A major application of tumor biomarkers is in serial monitoring of cancer patients, but there are no published guidelines on how to evaluate biomarkers for this purpose. The European Group on Tumor Markers has convened a multidisciplinary panel of scientists to develop guidance on the design of such monitoring trials. The panel proposes a 4-phase model for biomarker-monitoring trials analogous to that in use for the investigation of new drugs. In phase I, biomarker kinetics and correlation with tumor burden are assessed. Phase II evaluates the ability of the biomarker to identify, exclude, and/or predict a change in disease status. In phase III, the effectiveness of tumor biomarker–guided intervention is assessed by measuring patient outcome in randomized trials. Phase IV consists of an audit of the long-term effects after biomarker monitoring has been included into standard patient care. Systematic well-designed evaluations of biomarkers for monitoring may provide a stronger evidence base that might enable their earlier use in evaluating responses to cancer therapy.

© 2012 American Association for Clinical Chemistry

Although serologic tumor biomarkers are primarily used to monitor the response of cancer patients to therapy and to provide early indication of changes in tumor burden (1–4), no consensus yet exists about the optimal design and conduct of longitudinal trials involving serial monitoring of tumor biomarkers. The website of the Enhancing the Quality and Transparency of Health Research (EQUATOR) initiative (5) provides a comprehensive catalogue of recommendations for promoting transparent and accurate reporting of medical research studies, as does an excellent complementary review (6). Few of the publications cited specifically refer to tumor markers, however, and those that do have generally focused on the diagnostic, prognostic, or predictive use of measurements made at a single time point (7, 8), rather than by serial monitoring (9).

To address this issue, the European Group on Tumor Markers (10) has established a multidisciplinary international panel, the MONITOR Working Group, with the aim of developing recommendations to stimulate improvements in the design of trials that assess the use of serial tumor biomarker measurements. Many previously reported monitoring trials were “fishing expeditions,” rather than well-designed, hypothesis-driven investigations. Consequently, it has often been difficult to conclude whether observed changes in the biomarker investigated influenced clinical decision-making, e.g., by prompting early requests for additional diagnostic tests (including imaging) or by a change in therapy. In the present report, the MONITOR Working Group proposes an approach to the design of longitudinal tumor biomarker trials that is analogous to that used to evaluate new therapeutic drugs (11). Although the present proposal is flexible and does not dictate a specific trial design or analytical strategies, it is likely to be most relevant to reasonably well-characterized serum biomarkers.

1 Department of Clinical Biochemistry, Hillebrand Hospital, University of Copenhagen, Hillebrand, Denmark; 2 Department of Pathology and Laboratory Medicine, St. Vincent’s University Hospital and UCD School of Medicine and Medical Science, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland; 3 Breast Oncology Program, University of Michigan Comprehensive Cancer Center, Ann Arbor, MI; 4 Department of Clinical Biochemistry, Royal Infirmary of Edinburgh, Edinburgh, UK; 5 Hadassah-Hebrew University Medical Center, Jerusalem, Israel; 6 Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, Amsterdam, the Netherlands; 7 Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; 8 Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; 9 Unit of Laboratory Medicine, Centre for the Study of Biological Markers of Malignancy, Venice, Italy; 10 NOKIUS, Norwegian Quality Improvement of Primary Care Laboratories, Division for General Practice, University of Bergen, Bergen, Norway; 11 2nd Department of Medicine, LMU Klinikum Grosshadern, University of Munich, Munich, Germany; 12 Department of Oncology, Breast Cancer Unit, Herlev Hospital, University of Copenhagen, Herlev, Denmark; 13 Siemens Healthcare Diagnostics, Gwynedd, UK; 14 Department of Oncology, Gynecologic Oncology Unit, Uppsala University Hospital, Uppsala, Sweden; 15 Department of Clinical Chemistry, Netherlands Cancer Institute, Amsterdam, the Netherlands.

* Address correspondence to this author at: Department of Clinical Biochemistry, Royal Infirmary of Edinburgh, Edinburgh EH16 4SA, UK. Fax +44-131-242-6882; e-mail C.Sturgeon@ed.ac.uk.

