

# CRITICAL NATIONAL NEED IDEA

## From Genomics to Personalized Medicine with Advances in Measurement and Experimental Design

Submitting Organization

### DataSpeaks, Inc.

Curtis A. Bagne, Ph.D.

2971 Vineyards Drive

Troy, Michigan 48098

(248) 952-1968

[cbagne@DataSpeaks.com](mailto:cbagne@DataSpeaks.com)

**Abstract:** At least eight societal challenges involving medical diagnoses, drug development, drug regulation, health information services, and the cost of healthcare can be addressed more effectively with new ideas, more rigorous science, and high-risk, high-reward research together with Dr. Hood's personalized (P4) approach to medicine. However, rapid progress is being impeded by six major technical hurdles that involve the types of measures used in medical diagnoses in combination with the use of parallel-group randomized controlled trial (RCT) designs to assess causality. In addition, excessive reliance on statistics to account for measurement error leads to the collection of essentially timeless cross-sectional data instead of time ordered data with more repeated measurements. These hurdles lead to four types of confounding in parallel-group RCT designs, which do *not* assess causality for individuals, thereby diminishing the value of their results for P4 medicine, public health, and public policy.

The proposed technical solution is to apply a computational algorithm to time ordered data to obtain reliable, valid, specific, detailed and comprehensive measures of interactions-over-time between and among independent and dependent variables. The resulting measures describe and help predict how individual living systems work over time – function internally, respond to their environments including treatments, and act as agents. Given such new phenotypic measures together with genetic information from each of many living systems, it will become more feasible to apply statistics to identify genomic markers or profiles that correlate with specific clinical phenomena such as susceptibility to particular diseases and responses to particular treatments. Then we will be better able to target “the right dose of the right drug to the right patient at the right time.”

However, these measures and the innovative experimental designs that these measures enable challenge current scientific standards of peer-review as well as standards for government regulation as by the FDA. Such standards are promulgated by the Consolidated Standards of Reporting Trials group and are largely supported by academics, publishers, and those who fund research world wide.

Accordingly, this white paper encourages NIST to organize, catalyze, and support high-risk, high-reward research aimed at further developing, validating, publishing, disseminating, applying and otherwise advancing the proposed measurement technology. Doing so would challenge the status quo as necessary to yield transformational results with respect to all eight societal challenges.

**Key words:** Measurement, Technology, Standards, Interactions, Personalized Medicine, Complex Systems

	Page
1. Introduction to Dr. Hood's Concept of P4 Medicine and Overview of Technical Solution	2
2. Personalized (P4) Medicine – A Critical National Need that Addresses Eight Societal Challenges	3
2.1. Six Key Specific Technical Hurdles to Personalized (P4) Medicine	4
3. The Need for Government Support – Scientific Standards for Peer-review and Government Regulation Must Change	8
4. The Potential Measurement and Technical Solution	11
5. Interested Stakeholders and Participants	14
6. How This Technical Solution Addresses the Eight Societal Challenges and Conclusions	14
7. References	16
8. Appendix: Elucidating Genotype/Phenotype Relationships for Personalized (P4) Medicine and Overview of Algorithm for Computing Benefit/Harm Scores	18

## CRITICAL NATIONAL NEED IDEA

### 1. Introduction to Dr. Hood's Concept of P4 Medicine and Overview of Technical Solution

Personalized medicine is the concept that managing a patient's health should be based on the individual patient's specific characteristics, including age, gender, height, weight, diet, environment, and now, increasingly, genetic testing. Although personalized medicine promises vast healthcare improvements in the future, this white paper identifies six key technical hurdles that stand in the way of rapid progress towards personalized medicine. These technical hurdles are *not* widely recognized and have *not* been effectively addressed.

This white paper introduces an innovative computational *measurement* algorithm that addresses all six of these key technical hurdles as a set to help researchers, clinicians and other stakeholders make dramatic progress in the development of genetically-based personalized medicine. This algorithm can be embodied as software.

Dr. Leroy Hood, President of the Institute for Systems Biology, helped lay the foundation for this white paper during his presentation at NIST on 9/24/07. He cast personalized medicine in the broader context of systems biology and P4 medicine. The title of Dr. Hood's presentation was "Systems Biology and Systems Medicine: Predictive, Personalized, Preventive and Participatory (P4) (1)". This white paper applies in Dr. Hood's context of systems biology and P4 medicine.

Among other things, the software introduced in this white paper facilitates *scientific assessments of causality over time for individuals*. Causality can now be assessed under a wide range of preferably experimental but also non-experimental conditions to an extent *never before achieved*. The rapid development of P4 medicine largely depends on improved capabilities for assessing causality over time for individuals. The capability of assessing causality and mining information and knowledge from time ordered data for individuals is crucial to speed progress towards P4 medicine that is data-driven, scientific, genetically-based and large scale. Large-scale P4 medicine applies systematically across many types of diseases and many types of treatments.

The software introduced here will help make it possible to build P4 medicine and systems biology on the foundation of the more reductionistic omic or parts list sciences to advance scientific understanding of how particular individuals *work over time*. Work is defined as how individual systems such as patients or cells:

- Function internally – how one or more independent or predictor variables that are internal to the individual interact, correlate or are associated over time with one or more internal dependent or predicted variables with all variables being time ordered, very preferably time series
- Respond to their environments including treatments – how one or more external independent variables such as drug dose interact over time with one or more dependent or health variables that are internal to, characteristics of, or behaviors of the individual
- Act as agents – how one or more independent variables characterizing the individual's behavior interact over time with one or more external dependent variables in the individual's environment.

How individual patients work over time is a critical aspect of their phenotypes that is distinct from but complementary to their genotypes and structures. Although an individual's cells can mutate, genotype is static and can be quite adequately assessed with essentially timeless cross-sectional data. Structures that develop and age can be assessed through trends in longitudinal data. In contrast, the *dynamics* of how individuals work are assessed best with time series data for both independent and dependent variables. This paper addresses the great paucity of scientific methods for processing time series data with both independent and dependent variables about one individual.

The following overview highlights one particular type of application of this software to illustrate how its capabilities can yield *transformational results* for much of drug development, drug regulation and healthcare. This example involves evaluations of how *patients respond* to drugs that can vary in dose over time as used to manage or control chronic health problems as assessed with dependent health variables that also can vary and fluctuate in level over time.

Now it is possible to design and conduct randomized controlled trials (RCT) that *measure and test the benefit/harm* of treatment in a manner that *has never been done before*. The software described in this paper would be used to compute benefit/harm scores from the time ordered or time series data for each patient. The benefit/harm scores for *each individual* patient can be:

## CRITICAL NATIONAL NEED IDEA

- Reliable – the reliability of measuring benefit/harm can be increased almost indefinitely for each patient by increasing the number of repeated determinations of dose, an independent variable time series, and measurements of health as dependent variable time series.
- Valid – the validity of measuring benefit/harm can be increased almost indefinitely by randomizing *two or more* different doses to additional periods of time for each patient. In other words, randomized experimental control would be exercised over time for each individual patient. Causality would be assessed at the level of each individual to enable *single group RCTs*.
- Detailed – to illustrate, beneficial and harmful treatment effects can be profiled across a multitude of health variables for each patient. Each health variable can be differentially weighted in accord with differences in clinical significance and patient preferences. Benefit/harm can be investigated as a function of dose for each health variable and across all health variables combined. Benefit/harm can be investigated as a function of any delay of treatment effects.
- Comprehensive – beneficial and harmful treatment effects can be summarized by a single score for each patient across a multitude of all health variables that can be monitored over time for that individual. Accordingly, *safety and effectiveness evaluations can be integrated* with respect to all such health variables and summarized by a single computed score for each patient.

Benefit/harm scores from two or more patients can be analyzed statistically to describe groups, test hypotheses, make inferences from samples to populations and, when patients are genotyped, to identify genetic predictors of differential responses and optimal minimal doses. However, causality would be assessed through the exercise of randomized experimental control and measurement of apparent benefit/harm *before* any statistical analyses of data from two or more patients.

Scientific assessments of causality for individual patients in clinical care would make much of clinical care more scientific, ethical, efficient, healthful, and cost-effective. Benefit/harm scoring also has the potential to foster individual responsibility and accountability.

Coordination of such causality assessments across patients would facilitate development of cumulative bodies of scientific knowledge directly from clinical practice and obviate some need for translational medicine that currently derives from the fact that clinical research and important aspects of clinical practice are based of two different types of evidence for treatment effects. To illustrate, clinical research tends to rely on group comparisons to assess causality. In contrast, clinicians often are expected to form subjective impressions about adverse drug reactions by assessing response to drug challenge, drug de-challenge, and drug re-challenge with two or more doses. Software can enable use of both of these two different types of evidence scientifically and in an integrated manner.

