Journal of Immunological Methods 433 (2016) 69–76

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

Research paper

Variability of the IFN-γ ELISpot assay in the context of proficiency testing and bridging studies

Wes Rountree [*], Mark Berrong, Ana M. Sanchez, Thomas N. Denny, Guido Ferrari

Duke Human Vaccine Institute, Duke University Medical Center, Durham, NC, USA

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form 29 January 2016
Accepted 18 March 2016
Available online 25 March 2016

Keywords:
ELISpot
Variability
Poisson
Bridging study
Proficiency testing
EQAPOL

ABSTRACT

Assays that assess cellular mediated immune responses performed under Good Clinical Laboratory Practice (GCLP) guidelines are required to provide specific and reproducible results. Defined validation procedures are required to establish the Standard Operating Procedure (SOP), include pass and fail criteria, as well as implement positivity criteria. However, little to no guidance is provided on how to perform longitudinal assessment of the key reagents utilized in the assay. Through the External Quality Assurance Program Oversight Laboratory (EQAPOL), an Interferon-gamma (IFN-γ) Enzyme-linked immunosorbent spot (ELISpot) assay proficiency testing program is administered. A limit of acceptable within site variability was estimated after six rounds of proficiency testing (PT). Previously, a PT send-out specific within site variability limit was calculated based on the dispersion (variance/mean) of the nine replicate wells of data. Now an overall 'dispersion limit' for the ELISpot PT program within site variability has been calculated as a dispersion of 3.3. The utility of this metric was assessed using a control sample to calculate the within (precision) and between (accuracy) experiment variability to determine if the dispersion limit could be applied to bridging studies (studies that assess lot-to-lot variations of key reagents) for comparing the accuracy of results with new lots to results with old lots. Finally, simulations were conducted to explore how this dispersion limit could provide guidance in the number of replicate wells needed for within and between experiment variability and the appropriate donor reactivity (number of antigen-specific cells) to be used for the evaluation of new reagents. Our bridging study simulations indicate using a minimum of six replicate wells of a control donor sample with reactivity of at least 150 spot forming cells per well is optimal. To determine significant lot-to-lot variations use the 3.3 dispersion limit for between and within experiment variability.

1. Introduction

External Quality Assessment (EQA) or Proficiency Testing (PT) has played a role in laboratory medicine for over 65 years (Belk and Sunderman, 1947; Wootton and King, 1953). The application of PT programs helps to assure that independent laboratories testing the same sample will yield comparable results. The Duke University Human Vaccine Institute (DHVI) has been the central laboratory of the National Institute of Health/National Institute of Allergy and Infectious Diseases (NIH/NIAID) External Quality Assurance Program Oversight Laboratory (EQAPOL) through a Department of Health and Human Services contract since September 2010. As part of this program, EQAPOL has developed and run PT programs for Interferon-gamma (IFN-γ) Enzyme-linked immunosorbent spot (ELISpot), multiparameter Intracellular Cytokine Staining (ICS) Flow Cytometry and Luminex bead-based multiplex cytokine assays (Lynch et al., 2014; Rountree et al., 2014; Sanchez et al., 2014; Staats et al., 2014).

Both IFN-γ ELISpot and ICS assays, performed to evaluate antigen-specific immune responses, have been subjected to validation procedures (Russell et al., 2003; Horton et al., 2007) according to Good Clinical Laboratory Practice (GCLP) guidelines, and the assays have been performed by EQAPOL laboratories accordingly (Sanchez et al., 2014; Staats et al., 2014). In the germinal paper that compared the performance of the ELISpot assay conducted by several different laboratory sites (Cox et al., 2006; Sanchez et al., 2014), several reagents were found to be key in the overall performance of the assay across laboratories with fetal bovine serum (FBS) being one of them (Cox et al., 2006; Janetzki et al., 2008). The initial report, and subsequent experience among the laboratories, indicated that studies addressing between reagent lot-to-lot variation should be performed for cellular assays.

The Clinical and Laboratory Standard Institute (CLSI) released guidelines on how to perform between reagent lot variation experiments that are tailored to the characteristics of assays routinely performed by clinical laboratories (Wayne, 2013). Most of the assays rely on what can be considered "true values" due to the assessment of parameters such as levels of electrolytes or glucose in blood. On the other hand, assays that evaluate antigen-specific cellular responses do

[*] Corresponding author at: 2 Genome Court, MSRB II Building, Room 3013, DUMC, Box 103020, Durham, NC 27710, USA.
E-mail address: wes.rountree@dm.duke.edu (W. Rountree).

http://dx.doi.org/10.1016/j.jim.2016.03.004
0022-1759/© 2016 Elsevier B.V. All rights reserved.

not usually have the ability to rely on a “true” value because of the variability of cellular subsets and because of the within and between variation of antigen-specific cell frequency among donors. The lack of such a “true” value to standardize cellular-based assays that measure immune responses poses a problem when assessing between reagent lot variations, which can impact the quality control of key reagents.

In this manuscript we describe how the evaluation of the results obtained from two NIH-sponsored PT programs conducted by Seracare and EQAPOL (Sanchez et al., 2014) was used to further understand IFN-γ ELISpot assay within experiment (precision) variability and determine a limit of acceptable variability. Furthermore, this limit of acceptable variability was applied to between reagent lot variation (i.e., bridging studies for changes in reagent lots). In order to derive a limit of variability it is necessary to apply a statistical distribution to the experimental data. Distributional assumptions are key to statistical analysis and multiple research articles have described ELISpot data as following a Poisson distribution and have developed methods for positivity criteria (Hudgens et al., 2004; Moodie et al., 2006; Moodie et al., 2010).

