Data sharing: The roles and policies of journals and funders

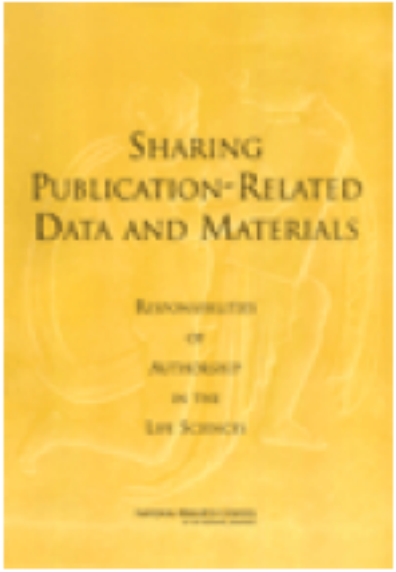
IOM Conference on Data Sharing
Oct. 4th and 5th, 2012

Steven Goodman, MD, MHS, PhD
Associate Dean for Clinical and Translational Research
Stanford University School of Medicine

Data sharing and reproducible research: The roles and policies of journals and funders

IOM Conference on Data Sharing
Oct. 4th and 5th, 2012

Steven Goodman, MD, MHS, PhD
Associate Editor, Annals of Internal Medicine
Editor, Clinical Trials



Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences

Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council

ISBN: 978-0-309-08859-6, 120 pages, 7 x 10, paperback (2003)

The uniform principle for sharing integral data and materials expeditiously (UPSIDE)

"Community standards for sharing publication-related data and materials should flow from the general principle that the publication of scientific information is intended to move science forward. More specifically, the act of publishing is a *quid pro quo* in which authors receive credit and acknowledgment in exchange for disclosure of their scientific findings. An author's obligation is not only to release data and materials to enable others to verify or replicate published findings (as journals already implicitly or explicitly require) but also to provide them in a form on which other scientists can build with further research. All members of the scientific community—whether working in academia, government, or a commercial enterprise—have equal responsibility for upholding community standards as participants in the publication system, and all should be equally able to derive benefits from it."

+ 5 additional principles

Purposes of data sharing

- To discover new things.
- To assure that study was correctly analyzed and interpreted.
  - To provide assurance that chain of scientific custody, from protocol→ conduct → data → analysis →published result is sound.

The New England Journal of Medicine

VOLUME 343 OCTOBER 12, 2000 NUMBER 15

LONG-TERM EFFECTS OF BUDESONIDE OR NEDOCROMIL IN CHILDREN WITH ASTHMA

THE CHILDHOOD ASTHMA MANAGEMENT PROGRAM RESEARCH GROUP*

ABSTRACT

Background Antiinflammatory therapies, such as inhaled corticosteroids or nedocromil, are recommended for children with asthma, although there is limited information on their long-term use.

Methods We randomly assigned 1041 children from 5 through 12 years of age with mild-to-moderate asthma to receive 200 μg of budesonide (311 children), 8 mg of nedocromil (312 children), or placebo (418 children) twice daily. We treated the participants

ASTHMA is a disease of chronic airway inflammation characterized by reversible airway obstruction and increased airway responsiveness.[1-3] Recent studies have demonstrated that asthma can be associated with impaired lung growth during childhood and with a progressive decline in pulmonary function in adulthood.[4-11] Clinical practice guidelines recommend antiinflammatory medication for the long-term control of per-

CAMP slides: Courtesy J. Tonascia

Data forms: CAMP Study

- No. of forms: 72
- No. of form revisions: 109
- No. of data entry forms: 41
- Total forms in the database: 293,761
  - Diary cards: 129,109
  - Other forms: 164,652

Database transactions

Transaction	No.
Add a form	297,649
Change a form	35,429
Delete a form	3,888
Total transactions	336,966
Total forms	293,761

CAMP: Manuscript

- No. of supporting manuscript tables and figures:
  - 73 and 9
- Number of revisions:
  - 40
- Published manuscript tables and figures:
  - 3 and 2
- Published, NEJM, October 2000.

The result

Attributed to Jon Claerbout –

- Most published papers are an *advertisement* for the research rather than a comprehensive report on the research itself.

Survival in Academy Award–Winning Actors and Actresses

Donald A. Redelmeier, MD, and Sheldon M. Singh, BSc

Background: Social status is an important predictor of poor health. Most studies of this issue have focused on the lower echelons of society.

Objective: To determine whether the increase in status from winning an academy award is associated with long-term mortality among actors and actresses.

Design: Retrospective cohort analysis.

Setting: Academy of Motion Picture Arts and Sciences.

Participants: All actors and actresses ever nominated for an academy award in a leading or a supporting role were identified ($n = 762$). For each, another cast member of the same sex who was in the same film and was born in the same era was identified ($n = 887$).

Measurements: Life expectancy and all-cause mortality rates.

