A computational solution to improve biomarker reproducibility during long-term projects

Feng Feng[1¶], Morgan P. Thompson[2¶], Beena E. Thomas[2#], Elizabeth R. Duffy[3], Jiyoun Kim[3], Shinichiro Kurosawa[3], Joseph Y. Tashjian[2,3], Yibing Wei[2,3], Chris Andry[2,3], D.J. Stearns-Kurosawa[3*]

[1] Department of Microbiology, Boston University School of Medicine, Boston, Massachusetts, United States of America

[2] Department of Pathology and Laboratory Medicine, Boston Medical Center, Boston, Massachusetts, United States of America

[3] Department of Pathology and Laboratory Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America

[¶]These authors should be considered equal first authors

*Corresponding author:

Email: dstearns@bu.edu (DSK)

[#]Current Address; Beth Israel Deaconess Medical Center Boston, MA

Abstract

Biomarkers are fundamental to basic and clinical research outcomes by reporting host responses and providing insight into disease pathophysiology. Measuring biomarkers with research-use ELISA kits is universal, yet lack of kit standardization and unexpected lot-to-lot variability presents analytic challenges for long-term projects. During an ongoing two-year project measuring plasma biomarkers in cancer patients, control concentrations for one biomarker (PF) decreased significantly after changes in ELISA kit lots. A comprehensive operations review pointed to standard curve shifts with the new kits, an analytic variable that jeopardized data already collected on hundreds of patient samples. After excluding other reasonable contributors to data variability, a computational solution was developed to provide a uniform platform for data analysis across multiple ELISA kit lots. The solution (*ELISAtools*) was developed within open-access R software in which variability between kits is treated as a batch effect. A defined best-fit Reference standard curve is modelled, a unique Shift factor "S" is calculated for every standard curve and data adjusted accordingly. The averaged S factors for PF ELISA kit lots #1-5 ranged from -0.086 to 0.735, and reduced control inter-assay variability from 62.4% to <9%, within quality control limits. S factors calculated for four other biomarkers provided a quantitative metric to monitor ELISAs over the 10 month study period for quality control purposes. Reproducible biomarker measurements are essential, particularly for long-term projects with valuable patient samples. Use of research-use ELISA kits is ubiquitous and judicious use of this computational solution maximizes biomarker reproducibility.

Introduction

In virtually every research project with real or potential clinical application, biomarkers provide valuable data to monitor presence or progression of disease, as well as therapeutic susceptibility or efficacy. Biomarker data monitor defined outcomes and there is considerable discussion about whether investigators should disclose incidental research findings to study participants [1-3]. Intrinsic to this discussion is the need for reproducible study data and this presents challenges, particularly with long-term studies. Data rigor and reproducibility is a systemic problem [4] and a priority issue for the NIH. It is an analytic challenge for long-term studies because research laboratories often do not have standardized operations nor validated biomarker assay reagents that adhere to the quality assurance and quality control standards for diagnostic use as required by the Clinical Laboratory Improvement Act (CLIA; www.cdc.gov/clia).

Protein biomarkers are measured frequently in plasma, serum or other matrices by solid phase Enzyme-Linked Immunosorbent Assay (ELISA) methods in which the antigen of interest in the sample is bound by antibodies and the amount of bound antigen is proportional to the signal strength that develops in the assay. There have been only 136 *in vitro* diagnostic ELISA kits or kit components cleared or approved by the FDA since 2000, but there are hundreds of commercially available ELISA kits labeled for “research purposes only” from dozens of vendors. Typically these are vetted by the manufacturer for sensitivity, selectivity, intra/inter-assay variability, stability and storage needs, but they are not required to adhere to federal CLIA guidelines. Lot-to-lot variability between ELISA kits is either not relevant or is manageable for short-term projects, but challenges arise for quality assurance when multiple lots of research ELISA kits are used in long-term

projects. A research laboratory may operate under NIH biomarker recommendations [5] and Good Clinical Laboratory Practice guidelines with appropriate training, auditing, assay validation and proficiency testing [6], but it does not have jurisdiction over kit reagents controlled by the manufacturer. In general practice, expected inter-assay coefficients of variation (CV) for ELISA standard curves will be within 10~20%, but if a commercial ELISA standard curve suddenly shifts significantly with a new kit lot and validated control data shifts outside limits, then the current patient biomarker concentrations cannot be compared with those quantified months earlier. An experienced laboratory can rescue data from one ELISA with a failed standard curve [7], but long-term quality assurance poses other challenges. Either all the samples have to be re-assayed, which introduces variables such as freeze/thaw issues, or the data discarded, all of which wastes precious patient samples and resources.

This was encountered by our research group with one commercial ELISA kit during an on-going two year project to evaluate effects of pre-analytic variables on plasma thrombosis biomarkers in patients. The project developed thirty-five standard operating procedures (SOPs) that define and document operations from blood acquisition to transport, processing, assay and storage. Nine biomarkers are quantified by research-use ELISA kits. For one biomarker, biomarker “PF”, after months of reproducible assays, we observed a significant shift of our standard curves and internal control results when assayed with new kits lots. The scope of the problem was revealed by review of data from 5 kit lots over 10 months and 65 ELISA plates. Rigorous review by the quality assurance team did not identify laboratory or operational pre-analytic contributions and similar changes were not observed with the other ELISAs. The project had encountered an unexpected analytic variable and the data from 420 plasma samples collected over 10

95 months could not be compared. The manufacturer was responsive but ultimately unable
96 to resolve the problem.

97 To rescue our patient data, we developed a computational solution with a
98 sufficiently generalized approach such that it may be used by others facing a similar
99 situation. In the solution, the lot-to-lot variability in ELISA kits is treated as a batch effect,
100 and a defined Reference standard curve is modelled with either a four- or five-parameter
101 logistic function. Based on this Reference curve, a Shift factor ("S") can be calculated and
102 applied retrospectively to every standard curve from every ELISA plate over many
103 months, and the biomarker concentrations for that plate are adjusted accordingly. In this
104 way, the data collected from many ELISA plates over many months can be compared on
105 a uniform platform. Once instituted, calculating the Shift factor for each ELISA standard
106 curve or each kit lot provides an expedient way to rapidly monitor standard curves as a
107 quality assurance metric and to facilitate data management.

108 # Materials and methods
109

110 ## ELISA kits
111

112 Biomarker ELISAs with at least two kit lots were analyzed for the current study.
113 ELISAs for human P-selectin/CD62P, human myeloperoxidase and human plasminogen
114 activator inhibitor-1/serpin E1 were provided by R&D Systems (Minneapolis, MN, USA).
115 The ELISA kit vendor for the biomarker of focus for the current study (biomarker "PF") is
116 not provided for discretionary reasons. Five lots of the biomarker PF ELISA kits were
117 received over a 10 month time period. All ELISA kits were a standard 96-well format,
118 sandwich antibody-based ELISA designated "for research purposes".