Our evaluation of the IFN-γ ELISpot assay variability is based on the Poisson distribution, which has the strict assumption that the variance and mean are equal and represented by λ. If the variance is greater than the mean, then the data are considered overdispersed by a dispersion factor ϕ (thus the variance becomes ϕλ). To determine a dispersion limit, the assumption of a Poisson distribution and an estimate of the dispersion factor are required. The key assumption that these data follow a Poisson distribution was addressed by analyses that modeled if an association exists between donor or reagent and the level of variability defined by dispersion (variance/mean). If no association was found, then the EQAPOL program could institute a non-PT send-out specific dispersion limit based on the observed dispersion factor in the PT data.

The establishment of an overall dispersion limit would provide an easy metric to calculate for within experiment variability. This first step for bridging studies is to determine whether or not the old and new lots have acceptable within experiment variability (precision). The second step is to determine how close the results obtained with the old reagent lot are to the results with the new reagent lot (accuracy). Overall, we asked the question whether or not the dispersion limit could apply to bridging studies for within and between variability to answer the questions of precision and accuracy, respectively. To address this overall question, data from an internal lab control sample were analyzed to compare the within and between experiment dispersion to the dispersion limit. The utility of the dispersion limit for between experiment (lot to lot) variability was subsequently explored via data simulations for comparing reagent lots in hypothetical bridging studies.

2. Methods

2.1. Proficiency testing programs

As previously reported, laboratories participating in the EQAPOL PT programs were required to test PBMC samples provided by EQAPOL in nine replicate wells according to their validated standard operating procedures (SOPs) to evaluate the presence of antigen-specific T cells (Sanchez et al., 2014). Methodology for ELISpot assay and reagents has been previously described (Currier et al., 2002; Sanchez et al., 2014). Data from two separate proficiency testing programs, SeraCare and EQAPOL, were used to establish a limit of variability based on the within donor by reagent (antigens used for cellular stimulation) dispersion (variance/mean). The SeraCare data set has two PT send-outs with each send-out including four donors, three reagents (CEF, CMV, DMSO), and ten sites. The EQAPOL data set used for analysis included five PT send-outs each with three donors, three reagents (CEF, CMV, DMSO), and nine sites. In both programs there were nine replicate wells for each donor by reagent combination and sites were instructed to plate cells at 2 × 10^5 cells per well.

2.2. Establishing the dispersion limit

We used data from PT send-outs 2–6 in the EQAPOL program and PT send-outs 7–8 in the SeraCare program for analysis in establishing the dispersion limit. These send-outs were selected because these were the only data available (SeraCare) and EQAPOL EP1 had higher variability attributed to the first run of the PT program. All data used from the PT programs were from the sites' analysis of their own validated ELISpot assay using PBMCs and peptides provided by the PT provider (i.e., EQAPOL or SeraCare). Sites that performed poorly in the programs, based on the EQAPOL grading criteria (see Table 1) were not included in this analysis because these sites would bias the results with higher variability than should typically be expected for a proficient site. Three sites from each program were removed due to general poor performance over time. Using data from proficient sites should provide a good estimate of the typical variability and be generalizable for sites running the IFN-γ ELISpot assay. These data provide an alternative method for assessing differences in antigen specific reactivity since there is a current lack of "gold standards" for the Elispot assay.

The within experiment dispersion was calculated for each donor by reagent combination using the nine replicate wells of data. For modeling purposes, these dispersions were square root transformed for variance stabilization and correction for moderately positive skewed data. Mixed effects models were used to assess the association between the donor and/or reagent with the corresponding dispersion for the EQAPOL data alone, the SeraCare data alone, and both data sets combined. The square root transformed dispersion was the outcome of the model with fixed effects for donor and reagent. Individual intercepts for PT send-outs by site were modeled as random effects to account for PT send-outs and site differences. First, an interaction term for donor by reagent was put in the model. If this was not significant at the alpha 0.05 level, it was dropped from the final model. If donor and reagent were not associated with the dispersion at the alpha 0.05 level, then the assumption of Poisson data would be reasonable to conclude.

2.3. Control data review

To assess the applicability of the dispersion limit for bridging studies, nearly 3 years (October 2011–August 2014) of control data (same donor samples tested across multiple dates) from the EQAPOL central laboratory were used to calculate the within and between experiment dispersions. There were 53 different experiments used for this analysis, and the between and within experiment dispersions were calculated using a mixed effects model with a random intercept for each experiment to estimate the between experiment variance and a repeated statement for experiment as well to estimate the within experiment variance. The model estimated conditional mean of all experiments was used as the denominator for the calculation of the dispersion.

2.4. Simulation studies

2.4.1. Estimate the number of replicate wells needed

Having enough information to determine whether a set of experimental data meets established criteria is very important. To determine the number of replicate wells needed for within experiment variability, we simulated ELISpot Poisson data with 1000 replicates under the following conditions:

- number of wells ranged from 3 to 18 by an increment of 3
- true dispersion factor ϕ ranged from 1.5 to 6.5 by an increment of 0.5
- average spot forming cells per well (SFC/well) was 100.

These count data were generated in SAS (Cary, NC) using the RAND Function for Poisson and Negative Binomial distributions (Rodríguez,

Table 1
ELISpot assay proficiency criteria and targets.

Criteria	Description	Target	Points
Timeliness	On time PT valid data and questionnaire upload	Per due date set by EQAPOL	• 10 points total
PBMC Processing	Viability: Pass/Fail per each donor sample Recovery: Pass/Fail per each donor sample	Donor must be >80% for D1 Donor must be between 70 and 120% for D1	• 9 points total • 1.5 points available per donor for viability • 1.5 points available per donor for recovery
Background	• Pass/Fail per each donor sample • Results from In-house assay only	Average for each donor must be <10	• 9 points total • 3 points available per donor
Accuracy	• Assessed for each donor sample for each stimulation condition (CMV and CEF) • Results from In-house assay only	Site's average for 9 wells must not be significantly different from consensus average	• 54 points total • 9 points deduction for each donor by stimulation condition significantly different from consensus average
Precision	• Assessed for each donor sample for each stimulation condition (CMV and CEF) • Results from In-house assay only	The dispersion must be less than/equal to 3.3	• 18 points total • 3 points deducted for each donor/stimulation outside of range

2007). The probability of exceeding the 3.3 dispersion limit under these various conditions was calculated (i.e., power calculations).

