Comment on NIST's Draft Document about Healthcare

The following is a solicited comment on the proposed topic *"Advanced Technologies for Proteomics, Data Integration and Analysis and Biomanufacturing for Personalized Medicine,"* which is available from NIST here.

1. Recommendation – NIST Needs to Open the Door to a Fundamentally New Kind of Ruler

The above named document can be improved in a manner that can help NIST make a dramatic step forward to advance its mission as the world's leader in measurement science. More specifically, this NIST document should explicitly *open the door* to the apparent need for a fundamentally new kind of ruler. This ruler is a major ***new category*** of measures that can dramatically accelerate progress towards basic and applied scientific understanding of complex systems. This includes applications of this new category of measures for proteomics, systems biology, healthcare, and personalized medicine.

More specifically, the above named document should open the door to a new category of computed or derived scientific measures that ***quantify how and to what extent the actions of interactants are coordinated over time***.[1] This measures would apply to many biological interactants.

In the context of this NIST document about healthcare and personalized medicine, ***coordination of action*** would be measured as an ***emergent*** property of ***individual*** living systems such as cells, tissues, organs, brains, and patients – all as whole individual complex systems living or *working* in their environments.

In the context of this NIST document, which focuses on proteomics, the primary interactants would be proteins. For interactants from the environment, this same NIST document mentions "the influence of environmental factors such as diet, exercise, exposure to toxins, and drug intake."

Interactants are *action variables* that can (i) vary and fluctuate in level over time for individuals and (ii) can be measured periodically to yield *periodic time-ordered data*.[2] Action variables are *distinguished from* germ-line genetic characteristics that help identify many living systems genetically.

The new category of measures called for in this recommendation appear to be the measures required to accelerate basic and applied scientific understanding of what genetic differences mean to help solve the genotype to phenotype problem. Values of the new measures would be obtained by applying a *computational algorithm to periodic time-ordered data for action variables*.

---

[1] Density, computed as weight per unit volume, is a simple example of a derived measure. Coordinated action can be measured by applying a computational algorithm to data collected over time to measure *interactions-over-time* for individuals. NIST's draft healthcare document appears to call for technologies to measure and monitor protein interactants in a three-dimensional "live tissue environment." This would be a huge step forward because the proteins would be localized in space thereby making it possible to measure and image coordinated action. The computational algorithm for measuring coordinated action uses probabilities to help account for stochasticity and measurement error.

[2] Periodicity of data is valuable because this makes it much easier to investigate temporal phenomena such as any episodes of events and any delays and/or persistencies in the actions of independent events on dependent events as well as the temporal criterion of causal relationships.

Measures of coordinated action can be used to describe how complex systems such as living biological systems ***work***.[3] "Work" has three major components:

1. *Internal function.* The term "function" applies when coordination of action involves two or more action variables that are internal to or characteristic of the behavior of the individual. At least one of these action variables would be designated to act as an independent or predictor variable and at least one to act as a dependent or predicted variable. (This component of this definition of "work" applies when proteins "function and interact" as with each other and with many other types of biomolecules.)
2. *Response* to environmental variables including treatments. "Response" applies when there is at least one independent action variable that is external to the individual and at least one action variable that is internal to or characteristic of the behavior of individual.
3. *Agency.* "Agency" applies when at least one independent action variable is characteristic of the behavior of the individual and at least one dependent action is external to the individual.[4]

Some portion of language to this effect should be added to material about "data integration and analysis" that appears in the current NIST document.

2. Background, Rationale, and Further Explanation of the Recommendation

NIST is world's foremost leader in measurement science. NIST has excelled in creating standard measures of time, length, and weight that have been crucial for advancing physical sciences and commerce.

Now NIST is reaching into complex systems. This includes biological science and applications of biological science such as healthcare, personalized medicine, and drug development. Biological sciences build on physical sciences such as physics and chemistry. To a large extent, much of the success of the omic sciences in identifying the so-called "parts lists" for biological systems can be seen as great achievements primarily of the physical sciences. This includes success in decoding the human genome. The current draft of NIST's healthcare document can be seen as a continuation and extension of this great tradition.

Now growing evidence suggests that much of the promise of decoding the human genome has been deferred. For example, the pharmaceutical industry is having great difficulty as illustrated by the drug development targeting and safety problems cited in NIST's draft healthcare document.