The software technology for doing all this was published in 1992 through a rather extraordinary application of *traditional peer-review* (2). However, this software is *not* in use today, apparently because it would disrupt the *status quo* with respect to standards for *traditional peer-review* for funding and publication as well as standards for government regulation as by the Food and Drug Administration (FDA). In contrast and currently, regulators and peer-reviewers demonstrate strong preference for parallel-group, placebo-controlled RCT designs that do *not* assess causality at the level of individual patients. In general, it *is* possible to assess causality over time when drugs are used to manage or control chronic health problems.

Current major scientific standards for peer-review and government regulation *must* change in order to speed progress towards personalized (P4) medicine and systems biology so that America can continue to lead and compete in the world.

### **2. Personalized (P4) Medicine – A Critical National Need that Addresses Eight Societal Challenges**

Personalized medicine is about “getting the right dose of the right drug to the right patient at the right time” (3).

According to administrative guidance, personalized medicine is a critical national need. The President’s Council of Advisors on Science and Technology (PCAST), a private sector advisory committee, issued a report in 2008 entitled “Priorities for Personalized Medicine” (4). This PCAST report addresses the “scientific opportunity and public health need represented by personalized medicine....” According to the

## CRITICAL NATIONAL NEED IDEA

Executive Summary of this report, the “current high level of interest in personalized medicine from a policy perspective is attributable not only to the promise of improved patient care and disease prevention, but also to the potential for personalized medicine to positively impact two other important trends – the increasing cost of health care and the decreasing rate of new medical product development.” This white paper is about “Priority Area 1: Technology and Tools” of the PCAST report.

In order to speed progress toward and harvest the fruits of P4 medicine, we need to address eight societal challenges with new ideas, more rigorous science, and high-risk, high-reward research. We must:

1. Develop diagnostic methods that are better because they are based on more reliable, valid and specific measures of how patients with chronic health problems work over time as living systems. “Work” was defined in the Introduction.
2. Develop a more innovative, efficient, productive and competitive pharmaceutical industry.
3. Develop a new system for regulating drugs that is more ethical, scientifically rigorous, protective of patient safety, and informative to decision-makers as well as speedier and less burdensome.
4. A health care system for patients with chronic health problems that integrates new gold standard methods for clinical practice with new gold standard methods for clinical research in a manner that obviates some need for translational medicine and speeds acquisition of cumulative bodies of scientific knowledge from both research and patient care.
5. Engineer and integrate a major new software application that can help drive medicine and healthcare into the information age by providing fundamentally new measurement-based information services required and demanded by clinicians and patients to provide quality and cost-effective patient care. (The term “killer app,” represented by spreadsheets and search, applies. However, “killer” has untoward meaning in the context of drug development and healthcare.)
6. Move beyond use of the current public health or reactive medicine version of EBM that is derived almost exclusively from group averages (measures of central tendency). We must advance to EBM that is based on P4 medicine and more often also assesses causality at the level of *individuals*.
7. Develop and provide a new generation of measurement-based information services that will empower individuals to take more responsibility for their own health and that of their loved ones.
8. Improve the cost-effectiveness of health care.

These eight societal challenges can be overcome. Overcoming these challenges will have transformational results based on scientific understanding. Overcoming these challenges will help America lead and compete in the world.

### **2.1. Six Key Specific Technical Hurdles to Personalized (P4) Medicine**

According to the PCAST report on personalized medicine “the limiting factor in clinical application of genomic information will be not the availability of patients’ genomes, but rather the lack of robust, clinically validated correlations between genomic markers or profiles and specific clinical phenomena such as susceptibility to disease or to the effects of a particular treatment” (5). This will be called the “correlation problem.”

Similarly, Dr. Christopher P. Austin, Senior Advisor to the Director for Translational Research at the National Human Genome Research Institute, described “the genome translation problem” in 2003 (6). Austin’s review mentions the “undulating course” of the “perceived value of the human genome sequence,” “ephemeral market success for companies that aimed to capitalize on the genome sequence,” and discusses “a potentially important role for chemical biology” in solving the genome translation problem. However broadly described, the correlation problem or the genome translation problem is the major technical hurdle to genetically-based personalized (P4) medicine.

There have been many noteworthy successes in solving the correlation problem or the genome translation problem. These successes generally used chemical or molecular approaches to solve the problem for particular diseases such as specific types of cancer or treatments such as warfarin. Nevertheless, it is still proving to be difficult to address the problem systematically and in large scale. It is still proving difficult to derive biologic insights and therapeutic benefits from the success of the Human Genome Project, other sequencing accomplishments, genotyping, and the omic sciences generally. It is still difficult to go from genotype to phenotype.

## CRITICAL NATIONAL NEED IDEA

This correlation problem or the genome translation problem appears to have three major discernable parts.

1. Means to genotype patients and perform genetic tests on patients
2. Reliable, valid, specific, detailed and comprehensive measures of how patients work – which includes their functional phenotypes
3. Means to correlate or associate the genetic measures with the phenotypic measures.

Of these three, means to genotype patients and perform genetic tests on patients exist and are developing rapidly. Similarly, statistics appears to be well suited to associate or correlate genetic characteristics with other person or patient characteristics across large groups and populations of patients and people. In contrast, the missing ingredient for solving the correlation problem or the genotype translation problem appears to be reliable, valid, specific, detailed and comprehensive phenotypic *measures* of how individual patients work over time as “work” was defined in the Introduction. This white paper addresses this measurement problem by identifying and solving a nexus of six interrelated but discernable technical hurdles.

In preview, the first of these hurdles involves the non-specificity of many medical diagnoses. Hurdles 2 through 4 address fundamental problems that derive from using parallel-group, placebo-controlled RCT designs to evaluate treatments. In partial summary, using group comparisons and statistics to assess causality in heterogeneous groups of patients makes it difficult to target “the right dose of the right drug to the right patient...” This difficulty in turn derives from using statistics to account for measurement error while assessing causal relationships – Hurdle 5. Hurdle number 6 is excessive reliance on essentially timeless cross-sectional data, which includes pre- and post-treatment change scores, to assess causality.

These technical hurdles can be overcome by making better use of information in time ordered or time series data combined with software technology that measures how living systems work. The software described in this paper for measuring interactions-over-time can account for measurement error and causality for individuals without any averaging. This software, illustrated for one application in the Introduction, enables new RCT designs that assess causality for individuals instead of for groups. These six technical hurdles can be overcome as a set in the context of diagnosing patients with chronic and complex disorders as well as using drugs to manage or control such health problems.

1. Diagnoses of patients with chronic health problems are not sufficiently specific. This yields heterogeneous diagnostic and phenotypic groups of patients with labels such as hypertension, type 2 diabetes, fibromyalgia, depression and Alzheimer’s disease. One problem is that such diagnoses are point diagnoses – diagnoses made from information gathered primarily at a particular point in time such as a clinic visit. As such, point diagnoses tend *not* to account for differences in dynamic mechanisms of how different patients work over time. To illustrate, patients can be hypertensive or depressed because of at least several different mechanisms or combinations of mechanisms. Similar point diagnoses that are due to different mechanisms could be expected to require treatments that work by different mechanisms. Increasing diagnostic group homogeneity will make it easier to identify specific genetic predictor profiles of various complex multigenic disorders. It appears as if progress in personalized (P4) medicine for some forms of cancer derives at least in part from improvements in diagnostic specificity.

2. Current first generation RCT designs assess causality by using statistical measures of central tendency and variability that are obtained from parallel groups. RCTs that use parallel groups to assess causality do *not* provide valid measures of how individual patients respond to treatments. Lack of measures of treatment response that are valid for individual patients makes it difficult to identify specific genetic predictors or predictor profiles of differential responses. Lack of valid measures of how individual patients respond to treatments is a technical hurdle when and to the extent that uncontrolled individual differences such as patient genotype, patient history and concomitant conditions such as additional treatments or other environmental exposures affect patient responses to treatments.

3. Current first generation RCT designs test hypotheses defined directly on health variables or changes in health variables. It is highly desirable for any particular RCT to test only one primary hypothesis with one statistical test in order to protect the meaning of the specified level of statistical significance. Testing primary hypotheses defined directly on health variables has a set of undesirable consequences. Such undesirable consequences have been illustrated by widely publicized drug development and regulatory

## CRITICAL NATIONAL NEED IDEA

failures such as Vioxx (Merck, pain), Bextra (Pfizer, pain) and torcetrapib (Pfizer, cholesterol management).

If primary hypotheses are defined on health variables to evaluate efficacy for a particular valued indication for drug use, then drug safety is neglected. If primary hypotheses are defined on health variables for safety, then efficacy and effectiveness are neglected. In contrast, integrated safety and effectiveness assessments in humans can begin from the first human tested.

If primary hypotheses are defined on specific health variables such as measures of pain or specific dependent laboratory variables such as cholesterol levels and treatments affect more than one specific variable, then treatment evaluations are not comprehensive of multiple treatment effects. If primary hypotheses are defined in terms of more comprehensive health measures such as those obtained with generic health status or quality of life questionnaires, then decision makers do not get the detailed information that they need to target drug development and use to the right patients.

