Assessing the Evidence Base for Cancer Screening as New Technologies are Developed

National Academy of Medicine March 2, 2020

David F Ransohoff MD

Departments of Medicine and Epidemiology
University of North Carolina at Chapel Hill

Disclosures

CONSULTING

- PastExact SciencesGRAIL
- PresentDelfi

EQUITY/OTHER

none

Assessing the Evidence Base for Cancer Screening as New Technologies are Developed

SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

CANCER

Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer

Huiyan Luo¹*, Qi Zhao¹*, Wei Wei¹*, Lianghong Zheng²*, Shaohua Yi³*, Gen Li⁴, Wenqiu Wang⁵, Hui Sheng¹, Hengying Pu¹, Haiyu Mo¹, Zhixiang Zuo¹, Zexian Liu¹, Chaofeng Li¹, Chuanbo Xie¹, Zhaolei Zeng¹, Weimin Li⁶, Xiaoke Hao⁷, Yuying Liu¹, Sumei Cao¹, Wanli Liu¹, Sarah Gibson⁴, Kang Zhang^{6,8}, Guoliang Xu¹, Rui-hua Xu^{1†}

Circulating tumor DNA (ctDNA) has emerged as a useful diagnostic and prognostic biomarker in many cancers. Here, we conducted a study to investigate the potential use of ctDNA methylation markers for the diagnosis and prognostication of colorectal cancer (CRC) and used a prospective cohort to validate their effectiveness in screening

Copyright © 2020
The Authors, some rights reserved; exclusive licensee
American Association for the Advancement of Science. No claim to original U.S. Government Works

Colon



Colorectal cancer screening with odour material by canine scent detection

Hideto Sonoda, ^{1,6} Shunji Kohnoe, ^{1,6} Tetsuro Yamazato, ² Yuji Satoh, ³ Gouki Morizono, ⁴ Kentaro Shikata, ⁵ Makoto Morita, ⁶ Akihiro Watanabe, ⁶ Masaru Morita, ¹ Yoshihiro Kakeji, ¹ Fumio Inoue, ⁴ Yoshihiko Maehara ¹

¹Department of Surgery and Science, Kyushu University at Fukuoka, Japan ²Department of Internal Medicine, Arita Kyoritsu Hospital at Arita, Saga, Japan ³St. Sugar Cancer Sniffing Dog

ABSTRACT

Objective Early detection and early treatment are of vital importance to the successful treatment of various cancers. The development of a novel screening method that is as economical and non-invasive as the faecal occult blood test (FOBT) for early detection of colorectal

Significance of this study

What is already known about this subject?

Canine olfactory detection of cancer has been

3/4/2020

Science

Cite as: J. D. Cohen *et al.*, *Science* 10.1126/science.aar3247 (2018).

Detection and localization of surgically resectable cancers with a multi-analyte blood test

Joshua D. Cohen, 1,2,3,4,5 Lu Li, ⁶ Yuxuan Wang, 1,2,3,4 Christopher Thoburn, ³ Bahman Afsari, ⁷ Ludmila Danilova, ⁷ Christopher Douville, 1,2,3,4 Ammar A. Javed, ⁸ Fay Wong, 1,2,3,4 Austin Mattox, 1,2,3,4 Ralph. H. Hruban, 3,4,9 Christopher L. Wolfgang, ⁸ Michael G. Goggins, ³,4,8,10,11 Marco Dal Molin, ⁴ Tian-Li Wang, ^{3,9} Richard Roden, ^{3,9} Alison P. Klein, ^{3,4,12} Janine Ptak, ^{1,2,3,4} Lisa Dobbyn, ^{1,2,3,4} Joy Schaefer, ^{1,2,3,4} Natalie Silliman, ^{1,2,3,4} Maria Popoli, ^{1,2,3,4} Joshua T. Vogelstein, ³ James D. Browne, ⁸ Robert E. Schoen, ^{1,5,10} Randall E. Brand, ¹⁵ Jeanne Tie, ^{17,18,19,20} Peter Gibbs, ^{17,18,19,20} Hui-Li Wong, ¹⁷ Aaron S. Mansfield, ²¹ Jin Jen, ²² Samir M. Hanash, ²³ Massimo Falconi, ²⁴ Peter J. Allen, ²⁵ Shibin Zhou, ^{1,5,4} Chetan Bettegowda, ^{1,2,3,4} Luis Diaz, ^{1,3,4} Cristian Tomasetti, ^{3,6,7}* Kenneth W. Kinzler, ^{1,3,4}* Bert Vogelstein, ^{1,2,3,4}* Anne Marie Lennon, ^{3,4,8,10,118} Nickolas Papadopoulos^{1,3,4*}

GRAIL Press Release

GRAIL Announces Promising New Data with Early Detection Blood Test to be Presented at 2019 American Society of Clinical Oncology Annual Meeting

May 15, 2019

Assessing the Evidence Base for Cancer Screening as New Technologies are Developed

Question

•How do we decide which technologies work? What kind of evidence is needed?