Assays and Equipment

ELISA kits were stored at 4 ± 2°C in a cooler (Helmer Scientific, Noblesville, IN, USA) equipped with alarmed wireless external temperature monitoring (SensoScientific, Inc., Simi Valley, CA). Temperature logs were reviewed and constant temperatures without drift were verified. All kits were used within the manufacturer's expiration date. Plasma samples stored at -80°C were thawed just before assay in a 37°C water bath for <10 minutes, gently mixed, and kept on ice. Plasma samples were assayed according to a detailed standard operating procedure (SOP) for each biomarker ELISA that includes the manufacturer's procedural steps. The SOP also included required documentation for every ELISA plate for operator, date, plate ID, critical reagents (date received, lot number, dilution/concentration, expiration date), incubation times (date, start/stop times, temperature), equipment (manufacturer, model, serial number) and a section to document any deviations from the SOP. Every plate included kit standards prepared according to the manufacturer's instructions. The standards were added to triplicate wells in the first three columns of the provided 96 well plate (columns A-C, rows 1-8 using standard plate designations).

For every biomarker and every plate, an internal spiked plasma-based control sample (BMC Control, see below) was included in triplicate wells. Samples were added to the plates in with calibrated pipettes, washing steps were performed with an automated plate washer (Biotek model Elx50, Winooski, VT) and developed color was quantified by measuring optical density (O.D.) at the appropriate wavelength with a microplate reader (VERSAmax; Molecular Devices, Sunnyvale, CA). The acceptable coefficient of variation (CV) of triplicate wells for each standard, control or unknown was ≤15%. For biomarker

PF, incubations with samples and detection antibody were done at 37ºC in a dry incubator per the manufacturer's instructions and perimeter wells were not used for patient samples to prevent possible evaporation complications due to the elevated incubation temperature. Other ELISAs were performed at room temperature. Initial data analyses were done with SoftMax Pro version 7 (Molecular Devices).

BMC Control Preparation and Storage

Every ELISA plate included a human pooled plasma control sample that had been spiked with supplemental biomarker and designated the BMC Control. For each biomarker, the BMC Control was made in bulk volume according to the respective ELISA SOP and stored at -80ºC in small aliquots for single use. For biomarker PF, lyophilized human pooled citrated plasma (Sigma Aldrich, catalogue P9523-5ML) was reconstituted with deionized water at room temperature for at least 15 minutes with gentle mixing, diluted with appropriate buffer to the same ratio as the patient samples and then spiked with reconstituted biomarker PF standard prepared from the same manufacturer's kit but purchased expressly for this purpose. On the assay day, a BMC Control aliquot was thawed just before assay in a 37ºC water bath for <10 minutes, mixed gently and added to each plate in triplicate wells. Two preparations of BMC Controls were made and aliquoted for storage: one in August 2017 (C1) and one in October 2017 (C2). Both preparations were made with PF standard from kit lot #1. The mean O.D. ± S.D. for C1= 0.759 ± 0.095 (CV=13%, n=26 plates) and 0.672 ± 0.062 (CV=9%, n=16 plates) for C2.

Data Analysis and Derivation of Shift Factor "S"

The approach was implemented in the statistical package R, an open software environment for statistical computing and graphics that accepts ELISA optical density

data and standard concentration data for calculation of a best-fit Reference standard curve. The lot-to-lot variability is modelled as a fixed batch effect calculated as the difference between each plate's standard curve and the Reference curve. This difference is designated the Shift factor "S". An adjusted plate standard curve is derived using the S factor and used to adjust the biomarker concentrations.

The Reference and standard curves are fitted with four- or five-parameter logistic functions (4pl, 5pl). These functions are well established models to relate analyte concentrations to their response signal intensities in immunoassays [8-10]. The 5pl has the form of:

$$Y = a + \frac{d - a}{\left(1 + e^{\frac{xmid - x}{scal}}\right)^{g}} \quad (1)$$

where Y is the signal intensity of measurement (OD in ELISA assays); x is the log-transformed concentration of analytes; a and d are the lower and upper asymptotes of signal intensity, respectively; $xmid$ is the x value of the curve's inflexion point; $scal$ is the scale parameter or the inverse of the slope of the curve at the inflexion point ($x = xmid$); g is the factor controlling the curve asymmetry. When g takes a value of 1, the 5pl becomes the 4pl function.

The 5pl could also be written equivalently as a non-logarithm or exponential form,

$$Y = A + \frac{D - A}{\left(1 + \left(\frac{x'}{C}\right)^{B}\right)^{g}} \quad (2)$$

where Y and g are identical to the parameters in eq. (1); x' is non-log transformed concentration and equal to e^{x}; A and D are identical to its equivalent lower-case letter

parameters in eq.(1); B and C are equivalent to $xmid$ and $scal$, but have an exponential value of them. In this study, the log-form equation is used for implementation.

To analyze data from multiple plates or lots of ELISA kits, batch effects have to be modelled and corrected for data comparison. Many biological or technical factors could impact immunoassay reproducibility and lead to batch effects [11-15]. Lot-to-lot variability between PF ELISA kits is among such factors. We proposed to model and correct it as a fixed batch effect. It first assumes the lot-to-lot differences mainly result from the variable quantitation of standard analyte concentrations, which can expressed as:

$$\begin{aligned} c &= c_0 \times N \\ \log(c) &= \log(c_0) + \log(N) \end{aligned} \tag{3}$$

where c_0 and c are the real and provided standard analyte concentrations from the manufacturer, respectively, and N is the fold difference between them. We can use x, x_0 and S to replace $\log(c)$, $\log(c_0)$, and $\log(N)$ and rewrite the above log-transformed equation as:

$$x_0 = x - S \tag{4}$$

Therefore the statistical model of the standard curves can be written as:

$$\begin{aligned} Y_{ijk} &= a + \frac{d-a}{\left(1+e^{\frac{xmid - x_{ij} + S_i}{scal}}\right)^g} + \varepsilon_{ijk} \\ i &= 1,2....n;\ j = 1,2...m; k = 1,2..l \end{aligned} \tag{5}$$

where a, d, $xmid$, $scal$ and g are the parameters for the 5pl as in eq.(1); Y_{ijk} is the observed signal intensity (measured OD in ELISA); x_{ij} is the log-transformed

concentration of the j^{th} standard analytes in the i^{th} batch; S_i, the Shifting factor, is the log fold difference between the known concentration and the true one; i and n are the i^{th} and total batch number, respectively; j and m are the j^{th} and total number of standards, respectively; k and l are the k^{th} and total number of measurements; ε_{ijk} is the random error for each measurements. This equation can be further rewritten into:

$$Y_{ijk} = a + \frac{d-a}{\left(1+e^{\frac{xmid_i - x_{ij}}{scal}}\right)^g} + \varepsilon_{ijk} \qquad (6)$$

where $xmid_i = xmid + S_i$ and all other parameters are the same as in eq.(5). As a result, the model indicates that the standard curves of the same batch differ from each other as a result of random errors of measurements, while differences between curves from different batches is ascribed to inaccurate quantitation of analyte concentrations. Furthermore, the standard curves from different batches follow the 5pl functions (or 4pl) with the identical parameters of a, d, $scal$ and g, but different $xmid$. The differences are defined by the Shift factor, S_i, which is estimated through the non-linear regression together with other 5pl parameters.

To do the batch normalization/correction, the analyte concentrations in unknown samples are first estimated based on unadjusted standard curves and then the Shift factor S of the batch is applied to obtain the final quantities:

$$\left(x_{ij}\right)_{adj} = \hat{x}_{ij} - S_i \qquad (7)$$

where x is the log-transformed analyte concentrations in the unknown sample, i and j are the batch and standard number as in eq.(5), respectively.