Furthermore, it is difficult to integrate results from different RCTs that test primary hypotheses defined in terms of different health variables used to evaluate safety and effectiveness. Important decisions such as whether or not a drug should be approved remain excessively subjective and consensual. Another problem is that defining primary hypotheses on the wrong health variables often results in drug development and regulatory delays and failures.

The designers of RCTs that chose or otherwise decide to test primary hypotheses defined directly on health variables often are forced to make decisions about what health variables to test before they have good information about what and how health variables are affected by treatments. Drug development becomes like an exercise in trying to use a funnel upside down. This makes it hard to target the right drug to the right patient.

Problems that derive from defining primary hypotheses directly in terms of health or other dependent variables have become a formidable hurdle to identifying profiles of genetic characteristics that can be used to target the right drug to the right patient.

4. Current first generation placebo controlled RCT designs typically randomize patients to groups defined by different doses of the same type of treatment. This includes placebo as dose zero. Such RCTs are not designed to provide reliable and valid measures that can be used to identify optimal minimal doses for individual patients. An optimal minimal dose for an individual is the smallest dose that provides the most benefit relative to harm across all health variables that can be assessed repeatedly for that individual. This technical hurdle is a problem to the extent that genotype and other factors might affect optimal minimal doses. Failure to have reliable and valid measures of optimal minimal doses makes it difficult to identify genetic predictors of optimal minimal doses. This in turn makes it difficult to target the right dose of a particular type of drug to the right patient.

5. Statistics is the primary computational tool that has been used to account for measurement error. Statistics as a discipline appears to be best suited for time invariant independent variables and measurements that are independent because they are obtained from different subjects. Accordingly, the primary means for increasing statistical power in parallel group RCTs has been to increase patient sample size requirements. This becomes a major technical hurdle for P4 medicine when there is need to access causal relationships as in RCTs. It becomes more difficult to obtain large patient sample sizes as diagnostic specificity increases. It appears as if reliance on statistics to account for measurement error has been a major technical barrier to assessing causality for individuals and making more use of time series data as in RCTs.

6. Many scientific investigations including many RCTs use essentially timeless cross-sectional data, which includes change scores. Cross-sectional data provide only a small amount of information per variable per individual. Information-poor cross-sectional data appears to be especially troublesome when there are large numbers of variables as with microarrays and there is need to assess how variables interact over time to understand how individuals work over time. In contrast, time ordered data or time series for *both* independent and dependent variables can provide orders of magnitude more information, compared to cross-sectional data. This additional information can be of great value when there is need to understand:

- Temporal dynamics, mechanisms, and how individual living systems work over time

## CRITICAL NATIONAL NEED IDEA

- Signaling networks
- Long term trends such as disease progression and spontaneous recovery and how to separate shorter term treatment effects from long term trends
- Temporal phenomena such as episodes of events as well as delays and persistencies as of treatment effects
- Other time dependent phenomena such as development, adaptation, and aging.

Furthermore, use of independent variable time series in combination with dependent variable time series is the basis for overcoming all five of the previous technical challenges. To illustrate, this makes it possible to:

- Improve diagnostic specificity of how individuals work over time
- Exercise randomized experimental control and assess causality over time for individuals
- Measure and test the benefit/harm of treatments, thereby reducing the dimensionality of treatment evaluation problems from one dimension for each health variable that might be affected by treatment in a beneficial or harmful manner to one dimension of overall benefit/harm
- Identify optimal minimal doses starting at the level of individual patients
- Increase the reliability of treatment evaluations for individual patients thereby increasing statistical power when there is need to make statistical inferences from groups to populations. Statistical power *can* be increased with established measures of health that have limited reliability without increasing the number of subjects and without averaging any of the data used to assess causality.

Time series data make it possible to compute more reliable, valid, specific, detailed and comprehensive measurements of how individual living systems work over time. This in turn will make it easier to identify “robust, clinically validated correlations between genomic markers or profiles and specific clinical phenomena such as susceptibility to disease or to the effects of a particular treatment” as described in the PCAST report on personalized medicine. This can make it easier to speed progress toward solving the genome translation problem and to translate genotypes into phenotypes.

Technical hurdles have created economic hurdles. This set of six technical hurdles accounts for much of the high cost of drug development. A commonly used figure for the cost of developing a single marketed drug is 802 million in 2000 dollars (7). More recent estimates are higher. Much of this cost appears to derive from trying to develop specific drugs for broad classes of heterogeneous patients in order to develop blockbuster products with annual sales of over \$1 billion in order to pay for the high costs of developing drugs with conventional measures and inefficient experimental designs.

The six technical hurdles identified in this section are *root causes* for many problems that society is having with drug development, drug regulation and healthcare. These root causes are based on fundamental technical limitations of conventional measures and experimental designs. However, these six technical hurdles are seldom recognized and have never been adequately addressed.

In contrast, factors such as drug utilization review, generic competition, managed care organization bidding, therapeutic substitution, and drug reimportation often are identified and presented as being major sources of difficulty for the pharmaceutical industry. In addition, there is increasing pressure for the pharmaceutical industry to deliver high-value therapeutic agents. However, all these appear to be reasonable economic conditions and expectations – not root causes for difficulty.

Much of this section can be summarized briefly – time matters. Cross-sectional data and change scores are old standbys. However, these simply do not provide enough of the right kind of information to understand how living systems work over time. Without such information, it has been difficult to identify genetic and other predictors of diseases, responses to different types of treatment and optimal minimal doses. This has been and continues to be a major impediment to P4 medicine.

All six technical hurdles are actionable now. However, action on such observations requires fundamental innovation in prevailing traditional scientific standards for peer-review and government regulation. Government action is needed now.

### 3. The Need for Government Support – Scientific Standards for Peer-review and Government Regulation Must Change

The PCAST report on personalized medicine refers to problems that affect policy such as “the increasing cost of health care and the decreasing rate of new medical product development.” Related problems involve drug induced patient morbidity and mortality, product failure, drug safety problems, legal liability, job losses, high and growing costs of drug development, economic disruption, and lost productivity that threaten loss of American leadership and competitiveness in the world. Data and other evidence involving these problems are striking, abundant, and well known. This evidence will not be detailed here.

In contrast, the root causes of these problems, which were identified and discussed in the previous section, apparently are not being recognized and addressed in accord with their importance. Instead, it appears as if current scientific standards for peer-review and government regulation are perpetuating and sustaining these problems. In general, these problems involving drug development, drug regulation and healthcare derive from excessive reliance on cross-sectional data and statistics to investigate causality and to account for measurement error.

This paper is about innovative measurement technology that can speed progress towards P4 medicine and help overcome the eight societal challenges identified in section 2. However, this measurement technology is disruptive. To illustrate, the FDA is charged with helping to assure that drugs are safe and effective. The FDA could require pharmaceutical companies to measure and test the benefit/harm of treatments used to manage or control chronic health problems. However, conventional RCT designs do *not* measure and test the benefit/harm of treatments. Instead, they typically perform statistical tests on health variables; changes in health variables; or dependent variables that are predictive of sentinel health events such as heart attack, stroke or death. As mentioned in the Introduction, the basic technology for measuring and testing the benefit/harm of pharmacotherapy was demonstrated and published in 1992.

Conventional parallel-group, placebo-controlled RCT designs are *not* beyond reproach. To illustrate, epidemiologists often address confounding. For example, it would *not* be acceptable to attribute lung cancer to vinyl chloride exposure without accounting for smoking history, a known risk factor for lung cancer. Confounding or mixing up the effects of vinyl chloride and smoking history might lead to erroneous conclusions. Science that confounds is *not* rigorous science.

RCTs are not always feasible or ethically acceptable as with the lung cancer example. However, RCTs are feasible and required in drug development and regulation. These RCTs do *not* commit the *particular type* of confounding that was just illustrated for lung cancer because RCTs exercise randomized experimental control over independent variables.

However, conventional RCTs do commit four other types of confounding listed below. These four types of confounding derive from the fact that conventional RCTs exercise randomized experimental control *across individuals* in groups with cross-sectional data. These four types of confounding can be avoided when drugs are developed and used to manage and control chronic health problems because randomized experimental control *can* be exercised instead *over time with time ordered data about particular individuals*. This difference makes a difference for P4 medicine.

The convention of assessing causality with randomized parallel groups has strong support from the Consolidated Standards of Reporting Trials (CONSORT) group (8). The CONSORT group standards are largely supported by academics, publishers, and those who fund research world wide. These standards are “aimed at first reports of two-group parallel designs.” The CONSORT group does focus on preparing and reporting for RCTs. Apparently, it is presumed that “two-group parallel designs” represent the best that science can offer.

However, conventional parallel-group RCT designs lead to four types of confounding that are different from the type of confounding that was illustrated above for lung cancer. These four types of confounding derive most directly from three of the six key technical hurdles identified in section 2.1.