---

[3] "*Measurement Challenges to Innovation in the Biosciences: Critical Roles for NIST,*" which is reference 27 in NIST's draft healthcare document, refers to "how biological systems operate." It appears as if the terms "work" and "operate" are being used in similar ways. The recommendation in this comment identifies three specific aspects of "work" that can be addressed now with an operationally defined *measurement* solution.

[4] The current draft of NIST's healthcare document does not appear to recognize or distinguish the role of "agency" for living systems from "function" and "response." However, "agency" might be of interest to NIST as for measures to investigate human agency with respect to the environment as in global warming. Furthermore, Earth's biosphere is one large, unique, and inclusive individual that needs to be investigated scientifically as a whole much as individual patients often need to be investigated scientifically as unique and whole individuals in order to accelerate progress towards personalized medicine.

This comment is based on the hypothesis that the omics "parts lists" might be *necessary but not sufficient* to capitalize on the potential of genomics to improve health and commerce. Furthermore, this comment offers a specific recommendation with transformative potential to solve a root cause scientific problem that has deferred the promise of genomics. This problem is a lack of appropriate measures.

The recommendation calls for a new ruler – a *measurement solution*. Perhaps no one is better prepared than NIST to appreciate the potential of new measures to accelerate science, technology, and commerce.

2.1. Emergence

The recommendation would open the door to a new category of measures that quantify coordinated action as an ***emergent*** property of living systems. Recognition of the importance of emergent properties has good precedent at NIST. Witness Slide 5 of Dr. Leroy Hood's presentation at NIST, which is available here. A recent search of the term "emergent" on the NIST website yielded about 455 hits. The current draft of NIST's healthcare document does *not* use the word emergent. The recommendation about emergent properties is offered in the spirit of opening the door to a new line of "high-risk, high-reward" research in measurement science. The biggest risk in this research might be the risk of deviating from the status quo.

Proteomics is the first major thrust of NIST's draft document on healthcare. The current draft of this NIST document does an admirable job. It calls for "Real-time, non-invasive technologies for evaluating and measuring proteomics in live tissues…in their natural environment." Such technologies will be a valuable step forward. Perhaps most valuable of all, such technologies have the potential to gather periodic time-ordered and spatially localized data about protein levels in living tissues as cells and tissues *work* in their environments. Protein levels from blood can be time-ordered but are not well spatially localized.[5]

"Data integration and analysis" is the second major thrust of this NIST document about healthcare. This material can be improved dramatically by explicitly opening the door to the recommended major new category of measures that quantify emergent properties that characterize individual living systems.

The following *extreme* example is intended to help illustrate how opening the door to the recommended measures has potential to add value.

Technologies are being developed that can measure hundreds to thousands of proteins in a single drop of blood. Blood is being used as a window to the body. Protein fingerprints and disease signatures have potential to identify diseases in many organs and tissues of the body and to identify diseases early. This will be a huge advance towards preventive medicine.

One fundamental limitation of protein fingerprints, disease signatures, and biomarkers collected at a particular point in time will be illustrated as follows. Suppose that we collect a drop of blood both a minute before and a minute after death of a patient. Both

[5] The type of measures called for in the recommendation can be applied to proteins measured periodically in blood. Blood in *living* systems can be seen as an important signaling and coordination system. The recommended measures have been validated in the context of reproductive endocrinology for ewes.

drops are apt to have similar protein fingerprints or disease signatures. But clearly something should change upon death. What has changed? And can it be measured?[6]

The recommendation in this comment would open the door to the possibility that we also might need to measure ***coordinated action*** as a dynamic and *emergent* property of *living* systems. It would open the door to the possibility that coordinated action between and among biological action variables might be a ***sign of life***.

Coordinated action is something fundamentally different from substances that can be measured at a particular point in time as with a laboratory measure or a fingerprint. This is the difference between proteins being present and proteins actually working as "work" is defined in the recommendation. One can fingerprint a dead body. But there would be little if any coordinated action in a dead body. Death might be indicated by the end of coordinated action.

Emergent properties can be measured and do have potential to help distinguish dead bodies, fixed tissues, and laboratory samples from living systems.

Time matters. Coordination becomes evident over time. It appears as if coordination can only be measured in data collected over time. The same measures can be used to measure how individuals adapt over time. Adaptivity is another characteristic of living systems.

Data collected at a particular point in time can be likened to a *data snapshot* in which each action variable is represented by a pixel. An ordered series of data snapshots or frames can be likened to a *data movie*. A data movie can provide orders of magnitude more information than a data snapshot not only because there are more frames but also because the frames are ordered in time. To help appreciate the value of temporal order, imagine trying to understand what is happening in a movie of a football game in which the frames are presented in random order. We can use additional information from temporal order to help *understand individuals scientifically* as required for scientific personalized medicine.