Results: All 1649 performers were analyzed; the median duration of follow-up time from birth was 66 years, and 772 deaths occurred (primarily from ischemic heart disease and malignant disease). Life expectancy was 3.9 years longer for Academy Award winners than for other, less recognized performers (79.7 vs. 75.8 years; $P = 0.003$). This difference was equal to a 28% relative reduction in death rates (95% CI, 10% to 42%). Adjustment for birth year, sex, and ethnicity yielded similar results, as did adjustments for birth country, possible name change, age at release of first film, and total films in career. Additional wins were associated with a 22% relative reduction in death rates (CI, 5% to 35%), whereas additional films and additional nominations were not associated with a significant reduction in death rates.

Conclusion: The association of high status with increased longevity that prevails in the public also extends to celebrities, contributes to a large survival advantage, and is partially explained by factors related to success.

Ann Intern Med. 2001;134:955-962. www.annals.org

For author affiliations, current addresses, and contributions, see end of text.

See editorial comment on pp 1001-1003.

Oscar® explanation

- Personal chefs, trainers, nannies, support “ideals of lifestyle”.
- Constant scrutiny and entourage invested in movie stars’ success leads stars to “avoiding disgraceful behaviors and maintaining exemplary conduct.”

Annals of Internal Medicine | ACADEMIA AND CLINIC

Do Oscar Winners Live Longer than Less Successful Peers? A Reanalysis of the Evidence

Marie-Pierre Sylvestre, MSc; Ella Huszti, MSc; and James A. Hanley, PhD

In an article published in *Annals of Internal Medicine* in 2001, Redelmeier and Singh reported that Academy Award–winning actors and actresses lived almost 4 years longer than their less successful peers. However, the statistical method used to derive this statistically significant difference gave winners an unfair advantage because it credited an Oscar winner's years of life before winning toward survival subsequent to winning. When the authors of the current article reanalyzed the data using methods that avoided this "immortal time" bias, the survival advantage was closer to 1 year and was not statistically significant. The type of bias in Redelmeier and Singh's study is not limited to longevity comparisons of persons who reach different ranks within their profession; it can, and often does, occur in nonexperimental studies of life- or time-extending benefits of medical interventions. The current authors suggest ways in which researchers and readers may avoid and recognize this bias.

Ann Intern Med. 2006;145:361-363. www.annals.org
For author affiliations, see end of text.

Life expectancy difference = 0.7 years, P = 0.16

Final note

"Of course, readers and commentators should be doubly cautious whenever they encounter statistical results that seem too extreme to be true."

