

For reprint orders, please contact reprints@future-science.com

Parallelism experiments to evaluate matrix effects, selectivity and sensitivity in ligand-binding assay method development: pros and cons

Parallelism is an essential experiment characterizing relative accuracy for a ligand-binding assay (LBA). By assessing the effects of dilution on the quantitation of endogenous analyte(s) in matrix, selectivity, matrix effects, minimum required dilution, endogenous levels of healthy and diseased populations and the LLOQ are assessed in a single experiment. This review compares and discusses all available approaches that can be used to assess key assay parameters for pharmacokinetic and biomarker LBAs, as well as the advantages and disadvantages of each approach. This review also summarizes a systematic approach that can apply to guide endogenous LBA method development and optimization with a suggested way to interpret parallelism data.

First draft submitted: 16 April 2017; Accepted for publication: 31 May 2017;
Published online: 21 July 2017

Keywords: biomarker • ligand-binding assay • method development • parallelism

Ligand-binding assays (LBA) quantitate macromolecules by comparing immunoreactivity of calibrators of known concentrations to the samples of unknown concentration. For a well-developed LBA with an appropriate logistic regression approach, the calibration curve should be parallel to support the assumption that the antibody-binding characteristics are similar enough to allow the determination of analyte levels in the diluted samples [1]. Two major factors that contribute to nonparallelism are: a difference between the immunoaffinity characteristics of calibrator reference material and unknown analyte, to the capture and detection reagents; and matrix effects variances among calibration curve matrix, quality control matrix and study population matrix (Figure 1).

For LBAs designed to measure an exogenous therapeutic protein to support pharmacokinetic (PK) studies, the reference standard is normally well characterized with a comprehensive certificate of analysis. The calibrator reference standard is the same as the analyte of interest. Quality controls spiked at vari-

ous concentrations are used to evaluate accuracy, precision, selectivity, sensitivity and dilutional linearity to support a definitive quantitative method validation. The detailed recommendations for those experiments are described in the US FDA draft guidance [2], EMA guidance [3] and industry consensus white paper [4]. Currently, those recommendations are widely adapted for PK LBAs. However, for development and qualification of LBAs measuring endogenous proteins such as biomarkers and 'free'/total' drug targets, the method development and qualification strategies will differ from the approach accepted for PK assays. Normally, endogenous reference materials are not available in a purified, fully characterized form and blank matrices free of analyte may not exist. Therefore, a relative quantitative approach is suggested [5–7].

Parallelism is a critical aspect to address relative accuracy by assessing the effects of dilution on the quantitation of analytes in biologic matrix. Essential characteristics of an assay including selectivity, matrix effects,

Jing Tu*¹ & Patrick Bennett¹

¹Biomarker Services, PPD Laboratories,
2244 Dabney Road, Richmond, VA
23230, USA

*Author for correspondence:

Tel.: +1 804 977 8225

jing.tu@ppdi.com

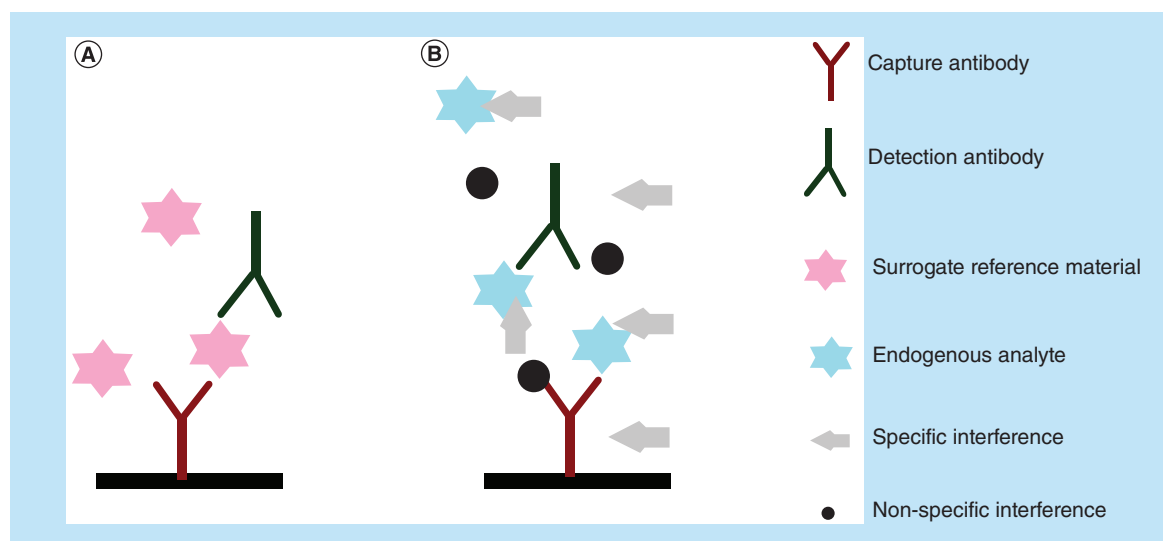


Figure 1. Major factors contributing to nonparallelism. (A) Calibration curves – surrogate reference material prepared in surrogate matrix; **(B)** natural samples – endogenous analyte in intact, unmodified matrix.

minimum required dilution (MRD), endogenous levels of health and diseased population, and the LLOQ can be evaluated [8–11]. It is recommended that parallelism should be evaluated during the early stage of endogenous LBA method development whenever possible.

Over the last decade, there have been many discussions published in the scientific literature regarding the conduct of parallelism experiments and the appropriate acceptance criteria to apply. In 2003, DeSilva *et al.* suggested using an incurred sample pool which was made from several C_{\max} study samples to support in-study validation for PK assays [4]. The 2011 EMA guidance suggested using a high concentration study sample (C_{\max} study sample) diluted to at least three concentrations to evaluate PK method parallelism. It also suggested calculating the precision between diluted samples to evaluate parallelism [3]. In 2006, in the first biomarker ‘fit-for-purpose’ white paper, Lee *et al.* emphasized the relative quantitative feature for most biomarker LBAs and suggested evaluating method parallelism using at least three individual intact samples during method validation. They also suggested using a tiered approach to validate both exploratory and advanced methods [6]. Since then, parallelism experiments for biomarker assays have been heatedly discussed and have been mentioned in numerous reviews and white papers [7,12–24]. Valentin *et al.* pointed out parallelism should be the most critical element for biomarker LBAs starting from the feasibility phase. Parallelism and selectivity should both be evaluated [16]. Ciotti *et al.* suggested to use at least ten normal and ten diseased samples (if available) with at least six dilutions to assess method paral-

lelism. They indicate that selectivity, sensitivity and MRD can be evaluated through parallelism experiments [8]. Stevenson and Puruhothama published a comprehensive paper to address the why, when and how questions for parallelism experiments [9]. Lee *et al.*, Khan *et al.* and Jani *et al.* published their white papers regarding method development and validation for ‘total’ and ‘free’ therapeutic antibodies and their targets [15], adapted commercial kits [22] and multiplex LBAs [24], respectively, in which parallelism experiments were all heavily mentioned.

In September 2015, the Crystal City VI workshop was organized by the American Association of Pharmaceutical Scientists in association with the FDA. A general understanding was reached that because biomarker assays and the data they provide are often very different from that of pharmacokinetic assays. Therefore, method validations may not always fit with the current pharmacokinetic-based bioanalytical guidelines. Parallelism was designated as key to biomarker LBAs [25,26]. Further articles were published to discuss the current status and applications of parallelism experiments [10,11,27–33]. By far the consensus regarding how to perform parallelism, how to evaluate parallelism data and what the acceptance criteria should be, have not been reached.

This article will review and compare the available approaches – including parallelism experiments as well as other traditional approaches that can be used to assess key assay parameters. The article also will discuss how to utilize parallelism data to direct method development and what are the advantages and disadvantages for each strategy. The author’s perspectives also will be included.