2.4.2. Donor reactivity (number of antigen-specific cells for a sample)

The level of reactivity could impact the determination of lot-to-lot differences for between experiment variability. To answer this question, simulations were performed with 500 replicates at reference (old) mean lot averages of 20 to 200 by increments of 10 with 9 replicate wells for following scenarios:

- Scenario 1 — true ϕ of 1.5 for reference and new reagents and true difference of ±25%, the standard for comparison
- Scenario 2 — true ϕ of 1.5 for reference and new reagents and true difference of ±20%, to assess impact of lower true difference
- Scenario 3 — true ϕ of 3.0 for reference and new reagents and true difference of ±25%, to assess impact of higher variability.

Simulations were also performed for two more scenarios, to further assess the number of replicate wells needed for within experiment variability along with reactivity, given a true ϕ of 1.5 for the reference and new reagents and true difference of ±25% with:

- Scenario 4 — 6 replicate wells
- Scenario 5 — 3 replicate wells.

These simulations assume that a 25% or 20% difference in reagent lots is a significant difference. Therefore, the probability/power of finding that new reagent lots differ from reference lots (accuracy of the new lot) was estimated using four metrics:

1. a two standard deviation boundary (2SD) based on the reference lot
2. a three standard deviation boundary (3SD) based on the reference lot
3. a 3.3 between lot (reference and new) dispersion boundary calculated using a mixed effects model (see Annex B of ISO 13528 (ISO, 2005) for calculation of non mixed effects model-based between lot variability)
4. a 95% exact Poisson confidence boundary based on the reference lot.

The exact Poisson 95% confidence interval (CI) was calculated using (Garwood, 1936)

$$\text{Lower bound} = \chi^2 \frac{\left(\frac{\alpha}{2}, 2x\right)}{2} \cdot \text{Upper bound} = \chi^2 \frac{\left(1-\left(\frac{\alpha}{2}\right), 2x+1\right)}{2}$$

where x = the count (i.e., average of the nine wells in this case) and $\alpha = 0.05$. The boundaries are based on a chi-square distribution where α defines the percentile and x the degrees of freedom.

3. Results

3.1. Establishing a dispersion limit

In the EQAPOL ELISpot PT program the grading criteria for precision was based on an upper limit of acceptable within lab dispersion for each donor by reagent combination. This limit of dispersion varied from 2.8–4.2 through five PT send-outs, and it was decided that enough information was available to determine if a non-PT send-out specific dispersion limit could be established for use in future PTs.

The donor reactivity (average SFC/well based on 2×10^5 cells per well) of the samples provided in the EQAPOL program is highly variable across the reagents and donors. The average site SFC/well, for donor by reagent combinations, ranged from 1 to 593, and the median values ranged from 1 to 609 (see Table 2/Fig. 1A and B). Despite the wide range of average SFC/well, the dispersion levels were rather constant across the reagents. The average site dispersion, for donor by reagent combinations, ranged from 0.88 to 2.01 and the median values ranged from 0.82 to 1.52 (see Table 3/Fig. 2A and B). These summary statistics suggested that the dispersion for a particular donor by reagent combination was not associated with the average SFC/well for that combination.

For each set of data analyzed, PT send-outs 2–6 in the EQAPOL program and PT send-outs 7 and 8 in the SeraCare program, there were no significant interactions at the alpha = 0.05 level between donor and reagent. The interaction terms were dropped for the final

Table 2
Summary of average SFC per well for PTID by reagent in EQAPOL 2–6 and SERACARE 7 & 8.

			Average SFC per well statistics		
Reagent	PTID	N	Mean	Median	Maximum
CEF	BX000155	176	2	2	11
	BX000169	176	9	8	31
	BX000187	86	334	325	795
	BX000199	90	520	522	716
	PPD32	396	20	19	43
	PPD79	387	287	298	426
	PPD90	396	5	4	39
CMV	BX000155	176	1	1	9
	BX000169	176	191	197	430
	BX000187	86	337	353	609
	BX000199	90	593	609	883
	PPD32	394	78	79	121
	PPD79	358	451	453	658
	PPD90	394	4	3	32
DMSO	BX000155	177	1	1	7
	BX000169	174	13	1	249
	BX000187	84	66	8	535
	BX000199	90	2	1	10
	PPD32	396	2	2	15
	PPD79	396	2	1	29
	PPD90	395	3	2	14

