
White Paper

Recommendations for Selection and Characterization of Protein Biomarker Assay Calibrator Material

Kyra J. Cowan,^{1,12} Lakshmi Amaravadi,² Mark J. Cameron,³ Damien Fink,⁴ Darshana Jani,⁵ Medha Kamat,⁶ Lindsay King,⁵ Robert J. Neely,⁷ Yan Ni,⁸ Paul Rhyne,⁹ Renee Riffon,¹⁰ and Yuda Zhu¹¹

Received 28 February 2017; accepted 20 June 2017; published online 2 October 2017

Abstract. As biomarkers continue to become an integral part of drug development and decision-making, there are increased expectations for reliable and quantitative assays. Protein biomarker assay results are directly influenced by the calibrator material. The selection of calibrator material presents many challenges that impact the relative accuracy and performance of the assay. There is an industry-wide challenge finding reliable and well-characterized calibrator material with good documentation. Several case studies are presented that demonstrate some of the challenges involved in selecting appropriate calibrators along with the resolutions that were ultimately applied. From these experiences, we present here a set of recommendations for selecting and characterizing calibrator material based on the intended purpose of the assay. Finally, we introduce a commutability approach, based on common clinical chemistry practices, which can be used to demonstrate interchangeability with calibrator materials across multiple lots and technology platforms for all types of protein biomarker assays.

KEYWORDS: biomarker assays; calibrator; reference material.

INTRODUCTION

Biomarkers and their corresponding assays have become a focal point for drug discovery and development. Reliable results from quantitative biomarker assays are essential to help ensure target engagement, assess pharmacodynamics in support of

dosing selection, evaluate efficacy, and monitor safety. There are many published white papers, reviews, and guidance documents that provide a framework on how biomarker assays should be evaluated and how to ensure the bioanalytical performance is adequate to support and generate scientifically meaningful data (“fit-for-purpose” or FFP) (1–13). Biomarker assay assessment, evaluation, and validation strategies based on FFP principles are used to ensure appropriate rigor surrounding the data used for science-based decision-making (14) during drug development. One of the most significant challenges with quantitative biomarker assays is finding calibrator material that is structurally and functionally representative of the endogenous analyte found in clinical samples. Biomarker assay calibrators are generally spiked into a relevant matrix; but these calibrators may be recombinant or synthetic and are not identical to the endogenous analyte (including, but not limited to, protein sequences, folding, glycosylation, and purity). Calibrators and endogenous analytes may have different binding characteristics with assay reagents, resulting in calibration curves not being fully representative of the endogenous biomarker measurement. A further challenge with obtaining a suitable calibrator is that endogenous levels of the biomarker may be present with a variety of post-translational modifications depending on disease-state genetics, environment, and matrix (15). Pharmacokinetic (PK) assays utilize a highly characterized calibrator supplied with a certificate of analysis listing purity and concentration. This material is considered equivalent to the primary active metabolite. Although the 2013 FDA draft

¹ Merck KGaA, Frankfurter Str. 250, F130/005, Darmstadt, 64293, Germany.

² Sanofi Genzyme, 1 the Mountain Road, Framingham, Massachusetts 01701, USA.

³ Beckmann Coulter, 3551 Lawrenceville Road, Princeton, New Jersey 08540, USA.

⁴ Janssen Research and Development, Inc., 1400 McKean Road, Spring House, Pennsylvania 19477, USA.

⁵ Pfizer, 610 Main Street, Cambridge, Massachusetts 02142, USA.

⁶ GlaxoSmithKline, 709 Swedeland Rd, King of Prussia, Pennsylvania 19038, USA.

⁷ Bristol-Myers Squibb, 3551 Lawrenceville Road, Princeton, New Jersey 08540, USA.

⁸ Regeneron, Tarrytown, New York, USA.

⁹ Q² Solutions, 5827 South Miami Blvd, Morrisville, North Carolina 27560, USA.

¹⁰ CiToxLAB North America, 445 Boulevard Armand-Frappier, Laval, Quebec H7V 4B3, Canada.

¹¹ Genentech/Roche, 1 DNA Way, South San Francisco, CA 94080, USA.

¹² To whom correspondence should be addressed. (e-mail: Kyra.Cowan@merckgroup.com)

bioanalytical validation guidance recommends that biomarker assays address some of the same validation questions as that of PK assays, alternative experiments and approaches need to be considered depending on the biomarker and suitability for the intended use of the assay.

The primary challenge for developing and validating any biomarker assay is ensuring that the calibrator material closely represents the endogenous form of the biomarker to be measured. For this, we provide practical solutions and best practices for the following main three topics:

1. Identification, selection, and characterization of appropriate protein and peptide calibrators, to support biomarker assay development,
2. Evaluation and maintenance of lot-to-lot consistency (i.e., the maintenance of assay system suitability during the period of its use after pre-study validation), and
3. Establishment of the relationship between calibrator material and endogenous analyte using actual samples in the framework of multiple assays by borrowing an approach commonly used in clinical chemistry known as commutability.

REGULATORY GUIDANCE DOCUMENTS AND RECOMMENDATIONS FOR THE BIOANALYTICAL COMMUNITY TO DATE

There are regulatory guidance documents on bioanalytical method validation from multiple regulatory agencies including the Food and Drug Administration (FDA) and European Medicines Agency (EMA) (1–3). These guidance documents are intended to assist sponsors of both human and animal studies for new drug applications and generally apply to bioanalytical procedures. These documents provide general recommendations for bioanalytical method validation (1). The EMA Guidance for bioanalytical method validation (3) refers only to PK assays and recommendations for biomarker method validation are out of scope for the current EMA guidance. The Japanese Pharmaceuticals and Medical Devices Agency (PMDA) does not include biomarker method validation in their guidance document. The FDA 2013 draft guidance has expanded sections that, in contrast to the previous 2001 version (2), include the type of biological matrix to be used for standard and QC preparation, the extent of the validation based on the intended use of the data, evaluation of the immunological activity differences between the calibrator from that of the subject samples, and lot-to-lot and reagent comparability assessment for critical reagents.

The approach for biomarker assay validation and use of the calibrator material is often extended from bioanalytical PK assay practices; whereas, in fact, most of these practices do not apply to biomarker assays. The 2013 FDA draft guidance for bioanalytical method validation stated that biomarker assays can be used for a variety of purposes during drug development and should incorporate a tiered validation approach. However, biomarker data that is to be used to support actions such as a pivotal determination of safety and labeled dosing instructions should be generated by a bio-

marker assay that is fully validated. The draft guidance document further states that method validation for biomarker assays should use the principles of PK validation as a starting point. The guidance document does not provide information on how to address the challenges associated with protein calibrators and the issues associated with achieving the acceptance criteria of PK assay validation (1). To address this gap, the Crystal City VI Workshop (14) focused on the many challenges encountered by bioanalytical scientists with regards to biomarker assays and provided some recommendations on addressing these key challenges and support for continued efforts to alleviate industry-wide concerns. But most significantly, biomarker assays almost never have authenticated analytical reference standards available and thus typically rely on calibrator material that is relatively uncharacterized by the provider. There is no guidance or requirement for calibrator material used for biomarker assays despite the fact these assays are utilized across the drug development continuum.

Several publications refer to the use of calibrator material for bioanalytical assays. However, these white papers mostly focus on PK assay standard and quality control material. Some of the key components of assays (including detection reagents, biological matrices, and calibrator material needed for biomarker assays) were not addressed. Recommendations for engineered protein characterization may be useful to consider for biomarker protein calibrators. The 2014 Global Bioanalysis Consortium (GBC) reagent white paper (5) on critical reagents used in assays, including biomarker calibrators, acknowledged that biomarker calibrator material is typically much less characterized than those for PK assays. Adding to the complexity, the National Institute for Standards and Technology (NIST) or the World Health Organization (WHO) reference materials are often not established making it challenging to control lot changes for calibrators. Most importantly, with regards to our efforts, the paper acknowledged the gap in expected practices around biomarker assay calibrators, and that specific guidance around selecting and characterizing biomarker calibrator materials is needed. Thus, it was suggested that biomarker calibrators could be treated in the same way as critical reagents in the interim (16).