1. RCT designs that randomize patients to different groups to assess causality as described for technical Hurdle 2 *confound individuality with measurement error*. In other words, such designs relegate individual differences, including differences that derive from genetic heterogeneity, into the error terms of statistical models. Genetic differences lead to noise. Mixing up genotype with measurement error can be expected to be a problem whenever there might be reason to suspect

## CRITICAL NATIONAL NEED IDEA

that genotype might affect treatment response. Conventional group RCTs are *not* designed to account for individuality. This becomes a major barrier to P4 medicine.

2. Placebo-controlled, parallel-group RCT designs *confound true responders to active treatment with patients on active treatment that would have responded to placebo*. This yields heterogeneous groups of responders. Again, it is difficult to identify specific genetic predictors of heterogeneous phenotypic groups.
3. RCT designs that randomize patients to different groups defined by different types of treatment in essentially the same way that patients are randomized to different types of treatment *confound dose with type of treatment*. This leads to the problems and missed opportunities identified in technical Hurdle 4. For example, it becomes difficult to get doses right.
4. RCT designs that perform statistical tests on health variables or changes in health variables *confound treatment effects and how they are valued* as identified in technical Hurdle 3. For example, this is a major cause of drug safety problems.

All four types of confounding impede scientific investigations of how genotypic differences affect and predict phenotypic responses to treatments. Such confounding contributes to the correlation problem or the genome translation problem. Such confounding makes it difficult to target “the right dose of the right drug to the right patient...”

The 1992 publication mentioned in the introduction did not commit any of these four types of confounding. This publication demonstrated a solution, albeit with mock data.

All four types of confounding can be overcome when drugs are developed, used, and evaluated for chronic health problems. This solution to these four types of confounding and the problems involving drug development, drug regulation and healthcare that were mentioned above does involve *more* use of:

- Data (time series for both independent and dependent variables to assess causality instead of or in addition to cross-sectional data and group comparisons)
- Randomized experimental control exercised over time instead of or in addition to across individuals
- Software that runs on computers and thereby embodies operational, transparent, and objective scientific definitions that need to be and can be specified in advance when testing hypotheses
- Measurement (benefit/harm scores and other measures of interaction-over-time).

This paper *is* about a scientific and technical solution. The primary hurdles to P4 medicine are no longer scientific or technical.

Speedy progress towards P4 medicine and success in meeting the eight societal challenges does appear to require that standards such as those advanced by the CONSORT group be challenged. The academic, industrial, venture capital, and economic development communities as well as the FDA have shown little interest in challenging the CONSORT group standards. It does appear as if government must mount this challenge as through high-risk, high-reward research that could be organized and supported through NIST and TIP.

The U.S. government has made considerable efforts to address certain problems involving drug development, drug regulation and healthcare that were referred to at the beginning of this section. However, none of these efforts appear to adequately recognize or address any of the six technical hurdles identified in section 2.1 or the four types of confounding in conventional RCT designs that derive from such hurdles and were identified above in this section. The following is a sampling of such government efforts, starting with the FDA.

FDA’s own Science Board said “FDA’s evaluation methods have remained largely unchanged over the last half century” (9). Current standards for RCT design largely predate modern genomics. It might *not* be possible to speed progress toward P4 medicine based on 21<sup>st</sup> Century genomics with 20<sup>th</sup> Century RCT designs.

The FDA has made considerable efforts to address certain problems. Some of its efforts derive from its own white paper: “Innovation or Stagnation – Challenge and Opportunity on the Critical Path to New Medical Products,” which was published in 2004 (10). In this document the FDA states: “Not enough applied scientific work has been done to create new tools to get fundamentally better answers about how the safety and effectiveness of new products can be demonstrated in faster time frames, with more

## CRITICAL NATIONAL NEED IDEA

certainty, and at lower costs.... A new product development toolkit – containing powerful new scientific and technical methods such as animal or computer-based predictive models, biomarkers for safety and effectiveness, and new clinical evaluation techniques – is urgently needed to improve predictability and efficiency along the critical path from laboratory concept to commercial product. We need superior product development science to address these challenges.” Although the FDA has made some progress since 2004, it apparently has yet to challenge the status quo with respect to RCT design in a manner that addresses specific technical hurdles and types of confounding identified in this paper.

The FDA established the Critical Path Initiative in response to its own white paper (11). The FDA released its Critical Path Opportunities List in March, 2006 (12). It has 76 projects in six broad topic areas. According to its press release, “FDA’s outreach efforts uncovered a consensus that the two most important areas for improving medical product development are biomarker development (Topic 1) and streamlining clinical trials (Topic 2).”

With respect to biomarkers, high or elevated blood pressure is a rather good and widely recognized predictor of increased risk of heart attack, stroke and death. Nevertheless, the three high profile product and drug development failures that were mentioned above (Vioxx, Bextra and torcetrapib) apparently involved, to some extent, increases in blood pressure. Unless new predictive biomarkers are used in better RCT designs, biomarkers alone are not apt to spur substantial progress along the critical path.

Topic 2 in FDA’s Opportunities List is “streamlining clinical trials.” Now the FDA is supporting the Clinical Trial Transformation Initiative (CTTI), which appears to be placing some effort on personalization (13). However, CTTI appears to focus on “streamlining.” This focus on streamlining does not appear to challenge the status quo or CONSORT group standards with respect to RCT design and reporting.

More recently, Janet Woodcock M.D., Director, Center for Drug Evaluation and Research of the FDA, presented “Personalized Medicine(s): Progress and Prospects” at a meeting entitled “Creating Value Through Personalized Medicine,” which was sponsored by the Personalized Medicine Coalition and held on January 27, 2009 (14). Dr. Woodcock identified many signs of progress, acknowledged that “biomedicine is now positioned to enter a new era,” noted that “the major mechanism for translating these insights into products that can provide health benefits is itself unhealthy,” and did note the requirement for “a change in culture of many participants.” However, Dr. Woodcock did not appear to address the six technical hurdles in section 2.1 or the four types of confounding in RCTs that were identified in this section.

Here are some initiatives outside the FDA that could affect drug development, drug regulation and healthcare. One of these has been the Small Business Innovative Research (SBIR) program (15). This program has funded much innovative research. However, it appears to use rather traditional peer-review as represented by the CONSORT group standards and does not appear to be intended to lead changes in scientific culture. Furthermore, the future of SBIR appears to be in jeopardy.

“The NIH Roadmap for Medical Research was launched in September, 2004 to address roadblocks to research and to transform the way biomedical research is conducted by overcoming specific hurdles... (16). However, the NIH Roadmap does not appear to address adequately any of the six technical hurdles identified in section 2.1.

The NIH has funded integrative biomedical informatics, which is highly relevant to this white paper (17). Data integration is important. However, integration alone might not be sufficient to solve the correlation problem or the genome translation problem as these problems were described at the beginning of section 2.1. In contrast, this might require more time series in combination with a measurement technology for converting data into scientific understanding of how living systems work. Data integration alone does not address the six technical hurdles or the four types of confounding.

On March 4, 2009 the NIH announced the NIH Challenge Grants in Health and Science Research as part of the American Recovery & Reinvestment Act of 2009 that would focus on “specific knowledge gaps, scientific opportunities, new technologies, data generation, or research methods...” (18). It is not yet clear if this program would be open to or most suitable for the type of measurement technology that is described by this paper.

Perhaps more generally than government involvement, the science community does not appear to recognize that there almost certainly are major problems with conventional parallel-group RCT designs.

## CRITICAL NATIONAL NEED IDEA

To illustrate, a recent **Science** magazine cover story entitled “Clinical Trials and Tribulations” shows a man in a white coat pushing a big pill up a big hill of paperwork in a manner reminiscent of the myth of Sisyphus (19). Paper work is a problem. However, the **Science** cover story does *not* seem to make any mention of fundamental methodological problems such as the four types of confounding in conventional RCT designs. This cover story does not challenge the status quo enough.

It appears as if standards such as the CONSORT group standards have created social inertia. The U.S. also has shown social inertia towards adopting the metric system despite considerable advantages to going metric. Social inertia might also affect the science community and make it difficult to recognize and appreciate innovative solutions that challenge traditional peer-review and deeply established government regulations.

There may be no good substitute for investigating individuals scientifically as proposed in this white paper, at least with respect to P4 medicine and the eight societal challenges identified in section 2.

U.S. competitiveness in the world is at risk. Key pieces of a solution are largely in place – the omic sciences, the internet, and substantial computing capacity. Furthermore, the 1992 publication and two issued software patents that address these eight societal challenges are public to the world.

#### 4. The Potential Measurement and Technical Solution

Measurement is of critical importance to science and commerce. Some measures such as weight, length, and time are fundamental. In contrast, derived measures such as density and electrical resistance are more abstract or conceptual. Derived measures are obtained from more fundamental measures as by computation.

Measures of interaction-over-time are an innovative class of derived measures that are computed from data obtained with other measures. Given the critical importance of measures, NIST has made great efforts to help assure that measures are clearly specified, precise, and standardized.

