The Journey to Regulation of Protein-Based Multiplex Quantitative Assays


BACKGROUND: Clinical proteomics presents great promise in biology and medicine because of its potential for improving our understanding of diseases at the molecular level and for detecting disease-related biomarkers for diagnosis, prognosis, and prediction of therapeutic responses. To realize its full potential to improve clinical outcome for patients, proteomic studies have to be well designed, from biosample cohorts to data and statistical analyses. One key component in the biomarker development pipeline is the understanding of the regulatory science that evaluates diagnostic assay performance through rigorous analytical and clinical review criteria.

CONTENT: The National Cancer Institute’s Clinical Proteomic Technologies for Cancer (CPTC) initiative has proposed an intermediate preclinical “verification” step to close the gap between protein-based biomarker discovery and clinical qualification. In collaboration with the US Food and Drug Administration (FDA), the CPTC network investigators recently published 2 mock submission review documents, first-of-their-kind educational materials that may help the scientific community interested in developing products for the clinic in understanding the likely analytical evaluation requirements for multiplex protein technology–based diagnostic tests.

CONCLUSIONS: Building on this momentum, the CPTC continues with this report its collaboration with the FDA, as well as its interactions with the AACC and the Centers for Medicare and Medicaid Services, to further the understanding of regulatory requirements for approving multiplex proteomic platform–based tests and analytically validating multiple analytes.

Protein biomarker panels for use in cancer diagnostics and therapeutics are currently of considerable interest in medicine. The discovery of proteins and peptides “leaked” by tumors into clinically accessible body fluids such as the blood and urine have led to the development of biomarkers for diagnosing cancer or monitoring the response to therapy. Today, we do not suffer from a lack of candidate protein biomarkers. More than 1200 protein biomarker candidates for cancer have been described in the scientific literature (1). Unfortunately, the rate of introduction of new protein analytes approved by the US Food and Drug Administration (FDA) has remained flat over the past 15 years, with an average of 1.5 new proteins cleared per year for all diseases (2). This vast discrepancy between discovery using proteomic technologies and the number of FDA-approved protein analytes suggests a deficiency in the effective translation of clinical proteomics within the biomarker pipeline. This state of affairs is possibly due to several factors (3–6): (a) technological variation within and across platforms; (b) an inability to credential biomarker candidates before costly and time-consuming clinical-qualification studies that use well-established methodologies; (c) the research community’s lack of knowledge about the
evaluation criteria required for these distinct processes in the pipeline and about regulatory science; and (d) the failure of biomarkers in clinical qualification. In this report, we provide the research community with a basic resource for defining and explaining the regulatory processes for translating biomarker candidates discovered in research laboratories into multiplex protein-based assays for clinical use. The assays we consider as examples include multiple reaction monitoring mass spectrometry (MRM-MS) on a triple-quadrupole mass spectrometer (TQMS) and its immunoaffinity version [i.e., stable isotope standards and capture by antipeptide antibodies (SISCAPA)] (7–10), and immunologic arrays (7, 11), both of which have been previously described (7).

**Regulation of Multiplex Protein-Based In Vitro Diagnostic Assays**

The Centers for Medicare and Medicaid Services, through CLIA, which was passed by Congress in 1988, establishes quality-testing standards for all laboratories to ensure the accuracy, reliability, and timeliness of patient test results (http://www.cms.gov/clia). CLIA defines a clinical laboratory as any facility that runs laboratory testing on human samples in order to provide information for the diagnosis, prevention, and treatment of disease, as well as to assess health. The FDA is responsible, under CLIA regulations, for categorizing commercially marketed in vitro diagnostic (IVD) assays on the basis of their technical complexity and requirements for operator training. These categories are waived tests, tests of moderate complexity, and tests of high complexity.