It might not be possible to know the precise concentrations of analyte in the standard samples (c_0 in eq.(3)) without other validation methods, such as proteomics [16]. Therefore, we designated one batch as the Reference batch, in which the analyte concentrations of the standard samples were treated as accurate, and from that we estimated the Shift factor S of all other batches relative to it. The data from 28 standard curves from PF ELISA kit lot #1 (batch #1) was used to calculate a 4pl reference curve for biomarker PF. For other biomarkers, at least four representative standard curves from at least two lots and three operators were chosen to model the best-fit 4pl Reference curve.

Statistical Analysis

Student's t-tests were performed for analysis of inter-assay differences. Analysis of variance between groups was performed with Bonferroni post-test. $P<0.05$ was considered significant.

Software Availability

An ELISA data analysis tool (*ELISAtools*) with the ability to correct batch effects has been implemented in the statistical R software (version 3.5.1) and is available freely for academic use at https://github.com/BULQI/ELISAtools. Instructions for calculation of S factors is provided in Supplemental Methods.

Results

Procedurally, the ELISA manufacturer's directions were followed for reagents, buffers, assay temperature, reagent incubations and wash times. Operationally, these instructions were supplemented with documentation for each plate that included operator, kit reagent lot numbers and clock times for reagent additions and incubation periods. Including a BMC control sample on each plate permitted comparison of data over months (Fig 1). Control optical density (O.D.) readings were consistent with time even with different ELISA kit lots (Fig 1A,B). Myeloperoxidase data (lots #1,2) is shown as a comparison with biomarker PF (lots #1-5). Observations were similar for the other seven biomarkers (data not shown). Myeloperoxidase O.D.s were slightly higher for lot #2, but calculated antigen concentrations were stable (Fig 1C). In contrast, calculated PF concentrations in the BMC controls (preparations C1 and C2) decreased by an average 62.4% between ELISA kit lot #1 and lot #5 over the time (Fig 1D), exceeding our quality control limits. This disconnect between O.D. readings and calculated PF biomarker concentrations over time raised problems for the 420 patient samples already analyzed.

Fig 1. Biomarker Controls with Time

ELISAs were completed over ~10 months as described in Methods for biomarkers myeloperoxidase and PF. (A,B) Optical density (O.D. at 450nm) readings and biomarker concentrations calculated from each plate's standard curve (C,D) for the BMC internal control samples are shown *versus* time. Two ELISA kit lots were used for myeloperoxidase, and five kit lots for PF. Two BMC control preparations were used for PF (C1 and C2), and one BMC control preparation for myeloperoxidase. OD readings for PF controls are reasonably constant with time, but unlike myeloperoxidase, PF concentrations decreased by 62.4% over the study period.

Loss of PF antigenicity in the BMC controls during freezer storage could be a contributor. BMC control preparation C1 was prepared in August 2017 and C2 in October 2017, and stored aliquots were used over the study period. The averaged optical density values for C1 were higher than C2 ($P<0.01$; C1 O.D.= 0.759 ± 0.095, CV=9%;n=26 plates; C2 O.D. = 0.672 ± 0.062, CV=13%; n=16 plates). However, O.D. values over time for each preparation were reasonably stable (Fig 1B), suggesting PF antigenicity did not change significantly during storage. Despite similar O.D. values, a plate with control C1 had a calculated PF concentration of 723.9 pg/mL (mean, triplicate wells) in October 2017 with kit lot #1, but the same control calculated as 238.5 pg/mL in June 2018 with kit lot #5 (Table 1). A similar change was observed for BMC control preparation C2 samples. Other than the kit itself, a detailed operations review did not identify significant changes in PF antigenicity, environment, equipment or operator contributions that could explain the large change in calculated PF concentrations observed in the BMC controls (data not shown). Notably, no similar changes were observed with the other eight biomarker kits (1-6 kit lots) over the same period.

Table 1. BMC Controls and ELISA Kit Lot number (#)

Assay Date	BMC Control C1	BMC Control C2	Kit Lot #	O.D. (450nm)*	PF Conc. (pg/mL)
10/11/2017	x		1	0.686	723.9
06/19/2018	x		5	0.638	238.5
02/19/2018		x	1	0.717	605.4
06/04/2018		x	5	0.684	225.1

* Mean O.D. ± S.D. is 0.759 ± 0.095 for C1 (CV=13%, n=26) and 0.672 ± 0.062 (CV=9%, n=16)

In contrast, comparison of PF standard curves from kit lot #1-5 showed a trend with time (Fig 2A) that paralleled changes observed with the BMC control concentrations. Expected variability between standard curves within the lot was observed (Fig 2B), but lot-to-lot variability showed a left-shift trend. The averaged standard curves for lots #1 and #2 (September, November 2017) were similar, but lots #3-5 (April, May 2018) curves had increasing O.D. at each standard concentration. The expected consequence of a left-shift in standard curves will be lower biomarker PF concentrations, which was observed (Fig 1, Table 1).

Fig 2. Lot-to-Lot Variability in PF ELISA.

(A) The averaged standard curves for each lot of PF ELISA as the mean optical density O.D. ± S.D. at each standard concentration. N= 28, 19, 8, 4, 9 curves for lots #1-5, respectively. (B) Variability of standard curve optical density at 450nm within and between PF ELISA kit lots #1 and #5.

To address this analytic variable and rescue our patient data, a strategy was developed such that each PF standard curve is compared to a best-fit Reference curve (O.D. *versus* PF concentration) and a Shift factor ("S") calculated that quantifies the difference between each plate's standard curve and the Reference curve. An adjusted plate standard curve equation is derived and the biomarker concentrations are re-calculated. Thus, every plate has a unique adjustment, based on the calculated S factor for that plate, and results are generated on a uniform platform. A small S factor value indicates the original standard curve on that plate is similar to the best-fit Reference curve. Conversely, a large S factor indicates a larger left- or right-shift relative to the best-fit

Reference curve. The PF Reference curve was fitted with a 4-parameter logistics (4pl) equation common to many ELISA analyses, but a 5-parameter logistics (5pl) curve fit option is also available (Supplemental Methods).

Assignment of biomarker PF data to calculate the Reference curve was based on our available data, the manufacturer's screening data, and judgment. We used data from 28 standard curves (PF kit lot #1) consistent over six months and three operators to calculate the (4pl) Reference curve. Fig 3 shows the Reference curve with unadjusted averaged standard curves for lots #2 and #5. Lot #2 curves had an average S factor of 0.0690 (n=19) indicating small differences from the Reference curve. In contrast, lot #5 curves had an average S factor of 0.6994 (n=9), indicating a substantial shift relative to the Reference curve.

Fig 3. Reference Curve for calculation of S factor.

The best-fit, 4-parameter logistic (4pl) Reference curve (closed circle, R, solid line) from twenty-eight PF ELISA kit lot #1 standard curves is shown as derived from the indicated equation and D= -0.01; A= 3.20; C= 1300.00; B= -1.30. The averaged standard curves for kit lot #2 has an S factor = 0.0690 (open square; n=19 curves) and S= 0.6994 (open triangle; n=9) for curves from kit lot #5.