American Journal of Epidemiology

Vol. 163, No. 9
DOI: 10.1093/aje/kwj093
Advance Access publication March 1, 2006

Commentary

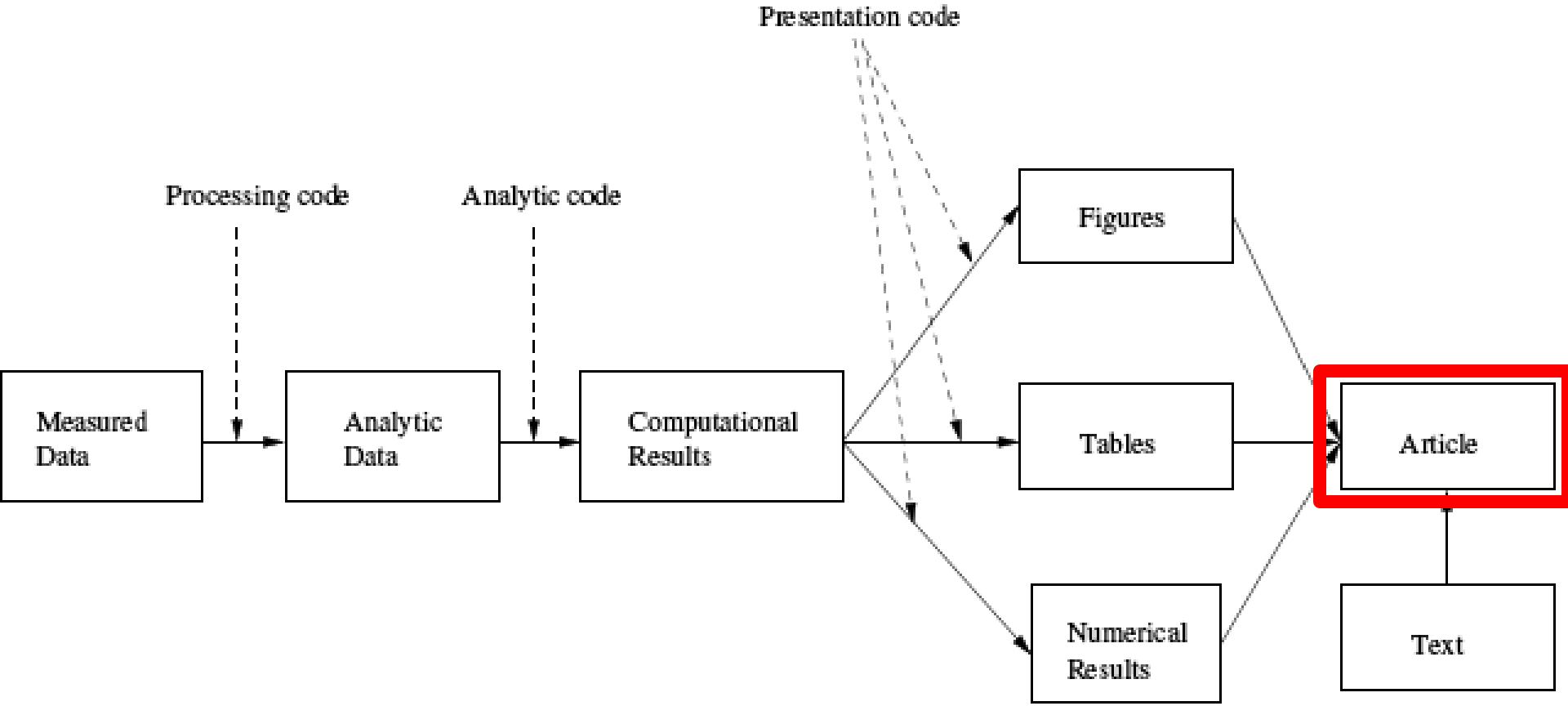
Reproducible Epidemiologic Research

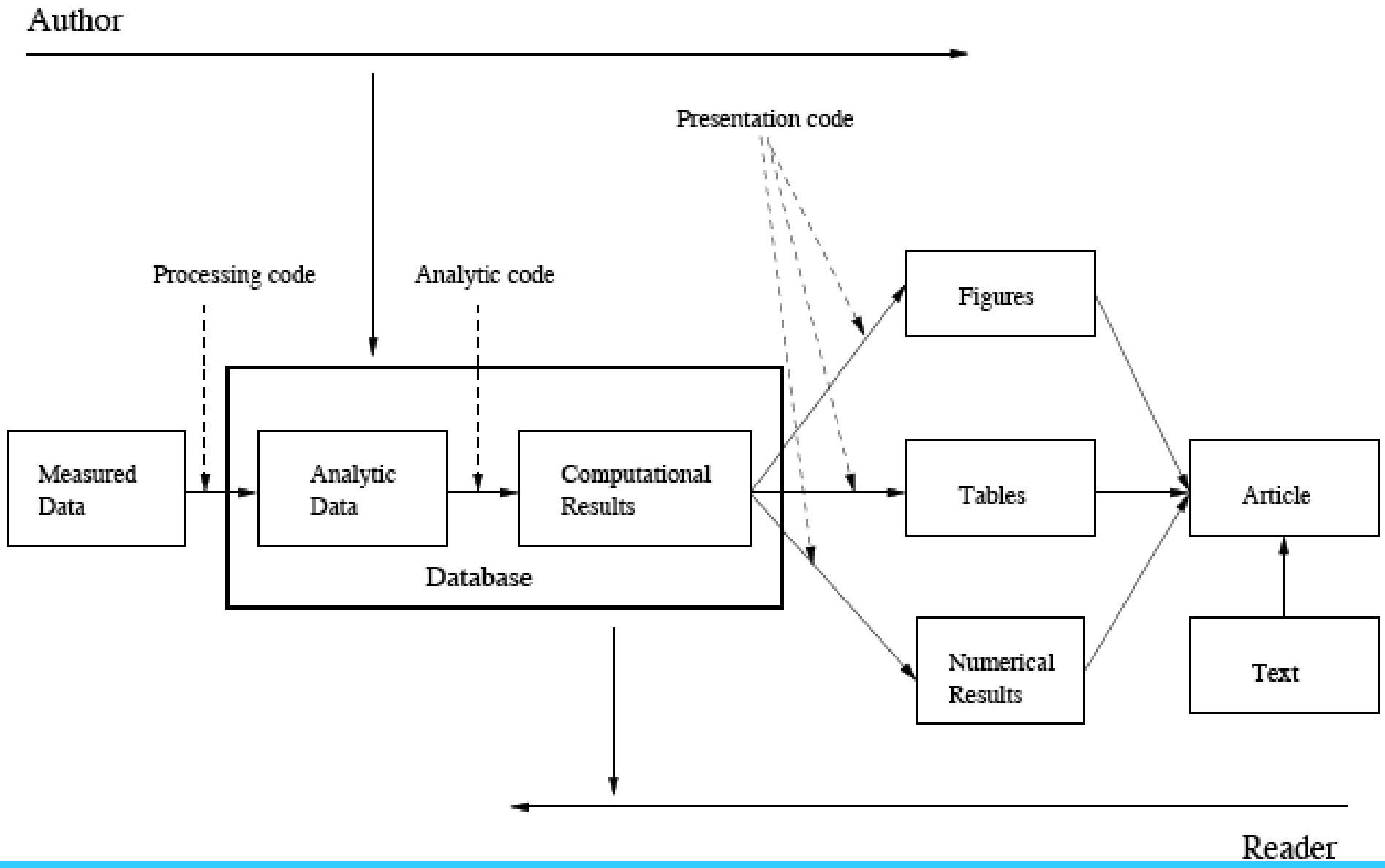
Roger D. Peng, Francesca Dominici, and Scott L. Zeger

From the Biostatistics Department, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Received for publication November 4, 2005; accepted for publication January 10, 2006.