Received December 9, 2011; accepted June 28, 2012. Previously published online at DOI: 10.1373/clinchem.2011.180778

16 Nonstandard abbreviations: EQUATOR, Enhancing the Quality and Transparency of Health Research; STARD, Standards for Reporting of Diagnostic Accuracy; CA125, cancer-associated antigen 125; SD, stable disease according to biomarker criteria; PD, progressive disease according to biomarker criteria.
Tumor Biomarker Monitoring Trials—4 Phases

The parallels with drug-development trials are readily apparent from Table 1, which presents a stepwise procedure for evaluating tumor biomarker–monitoring trials that mimics the well-established phases of drug evaluation. Each phase in the tumor biomarker–monitoring trial addresses questions analogous to those in therapeutic-drug trials:

- **Phase I**: Is the new monitoring method/drug safe?
- **Phase II**: Is the new monitoring method/drug effective?
- **Phase III**: Is the new monitoring method/drug superior to currently available methods/drugs?
- **Phase IV**: What additional information about the new monitoring method/drug is required?

Further refinement of these questions, perhaps by adapting the “population, intervention, comparison, and outcomes” approach favored by a number of authors (12, 13) and routinely used by UK National Institute for Health and Clinical Excellence guideline-development groups (14), is likely to be desirable. Data from phase I and II trials will usually complement available data about the clinical validity of the biomarker and its serial use, the focus of the subsequent phase III and IV trials (15, 16).

### Phase I Monitoring Trials—Correlation with Tumor Burden

Phase I monitoring trials should be undertaken for any biomarkers intended for monitoring and be similar to the therapeutic trials that evaluate the optimal dosing schedule and toxicity of a new drug or the novel use of an established drug. For drug trials, some preliminary evidence of clinical utility is essential (e.g., data on clinical sensitivity and specificity, biomarker distribution in healthy individuals and in those with benign and malignant disease). Such trials need to follow previously established recommendations for reporting, such as those of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative (17). Defining the minimum requirements for the clinical utility of a cancer biomarker is a complex task, however, because such requirements will depend on the malignancy, stage of disease, treatment options, and biomarker kinetics.

Phase I biomarker trials should examine the relationship between a change in tumor burden and a change in serially obtained biomarker concentrations in nonblinded investigations. The proposed phase IA, IB, and IC trials address 3 major questions (outlined below) and can be carried out either concurrently or sequentially, depending on what is most appropriate.
PHASE IA TRIALS: WHAT CONSTITUTES A SIGNIFICANT CHANGE IN THE BIOMARKER CONCENTRATION?

Well-designed phase IA monitoring trials should enable assessment of (a) the rate of biomarker change associated with changes in tumor burden, (b) the within-individual biological variation in the biomarker, and (c) the ability of the biomarker to identify changes in tumor burden during different treatment periods.

A. Estimation of rate of change of the biomarker during tumor shrinkage or growth. The rate of biomarker change is usually expressed as the velocity (i.e., a change in concentration per unit time) after chemotherapy or radiotherapy, the doubling time, or the half-life ($t_{1/2}$) after tumor removal: $t_{1/2} = dt/\log(tm_1/tm_2)$, where $tm_1$ and $tm_2$ are the tumor marker values at times 1 and 2, and $dt$ is the interval between the 2 dates. The doubling time is the interval required to double the serum concentration of the marker (18).

B. Estimation of within-individual biological variation. An estimate of the inherent within-individual biological variation (19), along with a realistic estimate of the analytical imprecision (20), is an essential prerequisite for assessing the clinical importance of observed changes in biomarker concentration. Data on biological variation can be derived from populations of healthy individuals, patients who have undergone curative surgery and are apparently disease free, patients with advanced but stable disease, or patients with benign disease (19); data involving many biomarkers are tabulated separately (21). The design of tumor biomarker studies that take account of inherent variation was recently considered in a report that reviewed 27 studies of prostate-specific antigen (22) and in publications of studies relating to cancer-associated antigens 15–3 and 125 (CA125), carcinoembryonic antigen, and tissue polypeptide antigen (23, 24).

C. Ability of the biomarker to identify changes in tumor burden. Provided an observed change in serial measurements of tumor biomarker concentrations exceeds the estimates of imprecision (biological and analytical), the change is likely to reflect a change in disease activity. The reference change value (RCV)—the difference between 2 results that must be exceeded before the change can be considered to be clinically important—provides a convenient means of assessing changes in disease activity:

$$ RCV = Z \times \sqrt{CV_A^2 + CV_{WS}^2}, $$

where $Z$ is the z score and $CV_A$ and $CV_{WS}$ are the analytical and within-individual CVs, respectively (19). In developing criteria for the interpretation of serial concentrations, data acquired to investigate rates of change and background variability can be modeled under standardized conditions in computer simulation models to inform assessment and selection of evaluation criteria in phase IB and IC trials (25, 26).

