MNI Open Science Symposium

John Wilbanks
“massive incrementalism”
Between 0.6 and 5 published papers per $100k funding
### Detailed Indexing Statistics: 1965-2017

MEDLINE® consists of completed citations indexed with MeSH® (Medical Subject Headings®).

<table>
<thead>
<tr>
<th>Fiscal Year (Oct. 1-Sep. 30)</th>
<th>Number of Journals Indexed in Index Medicus</th>
<th>Number of Journals in MEDLINE</th>
<th>Number of Citations in MEDLINE</th>
<th>Total Citations $^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>5,150</td>
<td>5,617$^3$</td>
<td>813,598</td>
<td>24,355,332</td>
</tr>
<tr>
<td>2016</td>
<td>5,136</td>
<td>5,623$^3$</td>
<td>869,666</td>
<td>23,531,948</td>
</tr>
<tr>
<td>2015</td>
<td>5,123</td>
<td>5,618$^3$</td>
<td>806,326</td>
<td>22,391,870</td>
</tr>
<tr>
<td>2014</td>
<td>5,118</td>
<td>5,647$^3$</td>
<td>765,850</td>
<td>21,582,742</td>
</tr>
<tr>
<td>2013</td>
<td>5,067</td>
<td>5,640$^3$</td>
<td>734,052</td>
<td>20,695,240</td>
</tr>
<tr>
<td>2012</td>
<td>5,025</td>
<td>5,633</td>
<td>760,903</td>
<td>19,974,272</td>
</tr>
<tr>
<td>2011</td>
<td>4,946</td>
<td>5,559</td>
<td>724,831</td>
<td>19,155,303</td>
</tr>
<tr>
<td>2010</td>
<td>4,866</td>
<td>5,484</td>
<td>699,420</td>
<td>18,340,055</td>
</tr>
<tr>
<td>2009</td>
<td>4,759</td>
<td>5,394</td>
<td>712,675</td>
<td>17,641,559</td>
</tr>
<tr>
<td>2008</td>
<td>4,660</td>
<td>5,319</td>
<td>671,904</td>
<td>16,888,840</td>
</tr>
<tr>
<td>2007</td>
<td>4,530</td>
<td>5,194</td>
<td>670,943</td>
<td>16,113,221$^2$</td>
</tr>
<tr>
<td>2006</td>
<td>4,416</td>
<td>5,020</td>
<td>623,089</td>
<td>14,103,589</td>
</tr>
<tr>
<td>2005</td>
<td>4,279</td>
<td>4,928</td>
<td>606,000</td>
<td>13,476,222</td>
</tr>
<tr>
<td>2004</td>
<td>4,189</td>
<td>4,839</td>
<td>571,000</td>
<td></td>
</tr>
</tbody>
</table>
propagation of uncertainty
within a statistical experiment...

### Example formulae

This table shows the variances of simple functions of the real variables $A, B$ with standard deviations $\sigma_A, \sigma_B$, **covariance** $\sigma_{AB}$ and exactly known real-valued constants $a, b$ (i.e., $\sigma_a = \sigma_b = 0$).