Research Pipeline




Author
Presentation code
Processing code
Analytic code
Figures
Measured Data
Analytic Data
Computational Results
Tables
Article
Database
Numerical Results
Text
Reader

Current model of data sharing

Share

Don’t share

Current data sharing model

- Authors
  - Just put stuff on the web
  - Journal supplementary materials
  - There are some central databases for various fields
- Readers
  - Just download the data and figure it out
  - Get the software and run it

Models of data sharing

- Based on intellectual property rights, software sharing agreements (from Creative Commons Project and Open Source Initiative). *Signed agreements.*
  - **Reproduction**. The data can be used for the purpose of reproducing the results in the associated published article or for commenting on those results via a letter to the editor. No original findings based on the data may be published without explicit permission from the original investigators in a separate agreement.
  - **Share alike.** The data can be used to produce new findings or results. Any modifications to the data, including transformations, additions, or linkages to other data, which are used to produce the new findings, must be made available under the same terms.
  - **Attribution**. The data can be used for any purpose, with authors cited.
  - **Full access**. The data can be used for any purpose.

Data sharing structures

- **Dyadic** (peer to peer)
- **Institutional** (funder supported)

  e.g. BioLINCC (NHLBI), caBIG (NCI))

- **Institutional**: (study-based)

  e.g. MESA, Framingham, Nurses Health Study

- **Non-profit institutional open access**, public data sharing (e.g. Dryad, iDASH).

iDASH

integrating Data for Analysis, Anonymization, and SHaring

Data Repository

An important part of the information infrastructure provided by iDASH is to provide a single, comprehensive set of facilities to explore, navigate, analyze, and combine different forms of information provided by different data sources, within the bounds of privacy restrictions. iDASH is designed to be scalable and extensible so that developers can integrate the heterogeneous data from the national biomedical, clinical, and informatics communities. Developed as an open, community-serving, crowd-sourcing resource, the iDASH team is collaborating with biomedical, behavioral, and quantitative researchers to establish the nation's most robust data repository for high-quality collections of data. This rich repository of medical data includes images and text accompanied by meta-data. The repository, based on the MIDAS platform (Kitware), requires registered credentials for access.

Public data repository

Non-PHI/PII, anonymized data for research

Private data repository

Sensitive, PHI/PII data (Two-factor authentication required)
(currently limited to UCSD)

iDASH Data Collections

Frequently asked questions about our data repository

Data Use Agreement (DUA) and procedure for data sharing

Repository Name	Type/Category	Data Set Description
Alzheimer's Images Data	Images	The UCSD Human Memory Laboratory uses functional and structural magnetic resonance imaging (MRI) to study memory processes in volunteers with healthy memory and in patients with memory difficulties, such as in Alzheimer's Disease (AD). This research focuses upon the medial temporal lobe (MTL), which shows selective damage early in the course of AD. The laboratory studies the contributions to memory that are made by distinct MTL substructures and the interaction of these structures with other brain regions. The overall goal of the work is to understand how the different parts of the brain work together to make and to retrieve memories. In the process, researchers plan to develop imaging techniques that can measure the location and severity of brain damage in diseases of memory impairment, thus offering improved diagnosis and treatment to patients with memory difficulties.
Cardiac Atlas	Images	The Cardiac Atlas Project seeks to establish a structural and functional atlas of the heart. The emphasis of this project to

Annals of Internal Medicine | ACADEMIA AND CLINIC

Reproducible Research: Moving toward Research the Public Can Really Trust

Christine Laine, MD, MPH; Steven N. Goodman, MD, PhD, MHS; Michael E. Griswold, PhD; and Harold C. Sox, MD

A community of scientists arrives at the truth by independently verifying new observations. In this time-honored process, journals serve 2 principal functions: evaluative and editorial. In their evaluative function, they winnow out research that is unlikely to stand up to independent verification; this task is accomplished by peer review. In their editorial function, they try to ensure transparent (by which we mean clear, complete, and unambiguous) and objective descriptions of the research. Both the evaluative and editorial functions go largely unnoticed by the public—the former only draws public attention when a journal publishes fraudulent research. However, both play a critical role in the progress of science. This paper is about both functions. We describe the evaluative processes we use and announce a new policy to help the scientific community evaluate, and build upon, the research findings that we publish.

Ann Intern Med. 2007;146:450-453. www.annals.org
For author affiliations, see end of text.

says "At 6'4", 220 pounds, Bob is a formidable man. But he's no match for something one millionth his size. A clot.") without educating consumers about the actual size of the risks they face. Exposed only to the ads, most control

misleading ads (28). But neither the guidance nor the legislation addresses the routine provision of efficacy data or standards for presenting side effect data in print ads. The drug facts box is a viable way to disseminate these data.

Although drug boxes could be produced by the proposed Center for Comparative Effectiveness Research (29) or existing, independent organizations (such as the Drug Effectiveness Review Project [30]), we believe that the FDA should produce and routinely update them.

Given its central role in summarizing drug information, the FDA is the most important leverage point in getting balanced drug information to physicians and consumers. Moreover, the FDA drug reviewers are uniquely suited to creating boxes: They are independent experts trained to assess drug performance, and no one knows new

Note: Drs. Schwartz and Woloshin contributed equally to this article. The order of authorship is arbitrary.