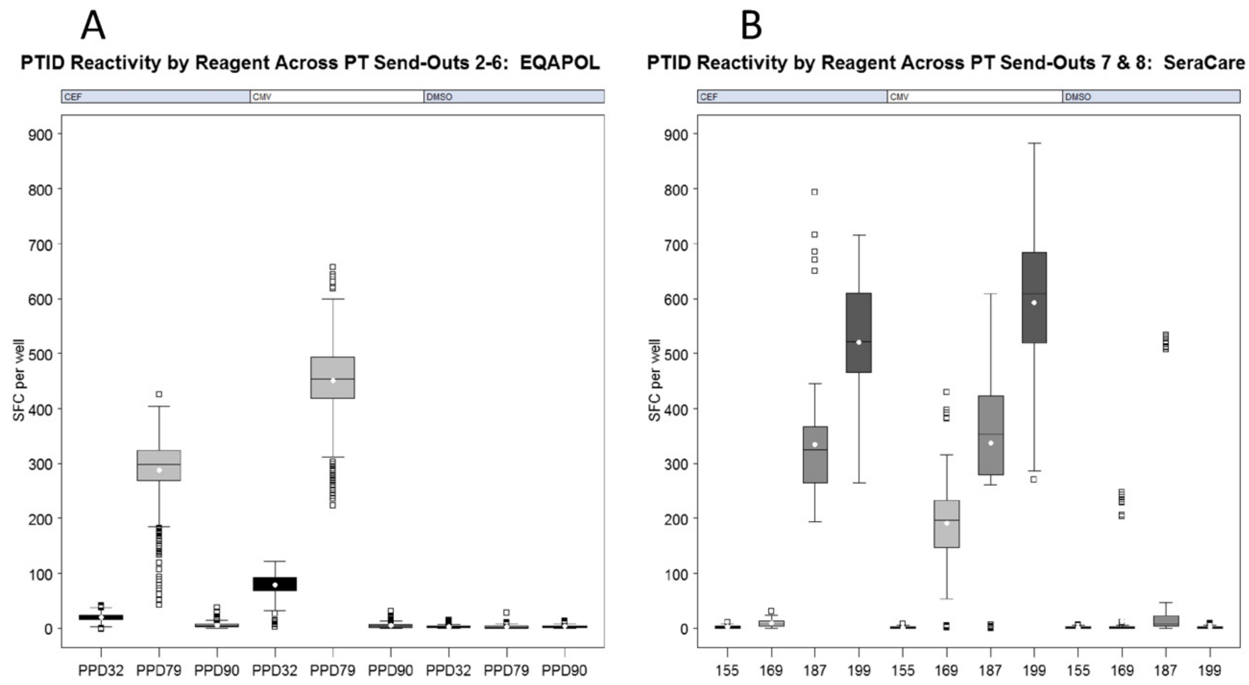

Fig. 1. PTID reactivity by reagent A) The average SFC/well for each donor/PTID across the three reagents: CEF, CMV, and DMSO for PT send-outs 2–6 for EQAPOL B) The average SFC/well for each donor/PTID across the three reagents: CEF, CMV, and DMSO for PT send-outs 7 and 8 for SeraCare.

models. For the EQAPOL data alone (N = 390), the p-value for the fixed effect of donor was 0.8410 and for the fixed effect of reagent it was 0.1373. For the SeraCare data alone (N = 172), the p-value for the fixed effect of donor was 0.4855 and for the fixed effect of reagent it was 0.3864. For the all data combined (N = 572), the p-value for the fixed effect of donor was 0.8029 and for the fixed effect of reagent it was 0.1605. This analysis did not prove that there is not an association between donor or reagent with the dispersion because the null hypothesis is not proven if the p-value is not significant. However, data from seven PT send-outs across two different programs were used, and no evidence of an association was discovered. Therefore, it did not seem unreasonable to assume a Poisson distribution.

Given that ELISpot data from these PT programs follow a Poisson distribution, the next step was to calculate the dispersion factor. Recall the Poisson assumption that the variance and mean are equal and represented by λ. If the variance is greater than the mean, then the data are considered overdispersed by a dispersion factor ϕ (thus the variance becomes $\phi\lambda$). The average dispersion factor was 1.35 using all data (for EQAPOL $\phi = 1.35$ and for SeraCare $\phi = 1.37$). The dispersion limit was calculated as an upper one-sided 95% confidence boundary of the dispersion factor. For the EQAPOL data alone this upper boundary was 3.3 and for the SeraCare data alone it was 2.9 and with all data it was 3.2. For the EQAPOL program the dispersion limit was set at 3.3 for the purposes of grading for within experiment precision of the nine replicate wells. This dispersion limit has been used for the last three EQAPOL PT send-outs (PT send-outs 7–9). The previous method of calculating the within experiment precision yielded a dispersion limit of 3.5 for PT send-out 8 and 3.5 for PT send-out 9.

Table 3
Summary of dispersions for PTID by reagent in EQAPOL 2–6 and SERACARE 7 & 8.

			Dispersion statistics		
ID	N	Reagent	Median	Mean	Maximum
BX000155	20	CEF	1.27	1.29	2.23
	20	CMV	1.24	1.26	2.55
	20	DMSO	1.05	1.03	1.88
BX000169	20	CEF	1.15	1.15	2.27
	20	CMV	1.52	1.65	4.19
	20	DMSO	1.22	1.61	4.41
BX000187	10	CEF	1.42	1.74	4.49
	10	CMV	1.01	1.35	4.33
	10	DMSO	0.82	0.88	2.00
BX000199	10	CEF	1.37	2.01	7.10
	10	CMV	1.18	1.32	3.71
	10	DMSO	1.07	1.15	2.31
PPD32	44	CEF	1.09	1.19	4.00
	44	CMV	1.01	1.39	6.98
	44	DMSO	1.19	1.35	2.75
PPD79	43	CEF	0.92	1.34	8.00
	40	CMV	1.02	1.71	9.11
	44	DMSO	0.99	1.17	6.66
PPD90	44	CEF	1.05	1.48	6.89
	44	CMV	1.13	1.50	6.19
	44	DMSO	0.99	1.05	2.84

3.2. Evaluation of control data

Next the within and between experiment dispersion for data from a control sample run by the EQAPOL Central Laboratory for the last 3 years was calculated to compare to the established 3.3 dispersion limit. This was done to determine if the dispersion limit could apply to between experiment variability for assessment of accuracy in bridging studies. The control sample was assayed with two reagents, CEF and CMV, during the time of evaluation. The estimated between experiment dispersion for CEF was 0.79, and the estimated within experiment dispersion was 1.27 with the model based average SFC/well at 148. The estimated between experiment dispersion for CMV was 1.23, and the estimated within experiment dispersion was 1.42 with the model based average SPF per well at 150. These dispersions were well below the 3.3 EQAPOL limit. However, this is not unexpected because these data all passed the laboratory standard acceptance criteria.