Part of the significance of NIST's draft healthcare document is that it calls for measures of proteins in "a live tissue environment." A primary "data integration and analysis" challenge that needs to be more adequately addressed is how to make sense of the resulting data.[7]

Coordination of action is not all or none.[8] Coordination can vary in type and by degree. Conceivably, many chronic health disorders are ***disorders of coordinated action***. If so,

---

[6] If drops of blood were collected and proteins were measured periodically both before and after death it would be possible to apply the recommended measures to monitor coordinated action by processing the resulting data iteratively over the repeated measurements (over time). My hypothesis is that the measures of coordinated action would deteriorate (trend towards zero) more rapidly than the proteins themselves would deteriorate after death.

[7] Ideally, technologies for measuring proteins in living tissues would yield periodic time-ordered data about protein levels. There is a lack of good quantitative methods for making sense of such data. This lack of methods can be illustrated in a different context. For example, there are huge amounts of time series data about individuals such as economies and capital markets. However, much of the information in such data is going to waste for lack of quantitative methods to investigate these systems scientifically and to assess predictive relationships. The measures called for in the recommendation also apply to economies and capital markets. These measures have potential to reduce economic difficulty and to help resolve differences in opinion about economic policy.

[8] Golf can be seen as a game of coordinated action of *motions* of various parts of a golfer's body and club as the golfer addresses the ball. Golf scores – here taken to be indicators of success in coordinating action –

we need to measure coordinated action as emergent properties of individuals in order to accelerate investigations and diagnoses of many chronic health disorders.

Measures of coordinated action have potential to increase the value of most technologies that measure the parts in the omic "parts lists" as well as many other measurement and monitoring devices used in healthcare.

2.2. Individuality

Each person is unique to some degree. We know that genetic differences and patient histories make a difference with respect to disease occurrence and responses to treatments. Then why do we do base so much science of living systems on group averages that *wash out* the effects of individual differences?[9] One possible explanation is excessive reliance on statistics to account for measurement error.

This problem of excessive reliance on group averages and statistics can be highlighted in the context of current gold standard randomized controlled trial (RCT) designs. The current convention is to randomize patients to different treatment groups and to assess causality by comparing group averages with statistics.[10] This particular convention mixes up or *confounds individuality with measurement error*. Such confounding, together with at least three other types of confounding, are major root causes of why the promise of genomics has been deferred.

Personalized medicine includes targeting the right drug to the right patient at the right dose. Genetic predictors and aspects of personal history have great potential for such targeting. Conventional RCT designs make such targeting almost impossible. Here are four reasons why that are stated in ways that highlight *measurement* issues.

1. Conventional RCT designs often use before and after treatment change scores that often are based on only two repeated measurements of noisy dependent variables. Such change scores do *not* provide reliable measures of treatment effect. Without ***measures*** *of treatment effect that are* ***reliable*** *for* ***individual*** *patients*, it is hard to identify genetic predictors of differential treatment effect.[11]
2. Conventional RCTs designs do *not* distinguish true responders to active treatment from responders on active treatment that would have responded to placebo. In

---

differ by golfer and over time. The coordination of an excellent golfer is highly repeatable under a given set of circumstances and highly adaptable to different sets of circumstances. Motions can be measured with motion capture technology and coordination of motion can be measured with the recommended measures. This concept of coordinated action is here being generalized from motion to other types of action as in protein levels. The same measures of coordinated action can be applied to brains using data such as that obtained by fMRI.

[9] Francis Collins has expressed concern about such washing out of individual differences in the context of comparative effectiveness research: http://www.genomeweb.com/print/926529?emc=el&m=532203&l=2&v=940d034888.

[10] Current gold standard RCT designs would continue to apply when either the independent variable, the dependent variable, or both variables are *not* action variables. As examples, a man having a prostatectomy is not an independent action variable for an individual man because it can happen only once. Similarly, death is not a dependent action variable for an individual patient because it can happen only once. The recommended new category of measures generally would enable fundamentally new RCT designs when drugs are used to manage or control chronic health disorders. For example, the new category of measures would enable *single group* RCT designs.

[11] Statistics is valuable in part because it can help account for measurement error. The proposed measures also can account for measurement error by using information from larger numbers of repeated measurements.

other words, conventional RCT designs to *not* provide valid measures of true treatment effect. Without ***measures*** *of treatment effect that are* ***valid*** *for* ***individual*** *patients*, it is hard to identify genetic predictors of differential treatment effect.