Disclaimer: The views expressed herein do not necessarily represent the

Reproducible Research Statement: *Study protocols:* Available at www .clinicaltrials.gov/ct2/showNCT00450931 and www.clinicaltrials.gov /ct2/showNCT00753857. *Statistical code:* Available from Dr. Woloshin (e-mail, steven.woloshin@dartmouth.edu). *Data sets:* Not available.

ported by t R01CA104721) and the Attorney Ge Education grant program.

Potential Finan est: Drs. Schwartz, Woloshin, and Welch are the a *Your Chances: Understanding Health Statistics* (Univ of r, 2008). They have received no royalties or other payments f book.

Reproducible Research Statement: *Study protocols:* Available at www .clinicaltrials.gov/ct2/showNCT00450931 and www.clinicaltrials.gov /ct2/showNCT00753857. *Statistical code:* Available from Dr. Woloshin (e-mail, steven.woloshin@dartmouth.edu). *Data sets:* Not available.

Requests for Single Reprints: Steven Woloshin, MD, MS, Veterans Affairs Outcomes Group (111B), Department of Veterans Affairs Medical Center, White River Junction, VT 05009; e-mail, steven.woloshin @dartmouth.edu.

Annals of Internal Medicine | ARTICLE

Differences in Control of Cardiovascular Disease and Diabetes by Race, Ethnicity, and Education: U.S. Trends From 1999 to 2006 and Effects of Medicare Coverage

J. Michael McWilliams, MD, PhD; Ellen Meara, PhD; Alan M. Zaslavsky, PhD; and John Z. Ayanian, MD, MPP

Background: Efforts to improve the care of cardiovascular disease and diabetes or expand insurance coverage for adults with these conditions may reduce differences in clinical outcomes.

Objective: To assess recent national trends in disease control, trends in sociodemographic differences in control, and changes in sociodemographic differences after age 65 years associated with near-universal Medicare coverage.

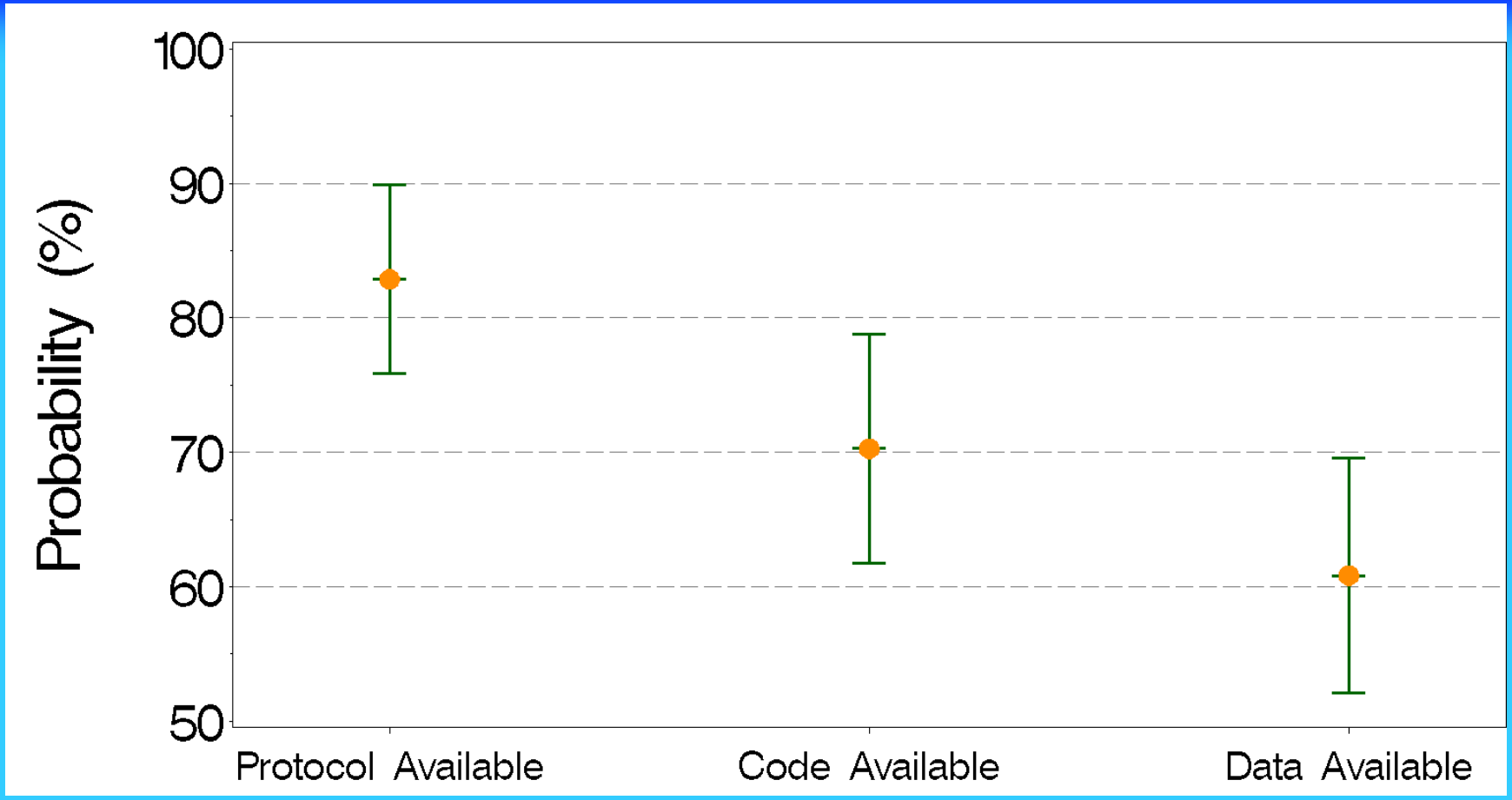
Design: Observational and quasi-experimental analyses of repeated cross-sectional data.

