Meeting Summary

Research Strategies, Study Designs, and Statistical Approaches to Biomarker Validation for Cancer Diagnosis and Detection

A National Cancer Institute Workshop

Gaithersburg Marriott
Washingtonian Center

July 28-29, 2004

Executive Summary

This NCI workshop, "Research Strategies, Study Designs, and Statistical Approaches to Biomarker Validation for Cancer Diagnosis and Detection," convened leaders from biostatistics, clinical oncology, and regulatory disciplines to discuss ways to enhance and optimize study designs that incorporate biomarkers for cancer diagnosis and early detection. During the course of two days of fruitful discussions, participants reviewed current study designs and suggested various approaches to address key problems in biomarker studies that use high-dimensional data, and multiple markers. Participants also offered recommendations regarding related study design issues, sample collection, processing and storage requirements, and regulatory guidelines for cancer-related markers and diagnostic tools.

In the upcoming weeks, the highlights of discussions and key recommendations from this workshop will be reviewed for publication and dissemination to the cancer research community.

Key Recommendations from this workshop include:

Biomarker Validation Study Design Considerations:

- Design the biomarker study upfront to achieve patient benefit
- Consider study designs that minimize bias
- Base the value of a biomarker on its positive predictive value (PPV) in a population that encompasses the marker's proposed clinical indication
- Use the true positive rate (TPR) and the false positive rate (FPR) as central guideposts to measure biomarker performance
- Avoid overfitting in multiple-marker studies by selecting a classification rule based on the training sample and estimating marker performance in a separate test sample
- Use biology, rather than technology, to determine parameters for "cutoff" points, especially in studies with high-dimensional data
- "Piggy-back" biomarker validation studies to ongoing prevention and treatment trials when possible
- For early detection studies, assess the marker as a trigger for early intervention using randomized clinical trials or observational studies.

Related Issues**:**

- Develop clear regulatory guidelines to approve cancer biomarker-related diagnostic tools
- Develop standards for tissue storage, handling, and preparation
- Create mechanisms to facilitate specimen distribution, sharing, and utilization

Welcome

Peter Greenwald, Dr.P.H.; Director, Division of Cancer Prevention, NCI

Dr. Greenwald welcomed participants to the workshop by emphasizing that "validation" can have different meanings and that the NCI Early Detection Research Network (EDRN) has given attention to the practical aspects of how the term is used. He commented that mathematics, rigorous criteria for evidence, and validation offer the best hope to solidify biomarker research into a hard science. Two years ago, NCI sponsored a workshop on the applications of bioinformatics to cancer detection (Kapetanovic IM, *et al. Ann NY Acad Sci* 2004;1020:1-9). Several points from the recommendations of this report and discussions with NCI colleagues inform the discussions of this workshop. First, there is a trend toward technology-driven, rather than hypothesis-driven research. Hypothesis development requires integrating the intellectual strength of numerous disciplines when thinking about the problem. Specific recommendations from the report that are relevant to this workshop's discussions include:

- Incorporate bioinformatics into experimental designs from their inception.
- Generate well-annotated data sets with functional categories that are related to cancer.
- Support the development and accessibility of computational tools for data analysis
- Encourage gold standards and reference laboratories for validation
- Provide cross-training opportunities for biology/medicine and bioinformatics.

Institute/Agency Perspectives

Greg Downing, D.O., Ph.D.; Director, Office of Technology and Industrial Relations, NCI
Theresa Mullin, Ph.D.; Assistant Commissioner for Planning, FDA

Dr. Downing noted that cancer research has reached a strategic inflection point, where researchers can generate a wealth of sensitive, specific data. The current challenge is to understand how these data and related technologies can be engaged to bring initiatives to the clinic. Since 1999, when the NIH convened a workshop on biomarkers and surrogate endpoints (Downing GJ, ed. *Biomarkers and Surrogate Endpoints: Clinical Research and Applications*. Amsterdam: Elsevier, 2000), there have been many advances in developing algorithms, statistical methods, and bioinformatics tools for preventive medicine and cancer chemoprevention. Realizing that these tools could benefit cancer patients, the NCI partnered with the US Food and Drug Administration (FDA) in April 2003 to create an Inter-agency Task Force and Working Group to:

- Provide a framework to exchange information between the NCI and FDA
- Develop strategic initiatives to facilitate regulatory aspects and the translation of diagnostics and drugs to the clinic
- Integrate with the NIH Roadmap
- Encourage cross-disciplinary training in clinical and regulatory sciences to hasten the translation of regulatory sciences into clinical settings.
- Establish the FDA as a site on the NCI Cancer Biomedical Informatics Grid (caBIG)

In addition, the NCI has been working with the American Association for Cancer Research (AACR), the American Society of Clinical Oncology (ASCO), and the FDA to evaluate

literature related to biomarkers of clinical benefit. The Institute has continued its emphasis on collaborations in proteomics, and has supported the development of collaborations with the FDA (regulatory issues) and the National Institute of Standards and Technology (NIST) (nanotechnology platforms). Dr. Downing noted that the NCI plans to develop new initiatives to incentivize collaboration, including public databases, a national infrastructure for biorepositories and specimens, and new strategies to use proteomics for biomarker discovery. He stressed that current needs arise in four areas:

1. Well-defined science, based on regulatory processes, that will enable technologies and their derived data to foster clinical applications
2. Standards to characterize data from contemporary technologies (e.g., proteomics, microarrays)
3. A Roadmap for new technologies (e.g., imaging, proteomics, genotyping)
4. Cross-cutting partnerships

Dr. Mullin noted that the NCI/FDA Inter-Agency Task Force addresses hurdles in infrastructure and product development. She noted that the two agencies share one goal—to deliver safe products to patients more quickly. She also discussed the FDA's Critical Path Initiative to reduce cost and uncertainty when developing cancer-related products. For product approval, innovators and researchers have to demonstrate safety, effectiveness, and commercial-scale manufacturing capability. Three types of uncertainty must be addressed:

- Technical—setting better science-driven standards
- Policy—streamlining and clarifying the regulatory pathway
- Market—determining what standards of performance are needed.

Dr. Mullin concluded by observing that the NCI plays a leadership role in developing data standards for application submission, and the agency's work with the FDA showcases the type of collaborations needed to realize the full potential of biomarkers in the detection and treatment of cancer.