S factors for each PF standard curve were calculated (Fig 4). Data from lot #1 was used to fit the Reference Curve, so the average S factor is close to 0 as expected. A shift from the Reference curve for lots #3-5 is shown by their higher average S factors. One lot #2 plate had an S = 0.7032, which is 20.8-fold higher than the average S factor from the remaining plates (mean S=0.0338 ± 0.158; n=18). Data from this plate is under

review. Average S factors for lots #3-5 are similar, but those from lot #3 have greater inter-assay variability. The CV for lot #3 S factors is 44.2% (n=8), compared to 16.7% (n=4) for lot #4 and 11.5% (n=9) for lot #5. Whether this is due to manufacturer's differences, sample size or laboratory-based variables is not known.

Fig 4. Calculated Shift Factors for PF ELISA Standard Curves.

A 4pl Reference curve was derived from PF kit lot #1 standard curves data and a Shift factor "S" was calculated to quantify the difference between each plate's standard curve and the best-fit reference curve. The S factors for each curve in lots #1-5 is shown with the mean (horizontal line) for each lot. S factors for lots #1 and #2 curves are not different, but those for lots #3-5 differ significantly from lot #1 (***P<0.0001, ANOVA with Bonferroni post-test).

The BMC control PF concentrations for preparations C1 and C2 were re-calculated using the S factor for each plate's standard curve. The reference curve was based on kit lot #1 data and C1 was assayed primarily with kit lots #1 and #2. Thus, the average PF concentration did not change significantly (P=0.843), but the variance was reduced (P= 0.002), as expected (Fig 5). C2 controls were included primarily on plates from kit lots #3-5, so the difference after correction with S factors is significant (P<0.0001).

Figure 5. BMC Control PF Concentrations.

A BMC control sample from preparation C1 or C2 was included on each ELISA plate for the 10 month study period and assayed with kit lots #1-5. C1 was used primarily on plates from kit lots #1 and #2. C2 was used primarily on plates from kit lots #3-5. Standard

curves from each plate were adjusted according to their calculated S factor and the control PF concentrations were re-calculated. The data shows the PF concentrations before and after correction with the S factors. ***P<0.0001; ns, not significant (Student's t test).

S factors for myeloperoxidase, soluble P-selectin and plasminogen activator inhibitor-1 were calculated (Table 2). Data for 4pl Reference curves was chosen from four standard curves performed by 3 different operators over at least 6 months to represent the composite data and inter-assay variability. One myeloperoxidase curve had an S = 1.099, and this data is under review. All BMC controls were within quality control limits. Overall, their calculated S factors agree with consistent BMC Control values over time and between kit lots.

Table 2. Average S Factors (min, max; N) for Biomarker Standard Curves from Kit Lots

Biomarker	Lot 1	Lot 2	Lot 3
Myeloperoxidase	0.041 (-0.352, 0.347; 10)	0.505 (0.085, 1.10; 35)	0.289 (0.231, 0.382; 3)
Soluble P-selectin	-0.032 (-0.294,0.156; 42)	0.105 (-0.153, 0.393; 4)	----
Plasminogen activator inhibitor-1	-0.226 (-0.220,0.107; 22)	-0.056 (-0.373,0.047; 22)	-0.008 (-0.043,0.039; 2)

Discussion

Acquiring ELISA data using commercial research kits over a prolonged period presents challenges for quality control needs. There is potential impact of multiple operators, antigen stability, environmental and equipment drift, lot-to-lot reagent variability, and lack of validated controls [15, 17-20]. It is virtually impossible to manufacture a new reagent lot that is identical in all respects to prior lots, and reagent variability occurs even with commercial diagnostic assay reagents [21]. Yet minimizing these influences is necessary so patient or other experimental data can be compared with confidence over the study period. In this study, changing BMC control concentrations for biomarker PF first raised the potential of excessive lot-to-lot variability in the ELISA kits. The optical density data did not change significantly, but calculated PF concentrations greatly decreased, and this disconnect triggered the comprehensive data review. After excluding other reasonable contributors, the S factor strategy was developed to provide a uniform platform for comparison of patient PF data collected over many months.

Calculation of S factors allowed retrospective adjustment of each plate's standard curve. The averaged S factors for PF kit lots #1-5 ranged from -0.086 to 0.735, and reduced the BMC control inter-assay variability to within our quality control limits (<9%). Review of the S factors for each standard curve was useful in that we could rapidly identify possible outliers. One curve each from PF ELISA lot #1 (high) and lot #3 (low), and one for myeloperoxidase (high) were sufficiently different and those plates were tagged for review. Unanticipated variability occurred even with SOPs, required documentation, equipment calibration and heightened operator awareness of pre-analytic variables. Controlling pre-analytic variables for quality assurance and method validation is difficult

[22] and identifying trends on a day-to-day basis is largely subjective. Once a Reference curve is established, calculating the S factor for each curve provides rapid and quantitative comparative data.

Few research-use ELISA kits include validated controls, so the early decision to include spiked plasma controls on every plate was advantageous. One limitation of laboratory-made controls is that each preparation will be slightly different and those differences may not be apparent until enough plates have been run to set the quality control ranges. Access to independently standardized and validated controls in a variety of matrices (plasma, serum, urine, etc) for established and emerging biomarkers would be a valuable resource for investigators and facilitate more consistent biomarker results between laboratories. This in turn supports the goals of improved data rigor and reproducibility, and the more ethical discussion regarding disclosure of validated data to study participants [23-25].

The problem of inter-assay variability is not new [26] and various approaches are proposed to quantify data that are acquired in batches using conversion of a signal from known samples into a meaningful value for unknown samples [15, 27]. For *in vitro* diagnostic ELISA kits used in a clinical laboratories, quality assurance is provided by a validated control sample(s) with defined value limits. Research-only ELISA kits do not have this foundation, yet are used universally. Calculation of S factors and retrospective re-analysis with adjusted standard curves is useful when quality control values exceed limits and environmental, procedural, equipment or operator contributions are ruled out.

This strategy should be used judiciously primarily because data choices for the Reference curve is partially subjective. We chose data from PF ELISA lot #1 based on volume of data, relative consistency over many months and similarity to the

manufacturer's data for that lot. We could have chosen standard curve data from lot #5 plates instead, but there were fewer curves over a shorter time period and our collective experience judged this to be a less favorable choice. We do not anticipate using this strategy for the other biomarkers. Their inter-assay ELISA variability is acceptable, their BMC controls remain within limits, and there is no justifiable need to adjust the curves. That said, we have found the S factors to be a useful monitor of ELISA outcomes to rapidly identify pre-analytic factors, such as operator differences, that otherwise may be difficult to identify by visual inspection of the data, particularly for long-term projects.

There are other approaches to adjust for batch effects in immunoassays similar to our current work. More complicated statistical approaches are employed, such as the mix-effect model [15] and the iterative maximum likelihood method [27]. They assume a linear relationship between the measured signals and the analyte concentrations, which is only an approximation and has its own limitations. Our implementation takes the form of a non-linear logistic function. Our proposed statistical model is simple and the assumption is appropriate for the observed lot-to-lot variability (Supplementary Figure 2).

We focused on developing an accessible and generalizable strategy to solve similar issues as long as the assumptions are met. The software is implemented in R and the data input uses a format familiar to those who use standard 96 well plates for ELISAs. It is written for either Mac or PC and a choice of 4pl or 5pl curve fitting is provided. The software is open source and in the public domain, with instructions for data input in the Supplemental Methods.