In addition to these aforementioned papers, a commentary paper from the GBC on PK reference materials (17) discussed the quality of reference material in a similar approach as the EMA guidance. The scope of the manuscript was limited to reference material and related key molecules such as metabolites and internal standards used in the support of regulated bioanalysis for new chemical entities (NCEs) and new biological entities (NBEs). The paper made several recommendations concerning stability, lot-to-lot variability, bridging material, and material purity.

For endogenous analytes, there are several international reference standard sources for their corresponding recombinant proteins such as the National Institute for Biological Standards and Control (NIBSC, www.nibsc.org/products/brm_product_catalogue.aspx), the European Directorate of the Quality of Medicines (EDQM, www.edqm.eu/en/ph-eur-reference-standards-627.html), the US Pharmacopeial Convention (USP, www.usp.org/reference-standards), NIST (www.nist.gov/srm/index.cfm), and WHO (<http://>

www.who.int/bloodproducts/ref_materials/en/). In general, the reference materials provided by these organizations have been developed for clinical diagnostic assays and typically involved contributions from numerous users in developing a consensus on the selection, testing, and use. For example, there is substantial literature on the development and characterization of reference materials for human troponin I (18–20). The first step in the characterization process for troponin I was to address the issue of standardization by assessing and then biochemically characterizing an appropriate candidate reference source. Multiple forms of troponin were available, so recombinant and purified sources were assessed for commutability, a property of the calibrator that shows it behaves similarly to samples in a defined setting that is described in much further detail later in this paper. Commutability is a property of the calibrator that shows it behaves similarly to samples in a defined setting and is described in much further detail later in this paper. Based on multiple rounds of testing in different labs and consensus within the industry, a certified standard reference material was developed by the NIST and made available to assay manufacturers for use in their assays.

Most biomarker calibrators in assays used by the pharmaceutical industry are not fully characterized according to the standards above. Therefore, we provide recommendations on how these calibrators can be selected and characterized by the assay user in the scope of drug development. Our goal is to provide the best practices and recommendations on assessing baseline characteristics for calibrator material, and how we can work towards the possible “gold standard” for select analytes.

CHALLENGES RELATED TO DIFFERENCES BETWEEN CALIBRATORS VS ENDOGENOUS BIOMARKER

For quantitative protein biomarker assays, it is often assumed that calibrator material is representative of the endogenous analyte of interest, similar to those used in PK assays (17). For protein biomarker assays, there is typically a lack of complete characterization of the endogenous analyte making it difficult to ensure that the calibrator is representative of the endogenous protein biomarker. This is further complicated when an endogenous analyte is not a single species, does not have the same amino acid sequence as the calibrator, does not possess a uniform physiochemical structure, or have fully described (or considered) post-translational states. Even with a relatively clear understanding of the biology of the endogenous analyte, it is often impossible to produce a calibrator material that is identical to its endogenous counterpart.

Thus, the most important starting point for protein biomarker assay development is a good understanding of the biology of a biomarker/analyte, including knowledge of normal or disease-state physiology, to select a representative calibrator.

Some key considerations in comparing the endogenous analyte to the protein calibrator include the following:

- The physicochemical properties of the endogenous analyte (i.e., whether the protein exists as a monomer, dimer, trimer, etc.),

- The relevant isoforms of the biomarker, post-translational modifications, cleavage, clipping, etc.,
- The recombinant calibrator material expression system (e.g., *Escherichia coli*, insect cells or eukaryotic expression system),
- The potential for misfolded proteins and consequences if they are not bioactive,
- The potential for lot-to-lot differences in calibrator material, and
- The potential for stability differences between the recombinant calibrator and endogenous analyte.

The following case studies illustrate the challenges related to the lack of a representative reference standard for protein biomarker measurements and potential impact on the biomarker measurements.

Case Study 1: Differences in the Physiochemical Structure of the Endogenous Biomarker and the Protein Calibrator

This example illustrates the importance of knowing the biology of the target analyte and specifically understanding differences between the physiochemical structure of the endogenous analyte and the calibrators planned for use. Soluble programmed cell death protein 1 (sPD-1) is a monomeric protein comprised of the extracellular domain of membrane-bound PD-1 (21). A commercial PD-1 ELISA kit used a dimeric sPD-1-Fc fusion as the calibrator material. Both monomeric sPD-1-His and dimeric sPD-1-Fc were available as recombinant calibrator material. During the development of a sPD-1 electrochemiluminescent (ECL) assay (22), surface plasmon resonance studies showed the capture antibody had a higher binding affinity for sPD-1-Fc than sPD-1-His (0.5 *versus* 2.8 nM). Consequently, there was approximately a 50-fold difference in concentration-dependent signal when the calibrator curves of both sPD-1-His monomer and sPD-1-Fc dimer were tested (Fig. 1). Ultimately, the authors concluded that the sPD-1-His protein most closely resembled the structure of the endogenous material and it was used as the calibrator material. Although the choice of calibrator affects the absolute value of the sPD-1 levels, the relative post-treatment percent change relative to baseline was not affected.

Case Study 2: A Case of Measurement of Multiple Isoforms of a Biomarker

This following case study illustrates challenges faced when extrapolating concentration information from activity information. The cytokine target of a therapeutic had multiple isoforms, and each isoform was available for purchase. The certificate of analysis provided the biological activity information for each isoform expressed in activity units per mass of material and as activity units per volume of material. Both sets of activity information were generated from two different assays with appropriate literature references. The vendor was consistent in selecting one assay for reporting activity, while also providing the results of the alternative method. Theoretically, assuming that the assays are appropriately comparable, dividing activity units per volume by activity units per mass should have yielded a mass/volume concentration.

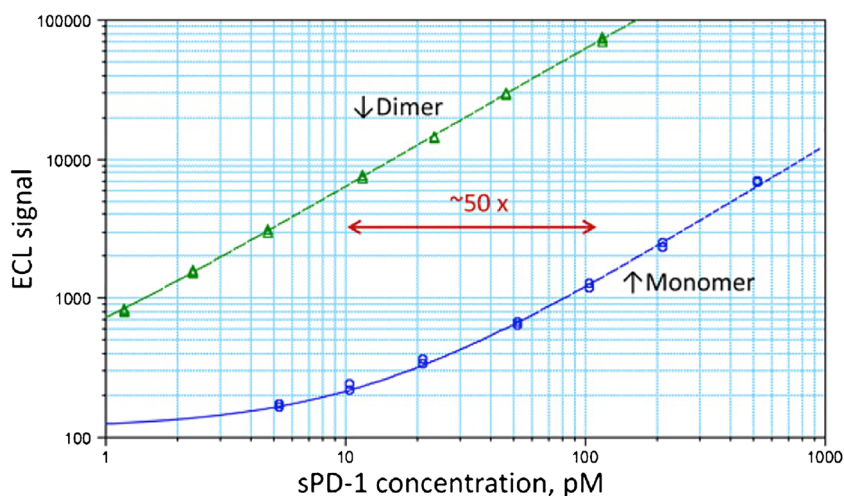


Fig. 1. Soluble PD-1 ECL assay standard curves using a monomeric (*green line*) or a dimeric (*blue line*) recombinant soluble PD-1 protein as calibrators. The capture antibody had a higher binding affinity for dimeric sPD-1-Fc than monomeric sPD-1-His. An approximately 50-fold right shift in the calibrator curve was observed when the sPD-1-His monomer was used as calibrator material. The commercially available assay using the dimeric sPD-1-Fc protein as calibrator material could underestimate amount of endogenous monomeric target

To determine the concentration of the commercially available materials, each isoform was spiked at two concentrations (high and low) and tested in three different ligand-binding assays. Assays 1 and 2 were internally developed assays (on two different platforms) able to measure all but two of the cytokine isoforms and utilized an internally produced version of the most common isoform as calibrator material. All the reagents used in each assay were identical and procedural steps were performed in a step-wise manner. For both assays, five isoforms recovered within 25% of the predicted concentration. Three isoforms recovered outside 25% but within 50% of the predicted concentration including the commercial version of the isoform used as the calibrator material. The last four isoforms were not measurable including the two isoforms that were not expected to be measurable based on reagent specificity. Assay 3 is a commercially available kit obtained from the producer of the commercially available isoforms while utilizing the same internally produced calibrator as assays 1 and 2 above. Only one isoform (the calibrator isoform) recovered within 25% of the predicted concentration. Four isoforms recovered around 10% of the predicted concentration. The remaining isoforms were not measurable including one isoform that was expected not to be measurable based on the assay product insert.