Parallelism & matrix effects

LBAs measure analytes of interest in the biological matrix without prior extraction [4]. In an ideal world, the capture/detection reagents and analyte of interest should bind and only bind to each other without cross-reacting with any additional compound(s). Unfortunately, this is often not the case. In the real world, endogenous compounds from the matrix frequently interrupt those bindings. The endogenous matrix interferences can specifically or nonspecifically bind to capture/detection reagents or the analyte of interest and lead to an increase or decrease of the signal generated. The most common nonspecific interferences are due to hemolysis, lipemia, interaction with anticoagulant [31] or other small-size organic or inorganic substances. The specific interferences for PK assays can be caused by catabolic species of the drug, analogs of the endogenous proteins, heterophilic antibodies, human antianimal antibodies, rheumatoid factors, soluble ligands, antidrug antibodies, high-dose hook effect, etc. [34–38]. For biomarker assays, specific matrix effects can be additionally caused by endogenous molecules with similar structure to the target analyte (e.g., homologous family members, isoforms and precursor proteins) [39] or their natural ligands and the analogs of the ligands. Typically, the potential interference compounds are not available in well-characterized forms and often the nature of the interference is not known, which prevents direct testing of all of the interferences.

Traditionally, matrix effects are evaluated by spike/recovery experiments during the method development stage. If standard curves are prepared in the same matrix as sample matrix (e.g., like most of the PK assays), matrix effects are normally similar between standards and samples. The matrix optimization is more for the purpose of increasing sensitivity instead of improving accuracy. For most of the endogenous LBAs where surrogate matrix is required, overspiked matrix control samples are calculated against the surrogate matrix curve to evaluate absolute accuracy. However, when a natural blank matrix is not available, there is a significant caveat. Because the endogenous concentration in natural matrix is measured using a surrogate matrix curve, there is no guarantee for the absolute accuracy for such measurement due to the potential matrix effect differences and potential assay sensitivity limitation. Therefore, the endogenous concentration for the matrix pool used for spiking needs to be assigned appropriately.

Parallelism experiments help to understand assay relative accuracy. They evaluate assay matrix effects through relative accuracy assessment by plotting signal against dilution or concentration. Several different

approaches have been used to process raw data. These approaches are discussed in the ‘Parallelism, Dilutional Linearity and Selectivity’ section. The final curve for each sample should reflect the combined binding affinity for the analyte and matrix interferences. For this reason, parallelism failure can be an indicator for the existence of matrix interferences.

Furthermore, by closely investigating the details and the trend of parallelism data, assay developers can gather clues regarding the type of interferences and subsequently narrow down the possible contributors for the matrix effects. For example, if nonparallelism can be mitigated by increasing dilution, nonspecific binding is likely occurring. Assay developers can either increase method MRD or optimize sample diluent (e.g., add blocking reagents, detergents, salts, etc.). In contrast, if matrix effects cannot be diluted out, specific interference is likely. Binding reaction inhibition or enhancement can further indicate the possible binding site(s) where interferences are taking place [39]. Assay developers can consider investigating new sources of critical material(s) that recognize a different epitope, add a sample pretreatment step (e.g., alkaline treatment, acid treatment or extraction) or add strong detergent(s) to disrupt such interference [11,15,40].

Parallelism & selectivity

Selectivity is the ability of an assay to measure the analyte of interest in the presence of other constituents in the sample [4]. A well-designed LBA should be able to accurately measure the analyte of interest from most of the study individuals without unique individual matrix biology interfering. According to the FDA draft guidance, EMA guidance and white paper, at least ten lots of the individual matrix for human (six lots for animal) should be spiked at or near the LLOQ to evaluate selectivity [2–4]. This approach currently is in use for validation of regulated PK assays. However, for biomarker assays the reference materials are often obtained from external vendors in various forms (e.g., heterogenetic forms, recombinant proteins, fusion proteins, etc.) that may be different from the endogenous analytes. By spiking a reference material with uncertain quality, the endogenous biomarker and spiked reference material may not always behave in an additive manner [16] and the spiked samples will not mimic the real-study samples. Ideally, all the protein standards used should be certified by qualified organizations. And for a particular protein LBA, the same reference standard should be used globally across different labs. However, currently there are only a few well-characterized protein standards available. The labs may also perform in-depth protein characterization according to their internal standard operating procedure (SOP), but the expense

has to be balanced against potential assay performance issues.

There are obvious advantages to performing parallelism experiments in the early stage of biomarker method development to appropriately evaluate method selectivity. Abundant information can be gathered through appropriate data interpretation. Parallelism failures can be viewed in two ways: the lack of parallelism among samples indicates matrix effects and selectivity issues; and the lack of parallelism between calibrators and samples indicates either matrix effects or reference material specificity issues or both (i.e., if all samples are parallel to each other but not parallel to the calibrator curve, the quality of the reference material is in doubt) [11].

Ideally, normal and/or diseased individual samples are used to perform parallelism experiments. The fresh samples obtained from a reliable source (e.g. samples obtained from hospitals) with full patient medical records and full sample storage records are always preferred. Alternately, the samples purchased from a commercial source can also be used. However, the stability and reliability of the purchased matrixes should be considered.

Generally, back-calculated concentrations of diluted samples are used to evaluate method parallelism [8–11]. Stevenson and Purushothama further suggested plotting dilution-adjusted concentration against sample dilution for multiple individuals to set the method MRD and evaluate method selectivity [9]. They suggested setting the MRD where most of the samples are measured in assay range and multiple dilutions beyond the MRD yield accurate results [9]. The European Bioanalysis Forum topic team-61 summarized alternative methods to process parallelism data. The approaches include plotting log in-well measured concentration against dilution factor, plotting log dilution-adjusted concentration against dilution factor, plotting dilution-adjusted relative error against dilution factor and plotting log in-well concentration against log 1/dilution [10]. These back-calculated concentration-based approaches allow assay developers to evaluate assay parallelism against predefined acceptance criteria.

Figure 2 demonstrates the data processed according to Stevenson and Purushothama's suggestion from a soluble BCMA (sBCMA) method case study [11]. In this case study, ten individual normal human serum samples were serially diluted with either 1% bovine serum albumin (BSA) phosphate-buffered saline (PBS) buffer or 1%BSA PBS plus 0.50% (v/v) Triton X-100 buffer. Calibration curves were prepared in either 1%BSA PBS buffer or 1%BSA PBS plus 0.50% (v/v) Triton X-100 buffer, respectively. Dilution-adjusted concentrations were plotted against 1/dilution to evaluate method par-

allelism. The data suggested that parallelism was not achieved for most of the subjects when using 1%BSA buffer as surrogate matrix (Figure 2A & B). However, parallelism was achieved for the samples diluted with 1%BSA PBS plus 0.50% (v/v) Triton X-100 buffer at 1:27-fold or greater (Figure 2C & D).

When a standard curve is not available or does not demonstrate parallelism to the samples, the back-calculated concentrations for neat and diluted samples are either not available or not accurate. Assessing parallelism among samples with the approaches that rely on back-calculated concentrations may lead to misleading conclusions. The authors recommend directly evaluating sample parallelism from the raw signal acquired from the instrument. For LBAs, the correlation between signal and concentration/dilution normally is not linear. Hence, linear plotting cannot accurately reflect the parallelism among samples. Four or five parameter logistic (4PL or 5PL) regression with or without weighting is suggested to achieve the best curve fit with the least variance [4,6]. The parallel relationship for each sample dilution curve can be seen directly from the plot and the parallelism among samples can be evaluated. One limitation is that it is not clear how to set up fixed acceptance criteria for the 'raw signal' approach. The 'B' parameter ('B' represents the Hill's slope of the curve for 4PL and 5PL) for each sample may be used to evaluate method parallelism. Further statistical approaches may be evaluated and applied for better data interpretation.

Figure 3 demonstrates the data processed according to the 'raw signal' approach using the same data from the sBCMA case study. Optical density was plotted against 1/dilution for ten individual samples with SoftMax[®] Pro software (molecular devices) using 4PL regression. The B parameter for ten samples ranged from 0.33 to 1.11 when using 1%BSA buffer as surrogate matrix (Figure 3A & B), and the B parameter ranged from 1.25 to 1.46 when using 1%BSA PBS plus 0.50% (v/v) Triton X-100 buffer as surrogate matrix (Figure 3C & D). The data suggested that adding 0.50% (v/v) Triton X-100 into the sample diluent mitigated matrix effects and improved method parallelism.