Laboratory-developed tests (LDTs), or “home-brew tests,” exist for at least 96 protein analytes for which there is no FDA-approved test. In general, LDTs have historically been the general subject of FDA enforcement discretion, although the FDA has been regulating some of them (e.g., tests used in blood banking). The FDA’s longstanding policy of enforcement discretion may undergo some major changes, as witnessed in the recent Federal Register notice (http://edocket.access.gpo.gov/2010/2010-14654.htm) and the FDA’s public meeting on the oversight of LDTs (http://www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm212830.htm). Proper clinical validation of multiplex proteomics-based tests used in a clinical setting might be beyond the scope of a single laboratory. In addition, these tests will likely require further simplification and analytical robustness in order to be used extensively in clinical applications. For this report, we define a multiplex protein assay as a device or test system in which one or more protein/peptide targets are detected simultaneously via a common process of sample preparation, measurement, and interpretation (Fig. 1A). Furthermore, an IVD multivariate index assay (IVDMIA) is a device that (a) combines the values of multiple variables by means of an interpretation function to yield a single, patient-specific result (e.g., a classification, a score, and so forth) that is intended for the diagnosis of disease or for the cure, mitigation, treatment, or prevention of disease (Fig. 1B); and (b) provides a result the derivation of which is nontransparent and cannot be independently derived or verified by the end user. Draft guidance on IVDMIA devices is available at http://www.accessdata.fda.gov/cdrh_docs/reviews/K081754.pdf. A recent community-wide research effort extensively evaluated current practices on the development and validation of classifiers and composite scores (12).
CLIA requires clinical laboratories to perform analytical validation for all tests prior to implementation. For FDA-cleared tests, this process involves postmarket verification of the manufacturers’ claims for the performance specifications, which helps to ensure that a test, when used in a clinical laboratory by testing personnel for its patient population, is performing as the manufacturer intended. For non–FDA-cleared tests (e.g., textbook protocols, LDTs), laboratories must establish the criteria for each of the performance specifications as a part of their analytical validation procedures. Although CLIA regulations do not specifically address clinical qualification, CLIA holds laboratories responsible for all aspects of their operation, which also include test selection.

As defined previously, multiplex and IVDMIA protein-based tests are intrinsically complex, involving a device or devices in which 2 or more protein/peptide targets are detected and collectively interpreted with the aid of software. It will likely be difficult for clinical laboratories both to validate these assays on their individual analytical platforms and to demonstrate clinical relevance by performing large-scale preclinical/clinical studies, which may need to involve multiple laboratories. Consequently, establishing a standardized evaluation paradigm should help ensure the highest level of performance within and across laboratories so that the tests provide patients with the most accurate results.

Furthermore, there is no mandatory system for reporting the adverse effects of LDT results on patient outcome. The FDA has encouraged healthcare professionals, consumers, laboratories, manufacturers, and others to voluntarily report adverse events and malfunctions associated with LDTs to its MedWatch system (http://www.fda.gov/Safety/MedWatch/default.htm). A separate product code for these tests is currently being implemented to facilitate voluntary reporting of adverse events associated with LDTs. The intention of this initiative is to provide a clear pathway and to encourage increased reporting.

The FDA’s Classification of IVD Assays

Mansfield et al. (13) has provided a thorough review of the FDA’s classification of IVD assays. In brief, the level of a device’s regulations required by the FDA depends on a device’s risk to patients’ health outcomes, which is based on device performance. The type of submission required depends on what class a device falls into and is highly dependent on the claim of the test (i.e., intended use and indication for use), not simply on the nature of the analytes.