The results are shown in Table I. This case study illustrates the possible pitfalls in extrapolating concentration information from activity information and for understanding the effectiveness of the protein biomarker assay to measure the isoform of interest. The performance using the in-house assays was closer to the predicted concentration than the vendor's assay emphasizing the need to independently characterize calibrators and commercial assays. Ultimately, the in-house calibrator material was used to prepare all calibrators and QCs for assay 1. If there had been no internal source of calibrator material and the vendor calibrator concentration was not available, the recommendation would

be to ensure consistent calibrator performance by assigning a concentration (i.e., value assignment; (23)) to new lots of materials against the original lot of material used to establish the assay.

Case Study 3: An Example of Higher Order Structure Differences Between Endogenous Analyte and the Protein Calibrator

The higher order structure between the endogenous and exogenous analytes could be different and is dependent on how they were generated. This could lead to differences in tertiary epitopes between the calibrator and the endogenous biomarker and have a direct impact on assay specificity and relative accuracy in relation to the endogenous analyte.

Interleukin 23 (IL-23) is a heterodimeric protein consisting of separately expressed p19 and p40 subunits. The p40 subunit is also found in the closely related interleukin 12 (p35 and p40 subunits). A commercially available Luminex-based kit used to measure IL-23 contained a calibrator produced with both the p19 and the p40 subunits fused as a single protein. The results were compared to an alternative, in-house developed assay specific for IL-23 which used an *ex vivo*-derived, non-fused IL-23. This kit gave substantially higher than expected levels of the cytokine, as compared to the literature and previous in-house results (Table II). For the in-house developed assay, most sample values were lower than detectable limits (<9.6 pg/mL). The same samples were measured using a third assay kit utilizing the *ex vivo*-derived calibrator material. These results verified the results generated with the in-house assay. Both assays that used the *ex vivo* calibrator corroborated the previously published results for IL-23.

In this case study, the calibrator material used in the Luminex assay did not fully represent the endogenous analyte. This calibrator was used to generate the antibodies used in the Luminex kit, leading to reagent antibodies with

Table I. Recovery of Isoforms

Percent of nominal spike recovered						
Cytokine isoform	Assay 1		Assay 2		Assay 3	
	High (%)	Low (%)	High (%)	Low (%)	High (%)	Low (%)
1	52	52	50	68	109	121
2	0	0	0	0	16	0
3	92	99	105	166	13	0
4	10	0	0	17	0	3
5	1	0	0	6	0	0
6	84	88	102	123	0	0
7	8	0	2	0	0	6
8	75	79	82	99	10	17
9	68	68	50	49	7	0
10	83	84	59	52	14	0
11	146	119	134	135	0	0
12	56	39	38	46	0	0

Concentrations for each cytokine isoform were calculated from units/volume and units/mass available in the certificates of analysis. Spiked isoform concentrations were prepared in the upper and lower half of the standard curve range (denoted as “high” and “low”). Results are expressed as empirical result divided by predicted result $\times 100$

potentially altered specificity that may have bound to some other proteins in the samples. Ultimately, specificity of the Luminex assay could not be confirmed and results generated in the assay contradicted established biology. The in-house assay was used in place of the Luminex kit.

Case Study 4: Lot-to-Lot Variability in the Calibrator Material

It is essential for a vendor (or user) to evaluate lot-to-lot variability. In this case, an MSD multiplex (4-plex) “research only” biomarker assay that was not in use for a period of time

Table II. Comparison of Endogenous IL-23 Concentrations from Three Different Assays

Sample	Luminex (pg/mL)	In-house (pg/mL)	Third-party vendor (pg/mL)
1	283	LLOQ	0.047
2	44.5	LLOQ	0.111
3	7004	13.4	0.084
4	LLOQ	LLOQ	0.033
5	102	LLOQ	0.035
6	396	LLOQ	0.065
7	60	LLOQ	0.116
8	1529	LLOQ	0.011
9	715	LLOQ	0.247
10	LLOQ	LLOQ	0.071

A Luminex kit and an in-house IL-23 assay measured substantially different levels of the analyte in the same set of samples. These samples were also measured with an established third-party vendor assay and the results were comparable with the in-house assay and expected biology
LLOQ lower than the limit of quantitation

was re-evaluated with newly purchased material from the vendor. During kit qualification, two out of the four analytes showed changes to their respective assay ranges—the anchor point and LLOQ for one analyte were shifted up decreasing sensitivity and half the dynamic range (with a decrease in sensitivity) was eliminated in the other analyte (Table III). Consequently, many of the samples for these two analytes were now below the LLOQ. The vendor was contacted; however, since they were not the original manufacturer of the analytes, they could not reproduce the material they had provided originally. Minimal technical support was offered. The assay had to be re-qualified to the revised range due to lack of suitable alternatives. This case study illustrates that when purchasing “research only” material as calibrators, it is important to be aware that they are not necessarily subjected to a highly standardized quality control process.

Case Study 5: Differences in Calibrator Material Performance from Multiple Vendors

Materials from multiple vendors may not perform similarly, as is the case of mouse TNF α calibrator material. Calibrator curves were prepared based on the vendor-supplied concentration values and were tested in two different assays (commercial and custom). In the commercially available assay, recombinant mouse TNF α obtained from vendors 1, 2, and 3 were diluted to generate identical calibrator curves. As indicated in Fig. 2a, the three source materials performed quite differently in the assay. The materials were also tested in the custom assay (Fig. 2b) using identical calibrator curves prepared based on the vendor-supplied concentration. Materials from vendors 1 and 2 were comparable, while signal from the calibrator curve generated from the vendor 3 material was higher. At the highest concentrations, the calibration curves are parallel suggesting the differences are due to the value assignment of the material. Additionally, mammalian cell line produced mouse

Table III. Case Study Results Comparing the Qualification of New Lots of Calibrator Material (A, Analyte 1 and B, Analyte 2)

Standard conc. (pg/mL)	Mean	%CV	%RE
A: Analyte 1-validated range: new lot performance			
1.00	0.41	73.8	-58.8
2.00	2.23	6.1	11.7
3.00	2.70	8.1	-10.2
10.00	11.69	3.2	16.9
50.00	54.43	2.4	8.9
250.00	236.97	2.0	-5.2
1000.00	958.55	1.4	-4.1
2500.00	2538.78	4.1	1.6
10,000.00	10,405.5	3.2	4.1
15,000.00	14,969.1	2.5	-0.2
20,000.00	19,874.8	4.2	-0.6
B: Analyte 2-validated range: new lot performance			
1.00	0.99	2.9	-1.4
2.00	2.01	3.5	0.4
3.00	3.06	6.1	1.9
10.00	10.09	2.1	0.9
50.00	50.84	1.8	1.7
250.00	240.68	2.3	-3.7
1000.00	1015.89	1.9	1.6
2500.00	2512.16	0.7	0.5
10,000.00	9,687.90	8.9	-3.1
15,000.00	13,901.9	17.2	-7.3
20,000.00	13,074.8	53.9	-34.6

Analyte 2 did not meet acceptance criteria. The dynamic range for the analyte was significantly different than the lot used for validation. Since the calibrators are purchased as a combination of all four analytes, it was impossible to replace Analyte 2. The manufacturer was contacted regarding the shift in performance; however, since the kit passed the manufacturer QC testing and it is “for research use only,” the discussions did not resolve the issue. Instead, the method had to be changed to further dilute and reanalyze samples that had concentration in the range of 10,000–20,000 pg/mL for more accurate concentration determination from the middle of the curve

TNF α was tested in the custom assay (identified as vendor 4). The mammalian expressed material performed similarly to the recombinant material from vendor 3. Performance of material from vendors 1 and 2 were different in commercial assays yet identical in the internally developed assay. Ultimately, the material from vendor 3 was selected to support further sample analysis because it exhibited consistent performance, comes from a qualified vendor (definition in the “Challenges Unique to Commercially Available Proteins” section) and performs similarly to the mammalian cell line produced material.