New measures have a history of spurring progress in science and commerce. This section introduces new measures that have the potential to speed progress in overcoming the eight societal challenges identified in section 2 by helping to enable and apply a P4 approach to medicine. More specifically, the measures introduced here are the “reliable, valid, specific, detailed and comprehensive measures of how patients work” that were identified in section 2.1 as being the “missing ingredient” for solving the correlation problem. These are the phenotypic measures that can be “correlated” with the genomic markers or profiles to solve the genome translation problem.

To illustrate, suppose that a distribution of benefit/harm scores for many patients has a multimodal distribution and that the patients have been genotyped. Statistics could be used to correlate patients in the different modes with potential genomic markers. (Benefit/harm scores were mentioned in the introduction and described more fully below.)

This white paper is about using time ordered data *collected over time* to measure *interactions-over-time* that describe and help predict how systems such as living systems *work over time*. Measures of interaction-over-time can be used to describe and help predict how systems work as “work” was defined in the Introduction.

The Appendix illustrates measurement of interactions-over-time in the context of P4 medicine together with an overview of the algorithm for computing such scores. The Figure in the Appendix is an extension of the figure, familiar in systems biology, which has nodes for elements and edges for interactions. Dr. Hood showed such a figure at NIST. This white paper is about methods and systems to apply computation to time ordered repeated measurements data about nodes to measure the edges or interactions.

Please note that the term “interaction-over-time” is being used to identify interactions over time that are being measured between independent or predicted variables and dependent or predicted variables. In contrast, “interaction” often is used to describe how two or more independent variables such as drugs or proteins interact. In contrast, this measurement system uses Boolean independent events defined across levels of two or more independent variables. Similarly, it defines Boolean dependent events to address phenomena such as syndromes.

## CRITICAL NATIONAL NEED IDEA

There appears to be *no other method or system* for measuring interactions-over-time that has the range of capabilities, applications, and uses as the system introduced here. These capabilities derive from what appears to be an important innovation. Information in each time ordered series of repeated measurements with more than two different values is converted into a set of dichotomous series before measuring interactions-over-time. Each dichotomous series can have a maximum of only two values, 0 and 1. It appears as if this conversion can be done without loss of information if the set of dichotomous series has a sufficient number of members. The resulting representation of information in the independent and dependent variables is digital.

According to Dr. Hood's presentation at NIST, digitalization of biology and medicine is a "revolution that will transform medicine even more than digitalization transformed information technology and communications." From this statement it is surmised that genomic information in DNA essentially is digital and that learning to use this digital information will transform biology and medicine.

This white paper takes the position that digitalization must go a step further. In order to capitalize on the digital information in DNA in a speedy manner, it might also be necessary to convert information in the environment of DNA into digital information – at least with respect to how individual systems work as "work" was defined in the Introduction. The inventor of this measurement technology and the first author of the 1992 publication, which was mentioned in the Introduction, happened upon this realization in a different context long before DNA was decoded.

The Introduction included this, which applies for each individual patient. "Benefit/harm can be investigated as a function of dose for each health variable and across all health variables combined. Benefit/harm can be investigated as a function of any delay of treatment effects." Such capabilities derive from digitalization. The current software can investigate up to eight so-called analysis parameters simultaneously for each pairwise interaction-over-time between one independent variable and one dependent variable. These are level of the independent variable and level of the dependent variable (both required) as well as two parameters to define episodes of independent events, two parameters to define episodes of dependent events, delay of action and persistence of action. Benefit/harm can be investigated as functions of any and all of them.

The proposed measurement system has many additional desirable features. Furthermore, the system is basically simple in that it is comprised of a set of old and well known computational tools used in an innovative manner. The system is fundamentally stochastic in that it uses probabilities. Moreover, the system is amenable to various extensions once the basic principles of the system are understood.

It is critically important to distinguish the innovative measures presented here from statistical measures of correlation or association. These innovative measures were invented specifically for time ordered data for two or more variables. In contrast, statistical measures tend to be best suited for cross-sectional data. Repeated measurements tend to exhibit auto-correlation or serial dependency in a manner that appears to be problematic for statistics. Although statistical measures have been applied to time series data, such measures appear to be severely limited at least in comparison to the capabilities of the current measurement system.

To distinguish further, statistics is taken to be best suited for describing groups and making inferences from samples of individuals to populations of individuals. In contrast, the proposed system is specifically designed for measuring interactions-over-time from data about only one individual. This measurement system can be used to investigate causality for individuals.

In contrast, statistics appears to have important limitations both for investigating causality for individuals and for heterogeneous groups such as groups made up of patients with different genotypes and histories. However, benefit/harm scores and other measures of interaction over time from two or more individuals appear to have good properties for statistical analysis of such scores.

Here is some additional information about the measurement solution to the genome translation problem. Measures of interaction-over-time are computed from time series data for at least one independent variable and one dependent variable about one individual. Types of individuals can vary hierarchically with respect to inclusiveness as from cells, organs, organ systems, people or patients, societies and economies, collective entities such as populations investigated as wholes without sampling of individuals that are part of the collective, and Earth's biosphere. Interactions can be investigated between sets of independent variables and sets of dependent variables to help address complexity.

## CRITICAL NATIONAL NEED IDEA

Time series are defined here to be comprised of two or more repeated measurements of all independent and dependent variables, generally all at the same times. In general, the measurement system introduced here would be inferior to change scores when there are only two repeated measurements. The advantages of this system tend to increase rapidly as investigators advance upward from three to more repeated measurements. The measurement system can accommodate at least hundreds of repeated measurements.

Repeated measurements should be equally spaced in time when there might be need to investigate temporal phenomena such as any episodes of events as well as any delays and/or persistencies of action of independent time series on dependent time series. Such capabilities are of particular value when there is need to investigate the temporal criterion of causal relationships with non-experimental data. The roles that time series play as independent and dependent variables can be switched for exploratory investigations and mining of non-experimental data for information. Temporal resolution can vary from fractions of a second to minutes, hours, days, months, years or more depending on data collection capabilities and the demands of the problem being investigated.

More specifically, measures of interaction-over-time measure the direction, amount and strength of evidence for interaction, association or correlation over time. Positive scores indicate that higher levels of an independent variable are associated with higher levels of a dependent variable. Negative scores are vice versa. Scores that quantify the amount of evidence for an interaction-over-time can increase in magnitude indefinitely with the number of repeated measurements.

Each score that measures the amount of interaction-over-time is one score from a distribution of potential scores, defined by the data in combination with a scoring protocol, which has a mean of 0 and a standard deviation of 1 unless 0 is the only potential score because of no variation in the independent time series, the dependent time series, or both. In general, amount type measures would be considered the primary scores in the measurement system. In contrast, measures of strength of interaction-over-time can be in the range of plus or minus 1 inclusive. Strength type measures are computed from amount type measures. Accordingly, the system can make distinctions such as having little evidence for a strong interaction-over-time, much evidence for a weak interaction-over-time, and much evidence for a strong interaction-over-time.

Benefit/harm scores, mentioned in the Introduction, are a subclass of measures of interaction-over-time that will become critical for much of drug development, drug regulation and healthcare as we move toward P4 medicine. Benefit/harm scores are measures of interaction-over-time for which the positive or negative signs are reversed if necessary in accord with the directionality of dependent or health variables. This is illustrated in the context of cholesterol management. If higher levels of LDL are considered to be unfavorable, then a negative measure of interaction-over-time between drug dose and LDL would become positive. In contrast, if higher levels of HDL were considered to be favorable, a positive interaction-over-time between drug dose and level of HDL would remain positive also to indicate benefit. The rule for negative benefit/harm scores is vice versa.

One way to help determine the directionality of dependent laboratory variables such as protein levels would be to measure interactions-over-time between laboratory variables and health variables. Another way is to investigate how measures of interaction-over-time that are obtained from many patients predict sentinel health events such as heart attack, stroke or death.

Benefit/harm is a *unitary dimension* that can be used to evaluate beneficial and harmful drug effects with respect to multitudes of health variables using the same metric. Thus the dimensionality of treatment evaluation problems can be reduced from one dimension for each health variable to just one dimension that unifies beneficial and harmful effects across many health variables. Safety evaluations and effectiveness evaluations can be integrated scientifically.

Benefit/harm scores can be computed using planned doses as randomized to different periods of time using intent-to-treat type analyses. However, benefit/harm scores also can be computed after substituting actual doses or blood levels of drug for planned doses in order to conduct exploratory analyses. This helps illustrate the capabilities of the measurement algorithm. The time series for both independent and dependent variables can vary and fluctuate in level over time in essentially any way. To further illustrate, this algorithm has been applied to economic time series.

## CRITICAL NATIONAL NEED IDEA

By extension from treatment evaluation, this scoring algorithm can be applied for problems involving medical diagnoses by measuring interactions-over-time between and among time series action variables such as laboratory measures of proteins or lipids, electrophysiological measures, measures of brain activity in particular brain regions, signs and symptoms of diseases, measures of mental and physical performance, etc.