<table>
<thead>
<tr>
<th>Function $f$</th>
<th>Variance</th>
<th>Standard Deviation $\sigma_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f = aA$</td>
<td>$\sigma_f^2 = a^2 \sigma_A^2$</td>
<td>$\sigma_f =</td>
</tr>
<tr>
<td>$f = aA + bB$</td>
<td>$\sigma_f^2 = a^2 \sigma_A^2 + b^2 \sigma_B^2 + 2ab \sigma_{AB}$</td>
<td>$\sigma_f = \sqrt{a^2 \sigma_A^2 + b^2 \sigma_B^2 + 2ab \sigma_{AB}}$</td>
</tr>
<tr>
<td>$f = aA - bB$</td>
<td>$\sigma_f^2 = a^2 \sigma_A^2 + b^2 \sigma_B^2 - 2ab \sigma_{AB}$</td>
<td>$\sigma_f = \sqrt{a^2 \sigma_A^2 + b^2 \sigma_B^2 - 2ab \sigma_{AB}}$</td>
</tr>
<tr>
<td>$f = AB$</td>
<td>$\sigma_f^2 \approx f^2 \left[ \left( \frac{\sigma_A}{A} \right)^2 + \left( \frac{\sigma_B}{B} \right)^2 + 2 \frac{\sigma_{AB}}{AB} \right]$</td>
<td>$\sigma_f \approx</td>
</tr>
<tr>
<td>$f = \frac{A}{B}$</td>
<td>$\sigma_f^2 \approx f^2 \left[ \left( \frac{\sigma_A}{A} \right)^2 + \left( \frac{\sigma_B}{B} \right)^2 - 2 \frac{\sigma_{AB}}{AB} \right]$</td>
<td>$\sigma_f \approx</td>
</tr>
<tr>
<td>$f = aA^b$</td>
<td>$\sigma_f^2 \approx \left( abA^{b-1} \sigma_A \right)^2 = \left( \frac{f \sigma_A}{A} \right)^2$</td>
<td>$\sigma_f \approx</td>
</tr>
<tr>
<td>$f = a \ln(bA)$</td>
<td>$\sigma_f^2 \approx \left( \frac{a \sigma_A}{A} \right)^2$</td>
<td>$\sigma_f \approx \left</td>
</tr>
<tr>
<td>$f = a \log_{10}(bA)$</td>
<td>$\sigma_f^2 \approx \left( \frac{a \sigma_A}{A \ln(10)} \right)^2$</td>
<td>$\sigma_f \approx \left</td>
</tr>
<tr>
<td>$f_e = e^{hA}$</td>
<td>$\sigma_f^2 \approx \left( \frac{e^{hA} \sigma_A}{A} \right)^2$</td>
<td>$\sigma_f \approx \left</td>
</tr>
</tbody>
</table>
at the “claim” level...
Search Results  7,501 items

Summative Effects of Vascular Risk Factors on the Progression of Alzheimer Disease.
Full text

Genetic and epigenetic study of an Alzheimer's disease family with monozygotic triplets.
Full text

Associations Between Midlife but not Late-Life, Elevated Coronary Heart Disease Risk and Lower Cognitive Performance: Results From the Framingham Offspring Study.
Full text

APOE ε4, white matter hyperintensities, and cognition in Alzheimer and Lewy body dementia.
Full text

Apolipoprotein E ε4 allele effects on longitudinal cognitive trajectories are sex and age dependent.
Full text

APOE genetic variants and apoE, miR-107 and miR-650 levels in Alzheimer's disease.
Open Science at The Neuro

In 1934, Dr. Wilder Penfield had the unique vision to advance medicine through patient-centred science. Today, we continue to lead scientific innovation as the first academic institute to develop and adopt Open Science.
University of California boycotts publishing giant Elsevier over journal costs and open access

By Alex Fox, Jeffrey Brainard | Feb. 28, 2019, 7:00 PM

The mammoth University of California (UC) system announced today it will stop paying to subscribe to journals published by Elsevier, the world's largest scientific publisher, headquartered in Amsterdam. Talks to renew a collective contract broke down, the university said, because Elsevier refused to strike a package deal that would provide a break on subscriptions.
Availability of data, materials, code and protocols

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature Research journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript.

After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the journal. In cases where editors are unable to resolve a complaint, the journal may refer the matter to the authors' funding institution and/or publish a formal statement of correction, attached online to the publication, stating that readers have been unable to obtain necessary materials to replicate the findings.