One further application of the 'raw signal' approach is to allow easy assay buffer optimization without including curves made in each different buffer. Figure 4 demonstrates the data processed for surrogate matrix optimization using the 'raw signal' approach. Three individual normal human serum samples were serially diluted with 1%BSA PBS buffer without Triton X-100 and with Triton X-100 added at 0.25, 0.50 and 1.00% (v/v), respectively. All of the diluted samples were analyzed on the same plate with no calibration curve added. The optical density versus 1/dilution curves

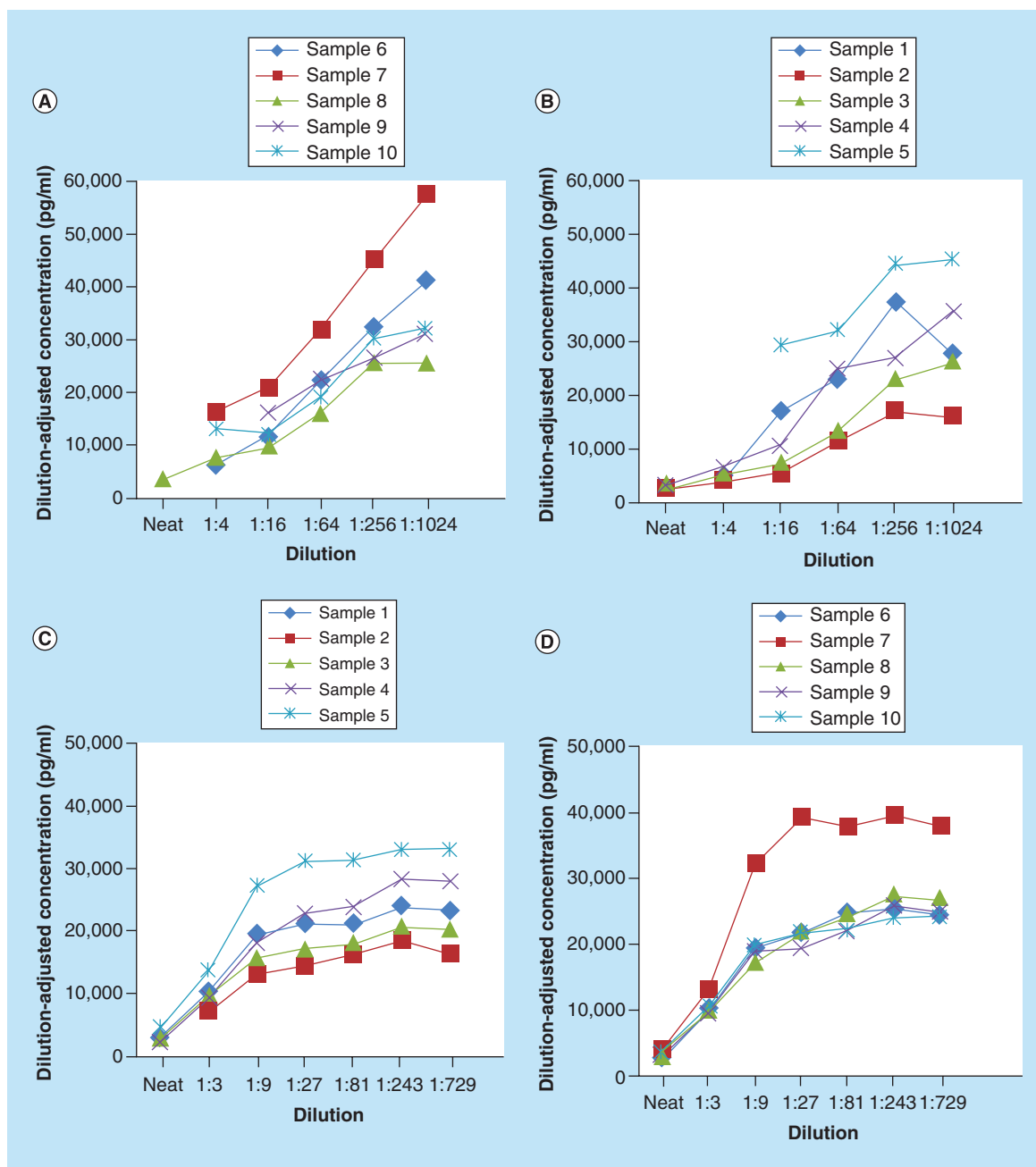


Figure 2. Soluble BCMA ELISA assay parallelism evaluation with ten individual normal human serum samples using 'dilution-adjusted concentration' approach. (A) Samples 1–5 (male) diluted with 1%BSA PBS buffer; (B) Samples 6–10 (female) diluted with 1%BSA PBS buffer; (C) Samples 1–5 (male) diluted with 1%BSA PBS with 0.50% (v/v) Triton X-100 buffer; (D) Samples 6–10 (female) diluted with 1%BSA PBS with 0.50% (v/v) Triton X-100 buffer. BSA: Bovine serum albumin ; PBS: Phosphate-buffered saline.

were used to evaluate method parallelism. The data suggested that parallelism was improved for all three groups with Triton X-100 added. Eventually, 1%BSA PBS plus 0.50% (v/v) Triton X-100 buffer was chosen as the surrogate matrix for better assay robustness.

Since it is not clear how to set up fixed acceptance criteria for the 'raw signal' approach, after a nonpar-

allel issue is resolved, the assay parallelism should be confirmed using the 'dilution-adjusted concentration' approach with a calibration curve prepared in the same buffer as sample dilution buffer.

Although the advantages are obvious, using parallelism experiments to evaluate method selectivity has clear challenges: Obtaining a decent number of samples that