Case Study 6: Stability Differences Between Endogenous Biomarker and the Protein Calibrator

In PK assays, it is standard practice to utilize matrix samples spiked with a reference material to evaluate short- and long-term stability of samples. However, due to the uncertainty in whether a reference material is identical to the endogenous analyte, caution should be exercised when establishing the stability of biomarker samples. There are a number of examples of biomarkers implemented during the

course of drug development where the stability of the endogenous biomarker was different than the recombinant/purified reference material spiked into matrix (10). We describe here the two scenarios where the stability of the calibrator material did not represent the endogenous biomarker stability. In the first case, long-term stability (over 1.5 years), for TGF β 1, was evaluated using spiked-in calibrator material. There was a need to repeat the measurement of baseline samples from the clinical study a year later. It was noted that the endogenous biomarker exhibited a decline in measured levels over time (~40% loss), behaving differently than the calibrator material (unpublished data). In another example, endogenous IL-13 in clinical samples was stable for 15 months while the recombinant calibrator was only stable for 4 months (10). This finding allowed the researchers to utilize and evaluate the analyte of interest in samples stored over a much longer period of time than that defined by the calibrator material stability. These two examples highlight the importance of utilizing endogenous samples for evaluating stability rather than relying on matrix spiked with purified recombinant protein calibrators. Individual or pooled patient samples have also been recommended in other biomarker publications as yielding more informative results (9,11). However, it should be noted that it is useful to characterize the stability of recombinant calibrators as this knowledge can aid in assay troubleshooting when inconsistencies arise between batches of sample analyses over a long period of time.

CHALLENGES WITH PROTEIN CONCENTRATION MEASUREMENT

A key characteristic of a protein biomarker assay calibrator is the manufacturer-assigned concentration. The determination of protein concentration is an important aspect in a wide range of scientific disciplines from basic academic research to biopharmaceutical lot release. Various platforms and methods are available to determine protein concentration (24,25). Provided here (Table IV) is a list of possible sources of functional and other information for proteins. Commercial vendors that offer proteins commonly advertise the concentration of the protein along with its purity (reported as a percentage; typically based on a silver stain or HPLC analysis), source or origin (e.g., *E. coli* derived), an accession number, formulation (e.g., phosphate buffered saline), storage, and stability information. Vendors may, or may not, provide the method(s) used to assign a concentration to their protein product. The types of assays that they use are typically plate-based colorimetric assays or UV absorption spectroscopy. As part of pre-purchase due diligence, information about how the vendor has assigned a concentration value to the protein produced should be gathered. Contacting technical support may be necessary to obtain this information. Traditional methods for the determination of protein concentration all have strengths and weaknesses and include absorbance at 280 nm, Lowry’s method, the bicinchoninic acid (BCA) assay, the Bradford method, and amino acid analysis (26,27). Understanding the reason the vendor selected a particular method over another may be worth discussing with technical support. If this information is not available, then the material may not be sufficiently characterized.

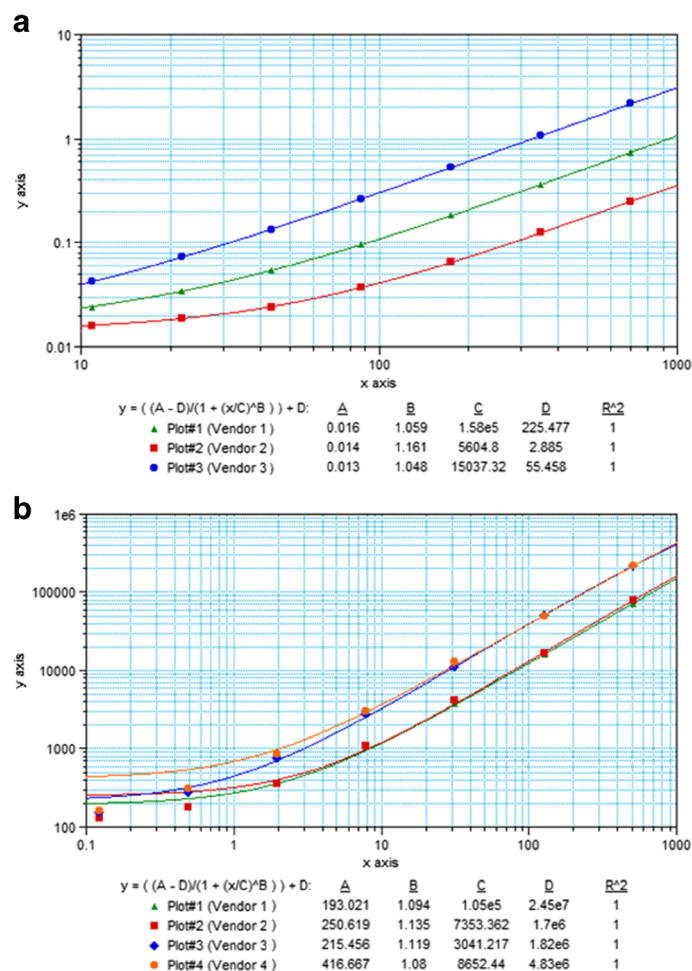


Fig. 2. Measuring multiple mouse TNF α materials in multiple methods. Each material was diluted to a common concentration (based on each manufacturer's product datasheet) prior to further twofold serial dilution. The three separate calibrator materials tested in the commercial assay (**a**) yielded three distinct calibration curves. The four separate calibrator materials tested in the custom assay (**b**) yielded two distinct calibration curves; as calibrators from vendors 1 and 2 performed similarly and calibrators from vendors 3 and 4 performed similarly. Vendors 1, 2, and 3 are identical in both graphs

In addition, control of lot-to-lot variability, details on the calibrator molecular weight (especially for fusion proteins), and accurate extinction coefficients for each protein are important to know prior to purchasing. Reliable protein concentration determination along with details of concentration determination is one of the most important features of a well-characterized material and part of the user due diligence in qualifying the vendor. In fact, the nominal concentration of the calibrator is the only “truth” of a quantitative assay—every other measure is inferred from the response and the model fit. It is important to note, that both the method used to determine calibrator concentration and the calibrator used in the protein assay should be known.

RISKS ASSOCIATED WITH UNREPRESENTATIVE CALIBRATOR AND ENDOGENOUS ANALYTE

Specific to the intended purpose of the biomarker assay results, the risk of calibrator material not representing the

endogenous analyte can not only have a significant impact on the interpretation of the data, but also decisions made based on that data, project timelines, and potentially resources. In extreme cases, the assays may not be measuring the intended analyte as was described by Prassas *et al.* (2013) for a CUZD1 assay that actually measured CA125 (28) and as was described earlier with our case study for IL-23.