Acknowledgements

The authors gratefully acknowledge Dr. Thomas Kepler (BUSM, Department of Mathematics and Biostatistics, Department of Microbiology) for strategic discussions and Dr. Daniel Remick (BUSM, Department of Pathology and Laboratory Medicine) for critical reading of the manuscript. We gratefully acknowledge the administrative assistance of Lindy Joseph (BUSM) and Dr. Jasmin Bavarva (Leidos Biomedical Research, Inc., Frederick, MD) who contributed to project set-up.

References

1. Wong CA, Hernandez AF, Califf RM. Return of research results to study participants: Uncharted and untested. JAMA. 2018;320(5):435-6. doi: 10.1001/jama.2018.7898.

2. Schmidt CO, Hegenscheid K, Erdmann P, Kohlmann T, Langanke M, Volzke H, et al. Psychosocial consequences and severity of disclosed incidental findings from whole-body MRI in a general population study. Eur Radiol. 2013;23(5):1343-51. doi: 10.1007/s00330-012-2723-8. PubMed PMID: 23239059.

3. Shalowitz DI, Miller FG. Disclosing individual results of clinical research: Implications of respect for participants. JAMA. 2005;294(6):737-40. doi: 10.1001/jama.294.6.737.

4. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov. 2011;10(9):712. doi: 10.1038/nrd3439-c1. PubMed PMID: 21892149.

5. Dancey JE, Dobbin KK, Groshen S, Jessup JM, Hruszkewycz AH, Koehler M, et al. Guidelines for the Development and Incorporation of Biomarker Studies in Early Clinical Trials of Novel Agents. Clinical Cancer Research. 2010;16(6):1745.

6. Sarzotti-Kelsoe M, Cox J, Cleland N, Denny T, Hural J, Needham L, et al. Evaluation and Recommendations on Good Clinical Laboratory Practice Guidelines for Phase I–III Clinical Trials. PLoS Medicine. 2009;6(5):e1000067. doi: 10.1371/journal.pmed.1000067. PubMed PMID: PMC2670502.

7. Natarajan S, Remick DG. ELISA rescue protocol: recovery of sample concentrations from an assay with an unsuccessful standard curve. Methods. 2013;61(1):69-72. doi: 10.1016/j.ymeth.2012.08.013. PubMed PMID: 22982520; PubMed Central PMCID: PMCPMC3604139.

8. Gottschalk PG, Dunn JR. The five-parameter logistic: a characterization and comparison with the four-parameter logistic. Anal Biochem. 2005;343(1):54-65. doi: 10.1016/j.ab.2005.04.035. PubMed PMID: 15953581.

9. Higgins KM, Davidian M, Chew G, Burge H. The effect of serial dilution error on calibration inference in immunoassay. Biometrics. 1998;54(1):19-32. PubMed PMID: 9544505.

10. Feng F, Sales AP, Kepler TB. A Bayesian approach for estimating calibration curves and unknown concentrations in immunoassays. Bioinformatics. 2011;27(5):707-12. doi: 10.1093/bioinformatics/btq686. PubMed PMID: 21149344; PubMed Central PMCID: PMCPMC3465100.

11. Fichorova RN, Richardson-Harman N, Alfano M, Belec L, Carbonneil C, Chen S, et al. Biological and technical variables affecting immunoassay recovery of cytokines from human serum and simulated vaginal fluid: a multicenter study. Anal Chem.

493 2008;80(12):4741-51. doi: 10.1021/ac702628q. PubMed PMID: 18484740; PubMed
494 Central PMCID: PMCPMC2646866.
495 12. Schrijver RS, Kramps JA. Critical factors affecting the diagnostic reliability of
496 enzyme-linked immunosorbent assay formats. Rev Sci Tech. 1998;17(2):550-61.
497 PubMed PMID: 9713894.
498 13. Tate J, Ward G. Interferences in immunoassay. Clin Biochem Rev.
499 2004;25(2):105-20. PubMed PMID: 18458713; PubMed Central PMCID:
500 PMCPMC1904417.
501 14. Wilson DH, Williams G, Herrmann R, Wiesner D, Brookhart P. Issues in
502 immunoassay standardization: the ARCHITECT Folate model for intermethod
503 harmonization. Clin Chem. 2005;51(4):684-7. doi: 10.1373/clinchem.2004.042358.
504 PubMed PMID: 15788785.
505 15. Whitcomb BW, Perkins NJ, Albert PS, Schisterman EF. Treatment of batch in the
506 detection, calibration, and quantification of immunoassays in large-scale epidemiologic
507 studies. Epidemiology. 2010;21 Suppl 4:S44-50. doi: 10.1097/EDE.0b013e3181dceac2.
508 PubMed PMID: 21422966; PubMed Central PMCID: PMCPMC3073366.
509 16. Hassis ME, Niles RK, Braten MN, Albertolle ME, Ewa Witkowska H, Hubel CA, et
510 al. Evaluating the effects of preanalytical variables on the stability of the human plasma
511 proteome. Anal Biochem. 2015;478:14-22. doi: 10.1016/j.ab.2015.03.003. PubMed
512 PMID: 25769420; PubMed Central PMCID: PMCPMC4492164.
513 17. Algeciras-Schimnich A, Bruns DE, Boyd JC, Bryant SC, La Fortune KA, Grebe SK.
514 Failure of current laboratory protocols to detect lot-to-lot reagent differences: findings and
515 possible solutions. Clin Chem. 2013;59(8):1187-94. doi: 10.1373/clinchem.2013.205070.
516 PubMed PMID: 23592508.

517 18. Potter DM, Butterfield LH, Divito SJ, Sander CA, Kirkwood JM. Pitfalls in
518 retrospective analyses of biomarkers: a case study with metastatic melanoma patients. J
519 Immunol Methods. 2012;376(1-2):108-12. doi: 10.1016/j.jim.2011.12.003. PubMed
520 PMID: 22210094; PubMed Central PMCID: PMCPMC3472364.
521 19. Kim HS, Kang HJ, Whang DH, Lee SG, Park MJ, Park J-Y, et al. Analysis of
522 Reagent Lot-to-Lot Comparability Tests in Five Immunoassay Items. Annals of Clinical &
523 Laboratory Science. 2012;42(2):165-73.
524 20. Teunissen CE, Willemse E. Cerebrospinal Fluid Biomarkers for Alzheimer's
525 Disease: Emergence of the Solution to an Important Unmet Need. EJIFCC.
526 2013;24(3):97-104. PubMed PMID: 27683444; PubMed Central PMCID:
527 PMCPMC4975183.
528 21. Thaler MA, Iakoubov R, Bietenbeck A, Luppa PB. Clinically relevant lot-to-lot
529 reagent difference in a commercial immunoturbidimetric assay for glycated hemoglobin
530 A1c. Clin Biochem. 2015;48(16-17):1167-70. doi: 10.1016/j.clinbiochem.2015.07.018.
531 PubMed PMID: 26187005.
532 22. Andreasson U, Perret-Liaudet A, van Waalwijk van Doorn LJC, Blennow K,
533 Chiasserini D, Engelborghs S, et al. A Practical Guide to Immunoassay Method
534 Validation. Frontiers in Neurology. 2015;6:179. doi: 10.3389/fneur.2015.00179. PubMed
535 PMID: PMC4541289.
536 23. Broadstock M, Michie S, Marteau T. Psychological consequences of predictive
537 genetic testing: a systematic review. Eur J Hum Genet. 2000;8(10):731-8. doi:
538 10.1038/sj.ejhg.5200532. PubMed PMID: 11039571.
539 24. Partridge AH, Wong JS, Knudsen K, Gelman R, Sampson E, Gadd M, et al.
540 Offering participants results of a clinical trial: sharing results of a negative study. Lancet.