Therefore, we evaluated the results from a control sample tested in each of the EQAPOL PTs, and the between and within PT dispersions for CEF were 3.3 and 1.6, respectively. For CMV the between and within PT dispersions were 4.1 and 1.6, respectively. The between PT dispersions were at (CEF) or slightly above (CMV) the EQAPOL dispersion limit, which implied that the dispersion limit might be applicable for accuracy via the between experiment variability. Therefore the use of

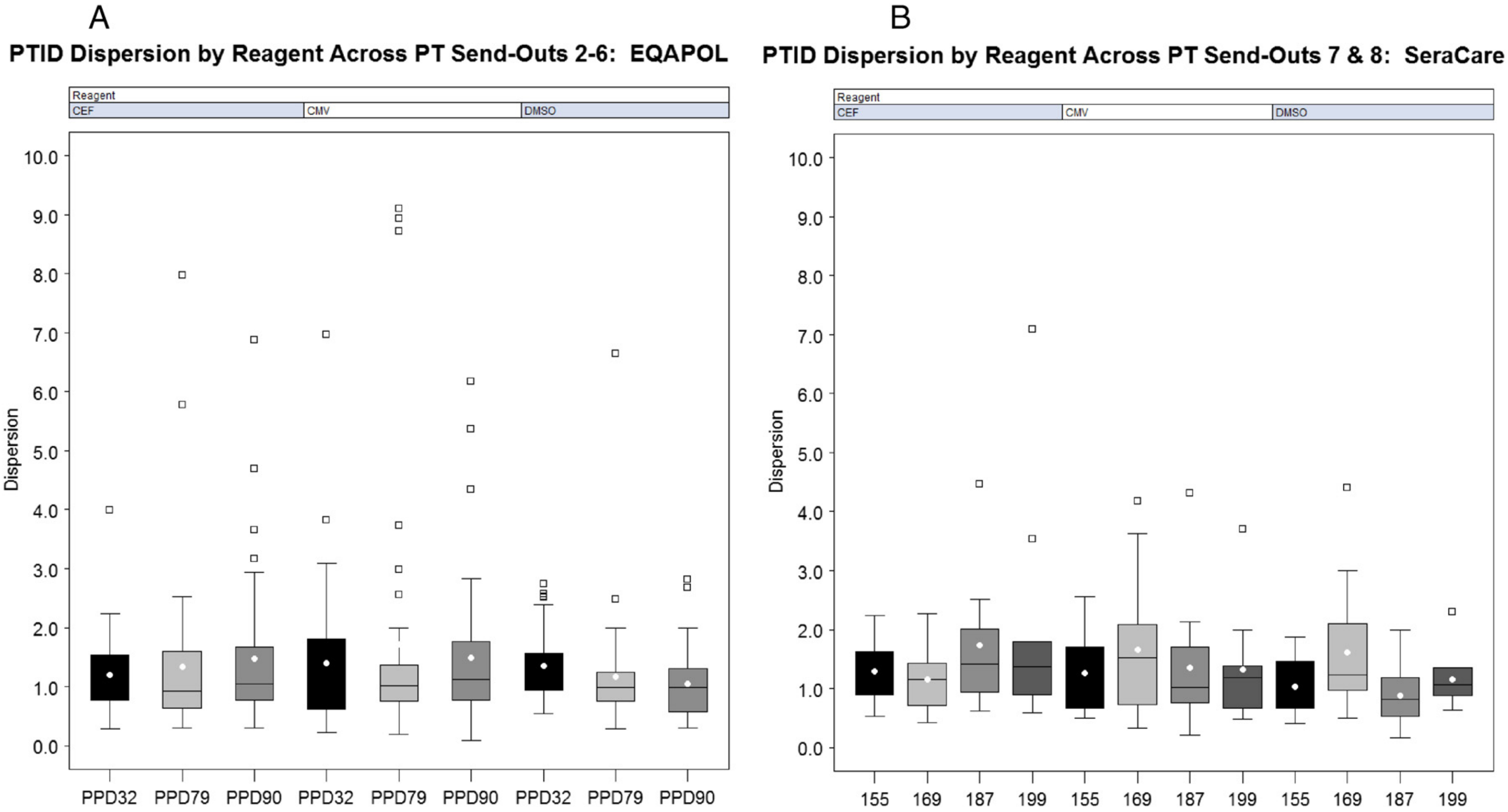

Fig. 2. PTID dispersion by reagent A) The dispersion for each donor/PTID across the three reagents: CEF, CMV, and DMSO for PT send-outs 2–6 for EQAPOL B) The dispersion for each donor/PTID across the three reagents: CEF, CMV, and DMSO for PT send-outs 7 and 8 for SeraCare.

this dispersion limit was further evaluated for between lot variability in bridging studies of new reagent lots under hypothetical scenarios.

3.3. Application to bridging studies

3.3.1. Number of wells for within experiment precision

The CLSI guideline I/LA26-A2 (CLSI, 2013) indicates that six replicate wells containing medium only should be used to increase the statistical power for positivity responses. However, there are no guidelines on the number of wells that should be used when conducting a bridging study nor are their guidelines for the number of wells to be used for wells containing stimuli. A survey conducted among the laboratories participating in the EQAPOL ELISpot PT program revealed that 2 to 4 replicate wells have commonly been used to evaluate the lot-to-lot variations for key reagents. Assuming that ELISpot bridging data should be similar to ELISpot PT send-out data and that a 3.3 dispersion limit is appropriate, the number of wells to be used can be evaluated via simulations.