RECOMMENDATIONS FOR BIOMARKER CALIBRATOR MATERIAL CHARACTERIZATION

Typical characterization parameters recommended for protein biomarkers are listed in Table V, showing a tiered approach for characterizing calibrator material for these assays. The recommendations take into account the intended use of the data (e.g., exploratory *versus* clinical decision-making *versus* regulatory authority review) regardless of whether the assay material is commercially available or generated in-house under

Table IV. Example of Commonly Used Protein Databases

Web address	Comments
http://web.expasy.org/docs/swiss-prot_guideline.html	Protein sequence database, which brings together experimental results, computed features, and scientific conclusions
http://www.hprd.org	The Human Protein Reference Database provides information about post-translational modifications and disease associations.
http://www.ncbi.nlm.nih.gov/protein/	A collection of protein sequences from several sources
http://www.uniprot.org/	A comprehensive and freely accessible resource of protein sequence and functional information
http://www.genecards.org/	A searchable, integrated database of human genes that provides comprehensive and user-friendly information on all known and predicted human genes
http://www.proteinatlas.org/	Protein information mainly provided by antibody interrogation
http://blast.ncbi.nlm.nih.gov/Blast.cgi	The Basic Local Alignment Search Tool (BLAST) finds regions of local amino acid sequence similarity.

The selected websites provide information regarding amino acid sequence, structure, function, binding partners, and post-translational modifications of many proteins

stringent conditions. One caveat for Table V is that the recommended amount of characterization is highly dependent on knowledge of the endogenous protein. The table provides a consistent approach in the selection and characterization of a calibrator; and it conveys the expectation for characterization from the scientific community. In addition to initial characterization rigor, the table recommendations were designed to ensure long-term assay performance during the life cycle management of the assay.

We categorized traditional biomarker assays into two categories in Table V: “relative quantitative assays used in exploratory setting” and “quantitative assays used to justify dose, to show efficacy, and/or to support drug registration”. Typically, the term definitive quantitative is applied when the calibrator material fully represents the endogenous analyte. We have categorized the use of biomarker assay data into these two columns knowing that most assays utilized during drug development are considered “relative quantitative” (12). In the table, we refer to the “product datasheet” although depending on the vendor, this could be called the certificate of analysis (C of A). We acknowledge that different vendors could have different terminology for the same type of document. The level of characterization shown in each row of Table V (identity, quantity or concentration, purity, etc.) is dependent on the intended use of the data.

The following characteristics are described in the context of the recommendations provided in Table V:

- Identity: for exploratory assays, we recommend that the scientist refers to the product datasheet regardless of the source of the calibrator material. However, for those calibrators in assays used for (or potentially so) treatment decision-making, we recommend additional characterization such as sequence confirmation, peptide mapping, mass spectrometry, or some form of comparison to other protein family members will reduce the risk of data misinterpretation.
- Quantity/concentration: for all protein biomarker assay calibrators, we recommend that the scientist refer to the product datasheet provided, reaches out to the vendor for additional information, and/or

confirms the protein concentration of the calibrator material in an assay if one already exists (and perhaps in an orthogonal method, if available). If the material is protein carrier-free, measurement of A_{280} using the proper extinction coefficient, BCA or Bradford, etc. can be used to confirm the information provided on the product datasheet. We recommend that analysts use the same method, and same protein calibrator, to confirm the protein concentration as the vendor. We acknowledge the challenge associated with trying to obtain the same calibrator, and use the same method as the vendor, and each analyst should understand the risks associated with not using these materials.

- Purity, impurities, contaminants, and physicochemical properties: we recommend relying on the product datasheet information, given it is provided from a “qualified” vendor (further description of a qualified vendor is below). Additional work is recommended for both levels of characterizations, as listed in Table V, but more rigor is required for treatment decision-making assays.
- Expression systems: mammalian systems are preferred in general to minimize differences in post-translational modifications, differing levels of glycosylation, differences in folding, and between species. Depending on the biology and availability, other systems could be appropriate.
- Biological activity: characterization by binding to the antibody reagents may be sufficient for an exploratory assay calibrator. Performance in an existing and orthogonal method, if applicable, is also suggested to support the treatment decision-making biomarker datasets.
- Stability of calibrator: for calibrators that are used in exploratory assays, we recommend that short-term stability tests be completed in-house. For treatment decision-making biomarker assay calibrators, long-term stability testing should also be completed (10) to establish the length of time a calibrator can be used for clinical trial sample analysis.

Table V. Tiered Approach for the Characterization of Biomarker Calibrator Material, Depending on the Intended Use of the Protein Biomarker Assay Data, and Impact Associated with Risk of Calibrator Not Representing Endogenous Analyte

Recommendations for biomarker calibrator material characterization		
Assay type/intended purpose	Relative quantitative assays used in exploratory setting	Quantitative assays used to justify dose, to show safety or efficacy, and/or to support drug registration
Identity	Product datasheet from producer (external or internal)	CoA plus additional internal characterization ¹
Quantity/concentration	Carrier-free calibrators—A ₂₈₀ ² , BCA, Bradford, performance in Assay ³ , or CoA from producer ⁴ Carrier-protein-containing calibrator—performance in assay ³ , or CoA from producer ⁴	
Purity, impurities, and contaminants Physicochemical properties ⁵	CoA from producer ⁴ CoA from producer ⁴	Electrophoresis, or SEC, or mass spec Electrophoresis, SEC, mass spec, MALDI-TOF, DLS, and/or DSC
Expression systems	Mammalian or other ⁶	Preferably mammalian expression in appropriate cell types ⁶
Biological activity/tertiary structure Stability of calibrator	Confirm activity ⁷ or binding to Ab reagents Short term (i.e., freeze/thaw, 4°C, room temperature, etc.)	Confirm activity in an existing method ⁷ Short and long term (multi-year -70°C)
Parallelism	Calibrator and endogenous should pass with few samples tested ($n \geq 3$)	Calibrator should represent endogenous analyte using more samples ($n \geq 10$) ⁸ , disease-state samples required
Lot-to-lot variability	Determine comparability of new lot to previous lot. Normalize (if needed) new material to old material or obtain a replacement lot	Test multiple lots from same producer to determine variability. Consider value assignment.
Risks/impact associated with unacceptable biomarker calibrator material Impact	Calibrator performance significantly different than endogenous biomarker, impacting data interpretation and decisions based on data, plus timelines, resources.	Data predicts incorrect dose Patient safety impacted Incorrect efficacy assessment Therapeutic approval affected

¹ Including, but not limited to sequencing confirmation, peptide mapping, mass spec, or comparison to other family members

² Ensure proper extinction coefficient used

³ Given there is one available; performance in orthogonal method optional

⁴ Consider internal characterization (e.g., SDS-PAGE/electrophoresis) for new vendors and for those products with incomplete characterizations

⁵ Molecular weight confirmation, post-translational modification, oligomerization, folding, etc.

⁶ Alternative expression systems may be deemed suitable depending on the biology of the specific biomarker

⁷ If available; performance in orthogonal method if available

⁸ If 10 samples are not available, scientists should consider the risks if fewer samples are tested. This number also applies to pre-clinical samples

- Parallelism: we highly recommend that a parallelism assessment be conducted between the calibrator and endogenous material during method development or with kit feasibility tests to ensure that the calibrator material is acceptable. This key parameter will confirm whether the calibrator material and the endogenous analyte are detected in the same manner by the assay. It also confirms if the assay is qualitative, relative quantitative, or definitive quantitative. Parallelism helps define the similarity between the endogenous analyte and the calibrator (29,30). We recommend that at least three individual samples containing quantifiable endogenous analyte levels are tested by diluting the samples serially and meet parallelism performance for exploratory assays. For treatment decision-making assays, the calibrator material should be tested in at least 10 individuals of the relevant disease-state matrix. Decisions concerning the suitability of the parallelism results will be dependent on the intended use of the biomarker assay data.
- Lot-to-lot variability: lot-to-lot variability helps to define the variation found in the calibrator material and can be done prospectively if multiple lots are available as part of vendor qualification. Similar to the comparability assessment, a calibrator curve should be prepared and a quality control sample concentration should be back-calculated from the calibrator curve. This should be repeated multiple times by multiple operators over several days for treatment decision-making assays. Lot-to-lot variability can be problematic if the assay has not been in use during a long period of time, the previous lot has been exhausted, or the lot has expired (see earlier case study depicted in Table III). In these

cases, parallelism should be reassessed and partial validation considered.