The Food, Drug and Cosmetic Act has defined 3 classes of devices. Medical devices must be manufactured under a quality-assurance program, be suitable for the claimed intended use, be adequately packaged and properly labeled, and have establishment registration and device-listing forms on file with the FDA (13). These 3 classes are as follows:

- Class I (low risk): Low-risk devices are generally placed in class I, and most are exempt from FDA premarket review. A few class I devices are additionally exempt from most good manufacturing practice (GMP) requirements; however, they are not exempt from other general controls, such as requirements for postmarket reporting.
- Class II (moderate risk): Moderate-risk devices generally require 510(k)-level review before their introduction to the market. A new 510(k) submission is also required for class II devices if a manufacturer or distributor makes changes to the intended use for a device already in commercial distribution (and the intended use remains moderate risk) or if a device is modified in a way that could affect its safety or effectiveness.
- Class III (high risk): High-risk devices have substantial importance for the prevention of impairment of health or have a potentially unreasonable risk of illness or injury for which general and special controls are considered inadequate (e.g., heart valves, pacemakers). These types of devices require a premarket approval (PMA) before commercial distribution.

By law, a device subject to premarket notification or 510(k) submission must be assessed for substantial equivalence to another legally marketed device, called a “predicate” (13). For this reason, 510(k) submissions for IVD assays usually include preanalytical and analytical data and in most cases include at least data from patient samples, with the extent of any further necessary clinical studies depending on preexisting information on the characteristics of the device/analyte combination. It is usually insufficient for a sponsor to demonstrate device performance in its own laboratory, because the data to establish the performance of the device should reflect its performance in routine clinical use. Hence, the FDA will require that certain studies for evaluating assay performance be conducted in representative clinical laboratories that reflect the expected users of the test system. In certain situations, adequate literature may substitute, or partially substitute, for extensive clinical studies, or foreign studies might suffice. Such decisions, however, are handled on a case-by-case basis.

When no predicate device is found to be substantially equivalent to a new device, a de novo review process may be indicated, depending on the risk of the device. The de novo downclassification process is intended to address low- or moderate-risk devices that have novel intended uses for which there is no legally
marketed predicate device. These devices are automatically assigned a class III designation (PMA), after which the FDA may choose to downclassify this device to class I or II if it believes the risk is sufficiently low given its intended use. The performance of the new device is assessed for safety and effectiveness as if it were the subject of a PMA.

Classification of Multiplex Protein-Based Assays

MRM-MS ASSAYS

Mass spectrometers for clinical use, as described under 21 CFR 862.2860, are considered class I devices (general controls) and are exempt from the 510(k) process unless they are used for the diagnosis, monitoring, and screening of neoplastic diseases, cardiovascular diseases, and diabetes (http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=862.9). Mass spectrometers under 510(k) exemption are still required to follow GMP guidelines (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/). Currently, 10 mass spectrometers are listed in the 510(k) database under this status (see Table I in the Data Supplement that accompanies the online version of this special report at http://www.clinchem.org/content/vol57/issue4). Only the Finnigan MAT TSQ-70 (Finnigan Corp.) is a TQMS that is no longer being manufactured. Class I devices should be registered and listed, regardless of whether they require a premarket review.

FDA regulations differentiate between this mass spectrometry instrumentation, which falls under the FDA’s exempt regulation requiring only general controls, and more complex instrumentation, such as instruments that include interpretive software and complex measurement functions, which may be regulated similarly to the multiplex instrumentation for nucleic acid assays (class II). Moreover, instrumentation used to run a specific assay takes on the classification of that assay when such an instrument is submitted for that particular intended use. Thus, a class III assay that uses mass spectrometry would involve evaluation of that mass spectrometer as a part of a class III assay for that intended use.

A multiplex proteomics-based MRM-MS assay simultaneously measures multiple peptides via a single HPLC injection. Hence, MRM-MS measurements of multiple peptides derived from proteolytic digestion of protein(s) to obtain “protein concentrations” in plasma or serum can be considered multiplex, regardless of whether these peptides have come from one protein or multiple proteins.