2005;365(9463):963-4. doi: 10.1016/S0140-6736(05)71085-0. PubMed PMID: 15766998.

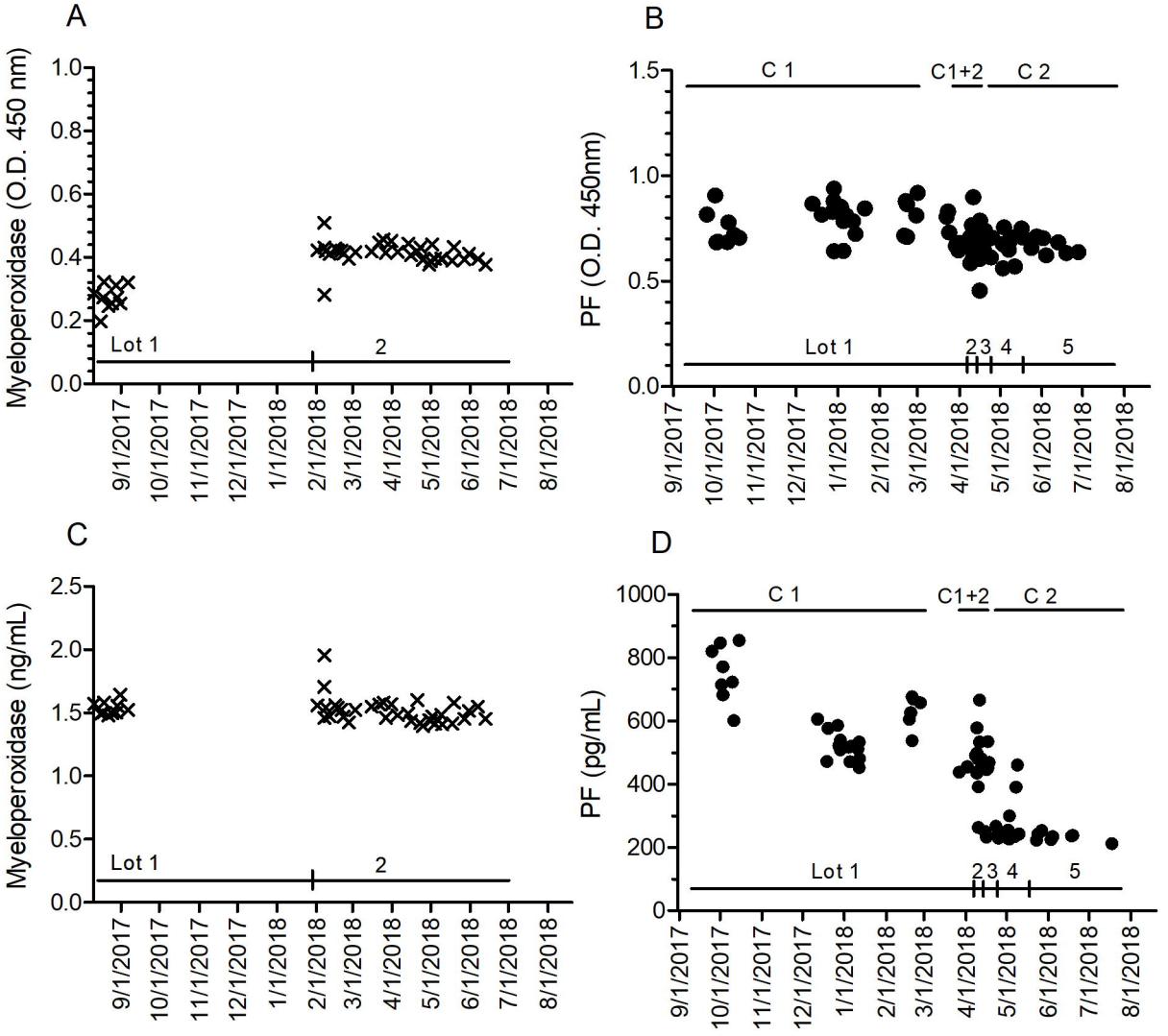
25. Gibson LM, Littlejohns TJ, Adamska L, Garratt S, Doherty N, Group UKBIW, et al. Impact of detecting potentially serious incidental findings during multi-modal imaging. Wellcome open research. 2018;2:114-41. doi: 10.12688/wellcomeopenres.13181.3. PubMed PMID: 30009267.

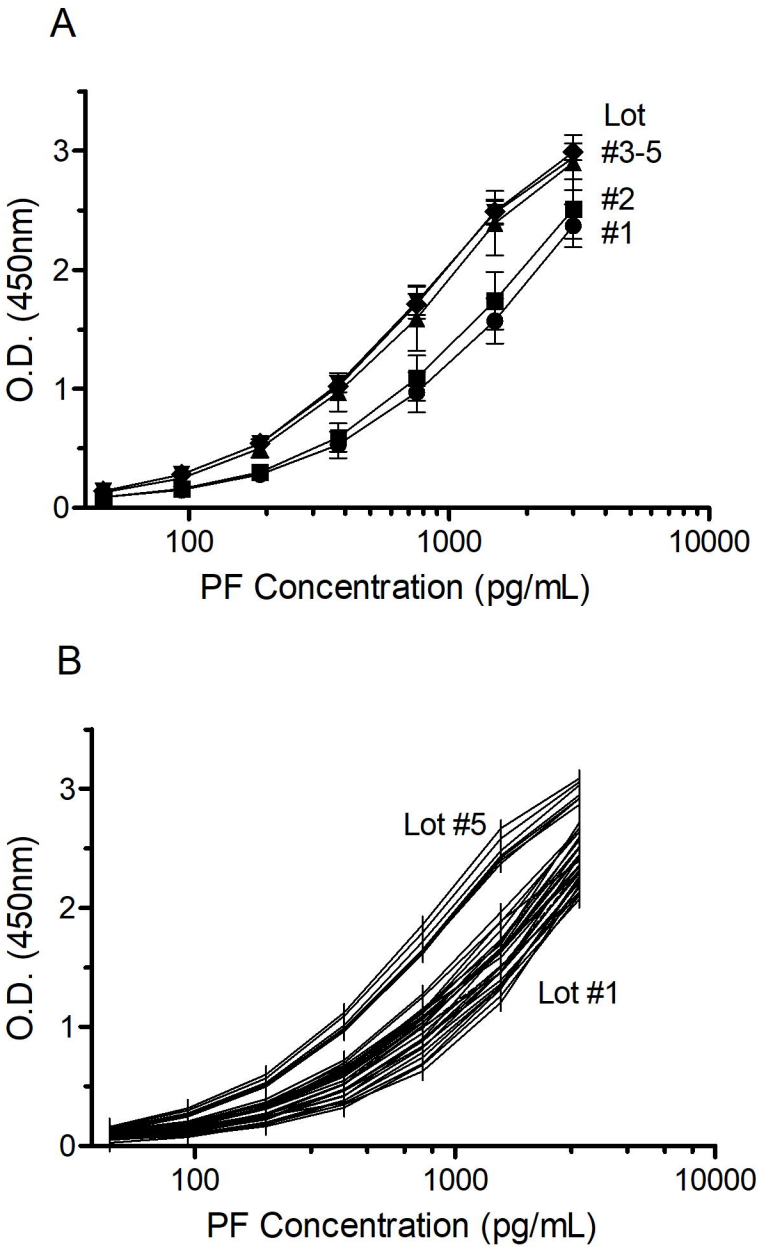
26. Liao JJ, Lewis JW. Qualifying ELISA data: combining information. J Biopharm Stat. 2000;10(4):545-58. doi: 10.1081/BIP-100101983. PubMed PMID: 11104392.