Workshop Goals and Objectives

Sudhir Srivastava, Ph.D., MPH; Chief, Department of Cancer Biomarkers, NCI

Dr. Srivastava echoed Dr. Greenwald's observation that the term "validation" has numerous meanings (see Ransohoff DE, *Nat Rev Cancer* 2004;4:309-314), thus making it difficult to determine when a biomarker has been validated. This lack of agreement has hampered the use of biomarkers in cancer research and development. Noting the EDRN's long-standing interest in biomarker validation, he highlighted the EDRN's five phases of clinical validation of biomarkers—preclinical exploration, clinical assay and validation, retrospective longitudinal analysis, prospective screening, and cancer control (*JNCI* 2001;93:1054-1061).

He noted also, however, that the rules of evidence for evaluating diagnostic and prognostic studies are not well developed, causing many studies to suffer from overfitting (when many variables are used to discriminate among relatively few outcome events) and bias (from small, nonrandomized or non-independent validation sets). Overfitting and bias ultimately create

substantial hurdles to reproducibility, including inherent cofounders and poor designs for proof-of-principle studies.

Dr. Srivastava noted that a consensus must be reached, with input from all members of the cancer community, to optimize current study designs and inform designs of the future. The goal of this workshop is to take one step toward that consensus by generating a series of recommendations for the NCI and the NCI-FDA Task Force regarding the study design and statistical approaches needed to validate biomarkers for cancer diagnosis. Specific areas of discussion include:

- Basic considerations that underpin the study design, statistical methodologies and validation approaches to the rapidly advancing field of cancer biomarkers
- Approaches to validate biomarkers for clinical utility in randomized clinical trial (RCT)-based and non-RCT based validation designs
- FDA guidelines for technology and biomarker evaluation, analytical performance characteristics, and regulatory aspects of biomarker validation
- Approaches to validate and analyze biomarker data
- Strengths and weaknesses of various types of study design
- Strategies to "piggy-back" validation studies to ongoing prevention and treatment trials
- Issues with case-control study designs based on completed trials
- The suitability of proteomic and genomic samples for use in biomarker validation studies

Biomarkers in the Clinical Trial Design for Diagnosis and Early Detection
Don Berry, Ph.D.; Professor and Chair, Biostatistics and Applied Mathematics, M.D. Anderson Cancer Center

When considering the appropriate uses of statistics in biomarker-based clinical trial design, Dr. Berry noted that numerous attitudes abound, including:

- Create entirely new approaches
- Discard the usual error types I and II cutoff points
- Recognize false-positive rates, multiplicities, and the need for randomization
- Recognize that modeling (molecular and for clinical endpoints) is essential to understand the relationship between biomarkers and disease
- Synthesize biological and empirical information using a Bayesian or other paradigm.

After providing some case studies that highlight the differences between using a marker for prognosis versus prediction and the use of retrospective studies to build and validate a successful statistical model, Dr. Berry discussed the uses of biomarkers in clinical trials. Citing the critical importance of modeling, he noted that biomarkers can be used as auxiliary variables and for seamless Phase II/III design. Conventional drug development paradigms contain a lag time of 9-12 months between Phases II and III. By contrast, a seamless design features a drug-vs-placebo Phase II study with few centers, each enrolling 10-20 patients/month. If predictive probabilities of biomarkers are encouraging, the trial is then expanded to Phase III, with many centers that enroll 40+ patients per month. In a single trial, survival data from both phases can be combined in the final analysis. Frequent analyses using predictive probabilities of statistical significance

allow judgments regarding accrual and continuation of the trial. Such an adaptive design allows fewer patients to be enrolled, enables a smooth transition between Phases II and III, and uses data from all patients to assess the Phase II endpoint and the relationship between the biomarker and survival. He commented that such a design allows for the possibility of up to 900 patients being enrolled, although the answer is usually known at a much earlier point.

A similar application is a dose/response study in which adaptive doses in the Phase II setting offer the efficient and rapid elucidation of the dose-response relationship. In this case, the biomarker is used as an auxiliary variable.

When using longitudinal markers (e.g., CA-125 in ovarian cancer), available data from the trial and previously generated data are used to model the relationship over time between the biomarker and survival, depending on therapy. By calculating predictive distributions for each patient and using covariates, the seamless Phase II/III model can be applied. Such an approach enables key decisions in trial design, including adding or discontinuing study arms or changing doses. As a result, the design offers earlier conclusions and greater precision.

Comments and Discussion:

It was noted that auxiliary variables are also useful in context (e.g., for a prostate cancer prevention trial with auxiliary variable of PSA level) to predict missing outcomes, such as patients who were not biopsied. Missing outcomes depend on the auxiliary variable to reduce bias. However, to use the auxiliary variable non-parametrically, a strong relationship between variables is necessary to yield useful results from intermediate information and final outcomes.

Biomarkers for the Early Detection of Cancer: Statistical Perspectives
Stuart Baker, Sc.D.; Division of Cancer Prevention, NCI

A biomarker for early detection is defined as a measure or indicator of a biological process in asymptomatic persons that predicts future occurrence of clinical cancer. Key questions for biomarker use include selection—What markers, if any, are promising for further study as triggers of early intervention?--and validation--If the marker is used to trigger early intervention, what are the harms and benefits?

For marker selection, the markers must yield a test (classification rule) that is either positive or negative. A “promising” marker suggests large benefits and few harms as a trigger of early intervention. Benefits should be related to the true positive rate (TPR) or the sensitivity, the probability of a positive test indicating cancer. Harms should be related to the false positive rate (FPR), the probability that a positive test does not indicate cancer. TPR and FPR should be estimated from patients with and without cancer, and confidence intervals are important for evaluation and sample size. Ideally, target values are based on a utility function using anticipated benefits and harms if the marker triggers early intervention. The overall goal is a high positive predictive value (PPV), the probability of cancer given a positive test.

Classification rules, based on single or multiple markers, allow the potential for better performance but offer a danger of overfitting. Overfitting occurs when the same data are used to

select a rule for a positive result and to evaluate its performance (via true and false positive rates). Due to overfitting, the true positive and false positive rates are "too good." To avoid overfitting, the training/test sample and the cross-validation must be considered. A potential marker should be selected and the true and false positive rates and confidence intervals should be estimated in the test sample. For cross-validation, data from some subjects should be left out successively, and true and false positive rates and confidence intervals should be estimated using left-out data.