Although the basic operating principles are the same, the designs of modern instruments from different manufacturers and with interfaced HPLC systems (e.g., nanoscale vs microscale flow rates, chip-based vs non–chip-based) can be different. Consequently, different instruments can potentially produce nonequivalent assay performances, thus requiring different specifications for each instrument. In light of this fact, recent results from interlaboratory studies by the National Cancer Institute’s Clinical Proteomic Technologies for Cancer (NCI-CPTC) network showed promise in obtaining reproducible MRM-MS data across instrument platforms and laboratories (11 target peptides from 7 proteins and their corresponding heavy isotope–labeled internal standards at 9 different concentrations) (14). In this study, the highest CV was ≤23% with a single MRM transition for all but one of the peptides, even in the most complex scenario in which multistep sample preparation was performed at individual sites. Analytical imprecision is expected to improve with reduced or more-streamlined sample preparation and robotics. Automation in conjunction with software development should eventually reduce labor-intensive work flows, variation between different instrumentation platforms, and the need for the high level of expertise currently required to perform this type of assays. Once different mass spectrometry platforms demonstrate equivalent analytical performance, separate regulatory evaluations of the same analyte(s) for each instrument may become simplified or unnecessary.

For FDA submission, performance data obtained with specific mass spectrometers in conjunction with any other instrumentation used as a part of the system (e.g., preanalytical, liquid chromatography, and so forth) to perform a specific assay should be provided as a part of the initial submission for a diagnostic assay. Adequate performance of an assay generally leads to approval only for the specific instrument used in evaluating performance, because assay performance on other platforms would be unknown. The initial submission can be followed by subsequent submissions that address any modifications to the assay, such as the addition of another instrument platform. For example, the most commonly used TQMS can first be cleared as a part of an MRM-MS assay measuring a protein analyte with a relatively low-risk intended use previously cleared with an immunoassay. That would make an instrument manufactured under the quality system available for clinical laboratory testing, and that can be followed by addition of TQMS instruments from other manufacturers. The evaluation of an instrument platform, however, should include all necessary components and accessories (e.g., nano-HPLC columns, traps, and tubing, electrospray source, predictive software). Any changes to this cleared platform would be reevaluated by the assay manufacturer and may or may not need subsequent regulatory submissions to address the safety and effectiveness of the changed platform.
Additional analytes or kits can subsequently be cleared for use on the same platform without additional platform-specific information, other than assay-specific components of the instrumentation.

**IMMUNOLOGICAL ARRAYS**

Immunoassays have been the most commonly used methods in clinical laboratory testing for detecting proteins, and the FDA has extensive experience in the science and regulatory processes for standard-format immunoassays. Prior clearance of multiplex immunological assays for measuring intact protein concentrations has been established (e.g., flow immunoassay for the qualitative detection of IgG antibodies of Epstein–Barr virus nuclear antigen, viral capsid antigen, and early antigen diffuse in human serum to aid in the diagnosis of infectious mononucleosis; http://www.accessdata.fda.gov/cdrh_docs/reviews/K062211.pdf). Similar to mass spectrometry, the status of exemption for immunological test systems (class I) is assay and system dependent. Although some systems are exempt from 510(k) review [e.g., the β-globulin immunological test system, a device that uses reagents and immunochemical techniques to measure β-globulins, including β-lipoprotein, transferrin, glycoproteins, and complement, in serum and other body fluids that are rarely associated with specific pathologic disorders (section 866.5160)], others require 510(k) submissions. Other examples of class I devices are the immunofluorimeter and immunonephelometer devices for clinical use, which are defined as readers to measure the fluorescence of fluorochrome-labeled antigen–antibody complexes and light-scattering from antigen–antibody complexes, respectively. Both kinds of devices are still subject to GMP regulation and limitations to exemptions. A tumor-associated antigen immunological test system for qualitatively or quantitatively measuring tumor-associated antigens in serum, plasma, urine, or other body fluids as an aid in monitoring patients for disease progress or response to therapy, however, is considered a class II device.

