Clinical Chemistry 64:6
885–891 (2018)

Q&A

Current and Emerging Multianalyte Assays with Algorithmic Analyses—Are Laboratories Ready for Clinical Adoption?

Moderators: Jessica M. Colón-Franco[1*] and Patrick M.M. Bossuyt[2]
Experts: Alicia Algeciras-Schimnich,[3] Christopher Bird,[4] Julia Engstrom-Melnyk,[4] Martin Fleisher,[5] Michael Kattan,[6] and Geralyn Lambert-Messerlian[7]

Over the past decade, there have been a rising number of clinically used tests that combine 2 or more biochemical or molecular assays, demographics, and clinical information into an algorithm to generate diagnostic, prognostic, or predictive information for a specific disease. The concept of multianalyte analyses is relatively new in the field of laboratory medicine. Dating back to the 1980s, prenatal screening for fetal abnormalities, such as Down syndrome, by use of maternal biomarkers is among the pioneer tests that use algorithmic analyses for risk assessment. Yet, the number of multianalyte algorithms used clinically remains modest. The American Medical Association provides current procedural terminology (CPT)[8] codes for 20 multianalyte assays with algorithmic analyses (MAAAs.) Among these, 9 consist of biochemical markers detected by immunoassay or mass spectrometry, with or without other clinical information; 11 use molecular genetics markers; and 1 is for generic use. In this Q&A, we refer to multivariable tests with risk scores as MAAAs, although not all of them have an associated MAAA CPT code.

Generally speaking, MAAAs aim at improving diagnostics for diseases in which single biomarkers have limited clinical validity. The Prostate Health Index (Beckman Coulter) and the 4Kscore® (GenPath) exemplify some of these strategies for prostate cancer detection. Likewise, multianalyte analyses such as the Risk of Ovarian Malignancy Algorithm (ROMA®) and OVA1® and its second-generation OVERA™ (Vermillion Inc.) improve upon the suboptimal performance of the tumor marker CA125 in the differential diagnosis and likelihood of ovarian carcinoma in women presenting with a pelvic mass. While both have Food and Drug Administration (FDA) clearance, ROMA is a nonproprietary algorithm cleared on various commercial immunoassays (Fujirebio Diagnostics, Inc.; Abbott Laboratories; Roche Diagnostics). Early identification of acute kidney injury is another area in which an FDA-approved multianalyte test, Nephrocheck® (Astute Medical), is diagnostically superior to single biomarkers. Moreover, multivariable risk score approaches are under clinical investigation in conditions like preeclampsia and sepsis in which single biomarker approaches have failed to meet diagnostic needs.

A growing interest in the clinical adoption of these risk scores is foreseen. While a few MAAAs are offered by specialized laboratories either as laboratory-developed or FDA-approved tests, the FDA has cleared a handful of these multianalyte tests and their respective algorithms for use in commercial analyzers. The latter scenario facilitates wide implementation of MAAAs across clinical laboratories. As this emerging field gains momentum, questions related to the path toward wide clinical adoption remain to be answered. These include the risk of deviating from the intended use, technical and implementation challenges, informatics needs, and the timeline for endorsement by professional guidelines.

Here, we discuss the growing field of MAAAs. The clinical advances and challenges of widely adopting emerging MAAAs and strategies for their successful implementation are debated among experts in the research, in vitro diagnostic, and clinical laboratory sectors.

[1] Department of Pathology, Medical College of Wisconsin, Milwaukee, Wisconsin; [2] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, University of Amsterdam, Academic Medical Center, Amsterdam, the Netherlands; [3] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota; [4] Medical and Scientific Affairs, Roche Diagnostics Corporation, Indianapolis, Indiana; [5] Department of Laboratory Medicine, Memorial Sloan-Kettering Cancer Center, New York City, New York; [6] Department of Quantitative Health Sciences, Cleveland Clinic Foundation, Cleveland, Ohio; [7] Department of Pathology and Laboratory Medicine, Women and Infants' Hospital of Rhode Island, The Warren Alpert Medical School of Brown, Providence, Rhode Island.