For cases involving multiple markers, two scenarios are possible--a small number of promising markers, and cases of high-dimensional data. For instances that feature a small number of promising markers, a possible classification rule can be created based on the marginal distributions of markers. For high-dimensional data, classification rules can be based on compound covariates (Rdmacher MD *et.al. J Comput Biol* 2002;9:505-511) or more complex strategies such as classification trees (Kapetanovic IM *et.al. Ann NY Acad Sci* 2004;1020:10-21). Study designs for marker selection include the preliminary performance study, retrospective performance study, and prospective performance study.

When validating early detection biomarkers as triggers of early intervention or for evaluation of cancer screening, it is necessary to estimate the harms (e.g., unnecessary biopsies) in a particular study or screening program. Benefits (e.g., cancer mortality) can be estimated using observational studies or randomized trials with a sample size of 50,000 or greater. In observational studies, the main challenge is self-selection bias; namely, that subjects who receive screening have differing risks of cancer. Case-control studies and mathematical models that combine parameters from different sources increase the potential for self-selection bias; periodic screening evaluation and a paired availability design decrease the potential for such bias.

Dr. Baker concluded by offering special considerations for a randomized trial, including:

- Use the cancer mortality endpoint rather than the overall mortality endpoint
- Use an adaptive approach to adjust for follow-up after screening stops, as overall mortality may be diluted by deaths from new cases of cancer after screening has stopped
- Adjust for refusers and immediate switching to estimate the effect of receiving screening (important for meta-analysis)

Discussion:

It was noted that validation of harms and benefits of a marker depends on treatment as well as on the marker itself; therefore, good disease classification is essential. As treatments evolve, validation will vary as the treatment becomes more effective. Thus, it may be prudent to define a variable that is inherent to the marker, rather than to treatment.

Some Aspects of the Use of High-Dimensional Single Nucleotide Polymorphism (SNP) Data for Cancer Risk Determination

Ross Prentice, Ph.D.; Women's Health Initiative, Fred Hutchinson Cancer Research Center

Several million SNPs have been identified across the genome. Substantial public and private efforts (e.g., HapMap) have been developed to identify tagging SNPs that will convey most genotype information (Gabriel SB *et.al. Science* 2002;296:2225-2229). Genotyping costs approximately $0.01/SNP, and this price may continue to fall. High-throughput capability is possible only by a few organizations that have the capacity to handle several hundred cases and controls, with 50,000 to 250,000 SNPs. SNPs can be used to identify persons susceptible to disease for targeted screening efforts and, in the longer term, could provide insight concerning related early detection markers.

Pooling DNA from cases and controls before genotyping provides an economical approach that retains reliable estimates of SNP allele frequencies (Sham P *et.al. Nat Rev Genet* 2002;3:862-871), as demonstrated by a recent test case comparing allele frequency estimates in a pool of 499 cases of type 2 diabetes, a pool of 182 unaffected spouse controls, and a pool of 228 elderly unrelated controls to corresponding measured allele frequencies from individual genotyping (Mohlke KL, *et. al. PNAS* 2002;99:16928-16933).

Several notes must be considered for pooled DNA analysis, including:

- There may be unequal PCR amplification at a given SNP, but analyses based on tests of $p_1/p_2 = 1$ could avoid the need for corrective measures.
- Power calculations based on odds ratios (ORs), rather than allele frequencies, would be of interest.
- Additional comparisons of pooled versus individual SNP allele frequencies are available (Le Hellard, *Nucleic Acids Res* 2002;30:e74).
- The major concern with the pooled DNA approach is the inability to study disease risk in relation to haplotype or other SNP/SNP interactions.

Dr. Prentice noted that a class of study designs worthy of further development involves the sequential exclusion of SNPs that do not satisfy a selected test-critical value at each stage. He then outlined the study design of the Women's Health Initiative (WHI) hormone trial and described several designs under consideration that use pooled DNA in observational studies.

In summary, Dr. Prentice noted that:

- High-dimensional SNP association studies are becoming practical and have the potential to identify susceptible persons and to help elucidate disease pathways.
- Large case-control data sets are needed from well-characterized cohorts, and small DNA volumes are required.
- Pooled DNA has some disadvantages but has major cost-reduction implications for large association studies.
- There is great potential for additional development of study design and analysis procedures, with implications for other types of high-dimensional data (e.g., proteomic data).

Discussion:

Several concerns were raised regarding DNA pooling strategies. For example, effects such as gene-dosing may be missed, and sets of 250,000 SNPs may be insufficient to provide all necessary genetic information in some cases.

Panel Discussion: Review and Weaknesses of Observational Data on Biomarkers' Utility in Cancer Detection and Diagnosis
Moderators: *Susan Ellenberg* and *Ross Prentice*

Strengths and Weaknesses of Observational Validation Designs for High Dimensional Data
Richard Simon, D.Sc.; Chief, Biometric Research Branch, NCI

Dr. Simon discussed common errors in study design and analysis of biomarker data, with emphasis on therapeutic trials and gene expression microarrays. These errors include a lack of a prospective statement of hypotheses and potential biomarker uses, lack of a clear protocol for patient selection and specimen collection, inadequate accounting of patients' heterogeneity, inattention to multiple testing aspects, and differential specimen handling of tissue assays on clinical outcome. He described the elements and value of proper cross-validation in the evaluation of biomarker indices, most notably that cross-validation is valid only if the test set is not used in any way to develop the model. With proper cross-validation, the model is developed from scratch for each leave-one-out training set. Dr. Simon also noted that, for smaller studies, cross-validation is preferable to split-sample validation; internal validation is limited by the precision in the estimated error rate and the data used for the developmental study (e.g., patient heterogeneity, confounding factors, failure to reflect sources of assay variability that will exist in broad clinical application). He described and illustrated the advantages of randomized clinical trials settings for biomarker development and validation, noting that the "gold standard" for design of prospective validation studies is to randomize patients to chemotherapy versus a classifier-determined therapy or to a more efficient design that involves the real-time use of an assay-based classifier.

Proteomics
Ziding Feng, Ph.D.; Division of Public Health Sciences, Fred Hutchinson Cancer Research Center

Dr. Feng discussed the utility of molecular profiling in the early detection of cancer, where single biomarkers are often inadequate. Using high-dimensional surface-enhanced laser desorption ionization (SELDI) mass spectrometric data for illustration, he described the sensitivity of case-control comparisons to variations in serum collection times, aspects of specimen handling, and the potential for model overfitting with this type of high-dimensional data. These issues have hampered replication of studies with some proposed disease classifiers, so that this important source of proteomic disease detection biomarkers has yet to yield a classifier that has proceeded to stage II of biomarker development. He advocated the use of better-quality specimens for epidemiologic cohort studies for the early stages of biomarker development, described aspects of statistical design and analysis to prevent bias and overfitting, and recommended that biomarker development in this area include identification of the proteins associated with diagnostic information.

