivity, specificity and reproducibility as well as good inter-laboratory concordance, although there is still a wide variation in the reported data. Further improvement in inter-laboratory concordance might be achieved by using the competitive assay format in all laboratories as well as by re-defining the threshold for positivity used in some assays. The introduction of a common IAA index based on the IAA-positive standard and negative control serum provided considerably improved the concordance of results between laboratories. The use of an appropriate dilution of an IA-positive standard with binding characteristics similar to the IAA-positive standard serum may achieve comparable concordance. These results are of great importance when comparing and interpreting data reported from different studies in type 1 diabetes research.

Acknowledgements The Diabetes Antibody Standardization Program is funded at the CDC by PL105-33, 106-310, 106-554, and 107-360 administered by the National Institutes of Health. We are most

grateful to the patients and clinicians who have donated the blood samples that enable the DASP workshops to take place.

Duality of interest statement The authors declare that there is no duality of interest associated with this manuscript.

References

1. Palmer JP, Asplin CM, Clemons P et al (1983) Insulin antibodies in insulin-dependent diabetics before insulin treatment. *Science* 222:1337–1339
2. Dean BM, Becker F, McNally JM et al (1986) Insulin autoantibodies in the pre-diabetic period: correlation with islet cell antibodies and development of diabetes. *Diabetologia* 29:339–342
3. Vardi P, Ziegler AG, Mathews JH et al (1988) Concentration of insulin autoantibodies at onset of type I diabetes. Inverse log-linear correlation with age. *Diabetes Care* 36:936–939
4. Verge CF, Stenger D, Bonifacio E et al (1998) Combined use of autoantibodies (IA-2 autoantibody, GAD autoantibody, insulin autoantibody, cytoplasmic islet cell antibodies) in type 1 diabetes: Combinatorial Islet Autoantibody Workshop. *Diabetes* 47:1857–1866
5. Williams AJK, Bingley PJ, Bonifacio E, Palmer JP, Gale EAM (1997) A novel micro-assay for insulin autoantibodies. *J Autoimmun* 10:473–478
6. Naserke HE, Dozio N, Ziegler AG, Bonifacio E (1998) Comparison of a novel micro-assay for insulin autoantibodies with the conventional radiobinding assay. *Diabetologia* 41:681–683
7. Mire-Sluis AR, Gaines Das R, Lernmark, Participants of the study (2000) The World Health Organization International Collaborative Study for Islet Cell Antibodies. *Diabetologia* 43:1282–1292
8. Bingley PJ, Bonifacio E, Mueller PW, Participating Laboratories (2003) Diabetes antibody standardization program: first assay proficiency evaluation. *Diabetes* 52:1128–1136
9. Törn C, Mueller PW, Schlosser M, Bonifacio E, Bingley PJ, Participating Laboratories (2008) Diabetes antibody standardization program: evaluation of assays for autoantibodies to glutamic acid decarboxylase and islet antigen-2. *Diabetologia* 51:846–852
10. Greenbaum CJ, Palmer JP, Kuglin B, Kolb H, Participating Laboratories (1992) Insulin autoantibodies measured by radioimmunoassay methodology are more related to insulin dependent diabetes mellitus than those measured by enzyme-linked immunosorbent assay: results of the Fourth International Workshop on the Standardization of Insulin Autoantibody Measurement. *J Clin Endocrinol Metab* 74:1040–1044
11. Mire-Sluis AR, Gaines Das R, Lernmark A (2000) The World Health Organization International Collaborative Study for Islet Cell Antibodies. *Diabetologia* 43:1282–1292
12. Achenbach P, Schlosser M, Williams AJK et al (2007) Combined testing of antibody titer and affinity improves insulin autoantibody measurement: diabetes antibody standardization program. *Clin Immunol* 122:85–90