See sections below for details on:

- reporting requirements
- availability of data
- availability of materials
- availability of computer code
- experimental protocols
- clinical trials
- further reading
The preregistration revolution

Brian A. Nosek1, Charles R. Ebersole1, Alexander C. DeHaven2, and David T. Mellor3

1Center for Open Science, Charlottesville, VA 22903; and 2Department of Psychology, University of Virginia, Charlottesville, VA 22904

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved August 28, 2017 (received for review June 15, 2017)

Progress in science relies in part on generating hypotheses with existing observations and testing hypotheses with new observations. This distinction between postdiction and prediction is appreciated conceptually but is not respected in practice. Mistaking generation of postdictions with testing of predictions reduces the credibility of research findings. However, ordinary biases in human reasoning, such as hindsight bias, make it hard to avoid this mistake. An effective solution is to define the research questions and analysis plan before observing the research outcomes—a process called preregistration. Preregistration distinguishes analyses and outcomes that result from predictions from those that result from postdictions. A variety of practical strategies are available to make the best possible use of preregistration in circumstances that fall short of the ideal application, such as when the data are preexisting. Services are now available for preregistration across all disciplines, facilitating a rapid increase in the practice. Widespread adoption of preregistration will increase distinctiveness between hypothesis generation and hypothesis testing and will improve the credibility of research findings.

methodology | open science | confirmatory analysis | exploratory analysis | preregistration

Progress in science is marked by reducing uncertainty about nature. Scientists generate models that may explain prior observations and predict future observations. Those models are approximations and simplifications of reality. Models are iteratively improved and replaced by reducing the amount of prediction error. As prediction error decreases, certainty about what overconfidence in post hoc explanations (postdictions) and inflate the likelihood of believing that there is evidence for a finding when there is not. Presenting postdictions as predictions can increase the attractiveness and publishability of findings by falsely reducing uncertainty. Ultimately, this decreases reproducibility (6–11).

Mental Constraints on Distinguishing Predictions and Postdictions

It is common for researchers to alternate between postdiction and prediction. Ideas are generated, and observed data modify those ideas. Over time and iteration, researchers develop understanding of the phenomenon under study. That understanding might result in a model, hypothesis, or theory. The dynamism of the research enterprise and limits of human reasoning make it easy to mistake postdiction as prediction. The problem with this is understood as post hoc theorizing or hypothesizing after the results are known (12). It is an example of circular reasoning—generating a hypothesis based on observing data, and then evaluating the validity of the hypothesis based on the same data. Hindsight bias, also known as the I-knew-it-all-along effect, is the tendency to see outcomes as more predictable after the fact compared with before they were observed (13, 14). With hindsight bias, the observer uses the data to generate an explanation, a postdiction, and simultaneously perceives that they would have anticipated that explanation in advance, a prediction. A common case is when the researcher’s prediction is vague so that many possible outcomes can be rationalized after the fact as supporting the prediction. For example, a biomedical researcher might predict that a
Vice President Joe Biden today released his vision for doubling progress against cancer over 5 years. It includes numerous policy recommendations and a laundry list of projects by the National Cancer Institute (NCI) and other federal agencies that would require additional funding.

Biden and his wife, Jill, have met with thousands of experts and patient advocates over the past year to craft the plan. It is expected to be a blueprint for how the US fights cancer in the next decade.
Expanding Public Access to the Results of Federally Funded Research

February 22, 2013 at 12:04 PM ET by Michael Stebbins

Summary: The Obama Administration is committed to the proposition that citizens deserve easy access to the results of research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than $100M in R&D expenditures to develop plans to make the results of federally funded research freely available to the public—generally within one year of publication.