For exploratory assays, the new lot of calibrator material should show similar results as the previous lot as suggested in Table V. In the event the performance of the new lot of material differs from the reference lot, we recommend obtaining another replacement lot (if available), normalizing the new lot concentration to the original lot using a correction factor (29), or value assignment (23,31,32). For treatment decision-making assays, multiple lots (a minimum of three) should be tested as early as possible and a normalization strategy via value assignment should be considered. Identifying confirmation experiments could also be used to compare lots.

CHALLENGES UNIQUE TO COMMERCIALY AVAILABLE PROTEINS

In this section, we discuss the challenges associated with procuring the appropriate calibrator material from a commercial source. Therefore, much of this section describes how to qualify a vendor so that the user has more scientific and technical confidence in a supplier of characterized calibrator material. Given that much of the exploratory assay calibrator information will be based on the product datasheet from the vendor (as recommended in Table V), it is important to understand what can be considered as a qualified vendor. Using protein product from a qualified vendor may reduce the risk associated with commercially acquired proteins. The term “qualified vendor” implies a history of protein product evaluation and use over an extended period of time. We define a qualified vendor as one that has consistently delivered quality protein product and is able to satisfactorily answer technical questions or resolve issues. The nature of biomarker immunoassay development projects may force the consideration of purchasing of a protein calibrator from a new vendor that has not been used previously. This circumstance represents risk that should be mitigated by gathering information about the company and about the protein product. Important information about a company may include the number of years it has been in business, the country in which the company is located, the number of publications the company is referenced in, the manufacturing status of the company (true manufacturer, manufacturer that contracts out to third party, or reseller), and the number of unique proteins in the catalog. Concerns may arise with a company that has many hundreds of protein products in the catalog but has only been in business for a few years. This may indicate a protein reseller and not a protein maker. The information gathered will allow a risk assessment with the protein product being considered. This risk assessment may prompt further characterization of the commercial protein before it is used in the assay.

RECOMMENDATIONS FOR READILY AVAILABLE AND COMMONLY USED PROTEIN BIOMARKERS

The standard practice for evaluating and purchasing commercial proteins to be used as calibrators may include the following:

1. Performing an assessment of the structure and function and potential metabolites/isoforms of the protein of

interest. Table IV contains links to useful protein databases.

2. Identification of multiple sources of the protein with the knowledge that a protein product may be sold by more than one vendor (multiple companies selling the exact same protein product from the same lot of material).
3. Conducting a review of supporting technical documentation (product datasheets) offered by supplier.
4. Requesting additional data from a supplier that may not be available in the product datasheet. Refer to Table V to determine the minimum expected level of characterization.
5. Identifying the method by which a supplier has assigned a concentration value to the protein product including the calibrator and kit used in that method whenever possible.
6. Assessing the availability of the protein for ongoing biomarker support including inquiries about the quantity of vials in the inventory may avoid bridging challenges. The vendor should be asked to hold any of the material in inventory while initial testing is conducted. This may require a strong relationship with the vendor(s). Even with an ample supply, lot-to-lot variability and multiple lots should be tested early on (and stability determined, if needed). Proactive negotiations with vendor regarding lot availability, inventory storage, technical support, and timely shipping will improve the likelihood of timely support of studies.

The recommendations above, along with the recommendations in Table V, suggest a process of due diligence when sourcing calibrator material.

RECOMMENDATIONS FOR ADDRESSING CHALLENGES UNIQUE TO NOVEL PROTEIN BIOMARKERS THAT ARE DIFFICULT TO BE EXPRESSED

Finally, we provide recommendations to those proteins that are difficult to express or unavailable commercially. For those calibrators that are not available for purchase, we highly recommend that the full-length protein be expressed in-house or through a contract research or manufacturing organization. If it is not possible to express recombinant, full-length analyte, we provide here a few guidelines for acceptable surrogates, and refer the reader to Table V for acceptable levels of characterization, depending on the intended purpose of the biomarker assay data.

When a full-length protein calibrator cannot be produced, a surrogate calibrator produced as a fusion protein or as a biomarker-derived peptide that contains the necessary epitope(s) for binding may be suitable. For example, a critical epitope may be produced as a single peptide sequence. To increase specificity of the surrogate calibrator, two or more epitopes may be produced and cross-linked using chemical linkers. Examples include, but are not limited to the following:

- A single peptide sequence with a linker,
- Two epitopes linked using a chemical linker with additional amino acids between linkers,
- Two epitopes using a chemical linker with polyethylene glycol (PEG) between linkers,
- Conjugation to HSA or other carrier protein.

However this recombinant surrogate calibrator is made, it must be shown to mimic the native endogenous form in samples as modeled by binding to the assay reagents. One method of doing this is to use the linear peptide sequence (alone or as part of the overall surrogate construct) to perform a competition experiment with the endogenous biomarker. By increasing the surrogate calibrator concentration and showing competition with the endogenous analyte from relevant samples, the suitability of the surrogate calibrator is supported in the assay conditions. Given the stark differences between these surrogate constructs and the endogenous biomarkers, demonstrating parallelism between them and endogenous biomarkers is crucial for confirming their use as calibrators in protein biomarker assays. The researcher must also account for molecular weight differences

between the endogenous material and the surrogate when assigning a concentration value to the surrogate.

MOVING TOWARDS A “GOLD STANDARD” AND COMMUTABILITY

As shown in our case studies, different sources of calibrator material of the “same” analyte can give different results when tested in one or more biomarker assay(s). Most of the time, there is no universally recognized “true” or “gold standard” to serve as a reference for biomarker assay calibrators to help define this or compare new sources or lots to. In the absence of reference material, there are many approaches showing that a calibrator material is comparable to a previous lot or source, assuming that all other variables of the assay are the same. Value assignment