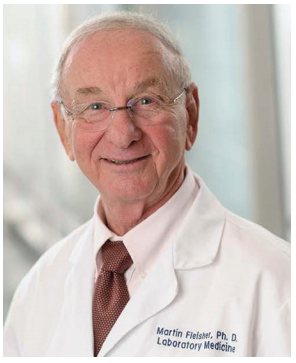
* Address correspondence to this author at: Medical College of Wisconsin, 9200 W. Wisconsin Ave., Department of Pathology, WDL Building, L58, Milwaukee, WI 53226. Fax 414-805-8458; e-mail jcolon@mcw.edu.

Received October 8, 2017; accepted December 12, 2017.

[8] Nonstandard abbreviations: CPT, current procedural terminology; MAAAs, multianalyte assays with algorithmic analyses; FDA, Food and Drug Administration; PSA, prostate-specific antigen; NT, nuchal translucency.

What is the clinical value of implementing risk scores generated from MAAAs as standard of care? Is there evidence of the benefits of using MAAAs over single markers to manage patients?

Alicia Algeciras-Schimnich: Over the years, various MAAA tests have been developed to refine or improve the use of single markers that lack the clinical sensitivity or specificity for disease detection. For example, prostate-specific antigen (PSA) has very poor clinical specificity in the context of prostate cancer. This low specificity of PSA has led to over-diagnosis of indolent disease and over-treatment, resulting in a high cost to manage patients with an increased PSA. In the PSA range of 4–10 ng/mL, approximately 75% of prostate biopsies are negative for prostate cancer. In this context, the development of MAAAs has been focused on refining the probability of detecting prostate cancer on biopsy in men with an increased PSA to spare patients with benign conditions or indolent prostate cancer. Both the Prostate Health Index and 4Kscore tests have been shown to reduce the number of unnecessary prostate biopsies. In the case of the Prostate Health Index, the number of unnecessary prostate biopsies could be reduced by as much as 30%. The downstream benefits of this model include a more individualized risk assessment for prostate cancer, sparing some men from an unnecessary biopsy, and the cost-avoidance of biopsy and potential downstream biopsy-related procedures.



Martin Fleisher: From my perspective, as a nonclinician, I would need to know specific information about the metrics of the MAAA before considering implementing a risk score that could change or markedly impact the care of the patient. I would ask what is known about the analytical validity of the test(s) on which the MAAA is based. How would the use of the MAAA affect patient care and, most importantly, clinical outcome? Before an MAAA was accepted as "standard of care," extensive clinical studies would be needed, and the data must support the efficacy of the algorithmic calculations. After the data is disseminated, there must be broad clinical acceptance of the MAAA and incorporation into clinical guidelines. For example, the Oncotype DX MAAA provides a genomic-based, comprehensive, individualized risk assessment for early-stage invasive breast cancer in adjuvant and neoadjuvant settings. The test provides information on the benefit from chemotherapy. This MAAA was introduced in 2004, but before it became "standard of care" the algorithm was extensively tested and evaluated. There were multiple publications in peer-reviewed journals and finally the MAAA was accepted in clinical guidelines from organizations such as the American Society of Clinical Oncology and the National Comprehensive Cancer Network. A long but thorough process, as it should be.

Michael Kattan: I think we should have MAAAs as standard of care in many areas of medicine. From a statistical point of view, taking more information into account (e.g., using an MAAA) will practically always out-predict the use of a single marker. I don't think there would be much debate over this issue, but realizing the benefits of using MAAAs requires more effort. MAAAs are generally more challenging to implement, and as such need to be shown to be worth the extra effort. If the use of MAAAs were free, everyone facing a difficult medical decision should prefer their use over single markers when trying to obtain predictions of benefits and harms associated with each choice.

Geralyn Lambert-Messerlian: Prenatal screening for Down syndrome has clearly proven the benefit of implementing risk scores based on multiple marker testing as standard of care. Initially, prenatal screening was performed by offering diagnostic testing to pregnant women 35 years of age or older. When first implemented, about 5% of pregnant women were over 35 years of age at delivery and therefore categorized as high risk. This program of maternal age-based screening led to prenatal detection of about 30% of pregnancies affected by fetal Down syndrome.