*Action variables*, unlike genetic variables, are variables that can vary and fluctuate in level over time for individuals. A common methodological limitation of systems biology appears to occur when action variables and genetic variables are investigated in essentially the same way with data collected at a particular point in time as when proteins are fingerprinted.

This illustrates how measures of interaction-over-time can be used for diagnoses. For example, it is conceivable that type 2 diabetes could be diagnosed in terms of ordered and disordered interactions-over-time between and among insulin, glucose and other substances. It is conceivable that drugs and other substances could be used as probes together with measures of interaction-over-time to help make such diagnoses.

Measures of interaction-over-time appear to measure mechanisms of normal and disordered workings as “work” was defined in the Introduction. Here is an example of how measures of interaction-over-time can be used to investigate mechanisms of drug action. Suppose that one of the biological action variables in the Appendix Figure is level of gonadotropin releasing hormone and a second biological action variable is luteinizing hormone. The computational algorithm can be used to measure the interaction-over-time between these two hormones. Such interactions-over-time can be described in great detail with large arrays of various measures of interaction-over-time for each particular patient. Differences across patients might have diagnostic value.

Furthermore, such measures of interaction-over-time could be measured over a period of time when a patient is on a placebo and another period of time when a patient is on a potential GnRH agonist or antagonist. Then the measures of interaction-over-time could be compared to help obtain a detailed descriptive understanding of how drug might up or down regulate interactions-over-time.

Hood’s Slide 5 from his presentation at NIST (20), includes this: “Interactions between/among elements give rise to the system’s **Emergent properties**.” This white paper proposes that measures of interaction-over-time between and among action variables are measures of an emergent system property – *coordination of action*.

To illustrate, Tiger Woods’ prowess in golf seems to have little to do with how much or little action he can generate in his knees, hips, shoulders, arms, his head, and the head of his golf club. His prowess in golf is more apt to be related to his ability to coordinate such motions or actions in a manner that is repeatable in particular situations and adaptable to a wide range of situations as indicated by excellent golf scores.

This idea of coordination of motion in golf is being extended to coordination of action between and among many different types of elements such as levels of transcripts, proteins, metabolites, symptoms, emotions, electrophysiological variables, cognitive performance, behaviors, social role function, nutrients, allergens, pollutants and drug doses as illustrated by the Figure in the Appendix.

*Many chronic health disorders appear to be disorders of coordination*. If so, it should prove to be valuable to measure the coordination of action.

Levels of thousands of action variables at any particular time can tell little about coordination of action as an emergent system property or of how a system works. This suggests that it might be critical to collect more time series data and actually measure interactions-over-time by computation. Doing so has the potential to help build a new kind of *data-driven* 21<sup>st</sup> Century science on 20<sup>th</sup> Century science.

Measures of interaction-over-time have the potential to become fundamental units of investigation for systems sciences much as molecules and cells are fundamental units of analysis for reductionistic science. Conceivably, attention will expand from measuring nodes to measuring edges when the nodes are action variables.

One can not measure edges without measuring nodes. Measurement of edges will build upon all the spectacular achievements in measuring nodes.

## CRITICAL NATIONAL NEED IDEA

Conceivably, measurement of interactions-over-time will contribute to the development of systems biology, systems medicine and P4 medicine as described by Dr. Hood at NIST, from the more reductionistic approaches of molecular biology and molecular medicine.

Systems science and P4 medicine has the potential to be a better guide to public policy.

### 5. Interested Stakeholders and Participants

NIST has an opportunity through TIP to catalyze and support coordinated action between and among a great variety of stakeholders to further research, develop, validate, publish, teach, apply and otherwise advance this innovative measurement technology. Doing so has the potential to advance personalized (P4) medicine and systems science generally.

This document already has mentioned some entities or groups that might be interested in developing proposal submissions, participating in funding, or otherwise participating in efforts to overcome the eight societal challenges identified in section 2. These include the National Human Genome Research Institute, the Personalized Medicine Coalition and its various members (21), the Institute for Systems Biology headed by Dr. Hood, NIH, and the FDA. Health information technology organizations such as the Regenstrief Institute might want to participate (22).

University participation is required in part because of the need for academic intellectual leadership. As described in section 3, important applications of this technology would challenge traditional peer-review standards such as those supported by the CONSORT group. Universities supply many of these peer reviewers. Universities could compete to further research, develop, validate, publish, teach, apply and otherwise advance this innovative measurement technology.

All stakeholders in any of the eight societal challenges identified in section 2 could be expected to participate. These include information technology and software companies that might be interested in societal challenge Number 5 and have demonstrated interest in health as by Microsoft HealthVault (23), Google Health (24), IBM healthcare and life sciences (25), GE healthcare (26), and SAS Institute (27).

### 6. How This Technical Solution Addresses the Eight Societal Challenges and Conclusions

Unfortunately, this measurement technology does not appear to have any simple  $E = mc^2$  type summary. Perhaps the most effective summary for this technology now is as follows in the context of evaluating how patients *respond* to treatments using a P4 medicine approach to drug development, drug regulation and healthcare.

RCTs **can** measure and test the benefit/harm of treatments over time and across multitudes of health variables starting at the level of individual patients. RCTs **can** assess causality for individual patients, enabling *single group RCTs* and often obviating the need for group comparisons. RCTs **can** yield reliable, valid, detailed, and comprehensive measures of benefit/harm that then **can** be used to help identify any genetic and other predictors of differential responses and optimal minimal doses. The FDA **can** help assure that drugs are safe and effective by requiring pharmaceutical companies to measure and test the benefit/harm of treatments. **None** of this is being done now. We **can** learn to target "the right dose of the right drug to the right patient at the right time."

Conversely, many of the problems in drug development, drug regulation and healthcare today derive from the fact that researchers, RCTs, and clinical caregivers have yet to measure the benefit/harm of treatments developed and used to manage or control chronic health problems. The basics of a technology for measuring and testing benefit/harm were published in 1992.

Section 2 included a list of eight societal challenges that must be addressed. The following presents some for-instances about how the measurement technology presented in this paper can help overcome these challenges. Most of these for-instances apply for multiple challenges.

1. "Develop diagnostic methods that are better because they are based on more reliable, valid and specific measures of how patients with chronic health problems work over time as living systems."

Many of the elements or parts identified by the omic sciences, with the exception of those from genomics, are action variables that can vary and fluctuate in level over time for individuals. This includes levels of transcripts, proteins, lipids, and metabolites. In addition, many signs and symptoms of diseases as well as measures of physical and mental performance, social role

## CRITICAL NATIONAL NEED IDEA

performance, and quality of life also can vary and fluctuate in level over time. Measures of interaction-over-time computed for such action variables elucidate how systems work over time in manners that might be ordered or disordered. Such measures have potential diagnostic value. More specific diagnoses can better guide treatment selection and form more homogeneous groups for scientific investigations. Investigations of how drugs affect such measures of interaction-over-time have the potential to elucidate mechanisms of treatment effect. All this has the potential to make it easier to identify genetic and other predictors of disease and treatment response.

2. "Develop a more innovative, efficient, productive and competitive pharmaceutical industry."

A more innovative pharmaceutical industry would try measuring and testing the benefit/harm of treatments. It would use information from more repeated measurements as in time series to obtain more *reliable* benefit/harm scores and improve efficiency of drug development by increasing statistical power to speed progress and cut costs. It would obtain *valid* benefit/harm scores by randomizing different doses of a particular type of treatment to different periods of time for each patient. It would get *detailed* information about benefit/harm as a function of dose for each health variable and across all health variables starting at the level of each individual patient so that the industry could get doses right. It would obtain *comprehensive* benefit/harm scores by first detailing beneficial and harmful treatment effects across multitudes of health variables starting at the level of each patient before combining these scores into one overall benefit/harm score using differential weights that account for differences in clinical significance and patient preferences. It would use one such overall score for each patient to test primary hypotheses about overall benefit/harm in groups.

Then, given that it has reliable, valid, detailed and comprehensive benefit/harm scores for each patient in addition to results from genotyping, it could identify genetic and other predictors to help target drug development toward patients most apt to benefit and away from patients most apt to be harmed. It would monitor patient safety for each health variable and across all health variables for each patient to avoid major safety problem surprises. A pharmaceutical industry that targets drug development more effectively, gets doses right and does more to protect patient safety starting from the first time that drugs are used in humans would be a more productive pharmaceutical industry. A more productive and efficient pharmaceutical industry would be a more competitive industry.

3. "Develop a new system for regulating drugs that is more ethical, scientifically rigorous, protective of patient safety, and informative to decision-makers as well as speedier and less burdensome."