Genomics
Yudong He, Ph.D.; Scientific Director, Informatics Department, Rosetta Inpharmatics

Dr. He noted that a combination of various types of biomarkers remains the best predictor of disease; any one type of biomarker (e.g., mRNA) is sub-optimal. He discussed the advantages of a high-dimensional parallel measurement system of genome-wide arrays for biomarker identification, in which mRNA profiling technologies suggest biomarkers that are then measured using other means, such as reverse transcription polymerase chain reaction (RT-PCR) or enzyme-linked immunosorbent assay (ELISA). He noted that such a strategy still contains artifacts and biases, some of which are environmental. He stressed that sample selection criteria for biomarker identification should include well-balanced groups for effective classifier training and cohorts for survival assessment and clinical use. Noting that validation performance differs from training performance, Dr. He offered these concluding observations about the use of genomics in biomarkers for early cancer detection:

- Medical, biological, and informatics practitioners must work together to optimize results.
- Expression-based methods, proteomics, and metabolomics are at different stages of development.
- A biomarker identification platform is not necessarily the optimal platform for diagnostic implementation.
- Various statistical methods are being developed for high-dimensional data, but validation remains the key challenge.

Discussion:

Participants suggested several options that may improve biomarker validation studies, including:

- Focus early on the medical context, then develop the classifier and undertake internal validation to determine whether classifier is valid within the specific context.
- Use large cohorts for biomarker discovery studies.
- Investigate high-dimensional metabolomics platforms using mass spectrometry and NMR.
- Reduce experimental bias in prospective studies though design strategies such as randomizing the order of performing assays and standardizing sample handling protocols.
- Split samples between validation and development when testing a model with high-dimensional data.
- Encourage openness among steering committees for the use and distribution of samples for pre-Phase III studies.

Panel Discussion: Strengths and Weaknesses of Longitudinal and Cohort-Based Designs: Piggy-Backing Approach through Treatment and/or Prevention Trials
Moderators: *Bob O'Neil* and *Richard Schilsky*

Investigating Biomarkers in Randomized Trials: Basic Concepts
Sylvan Green, M.D.; Professor, Arizona Cancer Center

Dr. Green reviewed basic concepts for biomarkers in clinical trials, noting that prospective randomized trials to investigate biomarkers have many strengths, including:

- The advantages of "piggy-backing" on randomized trails to obtain an unbiased estimate of the effect of an intervention on biomarkers
- The advantages of using the control group as a longitudinal cohort to get a prospective assessment of the relationship of biomarkers to subsequent disease
- The use of regression models to assess the relationship of biomarkers to the risk of disease
- The possible use of permutation tests to compare intervention groups with respect to multiple biomarker changes observed between groups

When designing any clinical study, the effects of chance and bias must be considered as they relate to patient heterogeneity. These can be addressed by enrolling adequate numbers of patients in the study and randomizing intervention assignments. Non-randomized controls generate a series of problems, including:

- The effect of unmeasured or unknown confounding factors
- Differential participant selection due to consent requirements
- Bias in intervention assignment
- Defining "time zero"
- Possible time trends in patient population, disease characteristics, diagnostic methods, and supportive care

Randomization provides many advantages, such as:

- Bias is avoided.
- Predictive factors (known and unknown) tend to be balanced between the intervention and comparison groups.
- A valid basis for statistical tests of significance is provided.
- A concurrent comparison group controls for time trends.
- Results are more likely to be convincing.

Dr. Green noted that factorial designs are applicable in some instances (e.g., the Physicians' Health Study). However, masking and blinding may include bias, so biomarkers must be assessed by a masked observer. Also, the experimental design should avoid bias from time-, laboratory-, or experiment-based factors. He noted that randomization and blocking can minimize most biases, citing the Prostate Cancer Prevention Trial as a good example of avoiding bias in assignment.

Dr. Green then discussed issues in case-control studies, including the definition and selection of cases, selection and matching of controls, the ascertainment of disease and biomarker status, and the potential for bias. He further recommended the use of regression models (e.g., logistic and survival models) for their advantages when assessing the relationship of biomarkers to the risk of disease. Conceptual frameworks for these models include association (biological, odds ratio,

logistic regression and proportional hazards models) and classification (sensitivity and specificity; ROC models) (Pepe *et.al. Am J Epidemiol* 2004;159:882-890).

Dr. Green described the uses of permutation tests in prospective randomized trials to investigate biomarkers. These tests are based on the randomization (permutation) distribution and conditional on the observed results for each individual (with observed correlations across measurements). Permutation tests require that the outcome measures are exchangeable between groups under the null hypothesis.

In conclusion, Dr. Green stressed the need to obtain unbiased comparisons and adequate sample size for reliable inferences when using randomized trials.

Utilization of Prevention and Treatment Trials for Biomarker Validation Studies
Donna Pauler Ankerst, Ph.D.; Cancer Prevention, Fred Hutchinson Cancer Research Center

Dr. Ankerst noted that prevention trials offer opportunities such as large sample sizes, asymptomatic populations, long follow-up times, baseline and multiple sera values, collected covariates, and otherwise unobtainable samples. Issues in these trials include the use of drug agents, blinding, sample availability at the end of the trial, fixed timing of measurements, analytical methods to accommodate bias, competition for samples, and a limited number of cases for rare diseases. She highlighted the Prostate Cancer Prevention Trial (PCPT), which followed 18,882 men randomized to finasteride or placebo for 7 years. At the end of the study, 60% of the men underwent a biopsy, providing a unique cohort that may function as a test set for new biomarkers. Interim serum samples, collected from each enrolled individual, may also be used.

She noted that treatment trials also offer unique opportunities, including a wealth of densely-spaced samples, independent confirmation of disease progression by sensitive means, markers that predict recurrence or progression early, existent funded trial machinery, collected covariates, and little competition for samples. Issues include short follow-up windows that are often censored by mortality, a heavy treatment influence on biomarkers, unequally spaced measurements, censoring, and a result that is often measured by time-to-event. In the PCPT, by using PSA profiles and recurrence times from 1011 patients treated over seven years, PSA was determined to be an early predictor of recurrence of prostate cancer in a design that had no verification bias.