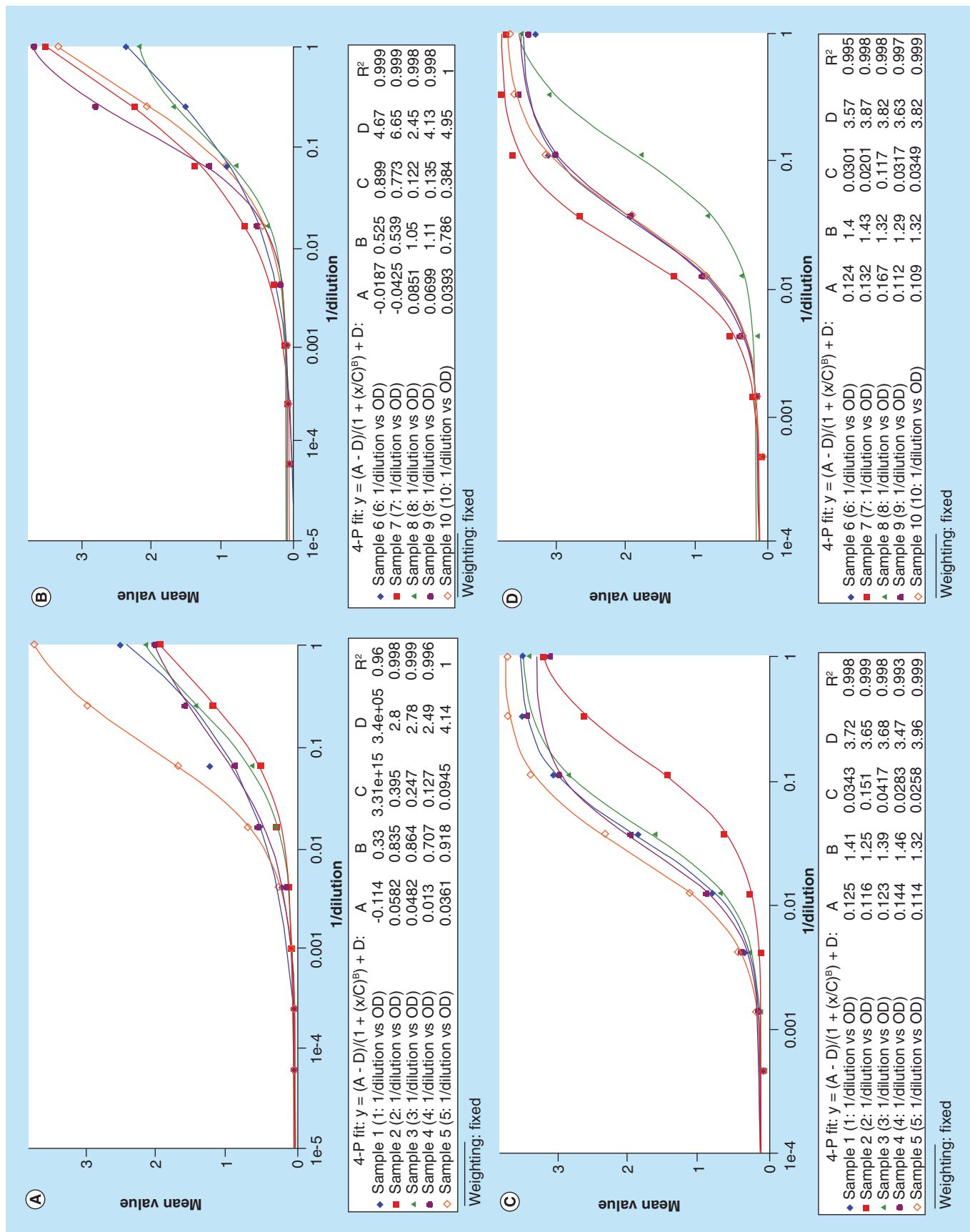


Figure 3. Soluble B-cell maturation antigen ELISA assay parallelism evaluation with ten individual normal human serum samples using 'raw signal' approach with four parameter logistic regression curve fitting (cont. from facing page). (A) Samples 1–5 (male) diluted with 1% BSA PBS buffer; (B) Samples 6–10 (female) diluted with 1% BSA PBS buffer; (C) Samples 1–5 (male) diluted with 1% BSA PBS with 0.50% (v/v) Triton X-100 buffer; (D) Samples 6–10 (female) diluted with 1% BSA PBS with 0.50% (v/v) Triton X-100 buffer. BSA: Bovine serum albumin; PBS: Phosphate-buffered saline.

have high-enough endogenous concentrations is not always possible. An alternative is to perform dilutional linearity experiments with several overspiked individual samples to evaluate method selectivity [6,16,24]. The normal concentrations for spiked samples can be set as the endogenous levels plus the overspiked concentration [24]. Alternatively, spiked concentrations can be used as nominal concentrations directly and correctly measured concentrations can be obtained by subtracting the endogenous concentrations for bias calculation [41–43]. This concept and data evaluation strategies should be the same as used in parallelism experiments.

Other challenges for parallelism experiments may include as follows: when special matrix types (e.g., cerebrospinal fluid, tissue homogenate) or matrix species (e.g., mouse) are required, the small sample volume available may not be sufficient to allow the bioanalysis of multiple dilutions. In addition, there is no unified way to set up nominal concentrations for each sample and to evaluate parallelism for the assay. The nominal concentration can be set as the results from either the MRD sample [8,9], from the greatest dilution that have results above LLOQ [11] or from the mean of all dilutions at which parallelism is shown. Parallelism can be evaluated by calculating the percentage bias from the nominal concentration for each dilution [8–11] or the percentage CV (%CV) among all the dilutions [3,32].

Parallelism & specificity

Specificity is the ability of assay-critical reagents (e.g., antibody) to distinguish between the analyte of interest and other components [4,6]. LBAs measure binding reactivity rather than a direct mass determination. Any compound that binds to critical reagents can generate a signal regardless of the specificity and affinity. This makes the specificity an inherent problem for all LBAs. The goal of most LBA specificity tests is not to demonstrate absolute assay specificity, but rather to provide information regarding what is being measured to fulfill the purpose of the intended application [13].

The specificity of LBAs is determined by the specificity of the critical reagents. The specific reagents are designed, confirmed and selected with the assistance of other technologies (e.g., western blots, surface plasmon resonance, HPLC, LC–MS/MS, etc.) [26,36]. The specificity test protocol currently in use is in a learn-and-confirm approach [25,26]. The method specificity is evaluated by calculating the accuracy of the samples

that have been spiked with various forms of potential cross-reactive materials [4]. However, this is with the assumption that the interferences are known and available within the form of the matrix or a purified form [26].

Parallelism experiments indirectly assess method specificity. The specificity of the assay critical reagents should be considered suspect if parallelism cannot be achieved within samples as well as between samples and calibrator curves after method optimization [11]. In other words, after ruling out nonspecificity-related matrix effects and reference material specificity issues, specificity may be the main contributor to assay non-parallelism. However, sample dilutions that are parallel with the calibrator curve do not guarantee assay specificity. In a method comparison experiment performed by our laboratory, a small-molecule biomarker showed no detectable levels for all the samples tested with a LC–MS/MS assay with assay sensitivity at approximately 20 pg/ml. However, when this was measured with a commercial ELISA kit assay, all samples showed measurable results and parallelism was achieved for six out of nine samples [11] (Table 1).

Unlike PK assays where the critical materials are usually made in-house under a controlled procedure or within a GMP environment, the critical materials for biomarkers are normally purchased from external vendors or are commercially available kits. Due to lack of control of the manufacturing processes of critical materials, biomarker methods suffer more from specificity issues than PK LBAs. The end users of these commercially available materials need to be cautious when applying the biomarker LBAs for their intended use. End users can also consider using LC–MS/MS technology to confirm LBA method specificity [13]. However, the time, expense and LC–MS/MS method sensitivity have to be balanced against the potential for specificity issues.

Parallelism & sensitivity

There are two terms related to assay sensitivity: LOD and LLOQ. LOD refers to a concentration resulting in a signal that is significantly different (e.g., mean background ± 2 or 3 standard deviation) from background [6], and it is often provided as sensitivity for biomarker research use only (RUO) or diagnostic kits [6,22]. LLOQ refers to the lowest concentration of analyte that has been demonstrated to be measurable