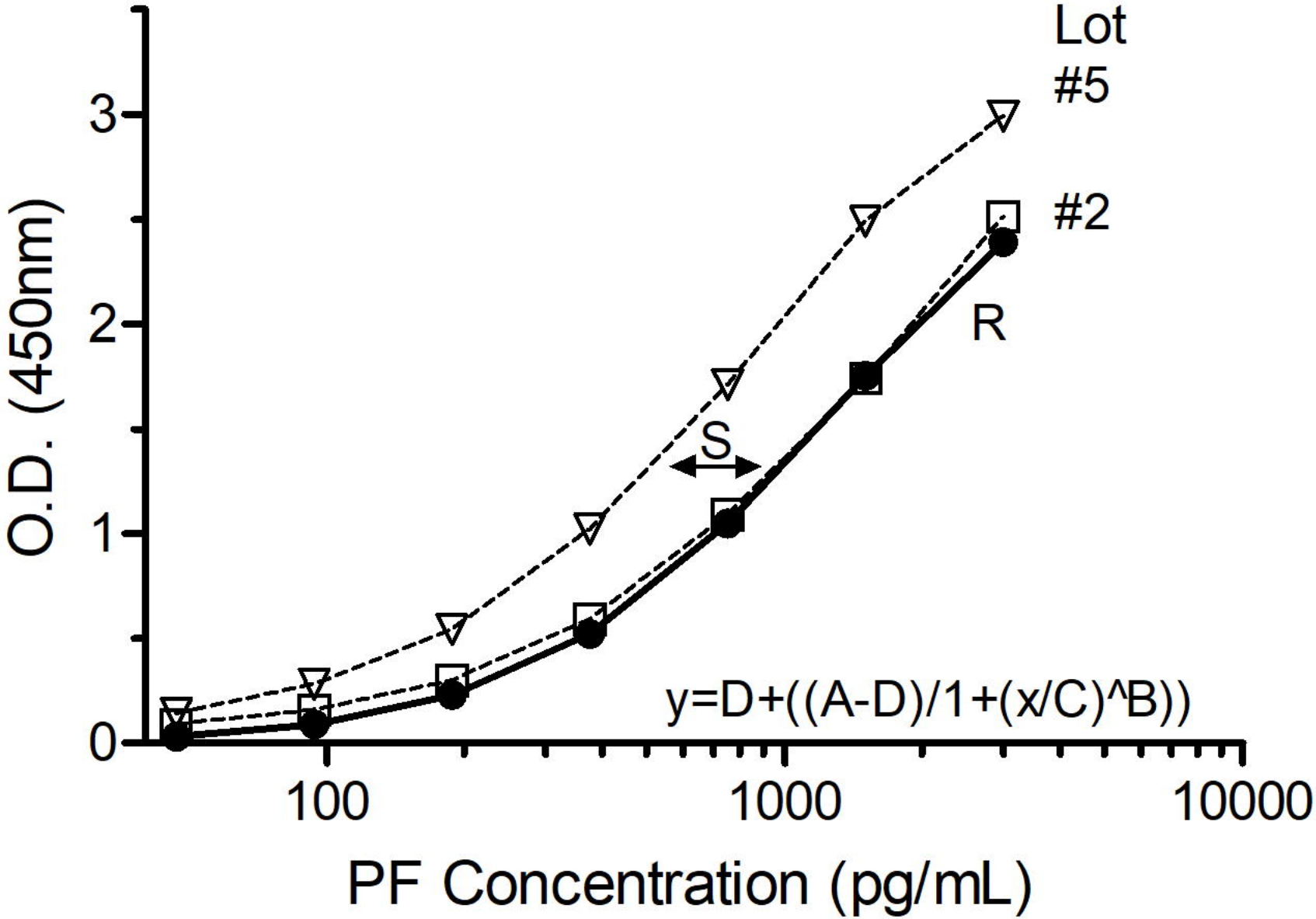
27. Andrews SS, Rutherford S. A Method and On-Line Tool for Maximum Likelihood Calibration of Immunoblots and Other Measurements That Are Quantified in Batches. PLoS One. 2016;11(2):e0149575. doi: 10.1371/journal.pone.0149575. PubMed PMID: 26908370; PubMed Central PMCID: PMCPMC4764494.