The results of these simulations are shown in Fig. 3. As the number of wells increases, the probability/power to detect dispersions over 3.3 also increases. If the true dispersion is only 1.5 then power is less than 2% with 9 or more wells. However, with only 3 wells there is an 11% chance of making a type I error of rejecting a true null hypothesis. At the other extreme, if the true dispersion is 6.5 then the power with 9 or more wells is over 80%, with 6 wells providing 78% power, and only 64% power with 3 wells.

3.3.2. Donor reactivity (average SFC/well)

Laboratories must switch reagents because their supplies have been exhausted or expired, and the new reagents should provide very similar reactivity as the old reagents for a control sample. In performing these tests, no guidelines exist as to what sample reactivity is best for detecting a true difference. Therefore, the relationship between reactivity and the probability of detecting true differences (i.e., power) for reference (old) and new reagent lots was assessed. If reactivity impacts the ability to detect true differences, then an appropriate reactivity should be used. For this analysis, 5 scenarios were evaluated (see Table 4). These scenarios assume a 25% difference, or 20% for Scenario 2, is significant and standard methods of comparing lots using a 2 or 3 SD boundary are presented along with the 3.3 dispersion limit and Poisson CI.

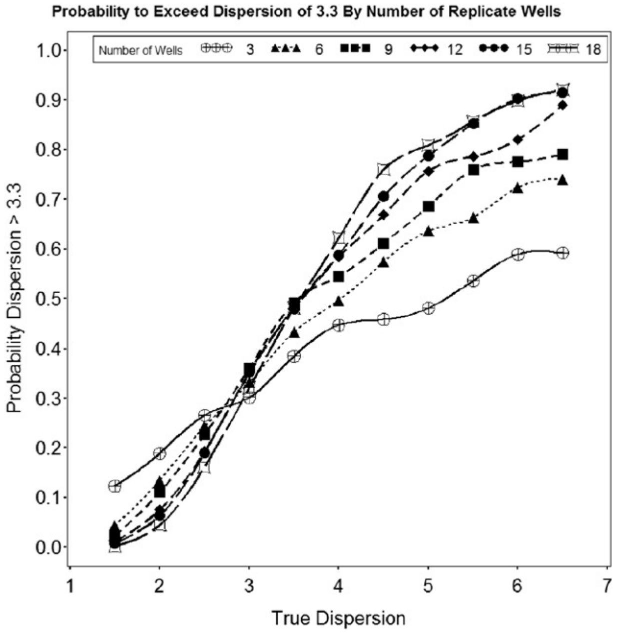

Fig. 3. Probability to Exceed the 3.3 Dispersion Limit for Number of Replicate Wells The power to detect dispersions of greater than 3.3 under various levels of true dispersion and for the number of replicate wells ranging from 3 to 18 by increments of 3.

Table 4
Summary of simulations for scenarios 1–5.

Scenario	True dispersion factor	Replicate wells	New lot change	SFC/well at 80% power or power at 200 SFC/well			
				Poisson 95% CI	3.3 Limit	2SD Limit	3SD Limit
1	1.5	9	±25%	110	150	150	50%
2	1.5	9	±20%	180	65%	72%	46%
3	3	9	±25%	130	190	57%	17%
4	1.5	6	±25%	120	170	170	54%
5	1.5	3	±25%	140	77%	77%	54%

Fig. 4A and B show the probability of exceeding the four separate between lot boundaries for Scenario 1 (the standard for comparison) when the new reagent lot is either 25% higher or 25% lower. The exact Poisson CI provides the highest power across the range of average SFC/well for the old reagent lots, regardless if the new lots are 25% higher or lower and reaches 80% power with an average SFC/well of 110 for the reference reagent lots. The 2SD boundary has higher power than the 3.3 boundary until the average SFC/well for the reference reagent lots is 150 and the power for either boundary is approximately 80%. Using a 3SD boundary provides little power with the highest at approximately 50% with average SFC/well for the reference reagent lots at 200.

Under Scenario 2 where the true difference is only 20% higher or lower for the new reagent lots, the Poisson CI still provides the highest power but does not reach 80% power until an average SFC/well is 180 for the reference reagent lots. The 2SD boundary has higher power than the 3.3 boundary with approximately 72% power versus 65% power with an average SFC/well of 200 for the reference reagent lots.

In Scenario 3, the true dispersion of the reference and new reagent lots is 3.0 instead of 1.5. With the increased variability the power of the 2SD and 3SD boundaries are diminished. If average SFC/well for the reference reagent lots is 200 the power for the 2SD boundary is approximately 57% and the 3SD only reaches 17%. The exact Poisson CI again provides the highest power and reaches 80% at an average SFC/well of 130 for the reference reagents regardless if the new reagents are 25% higher or lower. The 3.3 boundary reaches 80% power at an average SFC/well of 190 for the reference reagents (see Table 4).

Scenarios 4 and 5 were conducted to compare the four boundaries and also further address the question of whether the number of replicate wells for within experiment variability impacted the power to conclude the new and reference reagents were different given the conditions in Scenario 1. With only 6 wells the exact Poisson CI reaches 80% power with an average SFC/well of 120 for the reference reagent lots. The 2SD boundary has higher power than the 3.3 boundary until the average SFC/well is 170 for the reference reagent lots and the power for either boundary is approximately 80%. Using a 3SD boundary provides little power with the highest at approximately 54% with average SFC/well of 200 for the reference reagent lots. With only 3 wells the exact Poisson CI reaches 80% power with an average SFC/well of 140 for the reference reagent lots. The 2SD boundary has higher power than the 3.3 boundary until the average SFC/well for the old reagent lots is 200 and the power for either boundary is approximately 77%. Using a 3SD boundary provides little power with the highest at approximately 54% with average SFC/well of 200 for the reference reagent lots.