The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than $100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. OSTP has been looking into this issue for some time, soliciting broad public input on multiple occasions and convening an interagency working group to develop the final policy. The final policy reflects substantial inputs from scientists and scientific organizations, publishers, members of Congress, and other
Better Science Together

At Sage Bionetworks, we believe that we can learn more by learning from each other. We partner with researchers, patients, and healthcare innovators to drive collaborative data-driven science to improve health. We advance biomedicine by making science more open, collaborative, and inclusive.
Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines

Article by: Andrea Schlicker, Garry Benani, Christine M Chresta, Gael McWalter, Alison Pritchard, Susie Weston, Sarah Runswick, Sara Davenport, Kerry Heathcote, Denis Alferez Castro, George Orphanides, Tim French and Lodewyk FA Wessels

A colorectal cancer classification system that associates cellular phenotype and responses to therapy


Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value

Article by: Laliella Mancia, Aurélien de Reynies, Alex Doual, Janick Selves, Marie Pierre Gaub, Laura Vescovo, Marie-Christine Etienne-Gimard, Renaud Schirr, Dominique Guenot, Mire Ayadi, Sylvain Kirzin, Maurice Chazal, Jean-François Filou, Daniel Banchiol, Anne Berger, Ana Maude, Erwan Penczech, François Pard, Dominique Elia, Yann Parc, Sylvaine Olschewski, Gérard Milian, Pierre Laurent-Puig, Valérie Boîte and Jean Fréau

Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer

Article by: Eva Budinska, Vlad Popovici, Sabine Tejpar, Giovanni D’Aria, Nicolas Lap青海省, Katarzyna Otylia Skora, Antonio Fabio Di Narzo, Pu Yan, John Graeme Hodgson, Scott Weinrich, Fred Bosman, Arnaud Roth and Mauro De Lorenzi
CRC subtyping consortium

Expert team data subtype

A → 1
B → 2
C → 3
D → 4
E → 5
F → 6

...
## CRC Subtyping Consortium

<table>
<thead>
<tr>
<th>CMS1</th>
<th>CMS2</th>
<th>CMS3</th>
<th>CMS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSI immune</td>
<td>Canonical</td>
<td>Metabolic</td>
<td>Mesenchymal</td>
</tr>
<tr>
<td>14%</td>
<td>37%</td>
<td>13%</td>
<td>23%</td>
</tr>
</tbody>
</table>

- **CMS1 (MSI immune)**: 14% patients. Characterized by MSI, CIMP high, hypermutation.
- **CMS2 (Canonical)**: 37% patients. Characterized by SCNA high.
- **CMS3 (Metabolic)**: 13% patients. Mixed MSI status, SCNA low, CIMP low.
- **CMS4 (Mesenchymal)**: 23% patients. SCNA high.

- **BRAF mutations** in CMS1 (MSI immune).
- **KRAS mutations** in CMS2 (Canonical).

- **Immune infiltration and activation** in CMS1 (MSI immune).
- **WNT and MYC activation** in CMS2 (Canonical).
- **Metabolic deregulation** in CMS3 (Metabolic).
- **Stromal infiltration, TGF-β activation, angiogenesis** in CMS4 (Mesenchymal).

- **Worse survival after relapse** in CMS1 (MSI immune).
- **Worse relapse-free and overall survival** in CMS4 (Mesenchymal).

CIMP: CpG island methylator phenotype; MSI, microsatellite instability; SCNA, somatic copy number alterations.
accessibility

degrees of freedom

closed + restricted

model to data

sandbox

collaboration

consensus model

open

closed
public / private partnership between NIH, 10 biopharmaceutical companies and several non-profit organizations
accelerating medicines partnership

coordinate sharing of early-phase target identification insights
accelerating medicines partnership

Consortium Space

AMP-AD Collaborative Workspace

Individual Partner Workspaces

- Broad/RUSH
- U FL/ISB/Mayo
- Sage
- Emory
- Other Partners

AMP-AD Synapse
Project Structure

Public space

AMP-AD Data Portal

Quarterly Depositions
RESEARCH TOOLS

These research tools have been generated by AMP-AD, MODEL-AD, M2OVE-AD, and Resilience-AD teams. Follow the links in the summaries below to learn how to obtain these resources for use in your own research.