3. Conventional RCT designs do *not* provide detailed measures of treatment effect. One of several ways to illustrate this problem is with dose. Patients often are randomized to different groups defined by different doses including placebo as zero dose. This convention makes it needlessly difficult to identify optimal doses for individual patients.[12] Without ***detailed measures*** *of treatment effect as a function of dose for* ***individual*** *patients that help identify optimal doses for individual patients*, it is hard to find genetic predictors of optimal dosing.
4. Conventional RCT designs perform statistical tests on health variables. It is widely recommended by statisticians that any particular RCT should test one primary hypothesis. These two points taken together essentially require that any particular RCT be limited to testing treatment effects with respect to only one primary health variable – usually a dependent variable for some particular drug indication. The problem here is that most drugs have many effects, both beneficial and harmful, across many dependent variables. Conventional RCT designs do *not* provide ***comprehensive measures*** *of treatment effect for* ***individual*** *patients*. Without comprehensive measures of treatment for individual patients, it is hard to identify genetic predictors of optimal treatment.

In short, conventional RCT designs essentially are *designed to fail for personalized medicine*. There is little reason to think that genotyping patients in RCTs with conventional designs will add much value.

The four problems just introduced can be solved with the recommended measures – a new ruler. Core components of this solution have been published here. This includes how to compute "*within-patient* patient indicators of treatment effect that are more reliable, valid, comprehensive, and detailed."

Randomized experimental control can be exercised *across individuals in groups*. Randomized experimental control also can be exercised *over time for individuals*.

These four problems appear to be in the province of the FDA. NIST's draft healthcare document says that in "2004, FDA published a document titled *Challenges and Opportunities on the Critical Path to New Medical Products* to provide FDA's analysis of the pipeline problem relating to the recent slowdown in innovative safe and effective medical therapies reaching patients." There appears to be little if any recognition of the four problems just identified at the referenced website, let alone any solutions that are identified and actionable now. NIST needs to help address these problems because they are not being addressed by the FDA.

NIST could play an invaluable role by helping FDA to recognize that many RCTs for drugs used to manage or control chronic health disorders should measure and test the *benefit/harm* of pharmacotherapy.

NIST's draft healthcare document mentions "translation research." The recommended new category of measures can help obviate the problem of translating clinical research

[12] This common convention also appears to be unethical when it is not necessary and when it might be counterproductive for science and commerce and harmful to patients.

results into clinical practice by integrating new gold standard methods for clinical research with new gold standard methods of clinical practice. New initiatives in electronic health records are helping to make this possible.

We need to know more about drug mechanisms of action. This too is largely a measurement problem with a measurement solution. One solution can be illustrated in the context of proteins. This proposed solution is to measure coordinated action as described above in the section on emergence both while patients are on drug and while patients are off drug. Differences between measures obtained under these two conditions would indicate how actions of proteins on other proteins might be either up or down-regulated by drug.

NIST's draft healthcare document is about personalized medicine. The recommended new category of measures can help enable personalized medicine. Personalized medicine that improves individual health will improve group average or public health.

2.3. "Data Analysis and Integration"

There is *urgent* need for NIST to strengthen this section, "*Issues with Data Analysis and Integration for Development of Personalized Treatment Strategies,*" which starts on page 10 of its draft healthcare document and to act accordingly.

Here is the opening sentence from this section. "Wellness and disease states result from a complex interaction between an individual's genetic makeup and the influence of environmental factors such as diet, exercise, exposure to toxins, and drug intake." And here is the last sentence from this section. "There is a need for high-risk high-reward research for developing information technology platforms that could potentially enable integration and analysis of disparate data sets not only for personalized medicine but for other application areas such as food safety and biodefense." Both sentences are well stated and true. This section also refers to "Federal investments" in "data resources" and technologies for measuring various interactants and potential interactants such as proteins, "environmental toxins" and "dietary intake" as well behavioral characteristics such as "physical activity of individuals."

At issue is whether or not these data are being or will be collected and archived in a manner that is well suited for integration. There is good reason and evidence to suggest that they are not.

Data collection designs for biology and healthcare *research* tend to be driven by the capabilities of currently dominant data analysis capabilities as described above for RCTs. More specifically, statistics is much better suited for the independent measurements of cross-sectional data than for the often serially dependent measurements of the repeated measurements in time ordered data. Statistics is *not* well suited for assessing causality over time for whole individuals such as patients, economies, or Earth's biosphere.