Setting: National Health and Nutrition Examination Survey, 1999 to 2006.

Results: Disease control improved significantly between 1999 and 2006 for all 6 measures ($P < 0.001$). These trends did not differ by race or ethnicity or by education ($P \geq 0.185$ for group–time interactions), except that white–Hispanic differences in glycemic control widened ($P = 0.042$). Black–white differences in systolic blood pressure were smaller among adults age 65 to 85 years than among adults age 40 to 64 years (reduction in difference, 4.2 mm Hg; $P = 0.009$). Black–white differences in hemoglobin A_{1c} levels were also smaller after age 65 years (reduction in difference, 0.7%; $P = 0.005$), as were Hispanic–white differences (reduction in difference, 0.7%; $P = 0.007$) and differences between less and more educated adults (reduction in difference, 0.5%; $P = 0.033$).

Reproducible Research Statement: *Study protocol and statistical code*: Available from Dr. McWilliams (e-mail, mcwilliams@hcp.med.harvard.edu). *Data set:* Available at www.cdc.gov/nchs/nhanes.htm.

Results: Percent Willing to Share Protocol, Statistical Code, and Data



Requests for Materials w/in 3 Months of Publication

- E-mailed corresponding authors of 109 manuscripts reporting “materials available by contacting author”, 71 (65%) responded
- 11/71 reported ≥1 request for protocol
- 5/71 reported ≥1 request for statistical code
- 9/71 reported ≥1 request for data

Annals of Internal Medicine

Journal	Require Code shared publically ?	Require author statement?	Require public data sharing?	Description of policy
NEJM	No	No	No	No policy identified. Protocols posted
JAMA	No	No	No	If requested, author must provide data and code to editor(s) or their assignees for examination (4). Requires an independent academic statistician to corroborate analyses of industry funded studies.
Annals Int Medicine *BMJ*	No	**Yes**	No	Requires authors to state their willingness to share and any conditions for sharing 1) study protocol (original and amendments), 2) statistical code used to generate results, and 3) the data set from which results were derived.
Lancet	No	No	No	No policy identified.
Biostatistics	No*	**Yes**	No*	Papers "kite-marked D if the data on which they are based are freely available, C if the authors' code is freely available, and R if both data and code are available,
PLoS Medicine	**Yes**	**Yes**	**Yes**	"Publication is conditional upon the agreement of the authors to make freely available any materials and information described in their publication that may be reasonably requested by others for the purpose of academic, non-commercial research." "Authors must comply with current best practices for data sharing in their fields… Data for which public repositories have been established should be deposited before publication, and the appropriate accession numbers or digital object identifiers (DOIs) published with the paper. If an appropriate repository does not exist, data should be provided in an open access institutional repository, a general data repository such as Dryad, or as Supporting Information files with the published paper. If none of these options is practical, data should be made freely available upon request." "The conclusions of a study must not depend solely on the analysis of proprietary data. If proprietary data were used to reach a conclusion, and the authors are unwilling or unable to make these data public, then the paper must include an analysis of public data that validates the conclusions so that others can reproduce the analysis and build on the findings. Any restrictions on the availability or use of datasets might be judged to diminish the significance of a paper and may therefore influence the decision about whether a paper should be published. These policies have been developed in accordance with the principles established in Sharing Publication-Related Data and Materials (National Academies Press, 2003)."
Epidemiology	No	No	No	No policy identified

Biostatistics (2011), **12**, 1, *pp*. 33–50
doi:10.1093/biostatistics/kxq043
Advance Access publication on July 12, 2010

C

A multilevel model to address batch effects in copy number estimation using SNP arrays

ROBERT B. SCHARPF*

Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD 21205
rscharpf@jhsph.edu

INGO RUCZINSKI, BENILTON CARVALHO

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205

BETTY DOAN, ARAVINDA CHAKRAVARTI

Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205

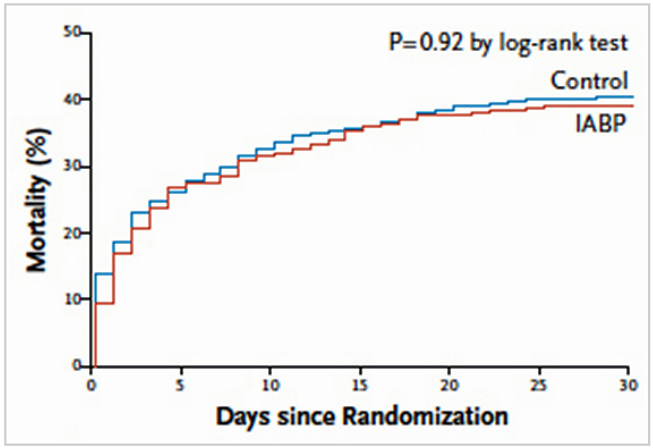
RAFAEL A. IRIZARRY

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205

SUMMARY

Submicroscopic changes in chromosomal DNA copy number dosage are common and have been implicated in many heritable diseases and cancers. Recent high-throughput technologies have a resolution that permits the detection of segmental changes in DNA copy number that span thousands of base pairs in the genome. Genomewide association studies (GWAS) may simultaneously screen for copy number phe-