Recombinant Adeno Associated Virus Vector (rAAV) Tools

Description: An extensive library of rAAV tools for the study of neurodegeneration.
Resource Contact: Todd Golde
Provided By: The UFL-Mayo-ISB AMP-AD Grant
More Information: This documentation provides an overview of promoters, capsids, fluorescent proteins, organelle markers and neurodegenerative disease relevant proteins that have been encoded in the available rAAV vectors and a list of the various rAAV constructs that have been generated. The documentation also describes any Material Transfer Agreements that are required to obtain the rAAV vectors.

Mouse Models: University of Florida

Description: Amyloid precursor protein (APP) and microtubule-associated protein tau (MAPT) mouse transgenic models.
Resource Contact: Todd Golde
Provided By: The UFL-Mayo-ISB AMP-AD Grant
More Information: See the table below for a list of models. RNAseq data and other results from the CRND8, PS1/APP, MAPT_P301L, and rTG4510 models are available through The TAUPPms study.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Model Name</th>
<th>Transgenes</th>
<th>Transgene Technology</th>
<th>Model Phenotypes</th>
<th>Contact/Access Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>APP Mouse</td>
<td>CRND8</td>
<td>APP695 SWE Indiana, Hamster Prion Promoter</td>
<td>Amyloid plaques by 2 months of age, Behr months</td>
<td>MTA from U. Toronto Contact Todd Golde UF or <a href="http://www.informatics.jax.org/allele/MGI:3589475">http://www.informatics.jax.org/allele/MGI:3589475</a></td>
<td></td>
</tr>
</tbody>
</table>
epistemically diverse findings

### Nominated Target List

AD researchers have nominated genes that may be good targets for new AD therapies. Select a gene from the list of nominated targets to view details and evidence.

1. **Learn more about Nominated Targets**

   - Search Gene Symbol or Ensembl ID

2. **Nominate a target**

### Table of Nominated Targets

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Nominations</th>
<th>Teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGF</td>
<td>3</td>
<td>MSSM, Emory, Broad-Rush-Columbia</td>
</tr>
<tr>
<td>C1QTNF4</td>
<td>2</td>
<td>Mayo-UFL-ISB, Broad-Rush-Columbia</td>
</tr>
<tr>
<td>INPP5D</td>
<td>2</td>
<td>Msyo-UFL-ISB, MSSM</td>
</tr>
<tr>
<td>ABCA7</td>
<td>1</td>
<td>Duke</td>
</tr>
</tbody>
</table>

*Screenshot*
triangulate to better bets

Alzheimer Centers for Discovery of New Medicines (U54 Clinical Trial Not Allowed)

U54 Specialized Center- Cooperative Agreements

New


RFA-AG-19-010

None

See Section III. 3. Additional Information on Eligibility.

93.866

This Funding Opportunity Announcement (FOA) invites U54 Cooperative Agreement applications aiming to establish multi-component Alzheimer Centers for the Discovery of New Medicines. The overarching purpose of this Centers program is to improve, diversify and reinvigorate the Alzheimer’s disease (AD) drug development pipeline by accelerating the characterization and experimental validation of next-generation therapeutic targets and integrating the targets into drug discovery and development.
Part 2. Full Text of Announcement

Section I. Funding Opportunity Description

Purpose
This Funding Opportunity Announcement (FOA) invites applications to establish multi-component Alzheimer Centers for Discovery of New Medicines. The overarching objective of this Centers program is to improve, diversify and reinvigorate the Alzheimer's disease (AD) drug development pipeline by accelerating the characterization and experimental validation of next generation therapeutic targets and integrating the targets into drug discovery campaigns. More specifically, each funded Center will 1) design, develop and disseminate tools that support target enabling packages (TEPs) for the experimental validation of novel next generation therapeutic targets, including those emanating from the NIA-funded target discovery programs such as AMP-AD/AMP-AD Wall of Targets, and 2) initiate early stage drug discovery campaigns against the enabled targets. To achieve these goals, it is expected that each Center will be staffed by a multi-disciplinary team of scientists with combined expertise in data science, computational biology, network biology, disease biology, structural genomics, biostatistics, assay development, medicinal chemistry, pharmacology, and clinical science. Central to this initiative is the open-access, rapid dissemination of data, methods, and computational and experimental tools generated by the Centers to all qualified researchers for their use in advancing AD drug discovery and AD disease biology.
New Open-AD Drug Discovery Center

SGC and collaborators to create open research tools for Alzheimer’s disease research funded by National Institute on Aging
field without built-in inertia
mPower helps decipher Parkinson’s disease.