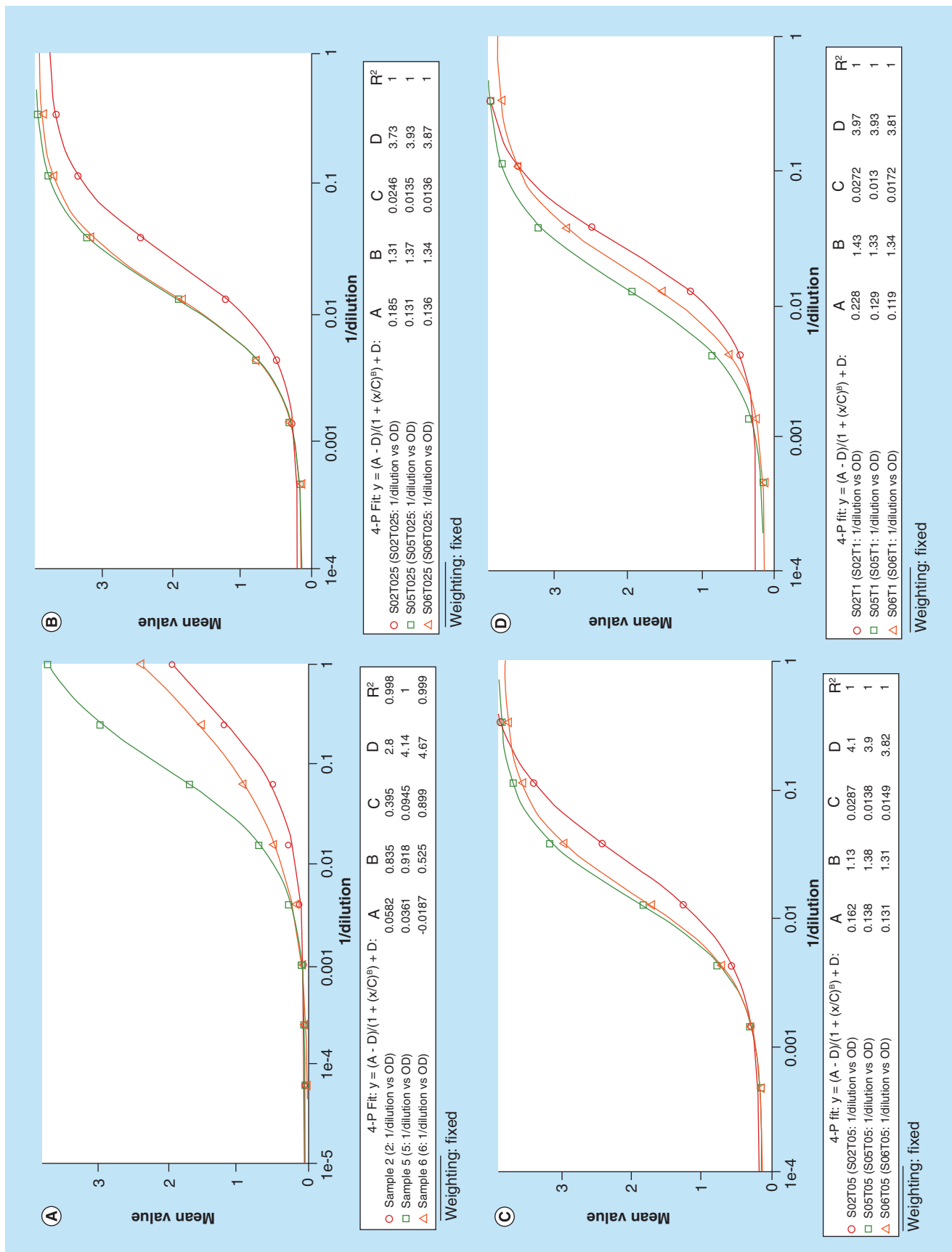


Figure 4. Surrogate matrix optimization using 'raw signal' approach with four parameter logistic regression curve fitting (cont. from facing page). (A) Samples diluted with 1% BSA PBS buffer (before optimization); (B) Samples diluted with 1% BSA PBS with 0.25% (v/v) of Triton X-100 buffer; (C) Samples diluted with 1% BSA PBS with 0.50% (v/v) of Triton X-100 buffer; (D) Samples diluted with 1% BSA PBS with 1.00% (v/v) of Triton X-100 buffer. BSA: Bovine serum albumin ; PBS: Phosphate-buffered saline.

with acceptable levels of accuracy, precision and total error [4,6]. LLOQ is normally greater than LOD.

Traditionally, the sensitivity (LLOQ) of a bioanalytical LBA is determined during the accuracy and precision runs with samples that were prepared by spiking reference standards into blank matrix [4]. Because blank matrix is usually not available for biomarker assays, spiking reference material into samples that have measurable endogenous level is also an option. The endogenous level is determined and the final concentration is the composite of the determined level and the amount of reference standard added. However again, the spiked reference material and endogenous biomarker may not behave in an additive manner. Even if they are additive, there are further concerns regarding this approach: what is the true LLOQ for the assay? Should the LLOQ be defined as the spiked concentration only or spiked plus endogenous concentration [11]? Does this approach demonstrate sensitivity for the endogenous analyte or reference material?

Parallelism experiments do not have the above concerns because they avoid the concept of absolute accuracy. Stevenson and Purushothama proposed to analyze the data from the same parallelism experiment but using the on-the-curve concentration (i.e., before dilution-adjusted concentration) to determine assay sensitivity for the endogenous analytes [9]. They suggested two different approaches to set the assay LLOQ. The 'common dilution method' is to set LLOQ at the highest concentration observed at the greatest dilution at which all individual samples give a parallel response [9]. One concern for this approach is that when at least one sample has an endogenous concentration that is significantly higher than the others, the determined LLOQ will be higher than the true assay LLOQ. The alternative way, 'common concentration method', is to set the LLOQ at the highest concen-

tration at which all of the samples had demonstrated parallelism [9]. In other words, any samples above the determined LLOQ concentration that demonstrated relative accuracy through parallelism experiments. The data from the sBCMA case study were processed with the 'common concentration method' for demonstration (Table 2) [11].

In summary, parallelism experiments are powerful tools which can be used to guide early method development and optimization. It directly evaluates analyte endogenous levels, assay matrix effects, selectivity and sensitivity. It is also an indirect indicator of method specificity. Figure 5 shows a decision tree regarding the systematic method development and optimization strategies using parallelism data.

Parallelism & adaptation of commercial kits

Commercially available kits can be categorized by quality level or intended use. If categorized by quality level, there are clinical-grade and research-grade diagnostic kits [44,45]. If categorized by intended use, there are *conformité européenne* (CE)-marked assays, *in vitro* diagnostic (IVD) assays, RUO assays and investigational use only (IUO) assays [13,22,46–48]. CE-marked assays and IVD assays are clinical-grade. CE-marked assays are the approved assays/medical devices that are sold within the European Economic Area for clinical purposes. IVD assays are approved by the FDA for clinical purposes. Normally RUO and IUO assays are research-grade, which vary in quality and have not been approved by any regulatory agencies. No matter whether commercial kits are adapted for laboratory-developed testing to support clinical diagnosis or for prognosis or for bioanalytical assays that support drug development, the goal for method development is the same – to establish reliable assays with sufficient accuracy and sensitivity that fit the intended use.

Table 1. Parallelism data for a small-molecule biomarker ligand-binding assay.

Sample dilution	Human urine sample, dilution-adjusted concentration (pg/ml); % bias from 1:8 or 1:4 dilution results)								
	Normal 1	Normal 2	Normal 3	Asthma 1	Asthma 2	Asthma 3	Asthma 4	Asthma 5	Asthma 6
1:2	225.8 (36)	120.9 (67)	205.8 (28)	233.1 (22)	20.3 (5)	1407.4 (117)	209.8 (72)	AQL	514.4 (36)
1:4	176.8 (6)	124.9 (73)	165.4 (3)	180.7 (-6)	19.3	905.3 (40)	150.9 (24)	865.4 (-2)	366.8 (-3)
1:8	166.4	72.4	160.6	191.3	BQL	649.6	121.9	879.1	377.3

All samples showed below quantifiable limit when measured by an LC-MS assay with assay sensitivity around 20 pg/ml.
AQL: Above quantifiable limit; BQL: Below quantifiable limit.

Table 2. Determining soluble BCMA method LLOQ using 'common concentration method'.

Sample dilution	Human serum sample, concentration after dilution (pg/ml; % bias from MRD 1:27 results adjusted by dilution factor)									
	1	2	3	4	5	6	7	8	9	10
1:27	783.384	536.263	634.695	847.745	1155.591	Bioanalysis 799.9	1458.221	806.082	717.324	801.086
1:81	262.771 (1)	201.742 (13)	221.117 (5)	294.52 (4)	387.206 (1)	306.138 (15)	Bioanalysis 7.762 (-4)	298.752 (11)	271.995 (14)	277.228 (4)
1:243	98.611 (13)	76.478 (29)	85.133 (21)	116.834 (24)	136.194 (6)	104.131 (17)	162.397 (0)	112.815 (26)	106.767 (34)	99.816 (12)
1:729	32.057 (11)	22.665 (14)	27.973 (19)	38.57 (23)	45.4 (6)	33.17 (12)	52.177 (-3)	36.917 (24)	33.987 (28)	33.138 (12)
1:2187	7.034 (-27)	BQL	5.425 (-31)	9.415 (-10)	13.371 (-6)	7.637 (-23)	[†] 14.076 (-22)	8.345 (-16)	BQL	7.261 (-27)

[†]Eight out of ten samples showed parallelism above 14.076 pg/ml. 15.63 pg/ml, the lowest standard curve point above 14.076 pg/ml, was chosen as assay LLOQ. BQL: Below quantifiable limit; MRD: Minimum required dilution.

Approaches for developing and validating assays using commercial kits have been discussed within the industry: the FDA draft guidance and EMA guidance indicated commercial kits used for drug development purposes need to be revalidated to ensure the reliability [2,3]. Different groups summarized the challenges and their recommendation of adapting commercial kits in their review or research papers [13,44,45]. In 2015, two white papers were published discussing the recommendations of method validation using singleplex and multiplex kits [22,24]. Kit end users also published some case studies demonstrating the strategies that they have used to adapt and qualify research grade commercial kits from different vendors for different purposes [31–33,44,49].

In short, the reference material and critical reagents need to be verified for quality and lot-to-lot variability, the assay LLOQ and MRD need to be confirmed or re-established, the buffers and diluents may need to be modified and assay steps may need to be optimized as necessary. The most efficient way to evaluate all of the above is through parallelism experiments.

Parallelism experiments are highly suggested at the early method feasibility stage. If nonparallelism is observed during the investigation, the same trouble shooting route is recommended (Figure 5).

Parallelism & multiplex LBAs

Compared with singleplex LBAs, multiplex LBAs appear to be more attractive when considering time and sample volume savings for sample analysis. Due to the extreme complexity of the assay environment, most of the multiplex assays cannot achieve the same level of quality when compared with singleplex assays [24,33]. Initially, it is suggested to use multiplex LBAs for exploratory biomarker screening purpose only [13,50]. After the

desired biomarker(s) is (are) chosen, singleplex or small multiplex (two or three) LBAs are preferred to support further drug development [13]. Although the path is not clear, assay developers continue their efforts to build reliable high-quality multiplex assays by improving method development and validation strategies.