In the late 1980s, second trimester maternal serum α-fetoprotein concentrations were found to differ in

Down syndrome and unaffected pregnancies. This was followed soon thereafter by the discovery of human chorionic gonadotropin, unconjugated estriol, and inhibin A as Down syndrome serum markers. The addition of maternal serum markers to the a priori risk derived from maternal age improved Down syndrome screening performance. A triple marker test combining maternal age with α-fetoprotein, human chorionic gonadotropin, and unconjugated estriol can detect 65% of fetal Down syndrome cases with a 5% screen-positive rate (5% of women are offered invasive diagnostic testing). Adding inhibin A for a Quad marker test can provide 80% detection of Down syndrome cases with the same 5% screen-positive rate.

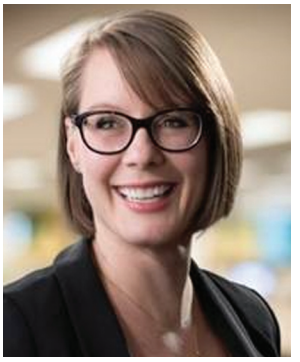
Prenatal screening evolved further in the 1990s when first trimester markers for Down syndrome, pregnancy-associated plasma protein A, and nuchal translucency (NT) ultrasound were validated. Today, all 6 markers can be combined with maternal age to provide up to 95% detection of Down syndrome by offering invasive testing to 5% of the screened population. Alternatively, one could change the risk cutoff level and identify 90% of Down syndrome cases by offering invasive testing to only 2% of the population.

Regardless of the marker combination, data are analyzed by software to generate a pregnancy-specific risk for each woman. If the risk is higher than a specified cutoff, she is considered at high risk and recommended for invasive diagnostic testing. The risks assigned to screened pregnancies have been proven to accurately reflect pregnancy outcomes. Thus, prenatal screening demonstrates that multiple marker testing can effectively improve patient care by greatly increasing the number of cases detected while maintaining or reducing the proportion offered invasive testing.

Are there specific challenges in developing MAAAs, compared to single markers?

Alicia Algeciras-Schimnich: Development and validation of MAAAs is more costly than single markers. A laboratory or manufacturer developing MAAAs will need to evaluate the analytical performance of each maker individually, but there is also the need to evaluate the performance of the combination of markers for the generation of the composite score. How the variation in preanalytical and analytical parameters might affect the generation of the composite score associated with the test will need to be considered and evaluated. In addition, the algorithms used in some of these tests are very complex and require personnel with strong expertise in information technology or bioinformatics. In cases of a proprietary algorithm, there is the issue of intellectual property and/or licensing that needs to be considered.

Chris Bird: MAAAs are difficult due to the requirement for increment predictive value for each additional biomarker. If a single biomarker is to be replaced by multiple biomarkers for screening purposes, the relative risk should be better managed for the additional biomarkers. The commercialization of MAAAs can be beneficial as well, given the ability to apply for unique CPT codes for an MAAA and also achieve protection from competition due to unique algorithms. Some manufacturers have achieved coverage for MAAAs despite not obtaining FDA approval first, or ever in some cases.



Julia Engstrom-Melnyk: To answer the question more specifically, if applied to a specific solution portfolio, the biggest challenge is the choice and the decision placed on the right combination of markers. In the case of cervical cancer screening, the concept behind this multianalyte algorithm is to first identify the presence of the infection followed by the assessment of the effect of this infection at the cellular level. While some manufacturers may choose to look at the activity of the virus, having the tools to monitor the development of oncogenic effects at the cellular level allows for a more tailored and well-timed intervention. The challenge, however, comes with knowing which biological marker to choose—it must be one that is consistently linked to oncogenesis, one that is ubiquitously observed (regardless of the human papillomavirus type leading to the transformation), and one that is predictable and whose changing expression is well characterized. Currently, there are multiple options for biomarker selection (e.g., methylation, translocations), but none have been rigorously tested on all 3 parameters and readily adopted into clinical practice like the tumor suppressor p16INK4A. This level of evidence is increasingly difficult to replicate because the rate of cervical cancers continues to decrease and women are managed before the ultimate diagnosis. Ultimately, the need for clinical evidence leading to rare outcomes is what presents the biggest challenge.