PHASE IB TRIALS: HOW WELL DO CHANGES IN THE BIOMARKER REFLECT CHANGES IN TUMOR BURDEN?

Phase IB trials use case control studies to address the relationship between changes in tumor burden and biomarker concentrations. These studies should establish whether changes in serial biomarker concentrations identify tumor shrinkage or growth during treatment and/or control periods in which there is no biomarker-initiated therapeutic intervention. This criterion requires establishing (during well-defined treatment periods) whether decreases in serial biomarker concentrations are more likely to occur in patients with tumor shrinkage and whether increases in serial biomarker concentrations are more likely to occur in patients with an increasing tumor burden, compared (in both cases) with patients with benign conditions, patients with stable disease, and/or healthy individuals.

PHASE IC TRIALS: HOW WELL DO SERIAL CHANGES IN THE BIOMARKER REFLECT CHANGES IN TUMOR BURDEN DURING CONSECUTIVE TREATMENT PERIODS?

Phase IC trials explore the ability of serially obtained concentrations of the biomarker to identify periods of tumor shrinkage and tumor growth in patients with established changes in their tumor burdens. Such trials are carried out during consecutive treatment periods with different therapies and during subsequent control periods (i.e., without biomarker-initiated therapeutic intervention).

Phase II Monitoring Trials—Clinical Performance

The purpose of phase II therapeutic trials is to determine whether a new drug is sufficiently effective against a specific disease to justify more extensive and costly trials. Analogously, phase II monitoring trials should validate the performance of the biomarker in other study cohorts by applying the evaluation criteria for response that the phase I trials had identified as optimal. Phase II monitoring trials thus reassess the serial biomarker measurements in terms of their clinical sensitivity (i.e., their ability to identify a change in disease status), specificity (i.e., their ability to exclude a change in disease status), and positive predictive value (i.e., their ability to predict a change in disease status earlier than conventional procedures).

Phase II biomarker-monitoring trials can be embedded most conveniently in studies of patient cohorts for which a change in tumor burden is likely, e.g., within phase II or III prospective clinical drug trials.
Both standard investigations (physical examination, imaging techniques, and routine biochemistry analyses) and tumor biomarker measurements should be evaluated in a blinded fashion and in accordance with guidelines, such as those of the National Comprehensive Cancer Network, which often incorporate useful decision trees.

A design in which the biomarker is measured in real time (as would be the case in clinical practice) is clearly ideal. Provided that considerable care is taken, an unbiased collection of archived serially obtained samples may provide a valid alternative, as has been advocated for studies of tissue biomarkers measured at a single time point. The major prerequisite for such a trial design, which could be termed “indirect prospective-retrospective,” is that an analysis plan and protocol fully documenting the entire study and its analysis plan be specified before the study begins. Attention to detail is essential. For example, if the samples are obtained from a biobank, individual samples from the same patient should be analyzed in separate runs, as would occur in routine practice. Rigorous quality-assurance procedures for the biobank—including its organization, management, and procedures for sample handling (e.g., time interval from sampling to analysis, storage conditions, number of freeze–thaw cycles), as well as comprehensive clinical data relating to the specimens themselves—must be in place, of course, and be documented appropriately.

**Phase III Monitoring Trials—Short- and Medium-Term Outcomes**

Whereas phase III drug trials assess whether a promising new drug is superior to the drugs in routine use, phase III monitoring trials establish whether a new biomarker has clinical utility. Such trials also assess whether the biomarker provides additional information that could improve patient care and/or outcome (i.e., whether patients who have their therapies changed on the basis of biomarker monitoring and routine surveillance fare better than patients whose interventions are based solely on routine surveillance without biomarker measurement).

**Phase III Trial Design**

The design of monitoring trials is inevitably much more complex than designs based on samples collected at a single time point, which may be appropriate for assessing such predictive or prognostic tissue biomarkers as epidermal growth factor or p53. Besides also being susceptible to the same confounding factors that have been identified for trials involving a single time point (e.g., lack of results for some patients, technical errors, and so forth), longitudinal monitoring studies might also show serum biomarker concentrations to be stable and then increase or decrease. In other words, a steady state cannot be assumed. We therefore propose 2 types of phase III study designs, IIIA (Fig. 1) and IIIB (Fig. 2), the selection of which will be informed by the results of the phase II trial. Randomization is performed as early as feasible for both designs to avoid losing lead time, because a delay might reduce the benefit of an early intervention demonstrated during the trial.