Jani *et al.* pointed out in their multiplex LBAs fit-for-purpose validation white paper, which in addition to the challenges that are shared with singleplex LBAs, multiplex LBA development needs to consider other unique challenges. These challenges include: assay range and MRD setting for all analytes; cross-reactivity among different antibody pairs and analytes; and cross-talk due to well-to-well or spot-to-spot physical 'carryover' ('bleed-over' or 'bleaching') [24].

Parallelism or spiked dilutional linearity experiments can address all three challenges. Sensitivity should be set for each analyte using the same approach as for singleplex assays. MRD should be selected as the greatest dilution at which all analytes achieve parallelism [24,49]. Cross-reactivity can be addressed by spiking different levels of each capture or detection antibody into different samples and then performing parallelism experiments. Cross-talk can be addressed by spiking a high concentration of each analyte into different samples, respectively, and then performing dilutional linearity (spiked parallelism) experiments. Alternatively, select samples (if available) that have one of the analytes with an endogenous concentration significantly higher than other analytes and then performing parallelism experiments.

Other approaches were also proposed to address cross-reactivity and cross-talk issues. 'Missing man' techniques (i.e., all critical reagents except one are added to the assay each time) can be used to evaluate reagent cross-reactivity. Varying the concentration

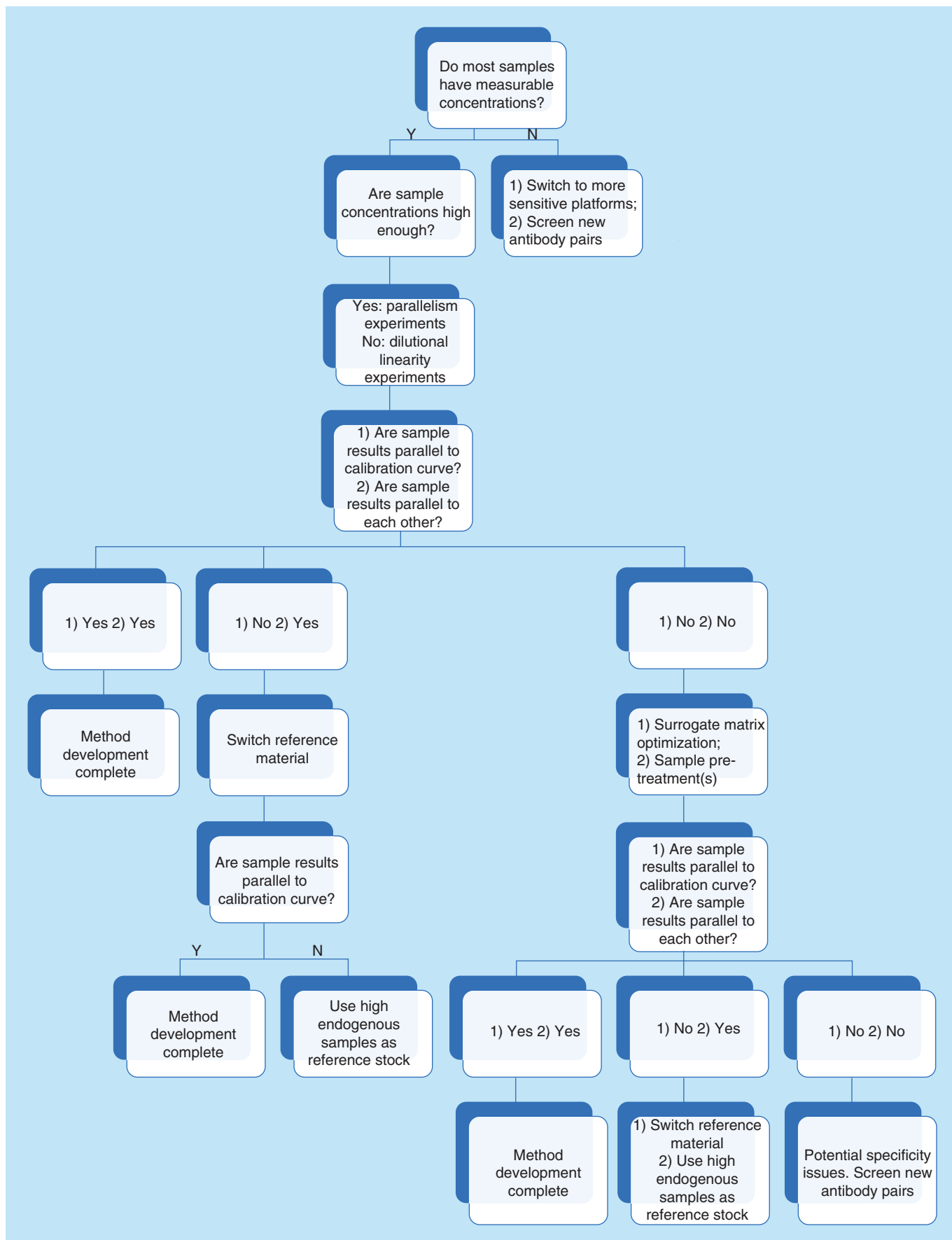


Figure 5. Method development decision tree using parallelism (dilutional linearity) experiments.

Table 3. Pros and cons of parallelism, dilutional linearity and spike/recovery and parameters evaluated.

Endogenous LBAs	Parallelism	Dilutional linearity (spiked parallelism)	Spike/recovery
Endogenous concentration	Yes Relative accuracy	Yes Relative accuracy	Yes No guarantee for accuracy
MRD	Yes For endogenous protein	Yes For reference material	Yes For reference material
Selectivity	Yes For endogenous protein	Yes For reference material	Yes For reference material
Reference material representative	Yes	No	No
Antibody pair specificity	No Provide indirect clues	No Provide indirect clues	Yes Only for the compounds tested
Sensitivity	Yes For endogenous protein	Yes For reference material	No
Cross-reactivity (multiplex LBAs)	Yes	Yes	Yes
Cross-talk (multiplex LBAs)	Yes For endogenous protein	Yes For reference material	Yes For reference material
Pros	<ul style="list-style-type: none"> • Relative accuracy for endogenous protein • Quick and easy all-in-one experiment • The closest mimic of real study samples 	<ul style="list-style-type: none"> • Relative accuracy for reference material • Quick and easy all-in-one experiment • Not restrained by sample endogenous concentrations 	<ul style="list-style-type: none"> • Absolute accuracy for reference material • Clear acceptance criteria setting
Cons	<ul style="list-style-type: none"> • Require samples with high enough endogenous concentrations • No guarantee for absolute accuracy • No guarantee for method specificity • No consensus regarding acceptance criteria setting 	<ul style="list-style-type: none"> • No guarantee reference material can represent endogenous protein • No guarantee for absolute accuracy • No guarantee for method specificity • No consensus regarding acceptance criteria setting 	<ul style="list-style-type: none"> • No guarantee reference material can represent endogenous protein • Time and material consuming • Cannot evaluate assay sensitivity without blank matrix • No guarantee for method specificity

LBA: Ligand-binding assay; MRD: Minimum required dilution.

of one analyte while keeping the other analytes at low concentrations is another method that can be used to address cross-talk issues. If commercial kits are used, due to the complexity and expense of reproducing those experiments, obtaining raw data copies from the kit manufacturer (if possible) is always recommended [24].

New technologies provide additional possibilities for developing well-characterized multiplex LBAs. One example is the cartridge-based ELISA platform ELLA (protein simple) [51,52]. ELLA fourplex cartridge is designed in a way that capture and detect antibodies that are physically separated in different microfluidic glass nanoreactors. The same sample then flows into four separated glass nanoreactors and

incubates with four different antibody pairs. Since different antibody pairs have never been physically mixed together, there should be no cross-reactivity and cross-talk caused by antibody pairs. Hence, the fourplex assay should perform the same as their singleplex counterparts. Fischer *et al.* published a case study that supported this assumption [33].