Martin Fleisher: There are 2 specific challenges an MAAA must address before it is accepted. The first challenge comes

from the analytical laboratory that must be convinced that the addition of an MAAA to the laboratory test menu is analytically accurate and valid. It is the laboratory, more specifically the laboratory director, that assumes full responsibility for the MAAA metric. Adding an MAAA to the laboratory's test order set may include the addition of 1 or more assays that were not performed before the use of an MAAA. This could impact the laboratory's workforce, space resources, and budget. It may also require culling information from the patient's chart to calculate the MAAA for which the laboratory information system may be adverse. The second challenge relates to regulatory and administrative issues. Assuming the MAAA is not FDA approved, and therefore a laboratory-developed test, its analytical performance characteristics must be established before results of the MAAA are placed in the patient chart and used for patient management, as required by the Centers for Medicare & Medicaid Services regulations through the Clinical Laboratory Improvement Amendments. This could be a difficult and costly process. And, finally, the laboratory director must be able to interpret the results of the MAAA and communicate the information to clinicians who may be somewhat leery of how to use the MAAA in clinical management.

Michael Kattan: Model building remains an art. As such, proper emphasis needs to be placed on techniques for measuring the accuracy of the MAAA. This literature is still very conflicting.

Geralyn Lambert-Messerlian: One challenge unique to developing an MAAA is that the relationship between markers must be carefully examined. Markers with a high correlation are unlikely to provide additive information about the condition of interest. Pair-wise correlation of markers in Down syndrome and unaffected pregnancies are included in the prenatal screening algorithm to account for the fact that markers are not completely independent predictors. An exception to this is that neither maternal age nor NT measurement is correlated with any of the biochemical markers.

Another issue that may arise in multiple marker testing is that data may be collected by different providers but must be combined using one algorithm. We have faced this in prenatal screening with the measurement of serum markers being performed in the laboratory while the ultrasound measurement of NT is performed at an obstetrical or radiology office. This situation introduces a unique challenge of quality control. How do we ensure that all necessary data is received in a timely manner for reporting? How do we normalize NT data collected at different sites and by different sonographers? How do we know that the NT measurement is being performed according to standardized techniques by qualified personnel?

To address this, the introduction of NT as a Down syndrome screening marker was accompanied by the development of training programs and credentialing agencies for sonographers. Furthermore, laboratories are required to monitor median NT data in the laboratory for each sonographer, in the same way as serum data are monitored. These efforts aid in assuring that high-quality risk assessment is provided for each patient. However, some gaps in managing this issue do persist. For example, it is unclear how to proceed when NT measurements from a particular sonographer do not conform to published expectations. Often, collegial discussion can lead to a solution, but not always. A laboratory may have to refuse data from a sonographer whose long-term performance does not meet expectations.

What criteria should institutions consider before adding an MAAA to their diagnostic test menu?

Alicia Algeciras-Schimnich: The number of MAAAs accessible for implementation in the clinical laboratory today is limited. Many of them are performed at a single laboratory that owns the algorithm. As a consequence, the laboratory considering adding the test to their diagnostic test menu might not have access to information regarding the analytical performance of the test or the reproducibility of the test overtime. It is important for the institutions to evaluate the peer-reviewed literature for the clinical validation of these assays and for whether the reported clinical performance will be applicable to their patient population. When reviewing the peer-reviewed literature, consider whether it was a single institution or multicenter study as well as the affiliation of the author with the company or laboratory that owns the algorithm. This will provide an idea of potential bias on the study findings. For example, in my institution the use of the Afirma® gene expression classifier, which is used to classify thyroid nodules with indeterminate cytology into benign or suspicious for malignancy, resulted in a lower than expected positive predictive value and surgery avoidance than published in the literature. Therefore, in our institution we have decided not to add this test to the menu.

Michael Kattan: Attention should be placed on the medical decision at hand. Any MAAA being considered for inclusion should include a discussion of how the prediction coming from the MAAA will be used. That discussion should actually occur before development of the MAAA to ensure proper resource allocation. Just having a prediction for prediction's sake is hard to justify, but still potentially valuable for patient counseling.

Geralyn Lambert-Messerlian: As enticing as it may be to offer an MAAA to improve patient care, it is essential to be sure that the test has had large-scale validation, preferably by multiple investigators at diverse geograph-

ical locations. This includes rigorous testing of the markers to be used, as well as careful evaluation of the software to be used for risk assessment and reporting. Multicenter trials are useful for this purpose. After the publication of a large clinical trial, a more limited study may be acceptable before the test is implemented locally.