She concluded by observing that opportunities are rich for attaching biomarker validation studies to prevention and treatment trials. However, there is a need for analytic methods to optimize inference that is subject to design constraints. Dr. Ankerst offered the following advice for "piggybacking" biomarker validation studies to other trials: Form relationships as early as possible, provide a strong justification for a particular study, embed correlate studies into the design phase, and assemble analytical techniques.

Discussion:

Attendees discussed several issues that influence the PCPT conclusions, including an elevated Gleason grade of subjects on finasteride. What is the best approach to determine prediction given

this data set, where pathology does not predict aggressiveness of the disease? Approaches are required to provide valid insight into prediction.

Some Considerations for Layering Biomarker Validation Research onto Prevention Trials and Large Cohort Studies: Examples from the Women's Health Initiative
Garnet Anderson, Ph.D.; Public Health Sciences, Fred Hutchinson Cancer Research Center

Dr. Anderson reviewed the phases of the development pipeline for early detection biomarkers: discovery, translation, assessment, and efficacy. She then discussed the design of the Women's Health Initiative (WHI), a randomized trial that examined the effects of dietary modification, hormone therapy, calcium and vitamin D on disease outcomes in more than 160,000 women. When this study was designed in 1992, specimens were to be used to explain intervention effects in the randomized clinical trial (RCT), examine disease mechanisms, identify/confirm biologic risk factors, develop risk strata, and describe the natural history of disease biomarkers. She outlined several sets of considerations for trial design as they relate to specimens (type, collection times, volumes), outcomes (adequate time to accumulate a sufficient number of events, quality of outcomes data), study population (relevance of biomarkers to a specific population), clinical practice (availability of screening, complementarity of a biomarker to existing modalities), and consent (changing rules, variability of requirements between sites). Dr. Anderson noted that 26 studies have been approved to date that use WHI blood specimens; seventeen of these with PIs who are not WHI investigators (although each is sponsored by a WHI investigator).

She offered the following recommendations:

- Develop state-of-the-art, standardized specimen protocols (emerging technology needs, standards for frequency of collection and processing and storing of samples)
- Develop appropriately targeted outcomes data collections
- Streamline human subjects processes (appropriate flexibility in research scope, coordination of multiple institutions)
- Develop an infrastructure to support research priorities that are ancillary to the main cohort study initiative

Discussion:

One attendee noted that receiver operating characteristic (ROC) curves are more appropriate for diagnostic application for use in prevention trials.

Participants discussed how to choose a series of time points for early detection if there is an observational cohort without screening. Although the "cutoff" time for early detection varies by disease, a prospective study should be designed with a legitimate screening protocol included.

One attendee noted that oncology drug development features an unusual, single arm response for Phase II. Differences between treatment and prevention trials are marked in cancer research, and this is often a problem in validation of biomarkers. The example of CD4 counts in early AZT

trials was provided; because results were noisy, CD4 was not originally considered as a possible marker for AIDS.

Panel Discussion: Experimental Designs and Analytical Methods to Support Validation of Biomarkers for Detection and Diagnosis
Moderators: *Lance Liotta* and *Sylvan Green*

Definitions of Risk, Early Detection, Diagnosis, and Prognosis for Biomarkers and Algorithms in Statistical and Clinical Contexts
Steven Skates, Ph.D.; MGH Biostatistics Center

Dr. Skates started the session by delineating the definitions for risk and early detection in a statistical context. These include:

Sensitivity—the proportion of cases that will be detected by the biomarker
Specificity—the proportion of healthy subjects that will be incorrectly detected
Positive predictive value (PPV)--the number of operations necessary to detect one case

He emphasized the need to use the PPV as a primary criterion for the fitness of a prospective biomarker when making clinical decisions. He cautioned investigators to clearly communicate reports of sensitivity and specificity in the contexts of the clinical study set used and the clinical indication (e.g., stage of development, line of test, and interpretation of test).

He noted also that the first-line test for a biomarker is often determined by factors such as tissue accessibility and incidence rates. He discussed the stages in biomarker development, noting that each stage has specific operating characteristics. Stages include:

1. Identify possible markers
2. Measure potential markers in cases compared to controls
3. Measure potential markers in biospecimens from biorepositories obtained as a result of screening trials

Flexible Study Designs and Power Considerations for Ongoing and Future Trials to Accommodate Emerging Biomarkers and Technology in Genomics Drug Trials
Sue-Jane Wang, Ph.D.; Senior Math Statistician, U.S. Food and Drug Administration

Dr. Wang discussed the need for flexible design of biomarker trials, particularly when the biomarker is being evaluated for its contribution to a decision about a choice of therapy. Often, multiple plausible models are available for consideration. For example, a genomic drug trial for biomarker validation may be split into two trials, either retrospective or prospective--one that explores the biomarker in the clinical context and one validates the biomarker's effect. Dr. Wang outlined several features of flexible genomic drug trial designs, including voluntary participation, model validation in an independent (prospective RCT) study, and risk/benefit analysis with the use of the genomic profile. She emphasized that a decision not to treat, based on expected toxicity predicted by the biomarker, was an important consideration for achieving true patient benefit.

Data Reduction Approaches for High Dimensional Data Derived From High Throughput Assays
Yu Shyr, Ph.D.; Professor, Biostatistics and Preventive Medicine, Vanderbilt-Ingram Cancer Center

Dr. Shyr discussed approaches and limitations of data reduction methods for high-dimensional data, a major challenge for high-throughput techniques such as microarrays and mass spectrometry. Reducing the dimensionality of the dataset allows the distribution of the data to be visualized and simplifies computation by reducing the size of the input space. He noted that the simplest approach is to identify important attributes based on input from domain experts. Other approaches to reduce high-dimensional data include principal component analysis, multidimensional scaling, self-organizing maps, and compound and weighted flexible compound covariate methods. Dr. Shyr cautioned that improper data reduction upfront can reduce regions of the data that have important diagnostic value.