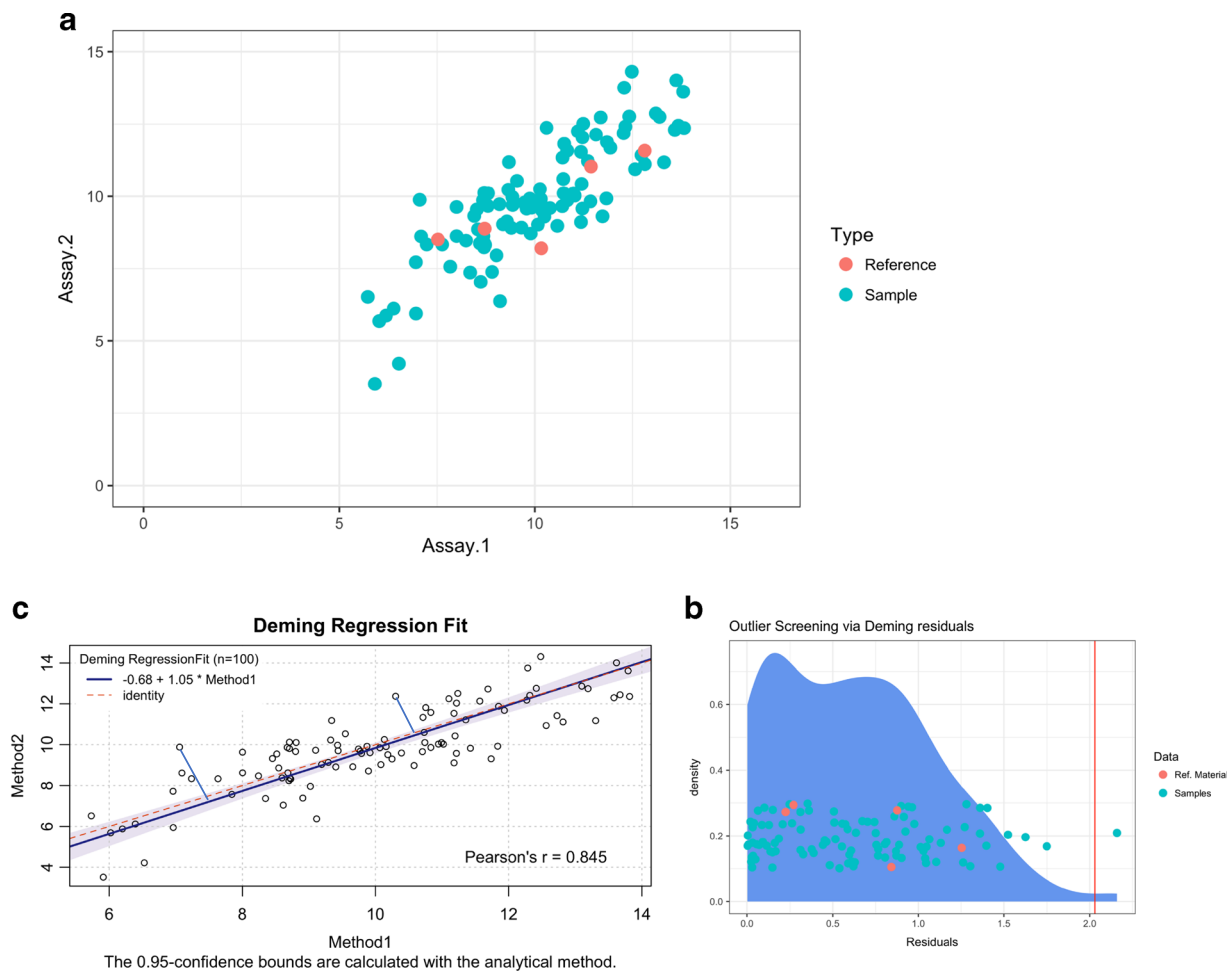


Fig. 3. Deming residuals. Simulated data is shown here representing 100 samples in blue with five new calibrator material samples assessed at select concentration levels in red (a). The axes represent the readouts from two different assays. The two different assays can represent any differentiation in the assay (such as different lots). The blue dots represent a sample space around which we expect typical samples should fall when measured across the two assays. Reference samples with the commutability criteria should similarly be found within this sample space implying similar characteristics as the samples. A Deming regression was fit and the Deming residuals, the perpendicular distances from the plotted data to the regression line shown in blue (b), are computed. These are used to describe the proximity and observed variability of any given data point to the modeled functional relationship between the two assays across all samples. Deming residuals far larger than those computed from the samples imply atypical behavior that does not mimic the behavior of samples across the two assays. A density plot of the computed Deming residuals is shown in c. An upper bound 99% quantile is computed based on the sample data as a cutoff for atypical behavior. Deming residuals of calibrator material larger than this value would not meet the commutability criteria. In this case, the calibrator, at different levels of concentration shown in red, all fit within the range of samples and the calibrator is considered commutable

can occur if there is a well-characterized lot of material used in an assay. Subsequent lots can then be compared to that first well-characterized lot and subsequent lot concentration can be reassigned against the first one to maintain value consistency (33). Commutability describes the mathematical relationship between a reference material and a set of samples across several different laboratories or other sources (34). It is a property of the calibrator and is used to describe the calibrator in relation to samples with respect to a defined set of assays and samples. In the field of clinical chemistry, commutability is used to compare calibrators to samples. Clinical chemists compare different lots of calibrators using a set of defined samples to determine if a new lot or source of calibrator is similar to the previous lot. Commutability is used for reference standard evaluation (18,19) as described earlier in the paper for troponin I.

The application of commutability in drug development may be most useful during the later stages of clinical use of a biomarker assay, for managing long-term application over multi-year clinical trials, or to support a biomarker assay on different platforms. Commutability also helps ensure that the reference calibrator behaves similarly to the actual samples being

measured. For commutability of biomarker assay calibrators at these later stages, we recommend using a mathematical relationship between different sets of assays across a predefined set of samples based on some classic statistical methods (35,36). Unlike the bridging approach which strives to compare the measurement bias and variance across two assays, this method models the overall distribution of samples across the two assays to determine if the calibrator falls within this distribution.

In Fig. 3, we show our recommendation for a commutability methodology using a Deming residuals statistical approach for long-term support of clinical assays. Deming residuals measure the perpendicular distance between the observed data and the modeled regression line. The distribution of these residuals provides a general idea of where samples measured across the two assays would typically fall. Deming residuals are ideally suited for method comparisons where the assignment of dependent (x) and independent variable (y) are arbitrary, i.e., interchanging (x) and (y) will arrive at the same conclusions. Sample data from two different assays were plotted with the new calibrator material prepared at five different concentrations. The

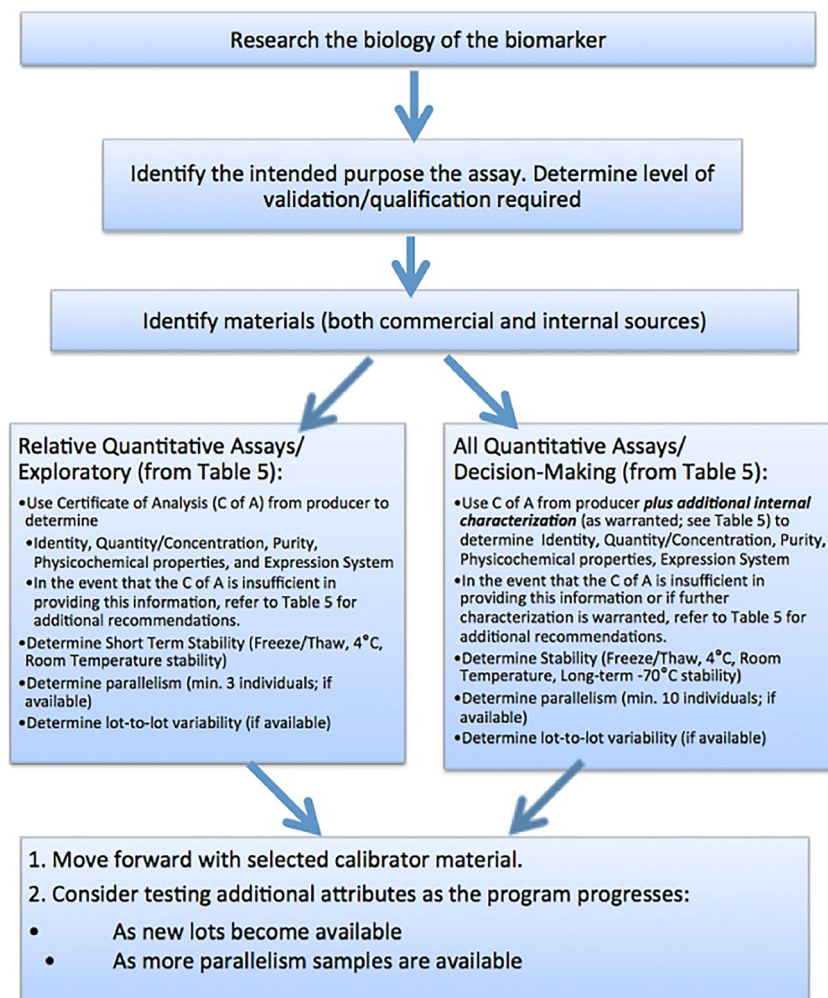


Fig. 4. Recommended workflow for selecting biomarker calibration material. This flowchart shows the set of decisions leading to these recommendations and can be used as a quick reference to our suggested best practices. These recommendations serve as a framework for the choice, evaluation, and appropriate use of calibrator material for protein biomarker assays that will result in more reliable and consistent assays

Deming regression was fitted and the Deming residuals, the distances from the plotted data to the regression line, were computed. Using these residuals, a 99% quantile was determined based on the sample data. The location of the calibrator material's Deming residuals was examined and, in this case, the calibrator at different levels of concentration fits within the range of samples. The calibrator is therefore deemed commutable to the compared samples across the two assays evaluated. In the event the test for commutability fails, the reference may not be similar to the intended samples across the space of the two assays. Care should be taken that use of this reference material may not represent behavior of the samples across the two assays.