Several ELISA-based solid-phase arrays that have been developed against peptide epitopes of multiple proteins are currently marketed for research-and-development purposes (15, 16), and similar multiplex microarrays have been cleared by the FDA as IVD assays [e.g., ENA IgG BeadChip™ Test System; BioArray Solutions (http://www.accessdata.fda.gov/cdrh_docs/reviews/K043067.pdf)]. Multiplex arrays designed to detect and quantify posttranslational modifications (PTMs), such as the array demonstrated by Wandall et al. (17), have also shown their utility in the analysis of sera from cancer patients for detecting cancer-associated IgG autoantibodies against different aberrant O-glycopeptide epitopes derived from mucin 1. Furthermore, Chen et al. have performed a multiplexed analysis of glycan variation on native proteins captured by antibody microarrays (18). Similarly, the immunological array described in the mock presubmission (7) that would detect hundreds of proteins and PTMs simultaneously, uses a peptide epitope in capturing all antigen forms (detected by spinning-disc interferometry) and subsequently uses laser-induced fluorescence to measure biomarker glycoforms presented on the antigens. The clearance/approval of microarrays targeting proteins and their attached PTMs would be expected to encounter analytical issues similar to those for bead-based multiplex immunoassays.

**Analytical and Clinical Validation to Fulfill FDA Requirements**

Device clearance or approval rests on the ability of the sponsor to provide analytical and clinical data demonstrating that device performance is adequate to meet its claimed intended use. The extent of the required data will largely depend on whether the measured markers have previously been established as clinically useful for the intended use of the assay. In cases of novel markers, simple analytical detection or quantification of an analyte is inadequate. The importance of measuring such novel markers for clinical management of patients must also be demonstrated, either through clinical data or, in some cases, through sufficient credible published information that supports the assay’s clinical use. A premarket submission document includes a description of the device’s intended use/indication for use, a description of the device (covering both the instrument and reagents), and descriptions of analytical and clinical performance studies that evaluated the performance of the device for its intended use.

In the analytical-performance section of a premarket submission, the performance of the device is described in terms of precision, accuracy, and performance around the cutoff point, along with other performance measurements as applicable, such as specificity, sensitivity, linearity, limit of detection, and limit of quantitation. A detailed description of appropriate internal and external controls and calibrators used in the assay should also be included. A multiplex assay requires that all analytes meet analytical-performance criteria, rather than extrapolating the performance of 1 analyte to all of the others. A multiplex proteomics assay should also address cross-reactivity or interference of analytes within and outside the panel.

MRM–MS assays for the quantification of proteins in serum or plasma should include controls for assessing the efficiency and variation of proteolytic digestion of proteins in different samples. Evidence exists in proteomic analysis for variation in trypsin digestion, and that issue would need to be addressed. If sample preparation in-
Involves an immunoaffinity enrichment step, an evaluation of the analytical recovery of proteins, peptides, and/or PTMs before MRM-MS analysis is also necessary (19). Additionally, the “overall measured protein concentration” from target peptides derived from the same protein would need to be reproducible from run to run. If one of the peptide measurements is consistently an outlier, sponsors need to understand why that occurred, what effect it would have on the assay, and whether the outlier and its effect would be recognized by the assay’s QC system. An outlier could stem from PTMs, interferences from other proteins, single-nucleotide polymorphisms, and so forth. In the mock presubmission, the PepCa10 assay measures 10 tryptic peptides from 5 cancer-relevant proteins (2 peptides per protein) in plasma to yield a single qualitative result. Immunoaffinity enrichment of target peptides and MRM-MS are used to measure each of the 10 analytes by a ratio in which the amount (peak area) of the analyte peptide (unlabeled) is divided by the peak area of its respective internal-standard peptide (labeled) derived from trypsin cleavage of a concatamer (7). Changes in the relative amounts of the 2 peptides for a given protein could indicate the presence of diagnostic alterations in the parent protein or could signal the effect of single-nucleotide polymorphisms or PTMs. The immunological array, on the other hand, must demonstrate high assay specificity of analyte(s) for a single or a very small number of disease-specific glycoforms, low cross-reactivity of capture antibodies, high disc-to-disc precision with spinning-disc interferometry, and so on (7).