The variability in Parkinson’s disease symptoms has left many questions unanswered. So the University of Rochester and Sage Bionetworks created the mPower app to precisely measure data such as dexterity, balance, memory, and gait. This information could help researchers better understand how various symptoms are connected to Parkinson’s disease. In turn, participants could start to recognize their own signs and symptoms.

First 6 months

- 16,585 participants consented
- 14,684 participants enrolled
- 9,520 agreed to ‘share broadly’
- 1,087 self reported PD diagnosis
dimensionality of lived experience

<table>
<thead>
<tr>
<th>Traditional Measures</th>
<th>First-order Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Taps</td>
<td>Number of taps, Mean tapping interval, Median tapping interval, Minimum tapping interval, maximum tapping interval, Standard deviation of tapping interval, Kurtosis of tapping interval, Interquartile range of tapping interval, Interquartile range of right button X, Range right button X, Standard deviation right button X, Interquartile range of left button X, Range left button X, Standard deviation left button X, Interquartile range of right button Y, Range right button Y, Standard deviation right button Y, Interquartile range of left button Y, Range left button Y, Standard deviation left button Y, Correlation X and Y, Skew tapping interval, No-button tapping frequency</td>
</tr>
</tbody>
</table>
invisible impacts made visible

Number of Taps
Mean Tapping Interval
Median Tapping Interval

Standard Deviation R Y
Range Right Y
Correlation X Y
the reality of personal health
benchmark as peer review
Accessibility

Degrees of Freedom

- Model to data
- Sandbox
- Collaboration
- Open
- Closed + Restricted
- Closed
Why a DNA data breach is much worse than a credit card leak

You can’t change your DNA

By Angela Chen | @changeling | Jun 6, 2018, 3:54pm EDT

Illustration by Alex Castro / The Verge
Genomic Justice for Native Americans: Impact of the Havasupai Case on Genetic Research

Nanibaa’ A. Garrison

1Center for Integration of Research on Genetics & Ethics, Stanford Center for Biomedical Ethics, Stanford, CA, USA

Abstract

In 2004, the Havasupai Tribe filed a lawsuit against the Arizona Board of Regents and Arizona State University (ASU) researchers upon discovering their DNA samples, initially collected for genetic studies on type 2 diabetes, had been used in several other genetic studies. The lawsuit reached a settlement in April 2010 that included monetary compensation and return of DNA samples to the Havasupai but left no legal precedent for researchers. Through semistructured interviews, institutional review board (IRB) chairs and human genetics researchers at US research institutions revealed their perspectives on the Havasupai lawsuit. For interviewees, the suit drew attention to indigenous concerns over genetic studies and increased their awareness of indigenous views. However, interviewees perceived no direct impact from the Havasupai case on their work; if they did, it was the perceived need to safeguard themselves by obtaining broad consent or shying away from research with indigenous communities altogether, raising important questions of justice for indigenous and minority participants. If researchers and IRBs do not change their practices in light of this case, these populations will likely continue to be excluded from a majority of research studies and left with less access to resources and potential benefit from genetic research participation.