CONCLUSIONS

It is important that bioanalytical scientists understand the biology of the biomarker of interest to ensure that the biomarker assay calibrator material is a suitable surrogate for the endogenous analyte. We have described the impact of having inadequately characterized biomarker assay calibrator material and provided calibrator characterization recommendations based on the intended use of the biomarker data. Table V provides these recommendations that apply to biomarker assay calibrators regardless of the source. All the attributes in Table V should be considered equally important. In support of these recommendations, it is important to understand the method used to determine biomarker assay calibrator material concentration and to know the calibrator used in the protein assay. This includes a greater industry-wide emphasis on assessing parallelism to ensure the calibrator material is detected in the same manner as the endogenous analyte in the assay used. A flowchart showing the set of decisions leading to these recommendations as a quick reference to our suggested best practices is shown in Fig. 4. These recommendations serve as a framework for the choice, evaluation, and appropriate use of for protein biomarker assay calibrators that will result in more reliable and consistent assays and biomarker data.

COMPLIANCE WITH ETHICAL STANDARDS

Competing Interests The authors declare that they have no conflict of interest.

Ethical Conduct of Research The authors state that they have obtained appropriate institutional review board approval or have followed the principles of the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

- Bioanalytical method validation, Draft guidance for industry In: CDER and CVM, editor. Rockville, MD.: Food and Drug Administration. 2013.
- Bioanalytical method validation, Guidance for industry. In: CDER C, editor. Silver Spring, MD. : Food and Drug Administration. 2001.
- European Medicines Agency Guideline on biomedical method validation. 2011.
- Jani D, Allinson J, Berisha F, Cowan KJ, Devanarayan V, Gleason C, *et al.* Recommendations for use and fit-for-purpose validation of biomarker multiplex ligand binding assays in drug development. *AAPS J.* 2016;18(1):1–14.
- Khan MU, Bowsher RR, Cameron M, Devanarayan V, Keller S, King L, *et al.* Recommendations for adaptation and validation of commercial kits for biomarker quantification in drug development. *Bioanalysis.* 2015;7(2):229–42.
- Cummings J, Raynaud F, Jones L, Sugar R, Dive C. Fit-for-purpose biomarker method validation for application in clinical trials of anticancer drugs. *Br J Cancer.* 2010;103(9):1313–7.
- Cummings J, Ward TH, Dive C. Fit-for-purpose biomarker method validation in anticancer drug development. *Drug Discov Today.* 2010;15(19–20):816–25.
- 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.* 2009;877(13):1259–71.
- Chau CH, Rixe O, McLeod H, Figg WD. Validation of analytic methods for biomarkers used in drug development. *Clin Cancer Res.* 2008;14(19):5967–76.
- Fraser S, Fleener C, Ogborne K, Soderstrom C. When close is not close enough: a comparison of endogenous and recombinant biomarker stability samples. *Bioanalysis.* 2015;7(11):1355–60.
- Lee JW, Devanarayan V, Barrett YC, Weiner R, Allinson J, Fountain S, *et al.* Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res.* 2006;23(2):312–28.
- Lee JW, Weiner RS, Sailstad JM, Bowsher RR, Knuth DW, O'Brien PJ, *et al.* Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. *Pharm Res.* 2005;22(4):499–511.
- DeSilva B, Smith W, Weiner R, Kelley M, Smolec J, Lee B, *et al.* Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res.* 2003;20(11):1885–900.
- Arnold ME, Booth B, King L, Ray C. Workshop report: Crystal City VI-bioanalytical method validation for biomarkers. *AAPS J.* 2016;18(6):1366–72.
- Walsh CT, Garneau-Tsodikova S, Gatto GJ Jr. Protein post-translational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed.* 2005;44:7342–72.
- King LE, Farley E, Imazato M, Keefe J, Khan M, Ma M, *et al.* Ligand binding assay critical reagents and their stability: recommendations and best practices from the Global Bioanalysis Consortium Harmonization Team. *AAPS J.* 2014;16(3):504–15.
- Bower JF, McClung JB, Watson C, Osumi T, Pastre K. Recommendations and best practices for reference standards and reagents used in bioanalytical method validation. *AAPS J.* 2014;16(2):352–6.
- Tate JR, Bunk DM, Christenson RH, Katrukha A, Noble JE, Porter RA, *et al.* Standardisation of cardiac troponin I measurement: past and present. *Pathology.* 2010;42(5):402–8.
- Christenson RH, Duh SH, Apple FS, Bodor GS, Bunk DM, Panteghini M, *et al.* Toward standardization of cardiac troponin I measurements part II: assessing commutability of candidate reference materials and harmonization of cardiac troponin I assays. *Clin Chem.* 2006;52(9):1685–92.
- Noble JE, Bunk DM, Christenson RH, Cole KD, He HJ, Katrukha AG, *et al.* Development of a candidate secondary reference procedure (immunoassay based measurement

- procedure of higher metrological order) for cardiac troponin I: I. Antibody characterization and preliminary validation. *Clin Chem Lab Med*. 2010;48(11):1603–10.
21. Zhang X, Schwartz JC, Guo X, Bhatia S, Cao E, Lorenz M, *et al*. Structural and functional analysis of the costimulatory receptor programmed death-1. *Immunity*. 2004;20(3):337–47.
 22. Ni YG, Yuan X, Newitt JA, Peterson JE, Gleason CR, Haulenbeek J, *et al*. Development and fit-for-purpose validation of a soluble human programmed death-1 protein assay. *AAPS J*. 2015;17(4):976–87.
 23. Handling and use of reference standards in the OMCL network 2012.
 24. The protein protocols handbook. Walker JM, editor 2009.
 25. Noble JE, Bailey MJ. Quantitation of protein. *Methods Enzymol*. 2009;463:73–95.
 26. Noble JE, Knight AE, Reason AJ, Di Matola A, Bailey MJ. A comparison of protein quantitation assays for biopharmaceutical applications. *Mol Biotechnol*. 2007;37(2):99–111.
 27. Knight MI, Chambers PJ. Problems associated with determining protein concentration: a comparison of techniques for protein estimations. *Mol Biotechnol*. 2003;23(1):19–28.
 28. Prassas I, Brinc D, Farkona S, Leung F, Dimitromanolakis A, Chrystoja CC, *et al*. False biomarker discovery due to reactivity of a commercial ELISA for CUZD1 with cancer antigen CA125. *Clin Chem*. 2014;60(2):381–8.
 29. Stevenson LF, Purushothama S. Parallelism: considerations for the development, validation and implementation of PK and biomarker ligand-binding assays. *Bioanalysis*. 2014;6(2):185–98.
 30. King LE. Parallelism experiments in biomarker ligand-binding assays to assess immunological similarity. *Bioanalysis*. 2016;8(23):2387–91.
 31. Wang J, Lee J, Burns D, Doherty D, Brunner L, Peterson M, *et al*. “Fit-for-purpose” method validation and application of a biomarker (C-terminal telopeptides of type 1 collagen) in denosumab clinical studies. *AAPS J*. 2009;11(2):385–94.
 32. Schimmel H, Zegers I, Emons H. Standardization of protein biomarker measurements: is it feasible? *Scand J Clin Lab Invest Suppl*. 2010;242:27–33.
 33. Thomas JB, Duewer DL, Mugenya IO, Phinney KW, Sander LC, Sharpless KE, *et al*. Preparation and value assignment of standard reference material 968e fat-soluble vitamins, carotenoids, and cholesterol in human serum. *Anal Bioanal Chem*. 2012;402(2):749–62.
 34. Vesper HW, Miller WG, Myers GL. Reference materials and commutability. *Clin Biochem Rev*. 2007;28(4):139–47.
 35. Deming WE. *Statistical adjustment of data*. 1st ed. New York: John Wiley; 1943.
 36. Bjerke M, Andreasson U, Kuhlmann J, Portelius E, Pannee J, Lewczuk P, *et al*. Assessing the commutability of reference material formats for the harmonization of amyloid- β measurements. *Clin Chem Lab Med*. 2016;54(7):1177–91.