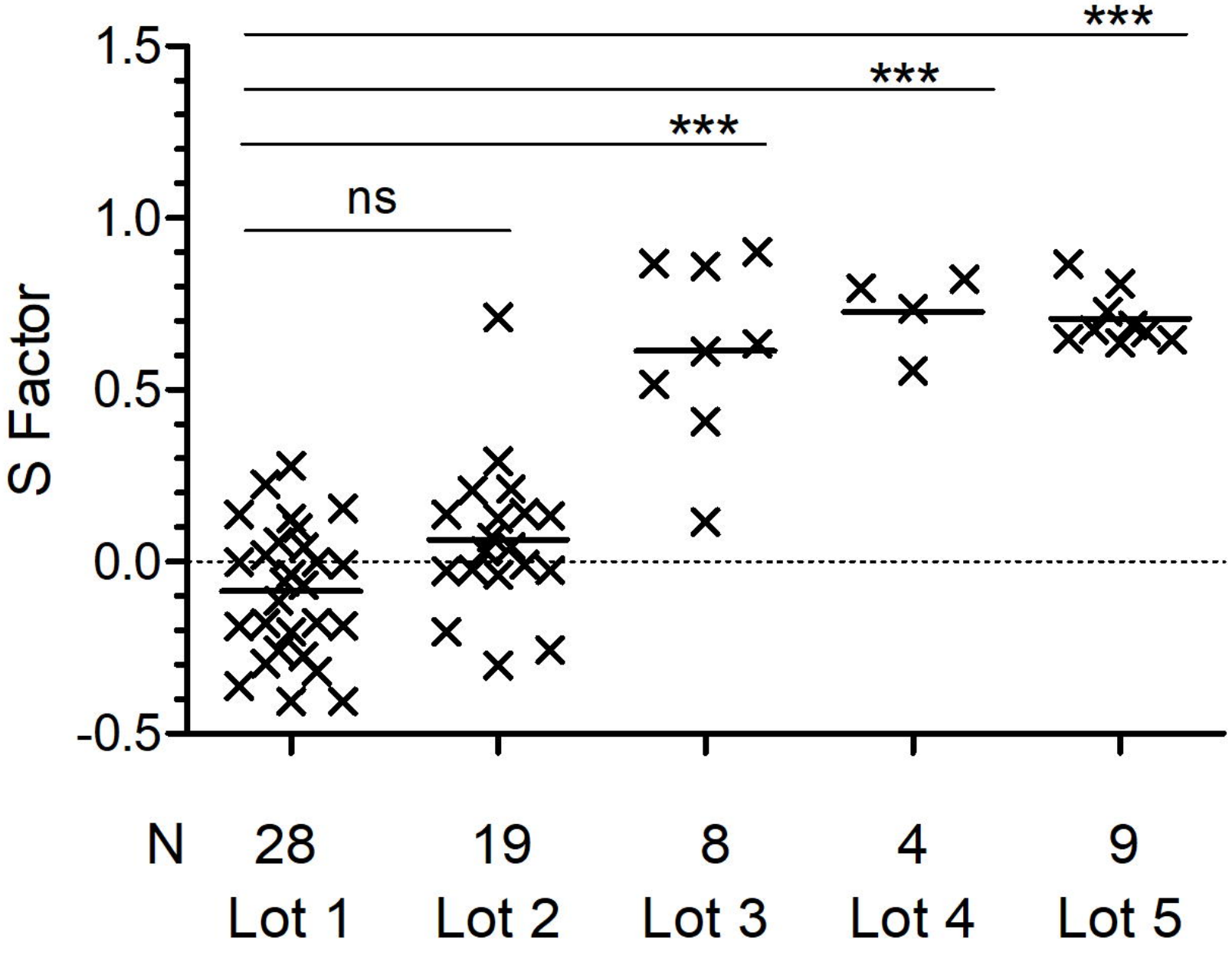
Supporting Information

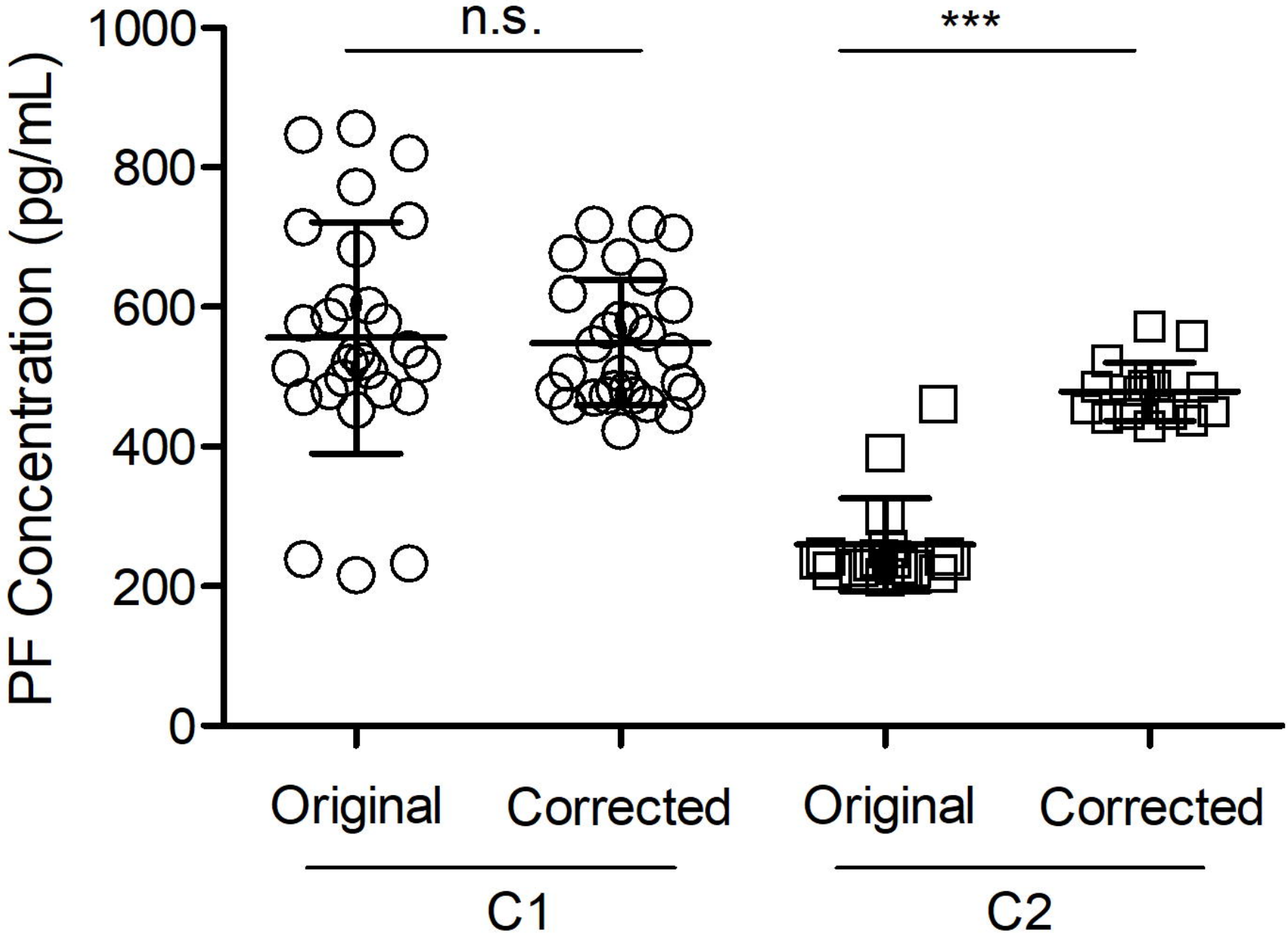
Supplemental Methods provides instructions on how to load and use ELISAtools for calculation of S factor(s).


A
Myeloperoxidase (O.D. 450 nm)
1.0
0.8
0.6
0.4
0.2
0.0
Lot 1
2
9/1/2017
10/1/2017
11/1/2017
12/1/2017
1/1/2018
2/1/2018
3/1/2018
4/1/2018
5/1/2018
6/1/2018
7/1/2018
8/1/2018
B
PF (O.D. 450nm)
1.5
1.0
0.5
0.0
C 1
C1+2
C 2
Lot 1
2 3 4
5
9/1/2017
10/1/2017
11/1/2017
12/1/2017
1/1/2018
2/1/2018
3/1/2018
4/1/2018
5/1/2018
6/1/2018
7/1/2018
8/1/2018
C
Myeloperoxidase (ng/mL)
2.5
2.0
1.5
1.0
0.5
0.0
Lot 1
2
9/1/2017
10/1/2017
11/1/2017
12/1/2017
1/1/2018
2/1/2018
3/1/2018
4/1/2018
5/1/2018
6/1/2018
7/1/2018
8/1/2018
D
PF (pg/mL)
1000
800
600
400
200
0
C 1
C1+2
C 2
Lot 1
2 3 4
5
9/1/2017
10/1/2017
11/1/2017
12/1/2017
1/1/2018
2/1/2018
3/1/2018
4/1/2018
5/1/2018
6/1/2018
7/1/2018
8/1/2018


A
Lot
#3-5
#2
#1
O.D. (450nm)
0
1
2
3
100
1000
10000
PF Concentration (pg/mL)
B
Lot #5
Lot #1
O.D. (450nm)
0
1
2
3
100
1000
10000
PF Concentration (pg/mL)


Lot
#5
#2
R
S
y=D+((A-D)/1+(x/C)^B))
O.D. (450nm)
0
1
2
3
100
1000
10000
PF Concentration (pg/mL)


***
***
***
ns
1.5
1.0
0.5
0.0
-0.5
S Factor
N
28
19
8
4
9
Lot 1
Lot 2
Lot 3
Lot 4
Lot 5


n.s.
***
1000
800
600
400
200
0
PF Concentration (pg/mL)
Original
Corrected
Original
Corrected
C1
C2