4. Discussion

Based on our analyses of SeraCare and EQAPOL PT data and the application of the 3.3 dispersion limit for EQAPOL PT send-outs 8 and 9, we determined that the dispersion limit represents a good boundary for within experiment precision given nine replicate wells of data in a

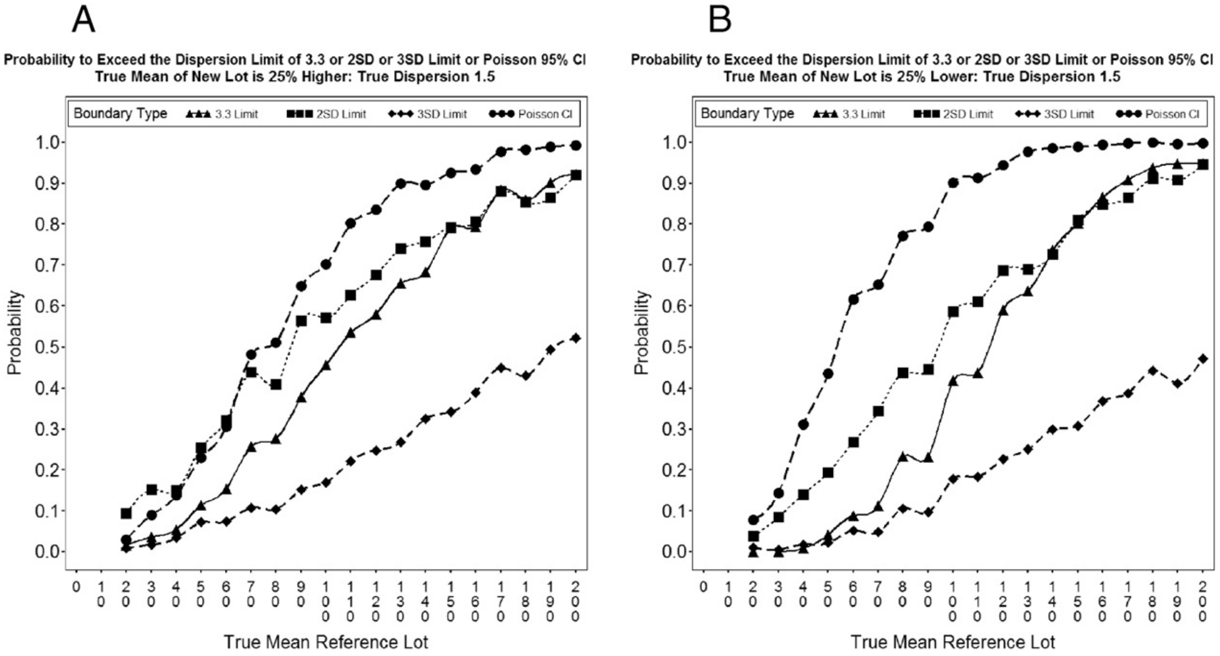

Fig. 4. Bridging study simulations A) Simulations with new reagent lot 25% higher, dispersion is 1.5, 9 replicate wells B) Simulations with new reagent lot 25% lower, dispersion is 1.5, 9 replicate wells.

proficiency testing setting. It was important to test the dispersion limit in other settings to explore if this boundary could have application outside of the PT setting. To this end, control data were reviewed and the within and between experiment dispersions were below the dispersion limit. The within variability explains the precision of the reagent lot and the between variability explains the accuracy of the new reagent lot to the reference. Therefore, the dispersion boundary was assessed via simulations for use in bridging studies to provide guidance on the number of replicate wells to use and the reactivity of the donor sample for comparing reference and new reagent lots.

Based on the probability of exceeding the dispersion limit of 3.3 for various numbers of replicate wells, using only three replicate wells of data does not provide enough information to accurately assess the within experiment precision. Therefore, it is recommended that at least six or more replicate wells of data should be used for assessment of within experiment precision using a dispersion limit of 3.3. Calculation of the within experiment dispersion is straightforward, and data from the EQAPOL and Seracare programs indicate that dispersion is not related to level of reactivity. This is not the case when using the coefficient of variation (CV), which does vary depending on the average SFC/well. It is more appropriate to use the dispersion of IFN-γ ELISpot assay data than the CV when making assessments of within experiment precision.

Since there is no gold standard for accuracy in the ELISpot assay, we compare the dispersion limit and Poisson 95% CI to common boundaries of 2 or 3 SD in simulated bridging studies. Data were simulated where the true variability (dispersion) and mean of the new and reference reagents was known and the data follow a Poisson distribution. Fig. 4A and B show that the exact Poisson 95% CI provides the most power given a 25% difference in the reference and new reagent means (average SFC/well). This was the case for all five scenarios that were evaluated. The 2SD boundary has slightly more power than the 3.3 dispersion limit. However, the power of these two boundaries converges at 80% with a mean of 150 for the reference reagents. Similar trends were found in Scenario 2 that had a true mean difference of 20% and Scenario 3 shows that as the within experiment variability increases the 2SD boundary gets wider and power declines. Thus the exact 95% Poisson CI or the 3.3 dispersion limit provides better power when there is more variability. Scenarios 4 and 5 support the findings of the within experiment precision that three replicate wells are not sufficient information since neither the 2SD or 3.3 dispersion limit reaches 80% power even with the means SFC/well at 200.