Future perspective

As the need for biomarker assays increases every day, the reliability of biomarker LBAs become more and more important. Parallelism experiments can be used to address matrix effects, selectivity and sensitivity concerns of the assay. However, as an inherited issue for

all LBAs, specificity cannot be addressed by parallelism experiments directly. But lack of parallelism in the assay after method optimization may be a sign for poor specificity. The specificity of the assay is determined by assay critical reagents (e.g., antibody pairs). Critical reagent specificity has to be evaluated either by the vendors or the end users to ensure the performance of endogenous/biomarker LBAs. The mathematical interpretation of parallelism data provides hints for LBA specificity characterizations. Furthermore, mathematical or statistical tools may be evaluated and used in order to fish out more details (regarding binding activities) from parallelism data. Other technologies such as western blots, surface plasmon resonance, HPLC and LC-MS/MS may also be used to help us determining the root causes for LBA nonparallelism. These technologies should be used as needed and as appropriated.

However, it is not clear what level of specificity characterization is adequate for an endogenous LBA. To determine the absolute specificity (which is not always needed) would require investing tremendous amount resources, which could have been used somewhere else. The ultimate goal for developing well-characterized endogenous LBAs is to make sure that they are sensitive and accurate enough to measure the changes it intend to measure. And at the same time, they measure the compounds as they supposed to measure.

Conclusion

Table 3 shows the summary of pros and cons of using parallelism, dilutional linearity (spiked parallelism) and spike/recovery methods to evaluate critical assay parameters for endogenous LBAs.

Executive summary

Understand nonparallelism

- Factors contributing to nonparallelism.
- Immunoaffinity differences between calibrator reference material and endogenous analyte.
- Matrix effect variances among calibration curve matrix, quality control matrix and study population matrix.
- Approaches of evaluating method nonparallelism.
- Back-calculated concentrations-based approaches.
- Raw signal-based approaches.
- Categories of nonparallelism and their root cause.
- Samples are parallel to each other, but samples are not parallel to calibrator curve.
- Root cause: The differences between reference material and endogenous protein.
- Samples are not parallel to each other.
- Root cause: The matrix effects of the samples and or the specificity issues of assay critical materials.

Suggested steps of using parallelism data to guide endogenous ligand-binding assay method development

- Assess assay sensitive and choose assay platform/antibody pair as appropriate.
- Assess endogenous level of the analyte and perform parallelism or dilutional linearity (spiked parallelism) experiments as appropriate.
- If samples are not parallel to the curve, use raw signal-based approach to evaluate the parallelism among samples.
- Determine the root cause of nonparallelism and optimize the assay accordingly.
- After method optimization, use back-calculated concentration-based approaches with fixed acceptance criteria to confirm assay parallelism.

Comparison of parallelism, dilutional linearity (spiked parallelism) & spike/recovery experiments

- Spike/recovery experiments demonstrate absolute accuracy. Parallelism and spiked parallelism experiments demonstrate relative accuracy.
- Spike/recovery and spiked parallelism experiments assess the performance of the spiked reference material. Parallelism experiments assess the performance of endogenous protein.
- Spike/recovery experiments cannot evaluate multiple assay parameters in one assay. Parallelism and spiked parallelism experiments can.
- None of the experiments can evaluate the specificity of the assay directly.

Future perspective

- The reliability of biomarker ligand-binding assays become more and more important, but assay specificity cannot be addressed directly.
- Furthermore, mathematical or statistical tools may be evaluated and used in order to fish out more details regarding binding activities.
- Other technologies (e.g., western blots, surface plasmon resonance, HPLC, LC-MS/MS, etc.) should be used as needed.

Unlike most exogenous drug assays, endogenous (e.g., biomarkers) LBAs often require surrogate reference standards and analyte-free surrogate matrices. The reference materials used for endogenous LBAs normally are neither highly purified nor well characterized and may not fully represent the endogenous proteins that are often heterogeneous and existing in multiple isoforms in the natural matrices. The matrix effects of the surrogate matrices are often different from the normal or target-diseased population matrices due to different matrix biology. All of the unique characters of endogenous LBAs make the assay often difficult to be developed and evaluated the same as traditional definitive quantitative PK assays. The foremost goal of well-characterized relative quantitative endogenous LBAs is to differentiate between diseased and normal population (i.e., for diagnosis/prognosis purpose) or drug treated and not treated individual samples (i.e., for drug development purpose) rather than measure the definitive concentration for each sample.

Spike/recovery experiments were demonstrated as the most effective and widely adapted approaches for method development and validation for definitive quantitative PK LBAs [4], which are not always needed for relative accuracy assays. The spiked dilutional linearity experiments can be used to demonstrate assay relative accuracy when samples do not have high-enough endogenous analyte levels to support parallelism experiment evaluation. However, the difference between spiked reference material and endogenous natural protein needs to be carefully tested and addressed.

Although the consensus regarding how to perform parallelism experiments and how to evaluate parallel-

ism data has not been reached, this should not affect the use of parallelism data for method development and optimization purposes. Parallelism experimentation is a powerful tool that can be and should be used to address the essential concerns of an endogenous LBA assay, with the exception of the specificity of the assay. It provides the closest imitation of the real study samples and provides sufficient information for the claim of relative quantitation and fit-for-purpose method design.

Acknowledgements

The authors thank R Jenkins, B Stouffer, D Zhang, R Kernstock and M Edwards for their critical review of this manuscript. The authors also thank J Hayes, C Lowery, C Wilson, L Yengi and C Horvath for their support to the case studies presented.

Financial & competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

- 1 Plikaytis BD, Holder PF, Pais SE *et al.* Determination of parallelism and nonparallelism in bioassay dilution curves. *J. Clin. Microbiol.* 32, 2441–2447 (1994).
- Discusses the causes of nonparallelism from mathematic and statistical perspectives.
- 2 *Draft Guidance for Industry, Bioanalytical Method Validation.* US Department of Health and Human Services, US FDA, Center for Drug Evaluation and Research, Center for Veterinary Medicine, MD, USA (2013). www.fda.gov/downloads/drugs/guidancecompliance
- 3 *Guideline on Bioanalytical Method Validation.* European Medicines Agency, London, UK. (2011). www.ema.europa.eu/docs/en_GB/document_library/
- 4 DeSilva B, Smith W, Weiner R *et al.* Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm. Res.* 20(11), 1885–1900 (2003).
- 5 Lee JW, Smith WC, Nordblom GD *et al.* Validation of assays for the bioanalysis of novel biomarkers. In: *Biomarkers in Clinical Drug Development.* Bloom JC, Dean RA (Eds). Marcel Dekker, NY, USA, 119–149 (2003).
- 6 Lee JW, Devanarayan V, Barrett YC *et al.* Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm. Res.* 23(2), 312–328 (2006).
- Proposes the fit-for-purpose approach for biomarker method development and validation.
- 7 Cummings J, Raynaud F, Jones L *et al.* Fit-for-purpose biomarker method validation for application in clinical trials of anticancer drugs. *Br. J. Cancer* 103, 1313–1317 (2010).
- 8 Ciotti S, Purushothama S, Ray S. What is going on with my samples? A general approach to parallelism assessment and data interpretation for biomarker ligand-binding assays. *Bioanalysis* 5(16), 1941–1943 (2013).
- 9 Stevenson LF, Purushothama S. Parallelism: considerations for the development, validation and implementation of PK and biomarker ligand-binding assays. *Bioanalysis* 6(2), 185–198 (2014).