Keywords
The Unbearable Requirement of Informed Consent

Ellen Wright Clayton, Vanderbilt University Medical Center

In the spirit of full disclosure, I have been a member of the Delphi panels discussed in this article (Beskow and Weinfurt 2019) since their inception and was one of the people who was recently interviewed for this study (not that I know which, if any, of the published comments were mine). I have thought about informed consent for a long time. I took Jay Katz’s year-long seminar on that topic during my last year in law school in 1978-1979 and was a research assistant for his book The Silent World of Doctor and Patient (Katz 2002). I have learned through experience that the kind of shared decision making he advocated is extremely difficult to achieve in both the clinic and research (Clayton 2006).

Nonetheless, reading Beskow and Weinfurt’s article in this issue has given me pause. Despite the authors’ best efforts in a prior study to create a truly effective consent process, it is not surprising that some of their participants could not demonstrate total comprehension on a quiz even after review and retesting (Beskow et al. 2015). After all, their subject population was representative of the U.S. population at large and hence was highly diverse. Most people from all walks of life had trouble getting a perfect score on their first try, and participants with low education, with low income, who lacked Internet access, and who were ethnic or racial minorities had more trouble. But these issues are complicated—even the highly educated participants in the Personal Genome Project have had a hard time passing that project’s test (Angrist 2009).

No, what was disquieting for me was reading the enormous diversity of the Delphi panelists’ responses—our responses—to the finding that many people could not thereby acknowledging them as potential participants. Maybe acceding to apparently poorly informed decisions to participate honors autonomy. Maybe we don’t see participants in a biobank as all that risky, especially in light of other safeguards, so it isn’t all that important to understand what is involved. Intriguingly, many of us would have allowed people who lacked complete understanding to enroll in Phase III oncology trials, a more risky endeavor. Maybe broad-based participation even by those who are less than fully informed is necessary for all of society to benefit from the fruits of research.

One might argue that the Delphi process and telephone interviews did not capture the full nuance of our thoughts, and that may be true to some extent. But one thing is quite clear from this study—for most of these respondents, truly informed decision making is not always ethically required for research participation. This finding stands in stark contrast to the literature over the last 50+ years that points to the pivotal role of informed consent. To make matters worse, the significantly increased disclosure requirements embodied in the recent changes to the Common Rule may well decrease comprehension even further.

This study has done a great service by highlighting some crucial questions. It is time to think more clearly about what we hope to achieve with informed consent. If something other than knowledgeable, well-considered agreement is required to participate, the goals of the process need to be specified. Perhaps different words that are more descriptive of what is at stake would encourage more honest discussion. Acknowledging the individual and his or her potential contribution to the research pro-
Welcome, All of Us Researchers.

The All of Us Research Program, part of the National Institutes of Health, is building one of the largest biomedical data resources of its kind. The All of Us Research Hub will store health data from one million or more diverse participants in the All of Us Research Program.

Here in the Research Hub, everyone can learn more about the types of data participants are providing and how approved researchers can use our data and tools to conduct studies that may speed up medical breakthroughs.

SEE THE DATA
Digital Mammography Dream challenge

Alternative models for sharing confidential biomedical data

Multiple Myeloma Dream Challenge

necessarily closed data
Traditional Data to Modeler

- Challenge participants (public)
  - Training data
  - Validation data
  - Challenge teams

  Model
  →
  Prediction

  Leaderboards & benchmarks

- Challenge cloud platform (private)
  - Predictions
  - Ground truth
  - Scoring

  Leaderboards & benchmarks

Model to Data

- Challenge participants (public)
  - Training data
  - Model

- Challenge cloud platform (private)
  - Model submissions
  - Validation data sets

  Leaderboards & benchmarks

- Contenized model

- Read-only access to directories
- No network access
- Time quota limits
- Size limit to log files
a. more massive incrementalism.
b. the stacks.
FIG. 1 - Centralized, Decentralized and Distributed Networks
FIG. 1 - Centralized, Decentralized and Distributed Networks
c. institutional change
FIG. 1 - Centralized, Decentralized and Distributed Networks
better science, together
thank you

@wilbanks
john.wilbanks@sagebase.org