Based on the data from the EQAPOL program, our overall recommendations for bridging studies of new reagent lots would be to use at least six replicate wells of data with a control sample with reactivity of at least 150 SFC/well. The use of a 3SD boundary for accuracy appears too lenient, whereas a 2SD or dispersion limit of 3.3 provides similar power when the average SFC/well is 150 or higher and the within experiment dispersion is around 1.5. If the within experiment dispersion is around 3.0, the dispersion limit of 3.3 should be used. Overall, the exact Poisson 95% CI is the most stringent and could be used if the new lot must be very close (<20%) to the old lot. The EQAPOL dispersion limit of 3.3 should be used to flag samples that potentially have too much variability and indicate that the experiment or bridging study be repeated.

Acknowledgments

This project has been funded with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract Nos. HHSN272201000045C. We are grateful to Jim Lane at NIAID/DAIDS for his leadership and guidance. We are grateful for the EQAPOL Scientific Advisory Board members Bob Coombs, David Barnett, Maurice R.G. O′ Gorman, Holden T. Maecker, and Marco Schito for their input. We also want to acknowledge the input of the EQAPOL ELISpot program-specific advisory committee members: Alexandra Schuetz, Josephine Cox, and Nicole Frahm.

References

Belk, W.P., Sunderman, F.W., 1947. A survey of the accuracy of chemical analyses in clinical laboratories. Am. J. Clin. Pathol. 17, 853–861.

CLSI, 2013. I/LA26-A2: performance of single cell immune response assays; approved guideline—second edition. Clin. Lab. Stand. Inst. 106.

Cox, J.H., Ferrari, G., Janetzki, S., 2006. Measurement of cytokine release at the single cell level using the ELISPOT assay. Methods 38, 274–282.

Currier, J.R., Kuta, E.G., Turk, E., Earhart, L.B., Loomis-Price, L., Janetzki, S., Ferrari, G., Birx, D.L., Cox, J.H., 2002. A panel of MHC class I restricted viral peptides for use as a quality control for vaccine trial ELISPOT assays. J. Immunol. Methods 260, 157–172.

Garwood, F., 1936. Fiducial limits for the Poisson distribution. Biometrika 28, 437–442.

Horton, H., Thomas, E.P., Stucky, J.A., Frank, I., Moodie, Z., Huang, Y., Chiu, Y.L., McElrath, M.J., De Rosa, S.C., 2007. Optimization and validation of an 8-color intracellular cytokine staining (ICS) assay to quantify antigen-specific T cells induced by vaccination. J. Immunol. Methods 323, 39–54.

Hudgens, M.G., Self, S.G., Chiu, Y.L., Russell, N.D., Horton, H., McElrath, M.J., 2004. Statistical considerations for the design and analysis of the ELISpot assay in HIV-1 vaccine trials. J. Immunol. Methods 288, 19–34.

ISO, 2005. Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons.

Janetzki, S., Panageas, K.S., Ben-Porat, L., Boyer, J., Britten, C.M., Clay, T.M., Kalos, M., Maecker, H.T., Romero, P., Yuan, J., Kast, W.M., Hoos, A., Elispot Proficiency Panel of the, C.V.C.I.A.W.G., 2008. Results and harmonization guidelines from two large-scale international Elispot proficiency panels conducted by the Cancer Vaccine Consortium (CVC/SVI). Cancer Immunol. Immunother. 57, 303–315 (CII).

Lynch, H.E., Sanchez, A.M., D'Souza, M.P., Rountree, W., Denny, T.N., Kalos, M., Sempowski, G.D., 2014. Development and implementation of a proficiency testing program for Luminex bead-based cytokine assays. J. Immunol. Methods 409, 62–71.

Moodie, Z., Huang, Y., Gu, L., Hural, J., Self, S.G., 2006. Statistical positivity criteria for the analysis of ELISpot assay data in HIV-1 vaccine trials. J. Immunol. Methods 315, 121–132.

Moodie, Z., Price, L., Gouttefangeas, C., Mander, A., Janetzki, S., Lower, M., Welters, M.J., Ottensmeier, C., van der Burg, S.H., Britten, C.M., 2010. Response definition criteria for ELISPOT assays revisited. Cancer Immunol. Immunother. 59, 1489–1501 (CII).

Rodríguez, G., 2007. Lecture Notes on Generalized Linear Models. (Available at) http://data.princeton.edu/wws509/notes/.

Rountree, W., Vandergrift, N., Bainbridge, J., Sanchez, A.M., Denny, T.N., 2014. Statistical methods for the assessment of EQAPOL proficiency testing: ELISpot, Luminex, and Flow Cytometry. J. Immunol. Methods 409, 72–81.

Russell, N.D., Hudgens, M.G., Ha, R., Havenar-Daughton, C., McElrath, M.J., 2003. Moving to human immunodeficiency virus type 1 vaccine efficacy trials: defining T cell responses as potential correlates of immunity. Asian J. Infect. Dis. 187, 226–242.

Sanchez, A.M., Rountree, W., Berrong, M., Garcia, A., Schuetz, A., Cox, J., Frahm, N., Manak, M., Sarzotti-Kelsoe, M., D'Souza, M.P., Denny, T., Ferrari, G., 2014. The External Quality Assurance Oversight Laboratory (EQAPOL) proficiency program for IFN-gamma enzyme-linked immunospot (IFN-gamma ELISpot) assay. J. Immunol. Methods 409, 31–43.

Staats, J.S., Enzor, J.H., Sanchez, A.M., Rountree, W., Chan, C., Jaimes, M., Chan, R.C., Gaur, A., Denny, T.N., Weinhold, K.J., 2014. Toward development of a comprehensive external quality assurance program for polyfunctional intracellular cytokine staining assays. J. Immunol. Methods 409, 44–53.

Wayne, P., 2013. User evaluation of between-reagent lot variation; approved guideline. CSLI CLSI document EP26-A.

Wootton, I.D., King, E.J., 1953. Normal values for blood constituents; inter-hospital differences. Lancet 1, 470–471.