The NEW ENGLAND JOURNAL of MEDICINE



ORIGINAL ARTICLE

Intraaortic Balloon Support in Cardiogenic Shock

October 4, 2012 | H. Thiele and Others

Published Online: August 27, 2012

In this trial, patients with acute MI and cardiogenic shock who were expected to undergo coronary revascularization were randomly assigned to receive or not to receive intraaortic balloon support. Balloon support had no effect on 30-day mortality.

CME | Comments

NEJM Protocol Posting

Protocol

This trial protocol has been provided by the authors to give readers additional information about their work.

Protocol for: Thiele H, Zeymer U, Neumann F-J, et al. Intraaortic balloon support for myocardial infarction with cardiogenic shock. N Engl J Med 2012;367:1287-96. DOI: 10.1056/NEJMoa1208410

IABP-SHOCK II Trial

Anhang Studienprotokoll

1 Beschreibung des Vorhabens

1.1 Thema

Prospektiv, randomisierte, multizentrische Studie zum Vergleich von intraaortaler Ballongegenpulsation (IABP) plus medikamentöser Standardtherapie im Vergleich zu einer

The previously durchgeführten, randomized trials comparing immediate PCI or aortocoronaren bypass surgery (ACB) and the first only with medical therapy any additional mechanical SUPPORT by an IABP could not have a unique Advantage for immediate reperfusion show. A study had statistically because no meaningful figures are canceled.

einem akuten Myokardinfarkt hoch. Die Mortalitätsraten liegen bei 45-70 %. Die IABP kann eine initiale hämodynamische Stabilisierung bewirken, aber auf der anderen Seite auch möglicherweise ischämische vaskuläre Komplikationen und ein "Systemisch-inflammatorisches Response Syndrom (SIRS)" fördern.[9]

Als Folge der fehlenden Evidenz wird die IABP in der klinischen Praxis zu selten benutzt trotz eindeutiger Richtlinienempfehlungen.[10, 11] Das kann in Teilen durch die höheren Kosten und die Unsicherheit bei der Implantation und dem Betrieb in weniger erfahrenen Zentren liegen.

Eine adäquat gepowerte klinische Studie, welche die Effekte der IABP auf die Mortalität untersucht, könnte daher die klinische Praxis nachhaltig ändern.

Science	**Yes**	**Yes**	**Yes**	"All data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of Science. All computer codes involved in the creation or analysis of data must also be available to any reader of Science. After publication, all reasonable requests for data and materials must be fulfilled. Any restrictions on the availability of data, codes, or materials, including fees and original data obtained from other sources (Materials Transfer Agreements), must be disclosed to the editors upon submission. … Fossils or other rare specimens must be deposited in a public museum or repository and available for research."
Proceedings of the National Academy of Sciences	**Yes**	**Yes**	**Yes**	"To allow others to replicate and build on work published in PNAS, authors must make materials, data, and associated protocols available to readers. Authors must disclose upon submission of the manuscript any restrictions on the availability of materials or information…. Before publication, authors must deposit large datasets (including microarray data, protein or nucleic acid sequences, and atomic coordinates for macromolecular structures) in an approved database and provide an accession number for inclusion in the published paper."
Nature	**Yes**	**Yes**	**Yes**	"Authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications in material transfer agreements. 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, including details of how readers can obtain materials and information. If materials are to be distributed by a for-profit company, this must be stated in the paper." "In rare instances, the journal reserves the right to require independent replication of findings prior to publication."(5)

Science proposal

- Deposit data in public repository.
- Place data in supplementary material
- Archiving agreement: institutional website + Science
- Code
- Include statement in paper about availability and curation of data.

What should journals do?
What can they do?