Non-Traditional Methods, Including Modeling for Biomarker Validation
Robert Boer, Ph.D.; Natural Scientist, RAND Health

Dr. Boer stressed that a cancer test must be viewed in the context of the time course of the disease. Early detection may lead to the false impression that survival is extended, when in fact the disease time course remains unaltered. He commented that screening is intended to prevent unfavorable future occurrences, and that early detection can be conceptualized as an unfavorable side effect of screening because it lengthens the disease burden. Dr. Boer discussed the microsimulation screening analysis (MISCAN) modeling strategy that simulates individual life histories and scales the results to representative population. The MISCAN approach combines individual patient data with screening data to predict incidence over time, life years with the disease, incidence following a negative screening, new cases detected by screening, and life years lost due to disease.

Combining Biomarker Panels from High Throughput Discovery
Martin McIntosh, Ph.D.; Department of Cancer Prevention, Fred Hutchinson Cancer Research Center

Dr. McIntosh compared statistical approaches for combining biomarkers into panels. He pointed out that the platform used for biomarker discovery frequently cannot be used for routine high-throughput clinical monitoring. He noted that many approaches are available to estimate a classifier, and all methods that claim to be optimal actually estimate a likelihood ratio. Any method that does so (e.g., binary regression or model based clustering) is therefore valid for combining biomarkers.

He suggested waiting until the end of the drug discovery process to determine the ideal combination of biomarkers, although validation of markers, panels, and algorithms should occur at several intermediate steps. He offered suggestions for combining a novel marker into an existing marker panel, noting that ROC criteria may reject useful, but redundant, markers. He recommended evaluating a marker based on stage, histology, and survival.

Discussion:

Highlights from the productive discussion are summarized below:

A biomarker study should be designed upfront to achieve true patient benefit, i.e. the biomarker should contribute to a clinical decision that improves survival or quality of life. The biomarker value must be viewed in the context of the PPV relating to the population encompassed by the clinical indication. With regard to ovarian cancer, panelists debated whether or not early detection of ovarian cancer recurrence had an impact on survival, and whether or not stage I disease was biologically different from stage III disease. Questions from the participants expanded on several themes, including:

Under what cases are research studies using Case/Control study sets statistically valid so that the research test can qualify for use in a large bank of precious clinical samples? Participants debated whether tests should be prioritized based on biologic relevance to the disease, and it was concluded that clinical benefit should be the deciding factor. One participant pointed out that test precision, reproducibility in the field, sensitivity and dynamic range should all be starting criteria before any test is considered for a planned clinical trial.

What areas can be improved in biomarker validation studies? In a continued discussion, investigators expressed that they needed help with the following: a) Antibody production and test development for their candidate serum biomarker, b) availability of serum standards and controls useful for comparison across institutions and platforms, and c) availability of qualified (accepted by the scientific community) clinical serum sets of true positive and true negative cancers versus controls.

FDA Guidelines for Technology and Biomarker Evaluation

Maria Chan, Ph.D.; Team Leader, Division of Immunology and Hematology Devices, FDA

Dr. Chan noted that the FDA as regulated medical devices (including *in vitro* diagnostic devices (IVDs)) since passage of the Safe Medical Device Amendments in 1976. This act set general controls for product registration and listing, good manufacturing practices, and post-market surveillance. Since implementation, the FDA has access to a menu of tests on the market, and tools to ensure quality production and problem identification. She discussed the avenues for pre-market review, the 510(K), and pre-market approval, the PMA. The 510(K) governs updated versions of existing devices; the PMA, new devices. She discussed FDA definitions of safety and effectiveness, as outlined in the FDA 21 CFR 860.7.

Dr. Chan discussed three core FDA review issues—analytical performance (e.g., accuracy/bias, precision, analytical specificity and sensitivity), clinical performance (how the analytical signal will be used, the clinical or diagnostic sensitivity and specificity, and the predictive value of positive and negative results), and labeling.

She reviewed the FDA's standardized review model for new devices, noting that FDA review is cost-blind, not outcome-oriented or evidence-based, and different from CMS assessment of

medical necessity. For diagnostic markers (e.g., PSA, PAP smear readers, HPV, mammography), she discussed issues such as the impact on the PPV, defining a "gold standard," and biases. She highlighted these issues in context of the approval of PSA and automated PAP readers. To resolve a discrepancy, the FDA does not allow re-testing of discrepants to lead to re-estimates of performance.

To establish performance of a biomarker, the FDA prefers a "yardstick of truth," ROC curve analysis, and the ability to understand behavior in intended use populations. When monitoring a biomarker, survival is the most definite endpoint. Challenges include the default from sensitivity and specificity to measures of agreement with non gold-standards, heterogeneous populations and difficulties in calibrating and standardizing novel tests.

Dr. Chan offered several suggestions for study design that uses predictive biomarkers. These include treating everyone as a control and treat the partitioned population (in which improved response is significant when the marker is present) as an experiment, treating if non-treatment is unethical, treating or avoiding treatment in a subset based on marker results, avoiding treatment if toxicity is unacceptable, results are incomplete, or there is a biologically plausible basis for curtailing the study. Alternately, one can treat positive cases only and randomize for control and treatment. She noted also that the time of the drug study is often a limiting step in the approval process.

With respect to established markers, Dr. Chan commented that the FDA is reluctant to have controlled studies repeated. For biomarker studies, the agency prefers a good upfront design, attention to data collection, use of clear statistics, and demonstration of good science.

Discussion:

Participants discussed labeling and the limitations of various approved biomarkers, such as HER2 and the PAP test. Biomarkers and devices may be approved conditionally, and the applicant must make specific recommendations of how the product will be used in context and to demonstrate that using a device impacts clinical outcome. However, the label will always clearly state the limitations of use and how the application will be used.

Report Presentation of Breakout Sessions

Breakout Group I: Analytical and Performance Characteristics
Co-Chairs: *Stuart Baker, Martin McIntosh, Steven Skates,* and *Yu Shyr*

This group discussed the use of statistics in the selection of promising biomarkers and highlighted the central role of true positive rate (TPR) and false positive rate (FPR) for measuring performance of the biomarker. FPR and TPR can be summarized using receiver operating characteristic (ROC) curves. When investigating the performance of a classification rule based on multiple markers, it is important to avoid overfitting, which can be accomplished by selecting the classification rule based on the training sample and estimating marker performance in a separate test sample. In multiple-marker scenarios, the classification rule involves either the joint distribution of a small number of promising markers or high-

dimensional data from gene expression arrays or proteomic studies. Three basic study designs, the preliminary, the retrospective, and the prospective performance study, were discussed. To validate a marker for the early detection of cancer, it is necessary to evaluate the marker as a trigger for early intervention, using randomized trials or observational study trial designs.