A more ethical regulatory system would *not* expect patients to be turned into subjects for the sake of research. A more ethical regulatory system would *not* expect the pharmaceutical industry to randomize patients to groups receiving only placebos, only potentially inadequate doses, or only potentially excessive or harmful doses when it is scientifically *disadvantageous* to conduct randomization in this particular manner. Instead, such a system could expect *increased* scientific rigor by demanding benefit/harm scores that are reliable, valid, detailed, and comprehensive for individual patients as just described for the pharmaceutical industry. A regulatory system more protective of patient safety would expect the pharmaceutical industry to monitor patient safety starting from the first human on drug and at the level of each individual patient also as described for the pharmaceutical industry. A more informative regulatory system would integrate safety and effectiveness evaluations in an objective and transparent manner and provide more detailed, comprehensive, and comprehensible information about treatment effects. A more informative regulatory system would expect results from comparative safety and effectiveness RCTs that randomize patients to groups defined by different types of treatment before doing most of the things described above. A regulatory system doing that which is described in point 4 below would be speedier, less burdensome, and get drugs to patients faster without compromising safety.

4. "A health care system for patients with chronic health problems that integrates new gold standard methods for clinical practice with new gold standard methods for clinical research in a manner that obviates some need for translational medicine and speeds acquisition of cumulative bodies of scientific knowledge from both research and patient care."

## CRITICAL NATIONAL NEED IDEA

A health care system that treats individual patients needs to be based on assessments of causality starting at the level of individual patients. This can be accomplished as described above for the pharmaceutical industry and its regulators. Doing so will obviate some need for translational medicine and speed acquisition of cumulative bodies of scientific knowledge from both research and patient care. Clinicians could provide better care and become more active participants in creating scientific knowledge. Budgets for conducting research could be more closely integrated with budgets for providing patient care. P4 medicine that improves individual health would improve group average or public health.

5. "Engineer and integrate a major new software application that can help drive medicine and healthcare into the information age by providing fundamentally new measurement-based information services required and demanded by clinicians and patients to provide quality and cost-effective patient care."

Is it ethical to prescribe drugs to patients without measuring the benefit/harm of those drugs when there is clinically significant uncertainty about their safety and effectiveness? Is it acceptable to spend scarce resources on drugs when there is substantial uncertainty about their cost-effectiveness? Should healthcare providers be held more accountable for the quality and cost-effectiveness of healthcare? Actual measurement of apparent benefit/harm with a new software application, starting at the level of individual patients, can help address such concerns.

However, much of the pressure for electronic medical records appears to be driven by concerns about improving the administrative business of healthcare – access to records, scheduling, ordering, billing, paying, etc. The administrative business of healthcare is important. However, it might be of more interest to administrators and payers than clinicians and patients. Some clinicians and patients have been known to resist electronic medical records.

The measurement technology described in part by this white paper can be used to measure the benefit/harm of treatments from data in suitable electronic medical records. Clinicians seeking to provide quality P4 medicine might need to monitor the benefit/harm of treatments by computation from data in order to make quality treatment evaluations and protect patient safety much as they need diagnostic tests and radiology to make quality diagnoses. Patients, payers and their attorneys might come to expect nothing less. If so, this could help drive medicine and healthcare into the information age.

This would be one application of a computational algorithm, embodied in software, with potential to improve health and wealth. It would need to be integrated with other software as for data collection and statistical analyses. All this needs to be made available through the internet. Some companies might be willing to compete for this opportunity.

6. "Move beyond use of the current public health or reactive medicine version of EBM that is derived almost exclusively from group averages (measures of central tendency). We must advance to EBM that is based on P4 medicine and more often also assesses causality at the level of *individuals*."

The measures and experimental designs introduced in this white paper helps make this possible as described for the other points.

7. "Develop and provide a new generation of measurement-based information services that will empower individuals to take more responsibility for their own health and that of their loved ones."

Health maintenance can become more difficult and expensive as people delegate responsibility to outside caregivers. Health information services that help gather data about individuals, process the data to measure apparent benefit/harm and assess causality for the individuals, and feed back the results to the individuals that provided the data, have the potential for people to learn about the health impacts of drugs, alternative and complementary therapies, nutrients, allergens and pollutants affect their own health and that of their loved ones. Packaging of some such materials could help do this under conditions of randomized experimental control. Such participatory medicine has the potential to guide and motivate corrective action. Conceivably, such results from many people could be accumulated and shared to generate hypotheses for more systematic and rigorous scientific investigations.

## CRITICAL NATIONAL NEED IDEA

### 8. "Improve the cost-effectiveness of health care."

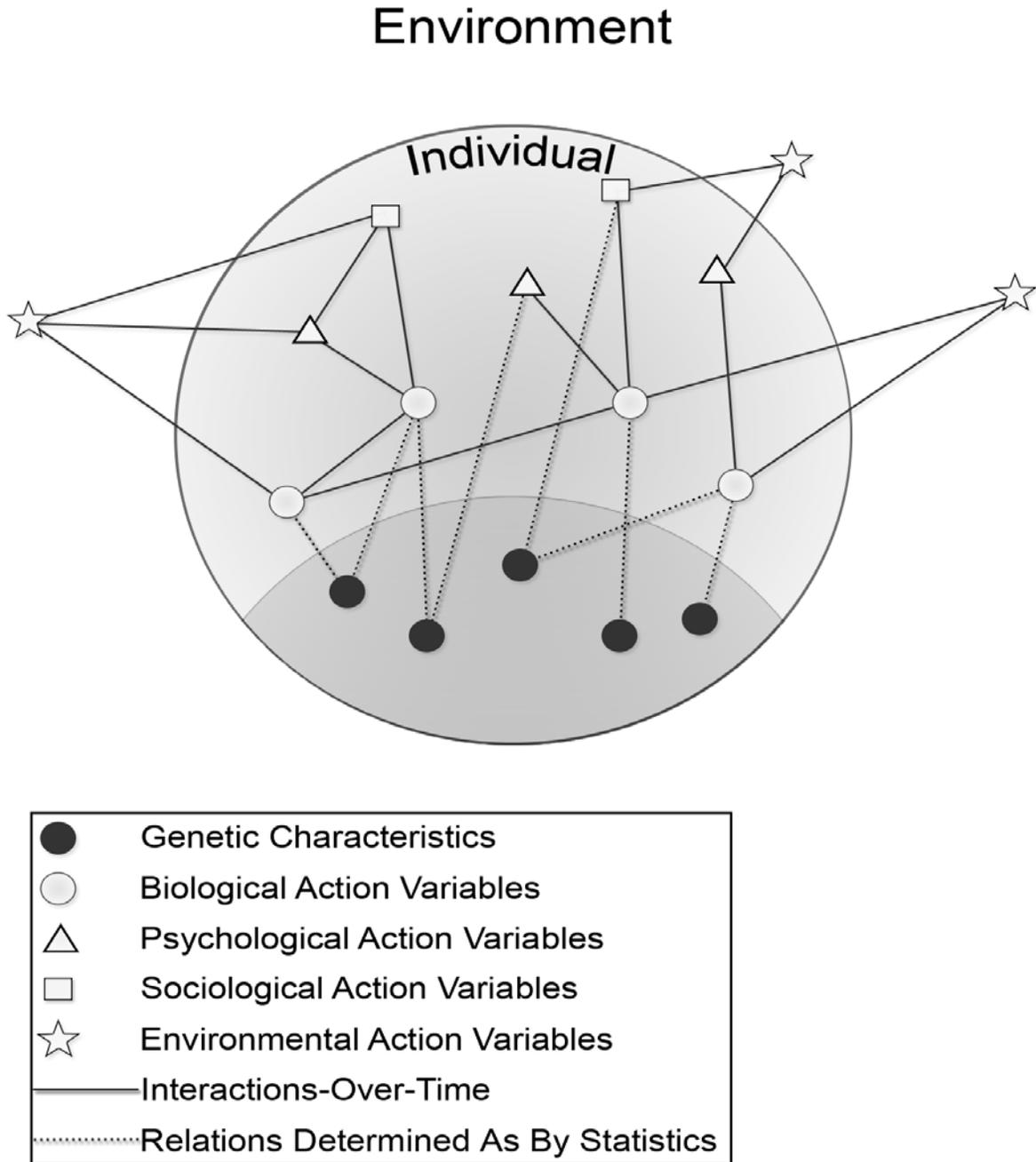
Please see all of the above. In addition, this measurement technology can help advance disease prevention through scientific understanding.

In conclusion, we must address these eight societal challenges by adopting Dr. Hood's concepts of systems biology and P4 medicine – medicine that is Predictive, Personalized, Preventive and Participatory. The innovative computational measurement technology described in this white paper could build on the spectacular achievements of the omic sciences. NIST can play a crucial role.