Of note is that all study arms of phase IIIA and IIIB trials include monitoring with standard methods (e.g., appropriate imaging methods) to avoid false-negative results (i.e., failure to identify progression) that could occur if the tumor(s) do not release the biomarker into the circulatory system at sufficient concentrations.
False-positive signals can occur in both phase IIIA trials (Fig. 1, treatment arm C) and phase IIIB trials (Fig. 2, treatment arm D), because increases in biomarker concentrations prompt an action without confirmation by standard diagnostic methods. Details of both trial designs are outlined below.

A. Phase IIIA trial design—single treatment and control period. The phase IIIA design (Fig. 1) is simpler and is most appropriate for short-term studies involving a single treatment and/or a relatively short time (i.e., several months to a few years). All patients are allocated to biomarker monitoring during a single treatment and control period of standard care (Fig. 1). Patients in treatment arm A (i.e., with stable disease according to standard criteria with no change in biomarker values: SDTM in Fig. 1) continue with current standard care until tumor growth is confirmed, when the next line of standard treatment is introduced. Patients in whom biomarker progression has been confirmed (PDTM in Fig. 1) are randomized into 2 groups: Patients in treatment arm B continue with standard care. When disease progression is noted with standard methods, the next line of treatment is introduced. Patients in treatment arm C are changed immediately to the next line of treatment, if available. The relevant primary end point is progression-free survival.

B. Phase IIIB trial design—consecutive treatment and control periods. In contrast, phase IIIB monitoring trials (Fig. 2) may span several consecutive periods of treatment and control. Enrolled patients are randomized for monitoring either via standard methods (pathway I) or via standard methods and biomarker measurement (pathway II). In pathway I, standard care is continued until tumor growth is identified, when the next line of standard treatment is begun. In pathway II, current care is changed to the next line of standard treatment on the basis of defined changes in biomarker values (PDTM in Fig. 2).

Blinding groups of patients in the biomarker-based monitoring strategy according to their biomarker results is unnecessary, because the patients maintain standard care until PDTM or until—in cases of PDTM absence—routine methods verify tumor growth. The situation is different for the patients randomized to the routine methods-based monitoring strategy, because the patients in these groups will not know their biomarker results. With a longer monitoring period and a larger number of patients, the relevant primary end point is overall survival.

ASSESSMENT OF OUTCOME THROUGH PHASE III TRIALS

Although consecutive treatments are initiated independently of the biomarker results, samples for biomarker measurements should be collected during the entire monitoring period for all patients. Such results enable the clinical outcomes of patients with PDTM for whom standard care is continued until tumor growth is verified by routine methods (Fig. 2, treatment arm B) to be compared with the clinical outcomes of patients for whom therapy is guided by early changes in the biomarker (Fig. 2, treatment arm D). Also feasible are comparisons of the outcomes for patients without PDTM who are monitored by standard methods (Fig. 2,
ADDITIONAL CONSIDERATIONS RELEVANT TO PHASE III TRIALS

If the data from phase II studies are sufficiently compelling to support the clinical utility of a biomarker for monitoring, the need for phase III trials is perhaps debatable. To reinforce points we have made earlier in this article and other authors have noted (20, 30), however, we note that there is substantial potential for bias in phase II studies. This conclusion is illustrated by results from a prospective randomized clinical trial in which outcomes for clinically disease-free ovarian cancer patients monitored routinely with CA125 were compared with patients who were not so monitored (31). Increasing CA125 concentrations in a patient with a history of ovarian cancer has a high positive predictive value for subsequent recurrence, with lead times of 2 to 6 months depending on the criteria used. Furthermore, the biomarker has been used for >20 years to monitor posttreatment relapse (4, 32); however, the results of the randomized clinical trial demonstrated no detectable difference between the 2 patient groups in either overall survival or quality of life (31). Serious flaws in the trial design, including the use of suboptimal algorithms to interpret serial changes in CA125 concentration, were subsequently suggested as possible reasons for this unexpected conclusion (33). The study spanned almost a decade, and the apparent failure of CA125 monitoring to affect survival may have been due to the inadequacy of the therapy for recurrent disease rather than to the information the biomarker results provided. These results highlight the urgent need for both effective alternative therapies and clear guidelines for the design of studies of biomarkers used for serial monitoring.