This data integration issue involves the classic distinction between extensive versus intensive data collection designs. Extensive designs tend to collect modest amounts of data from each of many individuals. Intensive designs tend to collect large amounts of data from particular individuals. (Data in health records of individual patients with chronic health disorders tends to be intensive.)

Some of the first complete human genome data, perhaps most notably data about the genome of Craig Venter, is an example of an innovative, useful, and productive intensive research design. One of the best ways to understand what genomic data means is to also collect intensive time ordered data on action variables so that it becomes possible to

measure the interactions-over-time that describe and help predict how living systems work as called for in the recommendation and as illustrated above for RCTs.

This issue of extensive versus intensive data collection designs can be illustrated as follows. Suppose that one has technology and money to measure 1,000 proteins in each of 1,000 drops of blood. One could collect one drop of blood from each of 1,000 individuals, 1,000 drops periodically at 1,000 times from one individual, or any other combination between. The best data collection strategy depends on the research question or hypothesis. The recommended measures help enable more intensive designs as often required for personalized medicine. Learning how one or just a few individuals *work* would be a huge step forward much as decoding the genomes of just a few individuals was a huge step forward.

Much data integration can be accomplished and applied specifically for personalized medicine by opening the door for the recommended measures. These measures describe and help predict how individuals work as "work" is defined in the recommendation. Obtaining such measures from each of many genotyped patients can enable statistical analyses that can help identify genetic predictors of diseases and responses to treatments.[13]

This section of the NIST document includes this. "Current approaches have been typically focused on studying a few factors at a time." Current RCT designs epitomize this problem when statistical tests are performed on health variables that represent particular indications. "Between 1995 and 2005, about 34 drugs were withdrawn from the market" because of safety problems. (Should we be surprised?) The recommended measures can help solve this problem because they provide a common metric for detecting and evaluating beneficial and harmful treatment effects across many dependent variables at one time. The recommended measures include additional features to account for more than "a few variables at a time" and to integrate scientific safety and effectiveness evaluations.

Many Federal and other investments in data collection technologies and data bases are apt to go largely to waste if NIST does not adequately and promptly address this issue of data integration and extensive versus intensive data collection designs. Many such databases collect a lot of data. However, they often are not designed to collect enough of the *right kind of data* for integration and personalized medicine.

2.4. Conclusion

Technology to address all of the problems identified in this comment does exist in the form of a key peer reviewed publication, issued U.S. software patents, demonstration software, and various demonstrations. This technology does need further research that would include and lead to additional development, validation, testing in practice, and dissemination as through publications. This technology does appear to be a prime candidate for "high-risk, high-reward research."

The solution to all of the problems highlighted in this comment is a ***measurement solution***. As such, this solution is in the province of NIST.

---

[13] NIST's healthcare document focuses on personalized medicine. There also often is need to personalize toxicology and nutrition.

NIST's draft healthcare document has great potential. It is apt to fall far short of this potential if NIST does not explicitly "open the door to a new category of computed or derived scientific measures that ***quantify how and to what extent the actions of interactants are coordinated over time***."

NIST does need to be explicit about this recommendation because the measurement solution does involve a fundamental change in scientific standards of excellence. It can be difficult to recognize fundamental changes as actionable possibilities. Established standards of excellence are understandably and appropriately difficult to change.

Here are three standards in potential need of change that have been touched upon above.

1. We diagnose chronic health disorders without explicitly measuring coordinated action when many chronic health disorders might well be disorders of coordinated action.
2. Current standards allow us to do drug development, drug regulation, and healthcare without explicitly and scientifically measuring the benefit/harm of pharmacotherapy for individual patients with chronic health disorders.
3. Current standards essentially demand that we to do RCTs using designs that confound individuality with measurement error when we know that genetic and other differences often make a difference. Such confounding is counterproductive for accelerating personalized medicine and for improving healthcare quality and cost.

All three of these examples involve chronic health disorders, which account for about 75% of U.S. annual healthcare expenditures of about $2.4 trillion. It might be worthwhile to open the door to the recommended measurement solution.

The proposed measurement solution would enable change towards more *rigorous, ethical, and informative science* for addressing the critical national needs and societal challenges already identified in the NIST document. Following the recommendation would help NIST achieve for basic and applied *biological* sciences what NIST has already achieved for basic and applied *physical* sciences.

NIST does need to open the door to a fundamentally new kind of ruler. As NIST quotes the President, we can "raise healthcare's quality and lower its cost."