## 7. References

1. Here are Dr. Hood's slides: <http://www.itl.nist.gov/Healthcare/conf/presentations/LH%20NIST%209-24-07.pdf>.
2. Bagne CA, Lewis, RF. Evaluating the effects of drugs on behavior and quality of life: An alternative Strategy for clinical trials. *Journal of Consulting and Clinical Psychology*, 1992 (60), 225-239. (<http://dataspeaks.com/resources/APA-JCCP-1992-Vol60-No2-P225-239.pdf>)
3. This quotation is from Dr. Felix Frueh: [http://www.fda.gov/fdac/features/2005/605\\_genomics.html](http://www.fda.gov/fdac/features/2005/605_genomics.html).
4. Here is the link to the PCAST report about personalized medicine: [http://www.ostp.gov/galleries/PCAST/pcast\\_report\\_v2.pdf](http://www.ostp.gov/galleries/PCAST/pcast_report_v2.pdf).
5. *ibid*.
6. Austin CP: "The completed human genome: implications for chemical biology." *Current Opinion in Chemical Biology* 2003, 7:511-515.
7. DiMasi J, Hansen R, Grabowski H (2003). "The price of innovation: new estimates of drug development costs". *J Health Econ* 22 (2): 151–85.
8. See: <http://www.consort-statement.org/?o=1001>.
9. See: [http://www.fda.gov/ohrms/dockets/AC/07/briefing/2007-4329b\\_02\\_01\\_FDA%20Report%20on%20Science%20and%20Technology.pdf](http://www.fda.gov/ohrms/dockets/AC/07/briefing/2007-4329b_02_01_FDA%20Report%20on%20Science%20and%20Technology.pdf).
10. See: <http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html>.
11. See: <http://www.fda.gov/oc/initiatives/criticalpath/>.
12. Here is the press release: <http://www.fda.gov/bbs/topics/news/2006/NEW01336.html>. Here is the Critical Path Opportunities Report: [http://www.fda.gov/oc/initiatives/criticalpath/reports/opp\\_report.pdf](http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_report.pdf).
13. More information about CTTI is available at these links: <http://www.fda.gov/oc/initiatives/criticalpath/clinicaltrials.html> and <https://www.trialstransformation.org/>.
14. See: <http://www.personalizedmedicinecoalition.org/>.
15. See: <http://www.sbir.gov/>.
16. See: <http://nihroadmap.nih.gov/>.
17. See: <http://www.med.umich.edu/opm/newspage/2005/biomedical.htm>.
18. See: <http://dpcpsi.nih.gov/osc/challengegrants/>.
19. *Science* 322 (5899): 10 October 2008.
20. Here are Dr. Hood's slides: <http://www.itl.nist.gov/Healthcare/conf/presentations/LH%20NIST%209-24-07.pdf>.
21. Here is the membership list for the Personalized Medicine Coalition: [http://www.personalizedmedicinecoalition.org/about/pmc\\_members.php](http://www.personalizedmedicinecoalition.org/about/pmc_members.php).
22. See: <http://www.regenstrief.org/>.
23. See: <http://www.healthvault.com/>.
24. See: <http://www.google.com/intl/en-US/health/tour/>.
25. See: <http://www-03.ibm.com/industries/healthcare/us/index.html>.
26. See: <http://www.gehealthcare.com/worldwide.html>.
27. See: <http://www.sas.com/>.

8. Appendix: Elucidating Genotype/Phenotype Relationships for Personalized (P4) Medicine and Overview of Algorithm for Computing Benefit/Harm Scores



## CRITICAL NATIONAL NEED IDEA

### Figure key:

- The Figure is an extension of Slide 5 that was presented by Dr. Hood at NIST.<sup>1</sup> Slide 5 presents “dynamic networks” using “nodes” to represent elements and “edges” to represent interactions.
- The big circle in the Figure represents the dividing line between that which is part of one individual system (the inner area of the circle) and that which is part of that individual's environment. An individual's behavior is considered to be an aspect of that individual. The individual's environment is outside the circle. An individual's environment can include other individuals.
- Although the Figure represents an individual person or patient as described here, various types of individuals can be represented by the circle in similar figures. As examples, the individual can be an individual cell, an individual organ, an individual organ system, an entire person or patient, or a group or population of individuals investigated as an entire whole. In the smaller extreme for living systems, almost everything appears to be in the environment of essentially digital DNA. Individuals can be nested within more inclusive individuals.
- The Figure represents five different types of nodes in two major classes that need to be investigated in two fundamentally different ways.
  - The first class of nodes, presented in the Figure by solid dots, represent genetic characteristics of an individual as indicated by the individual's DNA, genetic sequence or genotype. These nodes generally do not change over time for a particular individual and can be used as in forensics to help identify and distinguish individuals. The more darkly shaded area within the circle represents information in an individual's genomic information. The nodes in this area represent specific parts of this genetic information.
  - The second class of nodes, presented in the Figure by small open symbols, represent elements that are *action variables*. Action variables can vary and fluctuate in level over time. Information about action variables is captured by time series. The Figure represents four different types of action variables.
    - The small open circles in the Figure represent *biological action variables*. These include levels of molecules or substances such as transcripts, proteins and metabolites as well as action variables such as measures of electrophysiological activity.
    - The open triangles in the Figure represent *psychological action variables*. These include measures of reported symptoms such as pain, depression, and anxiety as well as measures of mental and physical performance.
    - The open squares in the Figure represent *sociological action variables* as from scales that measure social function and role performance. Together, biological, psychological and sociological action variables can be used to investigate biopsychosocial systems at and across different levels of analysis.
    - The open stars in the Figure represent *environmental action variables* such as levels of nutrients, allergens, and environmental pollutants as well as external stimuli. Environmental action variables include doses of drugs taken repeatedly.
- The solid straight lines in the Figure represent *interactions-over-time*. The heart of this white paper is an innovative computational algorithm that can be applied to time ordered repeated measurements data or time series for two or more action variable interactants to measure the positive or negative direction of evidence, the amount of evidence, and the strength of evidence for interactions-over-time. Measures of inter-action-over time describe and help predict how individuals *work* as “work” was defined in the Introduction.
- The dashed lines connecting nodes representing genetic characteristics to nodes representing action variables can *not* be measured with this technology because genetic characteristics are not action variables for individuals. However, measures of interaction-over-time can be used to elucidate what genetic differences mean in terms of how individuals work as illustrated with a multimodal distribution in section 4.
- Measures of interaction-over-time can be used to inform the development of mathematical models.
- Measures of interaction-over-time about two or more individuals can be analyzed statistically to describe groups, test hypotheses, make inferences from samples of individuals to populations, and to identify genetic predictors of phenotypic characteristics as measured with this technology.

## CRITICAL NATIONAL NEED IDEA

### Overview of Algorithm for Computing Benefit/Harm Scores – Eight Main Steps

The basic procedure for computing benefit/harm scores with software will be summarized in eight main steps with references to additional information.

1. Enter time series data for both the independent action variable, drug dose, and the dependent health variable.
2. Define a scoring protocol. A software system for measuring benefit/harm and other measures of interaction-over-time would include a number of options, typically menu driven. These include options for transforming and detrending the time series variables before measuring benefit/harm, identification of independent and dependent variables, selection of analysis parameters as described partially in section 4, the positive or negative directionality of dependent health variables, any differential weights for health variables that might differ in clinical significance and patient preferences, etc.
3. Digitalize both the independent and dependent action variable time series as mentioned in section 4. The 1992 publication<sup>ii</sup>, mentioned before, includes one particular simple way of converting time series with more than two different levels into a set of dichotomous series, apparently with no necessary loss of information.
4. Cross-classify each independent variable digital time series with each dependent variable digital time series to form an array of 2 x 2 tables. This process was clearly defined and illustrated in the 1992 publication.
5. Compute the raw benefit/harm score for each 2 x 2 table. This multi-step computation also was described and demonstrated in the 1992 publication.
6. Standardize each raw benefit/harm score so that it is one score from a distribution of potential scores that has a mean of 0 and a standard deviation of 1 unless 0 is the only potential score. This step was mentioned in section 4. It also is described and demonstrated in the 1992 publication.
7. Summarize each standardized benefit/harm score array as functions of all analysis parameters such as dose and identify the most extreme positive or negative benefit/harm score in the array as a single summary score. This too was demonstrated in the 1992 publication.
8. Each of the health variable specific benefit/harm scores in a profile of benefit/harm scores across two or more health variables can be differentially weighted before being averaged to compute an overall benefit/harm score for a particular patient as demonstrated in the 1992 publication.

Benefit/harm scores from two or more patients in one or more groups can be analyzed statistically as demonstrated in the 1992 publication and mentioned repeatedly above.

This overview provides only some of the most rudimentary aspects of the method and system for measuring interactions-over-time.

---

<sup>i</sup> See the following, which also was discussed in the Introduction:

<http://www.itl.nist.gov/Healthcare/conf/presentations/LH%20NIST%209-24-07.pdf>.

<sup>ii</sup> Bagne CA, Lewis, RF. Evaluating the effects of drugs on behavior and quality of life: An alternative Strategy for clinical trials. *Journal of Consulting and Clinical Psychology*, 1992 (60), 225-239.

(<http://dataspeaks.com/resources/APA-JCCP-1992-Vol60-No2-P225-239.pdf>)