The breakout group offered the following recommendations for consideration:

- If stored sera are available, skip the preliminary performance and prospective performance studies.
- For retrospective performance studies, it is important to investigate marker features that change over time and how easily these data could be collected in practice.
- Because of noise with high-dimensional data, one should first try to identify the top few features that differ between cases and controls and investigate classification rules for some combination of these features. For example, one could consider re-estimating the top 20 features in one application versus the top 20 that work together as joint behaviors.
- Final performance evaluation should be based on an external independent sample.
- In preliminary performance studies, one should investigate the longitudinal behavior of markers in controls to see if it is stable over time. This behavior is a good indicator for the retrospective performance phase.

Comments:

It was noted that there are points in the selection and validation processes in which decisions must be made regarding which markers continue into subsequent study phases. Biology should drive the choice at these points; statistically, there is no justification to optimize with fancy methods, when the function of these markers and methods will change downstream. One attendee commented that small changes may have large downstream impacts. If a selection decision is made too early, key markers may be missed. Five genes that change subtly may make more difference than one gene that undergoes dramatic alteration. Another attendee noted that the scale of a particular change is less important than the biological significance of the change. However, it was noted that a large sample size is necessary to identify multiple markers that work together (if they do not have individual impact).

It was also noted that, from a set of candidate genes, many will undergo only a small change. Therefore, the biologic relevance and pathway involvement of the genes must be understood. Therefore, a realistic number of genes should be selected that will be manageable in a clinical testing platform that can be envisioned.

Breakout Group II: Considerations for Biomarker Validation Regulatory Requirements for Commercialization
Co-Chairs: *Emmanuel Petricoin, Sudhir Srivastava,* and *Lakshmi Vishnuvajjala*

This breakout group began by characterizing the issues related to analyte-specific reagents (ASRs) and biomarkers for early detection of cancer, including:

- The definition of Class III ASR is confusing, and a better definition is needed in light of the diagnostic uses. For example, HIV and components of its tests are classified as Class III.
- Problems exist with reimbursement for Class III ASRs if they are for research or investigative use only. Clarification is needed on this issue.
- Other issues exist with labeling for ASRs and biomarkers (early detection is classified as Class III).
- Classification is determined at the FDA and may be done at the level of the Department of Health and Human Services; thus, uniform guidelines are needed on the requirements for different classes.
- FDA is just beginning to require that manufacturers provide information, and regulation is expected to increase in time.

From the perspective of the FDA, the following observations were made:

- Industry needs a clear idea of the intended use of a biomarker; changes in intended use reconfigure the requirements.
- Most examples are considered on a case-by-case basis.
- Industry requires robust and firm guidance for the FDA, which can be difficult in a rapidly-changing field.
- Industry must be transparent and decide upon its platform for biomarker evaluation.
- New, specific analysis is required of manufacturers.
- Applicants for device/drug or device/biologic combinations can receive help by contacting the FDA Office of Combination Products.

However, the industry considers the following perspectives:

- Expedited review processes are needed for markers that would benefit a small patient population.
- The FDA should consider multiplex testing and how it would regulate it.
- A *de novo* process can be used for a low-risk device if there is no predicate.
- Clear definitions are needed for the term “multiplex”; i.e. is an imaging/biomarker combination considered multiplex?
- Clinical benefit of the test must be demonstrated for any application.

The clinical diagnostic industry has the following questions:

- How will the FDA participate and interact in the clinical trial design and statistical analysis processes?
- What types of correlation data will be required for a biomarker discovered and validated on a non-clinical platform but transferred to a commercial IVD platform?
- Will pre-approval or an IDE be required for a biomarker to be used to guide a drug trial?

Recommendations for the FDA include:

- Develop clear guidelines to approve cancer biomarker-related diagnostics (e.g., ASR definitions, criteria for classification and demonstrating clinical utility of biomarkers that affect therapeutic decision making)
- Create roadmaps of composite markers that indicate areas of regulatory emphasis and priorities (e.g., multiplex testing, genomics, quality standards and benchmarks, combination products guidelines for device/drug or device/biologic applications)
- Establish an early dialogue between FDA and industry, starting with a narrow indication before expanding. Issues covered in the dialogue should include how to clearly define intended use, clear study designs that link the hypothesis, statistical analysis and what needs to be demonstrated), and issues for implants that monitor or deliver drugs.

Discussion:

One attendee noted that the FDA often receives requests for a "roadmap" for the evaluation process, and the agency has found it challenging to devise an all-purpose guide that is not restrictive. It was noted that communication is the key factor in all cases; although a roadmap is significant, the approval process remains a collaborative venture between numerous constituents.

Breakout Group III: Development of Alternative, Non-Traditional Approaches to Biomarker Validation
Co-Chairs: *Ziding Feng, Sue-Jane Wang* and *Ralph Kodell*

This group used the EDRN's five phases of biomarker development and validation to frame the following questions:

- At which phase should FDA approval be sought?
- What should be done after a typical case-control Phase I study?
- Can simulation models be used widely, even for early stages, in addition to evaluating ultimate impact?
- Is evaluation of high-dimensional biomarkers different? Should the mechanism be known?
- Should Phase III-IV specimens from large cohort studies be reserved only for Phases III and IV?

The group discussed several issues that relate to high-dimensional genomic and proteomic biomarkers, such as:

- Is a protein spectrum or set of genes one marker or many markers?
- Do we need to know the features and fully understand the biology to use these markers?
- These markers will be more difficult to validate if the mechanism is not fully understood.
- For development of high-dimensional markers, how data are collected is important, as measurements are limited and technology is changing.

The group noted that each phase of biomarker development requires validation, but true confirmation of patient benefit is established in Phase V. Because of the expense, however, phases IV and V can be completed for a select few biomarkers. Moreover, the necessary patient

cohort must be quite large (tens of thousands) for at least a decade of follow-up. Not surprisingly, sponsors wish to use retrospective data instead of conducting large-scale randomized clinical trials.