- Discusses the fundamental key questions around parallelism assessments: why, when and how.
- 10 Nelson R. TT-61: non-parallelism in biomarker assays. Presented at: *EBF 9th Open Symposium*. Barcelona, Spain, 17 November 2016.
- 11 Tu J. A soluble receptor (sBCMA) biomarker parallelism case study – using parallelism experiments to effectively evaluate matrix effects and selectivity in early stage LBA method development. Presented at: *EBF 9th Open Symposium*. Barcelona, Spain, 17 November 2016.
- Demonstrates how to use parallelism experiments to guide method development and optimization with a case study.
- 12 Chau CH, Rixe O, McLeod H *et al.* Validation of analytic methods for biomarker used in drug development. *Clin. Cancer Res.* 14(19), 5967–5976 (2008).
- 13 Lee JW, Hall M. Method validation of protein biomarkers in support of drug development or clinical diagnosis/prognosis. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 877(13), 1259–1271 (2009).
- 14 Lee JW. Method validation and application of protein biomarkers: basic similarities and differences from biotherapeutics. *Bioanalysis* 1(8), 1461–1474 (2009).
- 15 Lee JW, Kelley M, King LE *et al.* Bioanalytical approaches to quantify ‘total’ and ‘free’ therapeutic antibodies and their targets: technical challenges and PK/PD applications over the course of drug development. *AAPS J.* 13(1), 99–110 (2013).
- 16 Valentin MA, Ma S, Zhao A *et al.* Validation of immunoassay for protein biomarkers: bioanalytical study plan implementation to support pre-clinical and clinical studies. *J. Pharm. Biomed. Anal.* 55(5), 869–877 (2011).
- Discusses the recommendations and considerations of using parallelism and spiked dilutional linearity experiments to evaluate method selectivity.
- 17 Timmerman P, Herling C, Stoellner D *et al.* European bioanalysis forum recommendation on method establishment and bioanalysis of biomarkers in support of drug development. *Bioanalysis* 4(15) 1883–1894 (2012).
- 18 DeSilva B, Garofolo F, Rocci M *et al.* 2012 White paper on recent issues in bioanalysis and alignment of multiple guidelines. *Bioanalysis* 4(18) 2213–2226 (2012).
- 19 Houghton R, Gouty D, Allinson J *et al.* Recommendations on biomarker bioanalytical method validation by GCC. *Bioanalysis* 4(20), 2439–2446 (2012).
- 20 Stevenson L, Kelley M, Gorovits B *et al.* Large molecule specific assay operation: recommendations for the best practices and harmonization from the global bioanalysis consortium harmonization team. *AAPS J.* 16(1), 83–88 (2013).
- 21 Booth B, Arnold ME, DeSilva B *et al.* Workshop report: Crystal City V-quantitative bioanalytical method validation and implementation: the 2013 revised FDA guidance. *AAPS J.* 17(2), 277–288 (2015).
- 22 Khan M, Bowsher RR, Cameron M *et al.* Recommendations for adaptation and validation of commercial kits for biomarker quantification in drug development. *Bioanalysis* 7(2), 229–242 (2015).
- Summarizes recommendations for adaptation and validation of commercial biomarker kits.
- 23 Amaravadi L, Song A, Myler H *et al.* 2015 White paper on recent issues in bioanalysis: focus on new technologies and biomarkers (Part 3 – LBA, biomarkers and immunogenicity). *Bioanalysis* 7(24), 3107–3124 (2015).
- 24 Jani D, Allinson J, Berisha F *et al.* Recommendations for use and fit-for-purpose validation of biomarker multiplex ligand binding assays in drug development. *AAPS J.* 18(1), 1–14 (2016).
- Summarizes recommendations for multiplex biomarker assay validation.
- 25 Lowes S, Ackermann BL. AAPS and US FDA Crystal City VI workshop on bioanalytical method validation for biomarkers. *Bioanalysis* 8(3), 163–167 (2016).
- 26 Arnold ME, Booth B, King L, Ray C. Workshop report: Crystal City VI – bioanalytical method validation for biomarkers. *AAPS J.* 18(6), 1366–1372 (2016).
- The workshop report that summarizes recommendations emerging from the Crystal City VI workshop that has focused heavily on biomarker method validation.
- 27 Cowan K. Implementing fit-for-purpose biomarker assay approaches: a bioanalytical perspective. *Bioanalysis* 8(12), 1221–1223 (2016).
- 28 King LE. Parallelism experiments in biomarker ligand-binding assays to assess immunological similarity. *Bioanalysis* 8(23), 2387–2391 (2016).
- 29 Jenkins RG. Accuracy: a potential quandary in regulated bioanalysis of ‘endogenous’ analytes. *Bioanalysis* 8(23), 2393–2397 (2016).
- 30 Richards S, Amaravadi L, Pillutla R *et al.* 2016 White paper on recent issues in bioanalysis: focus on biomarker assay validation (BAV): (Part 3 – LBA, biomarkers and immunogenicity). *Bioanalysis* 8(23), 2475–2496 (2016).
- 31 Miller AM, Rutkowska A, Bahl JM *et al.* Multicenter immunoassay validation of cerebrospinal fluid neurofilament light: a biomarker for neurodegeneration. *Bioanalysis* 8(21), 2243–2254 (2016).
- 32 Cowan K, Gan X, Parab V *et al.* Fit-for-purpose biomarker immunoassay qualification and validation: three case studies. *Bioanalysis* 8(22), 2329–2340 (2016).
- 33 Fischer SK, Carrasco-Triguero M, Hong Kyu *et al.* Commercial biomarker assays: friend and foe. *Bioanalysis* 8(22), 2351–2362 (2016).
- 34 Tate J, Ward G. Interferences in immunoassay. *Clin. Biochem. Rev.* 25(2), 105–120 (2004).
- 35 Kricka LJ. Human anti-animal antibody interferences in immunological assays. *Clin. Chem.* 45(7), 942–956 (1999).
- 36 Lee J, Ma H. Specificity and selectivity evaluation of ligand binding assay of protein therapeutics against concomitant drugs and related endogenous proteins. *AAPS J.* 9(2) E164–E170 (2007).
- 37 Gorovits B, McNally J, Fiorotti C *et al.* Protein-based matrix interferences in ligand-binding assays. *Bioanalysis* 6(8), 1131–1140 (2014).

- 38 Hennig C, Rink L, Fagin U *et al.* The influence of naturally occurring heterophilic anti-immunoglobulin antibodies on direct measurement of serum protein using sandwich ELISAs. *J. Immunol. Methods* 235(1–2), 71–80 (2000).
- 39 Schwichart M, Vainshtein I, Lee R *et al.* Interference in immunoassays to support therapeutic antibody development in preclinical and clinical studies. *Bioanalysis* 6(14), 1939–1951 (2014).
- 40 Salimi-Moosavi H, Lee J, DeSilva B *et al.* Novel approaches using alkaline or acid/guanidine treatment to eliminate therapeutic antibody interference in the measurement of total target ligand. *J. Pharm. Biomed. Anal.* 51(5), 1128–1133 (2010).
- 41 Bastarache JA, Koyama T, Wickersham NE *et al.* Accuracy and reproducibility of a multiplex immunoassay platform: a validation study. *J. Immunol. Methods* 367(1–2), 33–39 (2011).
- 42 Stevenson L, Kelley M, Gorovits B *et al.* Large molecule specific assay operations: recommendations for best practices and harmonization from the Global Bioanalysis Consortium Harmonization Team. *AAPS J.* 16(1), 83–88 (2014).
- 43 Marcelletti JF, Evans CL, Saxena M *et al.* Calculations for adjusting endogenous biomarker levels during analytical recovery assessments for ligand-binding assay bioanalytical method validation. *AAPS J.* 17(4), 939–947 (2015).
- 44 Bowsher RR, Sailstad JM. Insights in the application of research-grade diagnostic kits for biomarker assessments in support of clinical drug development: bioanalysis of circulating concentrations of soluble receptor activator of nuclear factor κ B ligand. *J. Pharm. Biomed. Anal.* 48(5), 1282–1289 (2008).
- 45 Nowatzke W, Cole TG, Bowsher RR. Systematic analytical validation of commercial kits for the determination of novel biomarkers for clinical drug development. *Bioanalysis* 2(2), 237–247 (2010).
- 46 Wikipedia. CE marketing. http://en.wikipedia.org/wiki/CE_marking
- 47 US FDA. Overview of IVD regulation. www.fda.gov/medicaldevices/deviceregulation
- 48 *IVD Guidance: Research Use Only Products. A Guide for Manufacturers and Notified Bodies.* European Commission Enterprise Directorate-General (2004). www.emergogroup.com/sites/default/files/file/europe
- 49 Eastman PS, Manning WC, Qureshi F *et al.* Characterization of a multiplex, 12-biomarker test for rheumatoid arthritis. *J. Pharm. Biomed. Anal.* 70, 415–424 (2012).
- 50 Ray CA, Bowsher RR, Smith WC *et al.* Development, validation, and implementation of a multiplex immunoassay for the simultaneous determination of five cytokines in human serum. *J. Pharm. Biomed. Anal.* 36(5), 1037–1044 (2005).
- 51 Protein Simple ELLA. www.proteinsimple.com/ella.html
- 52 Aldo P, Marusov G, Svancara D *et al.* Simple Plex™: a novel multi-analyte, automated microfluidic immunoassay platform for the detection of human and mouse cytokines and chemokines. *Am. J. Reprod. Immunol.* 75(6), 678–693 (2016).