The NEW ENGLAND JOURNAL of MEDICINE

Clinical Trial Registration — Looking Back and Moving Ahead

Christine Laine, M.D., M.P.H.
Senior Deputy Editor, *Annals of Internal Medicine*

Richard Horton, F.Med.Sci.
Editor, *The Lancet*

Catherine D. DeAngelis, M.D., M.P.H.
Editor-in-Chief, *JAMA*

Jeffrey M. Drazen, M.D.
Editor-in-Chief, *New England Journal of Medicine*

Frank A. Frizelle, M.B., Ch.B., M.Med.Sc.
Editor-in-Chief, *The New Zealand Medical Journal*

Fiona Godlee, M.B., B.Chir., B.Sc.
Editor-in-Chief, *BMJ*

Charlotte Haug, M.D., Ph.D., M.Sc.
Editor-in-Chief, *Norwegian Medical Journal*

Paul C. Hébert, M.D., M.H.Sc.
Editor-in-Chief, *Canadian Medical Association Journal*

Sheldon Kotzin, M.L.S.
Executive Editor, MEDLINE, National Library of Medicine

Ana Marusic, M.D., Ph.D.
Editor, *Croatian Medical Journal*

Peush Sahni, M.S., Ph.D.
Representative and Past President, World Association of Medical Editors

: of Medical mary and partner registers that meet WHO-spec-
olicy requir- ified criteria.[4] Primary registers are WHO-select-
ation about ed registers managed by not-for-profit entities

Key Summary Points

- In addition to accepting registration in any of the five existing registries, the ICMJE will accept registration of clinical trials in any of the primary registers that participate in the WHO ICTRP. Registration in a partner register only is insufficient.
- The ICMJE will begin to implement the WHO definition of clinical trials for all trials that begin enrollment on or after July 1, 2008. This definition states that a clinical trial is "any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes."
- The ICMJE will not consider results posted in the same clinical trials registry in which the primary registration resides to be previous publication if the results are presented in the form of a brief (<500 words) structured abstract or table.

Why was this successful?

- Central repository
- Legislative mandate for registering some studies
- Journal collective action
- Deb Zarin

What can journals not do?

- Be effective acting alone.
- Be the custodians of all research data and protocols.
- Afford staff just for this purpose.
- Be the sole guarantors of scientific integrity.

Funder policies

- **Shelby amendment** – FOIA for any research used for public policy.
- **NIH** - Data sharing plan for research >$500K
- **NHLBI** - Deposit into BioLINCC 2 yrs after publication or 3 yrs after end of f/u
- **AHRQ, VA** – None specifically identified
- **NSF** – All grantees have data sharing plans, expected to share in a timely fashion and nominal cost.

Funders, cont.

- Howard Hughes: "Expected to make materials, data and databases, and software integral to their publication freely available for research use by other scientists and to handle request expeditiously."
  - 1) May not insist on collaboration, co-authorship or prior review of manuscripts as a condition of sending materials;
  - 2) May require recipients of materials to pay for costs
  - 3) should include all data in a publication or if not possible, freely available online with no restrictions on research use.
  - 4) provide an executable file and source code for any new software key to the research, which may be provided under a license agreement but at no cost to academic researchers.

Funders statement

- National Health and Medical Research Council, Australia
- Canadian Institutes of Health Research
- Economic and Social Research Council, UK
- NIH, CDC, AHRQ
- Substance Abuse and Mental Health Services Administration
- Deutsche Forschungsgemeinschaft
- World Bank
- Health Research Council of New Zealand
- Medical Research Council, UK
- INSERM, Paris, France
- Health Resources and Services
- Gates Foundation
- Hewlett Foundation

Panel: Joint statement of purpose—vision, principles, and goals

Vision

We intend to work together to increase the availability to the scientific community of the research data we fund that is collected from populations for the purpose of health research, and to promote the efficient use of those data to accelerate improvements in public health

Principles

Funders agree to promote greater access to and use of data in ways that are:

- *Equitable:* it should recognise and balance the needs of researchers who generate and use data, other analysts who might want to reuse those data, and communities and funders who expect health benefits to arise from research
- *Ethical:* it should protect the privacy of individuals and the dignity of communities, while simultaneously respecting the imperative to improve public health through the most productive use of data
- *Efficient:* it should improve the quality and value of research, and increase its contribution to improving public health; approaches should be proportionate and build on existing practice and reduce unnecessary duplication and competition

Immediate goals

- Standards of data management are developed, promoted, and entrenched so that research data can be shared routinely and reused effectively

(Continues in next column)

(continued from previous column)

- Funders and employers of researchers recognise data management and sharing of well-managed datasets as an important professional indicator of success in research
- Researchers creating datasets for secondary analysis from shared primary data are expected to share those datasets and act with integrity and in line with good practice, giving due acknowledgment to the generators of the original data

Longer-term aspirations

- Data collected for health research are made available to the scientific community for analysis which adds value to existing knowledge and which leads to improvements in health
- The research community, particularly those collecting data in developing countries, develop the capacity to manage and analyse those data locally, as well as contributing to international analysis efforts
- To the extent possible, datasets underpinning research papers in peer-reviewed journals are archived and made available to other researchers in a clear and transparent manner
- The human and technical resources and infrastructures needed to support data management, archiving, and access are developed and supported for long-term sustainability

The full statement is online: http://www.wellcome.ac.uk/publichealthdata. Other funding organisations are invited to join as signatories and partners in this work.

OK, let's slowly lower in the grant money.

New proposal for credit sharing

- Currently, there are 2 ways to use publications to measure academic success: authorship (+position) + citations.
- Citations (in the form of Impact Factor) are also used to evaluate *journals*.
- *3rd measure is needed*: Marker for use of data for original research published by others.
  - Includes meta-analyses, re-analyses, new analyses.
- Requires unique identifier for each (large) dataset, like the PMID we have for papers.