The breakout group's specific recommendations for drug trials that use genomic biomarkers include:

- *Use two trials to (1) identify a marker, and (2) conduct a prospective validation study.* However, all potential markers must be identified prospectively in the first trial—otherwise, the test set becomes the training set.
- *Reconfigure the current paradigm.* The current approach involves a single study at a single site, where a big effect observed (e.g., OR=30), followed by a Phase II study with a smaller effect (e.g., OR=3). A better approach is to conduct the first study as a multi-site trial with a larger population and a solid design, with a smaller OR (e.g., OR=10). The Phase II study can then have a smaller reduction in OR (e.g., OR=5).
- *Carry through to Phase II only those potential biomarkers that are identified in all participating studies, once reproducible measurement criteria have been established.*
- *Use this level of identification and validation before moving to Phase III.*

The group concluded by observing that replication is necessary in Phase I in order to move to Phase II, and this replication may involve multiple sites and studies. Furthermore, a large randomized clinical trial is needed in Phase V to confirm patient benefits, although realistically some tests will be in clinical use without an RCT. When validating via a simulation, it was noted that a model can never replace observation. True validation means that results can be reproduced by other labs in other settings. Thus, banking specimens for sharing is important.

Discussion:

One participant expressed concern over using OR as a criterion for moving between phases, and group members noted that OR was used only as an example of a quantitative criterion; others could include sensitivity and specificity.

Another attendee suggested that high quality specimens should be used for biomarker study rather than be reserved only for use in later phases of biomarker development.

Breakout Group IV: Clinical and Biological Challenges: Biological Perspectives
Co-Chairs: *William Grizzle, Dean Brenner* and *Jose Costa*

This group discussed how to refine the clinical questions that should drive the technology; defining the scientific and mechanistic constructs and how the technology can help to address these constructs. Also discussed were the scientific and technical aspects of sample management (collection, processing, storage, distribution). The group noted that future technology cannot be predicted, so pristine samples should be prioritized for discovery efforts. Members debated whether use should be established prior to selecting and testing the technology platform with heterogeneous samples for validation.

The group stressed the need for prospective standards for accepting data and progressing through subsequent phases of the validation pipeline. Progression should be based around standards for acceptance or rejection that are based on consequences.

The group made recommendations in three key areas, as detailed below.

1. **Standards for specimen collection, processing, and storage.** Because standards vary with application, a prospective approach is needed to collect samples. How samples are collected, processed, and stored determines the technologies that will be used. For samples to be useful in discovery studies, optimal storage conditions (e.g., storage at -80 °C) and data recording optimize future sample use. Thus, accurate sample conditions are necessary to understand the limitations that they impose on study and use parameters. The group recommended creating standard repositories that are useful for initial studies, standardizing collection and storage procedures to reap maximum benefit from the samples, and match sample analysis with appropriate technology. Also, academic labs must be encouraged to move toward industry standards for compliance and good laboratory practices.

2. **Selection criteria to choose and standardize technologies.** The group stressed that the choice of scientific question drives sample procurement and technology used for validation studies; for a validation study, the quality of the scientific question should inform study design. For situations in which several innovative technologies are being considered, but a finite amount of tissue samples is available, appropriate retrospective repositories should be considered.

3. **Validation criteria.** The group stressed that standards must be developed for validation, noting that these standards will vary between studies. They recommended using a cost-benefit analysis when developing approaches that reduce sensitivity and specificity, thereby enabling the future use of less expensive biomarkers. The National Comprehensive Cancer Network (NCCN) was suggested as one model approach.

Discussion:

It was suggested to recommend an agreed-upon specimen collection that could serve as a generic "starting set" for new investigators. Existing resources could be made available for discovery through a cooperative human tissue network, rather than though single banks. By allowing groups to use unique strategies to collect and utilize specimens, these networks could then be searched on the basis of how the samples will be used rather than how they were collected. An individual tissue bank within the network would receive credit in publications that use its tissues in a scientific manner.

Breakout Group V: Biological Specimens from Large Institutional Trials in Support of Biomarker Validation
Co-Chairs: ***Steven Hirschfeld, Ross Prentice*** **and** ***Padma Maruvada***

This group discussed the challenges and issues that affect the collection, distribution, and use of specimens acquired in large clinical trials. Group members noted the value of designing a resource that would catalogue tissue specimen collections, acting as a "tissue locator" for specimens that meet search criteria.

Top recommendations for specimen collection, distribution, and use include:

- Facilitate specimen distribution and utilization (user friendly platforms for access to specimen collections for early detection and therapeutic questions and establish a hierarchy of quality of specimens and an appropriate number of samples matched to a type of question)
- Develop or promulgate existing standards for completeness of specimen-associated data (including handling and preparation methods and quality to biomarkers of interest, associated clinical data, stability, and existence of associated specimens)
- Develop systematic monitoring of specimen quality over time—mechanisms to randomly select specific samples and perform other checks to test the integrity of a specimen bank
- Create a forum for collaborative biomarker research among major programs

Secondary recommendations from this group include:

- Articulate sample needs and establish a paradigm for sample needs in cancer detection studies
- Coordinate and promulgate standards for sample acquisition
- Clarify ownership and IP issues.

Discussion:

It was noted that several broad international efforts are underway to begin the process of standardizing and cataloguing tissue specimens. Moreover, forums such as the one chronicled here allow representatives from various cohort studies to detail their sample collection criteria, share ideas, learn needs, and establish collaborations. It was noted that studies of the magnitude of the WHI will not be sufficient on their own, and that another consortium oriented toward early detection oriented may be appropriate.

To stimulate sample exchange between specimen banks, it was suggested to create a special forum for information exchange on research applications and informational needs. Investigators from current joint efforts, such as the WHI, are well-attuned to sharing resources and should be included.

One attendee noted that standardization is important for the EDRN—if sites can set standards for methods of collection by developing common data elements (CDEs), perhaps standard operating procedures (SOPs) could also be developed at each epidemiologic center. To acquire sufficient numbers of samples for early detection studies, samples can be collected more rapidly to yield added outcomes. However, this approach carries the danger that collections will become too large for practical storage, necessitating the need to prune collections. Currently, there are no guidelines that survey issues of reducing collection size at national respositories. It was noted

that the NCI is developing the National Biospecimen Network (NBN), although this resource has not been developed completely. Stored samples are available at various locations throughout the country, although some of these samples are old, and the complete roster of sites has yet to be catalogued fully.

Workshop Conclusions and Next Steps:

Drs. Srivastava and Maruvada thanked participants for this invigorating workshop. They noted that a white paper will be prepared from these proceedings that will be submitted for publication in the *Journal of the National Cancer Institute*. Authors for the paper will be contacted to determine next steps for paper submission.

The workshop was then adjourned.