Following demonstration of test validity, the laboratory must be prepared to conduct in-house validation that meets all standard College of American Pathologists and Clinical Laboratory Improvement Amendments regulations. Markers may or may not be approved by the FDA. Those without FDA approval require extensive testing as a laboratory-developed test. Results of the software must also be checked against another computer system and/or with regard to patient clinical outcomes before use. Development of an external proficiency-testing program is also helpful.

What are the main challenges and limitations of implementing MAAAs in clinical laboratories? How can these challenges be mitigated for successful clinical adoption?

Alicia Algeciras-Schimnich: Given that many MAAAs are performed at a single laboratory, which owns the algorithm, clinical laboratories will have to depend on referring the testing to the performing laboratory. For MAAAs with nonproprietary algorithms, laboratories wanting to implement this testing need to consider instrumentation accessibility. Some of the tests, such as ROMA and the Prostate Health Index, require a specific combination of manufacturer's assays for the generation of the composite score. Laboratories could also have an active role in successful clinical adoption of the MAAA tests by ensuring that they are ordered in the correct clinical context. For example, in our laboratory, the Prostate Health Index score is only calculated when the total PSA is between 4 and 10 ng/mL. When a Prostate Health Index is ordered, the total PSA is performed first; only if the result is in the correct range are the other 2 components of the test performed and a composite score calculated.

Martin Fleisher: The challenges are stated above: impact on workforce, work flow, and budget. The impact on the laboratory budget could be a deal-breaker for some laboratories since the Centers for Medicare & Medicaid Services does not reimburse for most MAAAs even if FDA approved. The Centers for Medicare & Medicaid Services states that "if a procedure or device lacks scientific evidence regarding safety and efficacy because it is investigational or experimental, the service is noncovered as not reasonable and necessary to treat illness or injury." However, if the results of the MAAA prove to be clinically efficacious and patient outcome is improved, even budgetary concerns may be mitigated.

Geralyn Lambert-Messerlian: One substantial challenge in implementing MAAAs is having user-friendly and validated software for interpretation of results. Most of the available programs are not FDA-approved, and laboratory demands to prove validity can be considerable. As new markers and data emerge, software updates may be needed, and the program then must be revalidated. Furthermore, despite the time and effort required to be sure the software works correctly, the interpretive part of the multiple marker test is not a billable item.

The best solution for this problem is to have highly qualified on-site information technology services. Collaboration among different laboratories can also be helpful. Risk results can be compared between laboratories as one method of software validation.

What can be learned from the experiences of institutions that have already implemented MAAAs?

Alicia Algeciras-Schimnich: For MAAA tests implemented in our laboratory, we faced a few operational challenges worth sharing. Our first experience was with the implementation of maternal serum screening for fetal aneuploidy. This testing depends heavily on demographic information that needs to be incorporated into the risk calculation; however, from the referral laboratory side, compliance with submission of this information along with the serum sample was low. Often, we were unable to result the test in a timely manner. To get around this, we decided to continue requesting missing demographic information, but if information is not received within 5 days of receiving the sample, the test is reported as uninterpretable, which rapidly gets the attention of the provider. For the ROMA test, which requires the menopausal status of the patient, we calculate and report both composite scores (for pre- and postmenopausal) to get around the potential lack of compliance on providing this information. Another aspect that we considered when implementing ROMA and the Prostate Health Index was where to house the calculation of the composite score. In both instances, having the calculation performed by the laboratory information system rather than at the instrument level was a better fit for our operations given that testing of the single analytes might occur on different instruments. By choosing this path, the up-front time required to incorporate and validate the equation in the laboratory information system was substantial but worthwhile.

Martin Fleisher: The most important experience an institution can pass on to other clinical facilities, most likely through publication of the data in a peer-reviewed journal, is the clinical impact that MAAAs have had on patient care. If an institution has truly experienced a major improvement in patient care that is attributable to the

implementation of MAAAs, they would have documented the improvement and probably updated their clinical standard operating procedures to reflect this improvement. Publication of the data is the best format to disseminate the information and to respond to questions related to the clinical efficacy of the MAAA. This would be the most reliable evidence of MAAAs affecting patient care in a positive manner.