Another approach is to perform systematic reviews and/or metaanalyses as soon as results become available for a sufficient number of studies that follow a design similar to that of phase II or phase III. Introducing a biomarker into routine clinical practice may be appropriate if the evidence provided by such reviews is sufficiently strong and if such analyses are carried out rigorously and according to best practice (12). A good example of this approach is the American Society of Clinical Oncology recommendation that carcinoembryonic antigen be measured in patients with stage II or III colorectal cancer every 3 months, for at least 3 years after diagnosis (4). This recommendation is based on a metaanalysis demonstrating that only monitoring studies that included carcinoembryonic antigen had a significant impact on survival. Introduction of a new biomarker should be considered only if its performance is superior to that of the diagnostic tests (imaging, biochemistry, and so forth) it is replacing or supplementing (e.g., because it is less invasive, more convenient, and/or less costly) (20).

Finally, it is obvious that if the evidence for the effectiveness of a biomarker is sufficiently compelling, it may be ethically unacceptable to include a control arm in which the biomarker is not measured.

Phase IV Monitoring Trials—Long-Term Outcome

With respect to new drugs, which are monitored over extended periods of time to independently assess their long-term benefits and risks (i.e., postmarket surveillance), the clinical consequences of introducing any new monitoring procedure should be evaluated through careful auditing practices. A convenient means of doing so is by designing phase IV trials (Table 1). Such trials follow large patient cohorts to assess whether the biomarker-monitoring program mirrors the performance observed in phase III trials and/or supported in relevant systematic reviews. These studies should include consideration of the impact of the biomarker’s use for the areas outlined below.

A. PREANALYTICAL AND ANALYTICAL FACTORS

Transferring new biomarkers from the research arena to the routine clinical environment presents substantial challenges (20). The performance of the biomarker in routine practice should therefore be audited to ensure adherence to the preanalytical requirements identified before the phase I study (e.g., those related to sample shipment, storage, freezing/thawing, type of blood-collection tube, and time of sampling). Internal QC data for such relevant parameters as imprecision should also be examined, because they may influence the clinical accuracy of the biomarker. Evidence of the long-term stability of the assay should also be obtained.

B. POSTANALYTICAL FACTORS

The accuracy of the criteria used to interpret serial measurements obtained in routine practice needs to be checked against that predicted from the phase III studies.

C. CLINICAL EFFECTIVENESS

Clearly, the most important question to address is whether introducing monitoring with the new biomarker has a beneficial effect on outcome, such as overall survival. A related question is whether earlier detection of tumor growth leads to more long-term adverse effects of treatment. The potential of longer treatments to reduce the quality of life should be as-
sessed, as should any psychological effects of repeated testing.

D. Economic Impact
Whether the costs of introducing biomarker monitoring have been balanced sufficiently by measurable improvements in the quality of life should be assessed with established economic models. This balance of cost and benefit over the long term is often expressed as quality-adjusted life years (13). Closely related to this concept is the assessment of cost-effectiveness, i.e., whether the new monitoring procedure improves patient outcomes sufficiently to justify the costs associated with its use. Such assessment is also essential (13). In this respect, ensuring that a new biomarker replaces rather than adds to existing tests is also mandatory (20).

Conclusion
The proposed strategy for evaluating tumor biomarkers used in monitoring treatment response is based on a phased approach, analogous to the well-established procedures for evaluating new pharmaceuticals (11). This phased framework for designing longitudinal tumor biomarker–monitoring studies is generally applicable to any solid tumor. Of course, the details of trial design will be influenced by tumor type, clinical setting (e.g., disease stage), and phase of the trial. The appropriate number of patients for each trial phase will depend on the number of events (i.e., disease recurrences or deaths), the kinetics of the biomarker during remission and progressive disease, the effectiveness of the criteria used to detect biomarker increases and decreases, and the potential lead times the biomarker provides. The length of the monitoring period, the sampling interval, and the number of samples per patient will also depend on the tumor type. The phased designs we have proposed have intentionally been kept flexible to facilitate modification when appropriate.

In conclusion, a biomarker intended for monitoring may supplement or replace accepted procedures—such as imaging methods or other biomarkers—that have already been adopted into routine clinical practice, provided that a strong correlation exists between the changing biomarker concentration and tumor burden. A biomarker’s clinical effectiveness for monitoring is ensured only if it identifies changes in tumor burden earlier than other procedures and only when effective alternative treatment options are available. Establishing the clinical utility of a biomarker for use in monitoring requires carefully phased trial designs, such as those we have described.

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.

References