Another key consideration for multiplex assays or IVDMIAAs is the use of software to arrive at a patient-specific result that aids in clinical decision-making (20, 21). The FDA generally requires that software algorithms included in the assay for data interpretation be prespecified before study data are analyzed. Alteration of the algorithm to better fit the data after the study is performed is generally unacceptable.

The Role of Clinical Chemists

Under CLIA, clinical chemists are required to routinely verify previously FDA-cleared/approved tests in their facilities. Verification is performed as the QC process to evaluate whether a previously cleared test (e.g., instrument, controls, and reagents) complies with regulations, specifications, or conditions. These verification studies typically involve precision, accuracy, linearity, and lower limits of detection and quantification. When laboratory directors set up a verification method for an FDA-approved multiplex protein assay by using a “score,” they should consider how their laboratories would perform verification studies. One approach might involve running an adequate number of samples from positively and negatively testing patients to assess the performance of the “score” in their diagnosis, compared with their medical charts and final clinical diagnosis. Again, a laboratory director should consider clinical validity when selecting a test.

Helpful Reference Documents for Analytical Validation

Clinical Laboratory Standards Institute (CLSI) documents are useful to sponsors and the FDA in the process of preparing and reviewing 510(k) and PMA submissions and are highly regarded by other organizations of clinical professionals. The CLSI aims to develop global consensus standards and guidelines for healthcare testing (industry,
government, and professional). A CLSI document goes through rounds of rigorous review prior to publication (http://www.clsi.org). These documents are developed and approved by a consensus of stakeholders in particular areas, which may include FDA representatives, and go through a public-comment phase. The FDA can recognize CLSI documents as standards either fully or partially, and compliance with the recommendations of CLSI documents may be accepted as evidence of the fulfillment of certain FDA analytical requirements. Even if the FDA does not recognize a document, it can still be referenced during review. EP17-A (Volume 24, Number 34, Protocols for Determination of Limits of Detection and Limits of Quantitation; Approved Guideline), for example, is a commonly referenced document. Although CLSI documents currently do not exist for multiplex proteomics assays, general guidance can be drawn from the nucleic acid–based multiplex world. An example of that is MM17-A (Verification and Validation of Multiplex Nucleic Acid Assays; Approved Guideline).

Lessons Learned by the Proteomics Community

Proteomic discovery of biomarker candidates does not necessarily lead to clinical utility because of gaps in the current biomarker-development pipeline. The implementation of a preclinical “verification” step that uses targeted proteomics in the NCI-CPTC pipeline has potential to reduce the costs and time associated with moving biomarker candidates into the clinic (Fig. 2), and such steps may lead to a more efficient product-development pipeline for IVD assays (Fig. 3). More importantly, communication among proteomic researchers, sponsors, the FDA, the Centers for Medicare and Medicaid Services, clinicians, and clinical chemists is critical in expediting the translation of these assays into the clinic to benefit patients for disease diagnosis, prognosis, and therapeutic monitoring.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: E.S. Boja, NIH; A.N. Hoofnagle, Clinical Chemistry, AACC; N.L. Anderson, Clinical Chemistry, AACC, and Anderson Forschung Group.

Consultant or Advisory Role: A.N. Hoofnagle, Thermo Fisher Scientific.


Honoraria: None declared.

Research Funding: A.N. Hoofnagle, Waters.

Expert Testimony: None declared.

Role of Sponsor: The funding organizations played a direct role in the preparation and final approval of the manuscript.

Disclaimer: Certain commercial equipment, instruments, or materials are identified in this paper to specify adequately the experimental detail and procedure. Such identification does not imply recommendation or endorsement by the National Cancer Institute, National Institutes of Health, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References