Michael Kattan: Taking a broad definition of MAAA (e.g., any multivariable model), we have multiple MAAAs in place that are automated within our electronic health record. The most difficult aspect of implementation seems to center around work flow and how exactly the MAAA will be used. These items require thoughtful discussion.

Geralyn Lambert-Messerlian: One important issue to consider is physician and patient education regarding multiple marker testing. It is not yet commonplace for physicians to order or to read and interpret multiple marker reports. Many multiple marker algorithms require detailed demographic information such as age, reproductive status, or body weight, among other variables. It is essential that the correct information has been incorporated into the algorithm, and the report should be checked for consistency.

Providers must understand the test's benefits and limitations and be able to explain the process to patients. We have found that developing written informational materials is an essential component of a successful program. The written material should be presented at an elementary reading level and be simple and clear. Diagrams may be helpful. Video or online education is also an attractive option for patients to learn more about testing.

What is the future of MAAAs? What will be their role 5–10 years from now? Do you anticipate this field will grow and at what pace?

Alicia Algeciras-Schimnich: The number of MAAAs in the market will likely continue to grow in the next 5–10 years. A few years ago, one of the hurdles with the implementation of MAAAs had to do with the CPT code assignment and reimbursement by the Centers for Medicare and Medicaid Services and not receiving payment for the algorithmic portion of MAAAs. Today, there are at least 23 MAAAs with CPT codes assigned. The scientific advances to identify hundreds of gene markers associated with complex illnesses and with a patients' ability to respond to treatments, alongside with improved technologies in molecular pathology and the introduction of mass spectrometry–based proteomics in clinical laboratories, will continue to facilitate the development and implementation of MAAAs.

Martin Fleisher: The projected use of MAAAs will be laboratory specialty-dependent. For example, genomic and proteomic MAAAs may well be transformed (or eliminated) by "Big Data" and the use of algorithmic medicine in the diagnostic application of disease management. If Moore's Law holds, algorithmic analyses by supercomputers using genomic data, clinical laboratory data, positron emission tomography, magnetic resonance imaging, and holographic imaging will transform diagnostic medicine in the very near future. With the rapid advancements in the generation of large-scale microarray gene expression data sets (note recent FDA approval of gene expression analysis), robust multigene expression signatures that are capable of guiding the use of specific therapies will be routinely implemented into clinical care. I foresee such an evolution with the introduction of new algorithms that will speed diagnosis. I consider most current MAAAs as stopgaps, or analytical bridges, that address the lack of clinical sensitivity in testing for specific analytes. In my opinion, many currently accepted MAAAs will give way to improved analytical methodologies that will change as soon as newer assay technologies become standard laboratory procedure and the application of highly specific diagnostic techniques become more widely available.

Michael Kattan: I think that as electronic health records become nimbler, easily implemented MAAAs shall proliferate. For challenging medical decision-making, their use makes a lot of sense, with the assumption that proper resources have been devoted to their construction.

Geralyn Lambert-Messerlian: I strongly believe, on the basis of our experience with prenatal screening, that MAAAs have the potential to improve clinical care and their use should be expanded in the future. There is increasing attention given to this topic, which suggests that expanded use will come into play in the near future.

One good example of this is the recent introduction of multiple marker screening for preeclampsia in Europe. Preeclampsia is a serious complication of pregnancy, with clinically significant maternal and fetal morbidity and mortality. First-trimester serum and ultrasound markers can be combined to identify pregnancies at high risk of preeclampsia. Importantly, new therapies are emerging such that, when coupled with a good screening program, the associated negative outcomes related to preeclampsia may be preventable. For example, recent data show that low-dose aspirin offered in the first trimester of pregnancy may reduce the

occurrence of preeclampsia in high-risk women. In the US, screening for preeclampsia still requires extensive testing and validation since there are no FDA-approved products, few women have the ultrasound testing required, and there is limited availability of the needed biochemical assays and software.

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 author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: C. Bird, Roche Diagnostics; J. Engstrom-Melnyk, Roche Diagnostics Corporation.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: G. Messerlian, institutional funding from Perkin Elmer, Ansh Laboratories.
Expert Testimony: None declared.
Patents: None declared.

Previously published online at DOI:10.1373/clinchem.2017.275677