Conceptual framework to assess evidence for screening

(from 1970s, by Nuffield Provincial Trust, Canadian Task Force, and US Preventive Services Task Force)

Four questions:

- 1. Does disease, untreated, cause bad outcome?
- 2. Can the screening test discriminate disease vs not?
- 3. Based on discrimination, does intervention cause improved outcome? Strong evidence, like RCT, is required; should be quantitative.
- 4. Is benefit greater than harm, quantitatively?

modified from Harris R. Am J Prev Med 2001;20 (Suppl):21

Assessing the Evidence Base for Cancer Screening as New Technologies are Developed

TODAY: EVIDENCE ABOUT "DISCRIMINATION"

1. Theme

An important step in conducting science - 'Asking what might be wrong' – is especially difficult in research about new technology

- 2. Problems
 - research design
 - data analysis
- 3. Lessons

Feynman says, "Ask what might be wrong" to avoid fooling yourself, which is easy.



"Details that could throw doubt on your interpretation must be given, if you know them.... [I]f you know anything at all wrong, or possibly wrong--to explain it."

Cargo Cult Science

by RICHARD P. FEYNMAN

Some remarks on science, pseudoscience, and learning how to not fool yourself. Caltech's 1974 commencement address.

Feynman. Engineering and Science 1974:10-13.

'Asking what might be wrong' is the *reason* some fields, like molecular biology of the gene, advanced faster than others

experimental test." Or "[o]n any given morning the blackboards of Francis Crick or Sidney Brenner... [will show] the hot new result just up from the laboratory or just in by letter or rumor. On the next line will be two or three alternative explanations, or a little list of 'what he did wrong.' Underneath... a series of suggested experiments or controls that can reduce the number of possibilities" [94]. Platt was saying that progress is based on considering alternative explanations and avoiding overinterpretation.

Assessing the Evidence Base for Cancer Screening as New Technologies are Developed

TODAY'S FOCUS: EVIDENCE ABOUT "DISCRIMINATION"

1. Theme

An important step in conducting science - 'Asking what might be wrong' – is especially difficult in research about new technology

- 2. Problems
 - research design
 - data analysis
- 3. Lessons

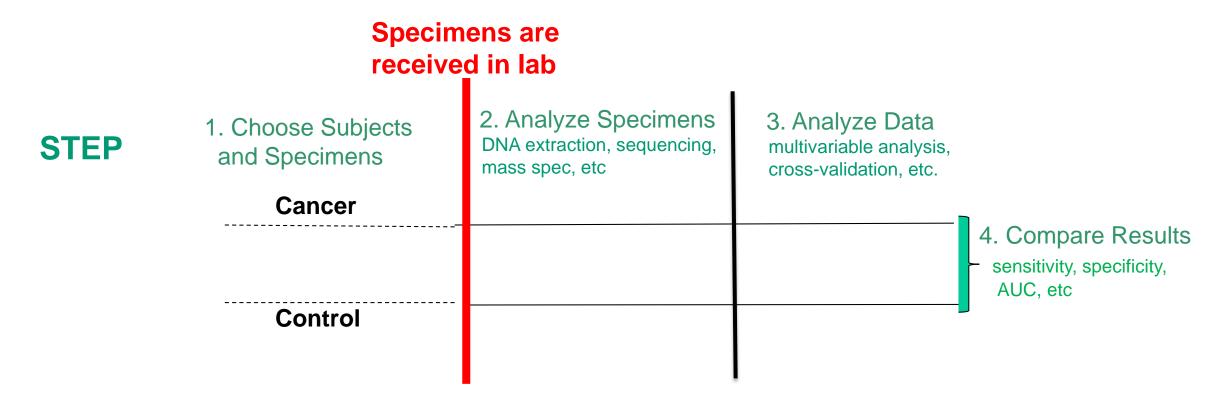
Challenge in research design about test discrimination: BIAS

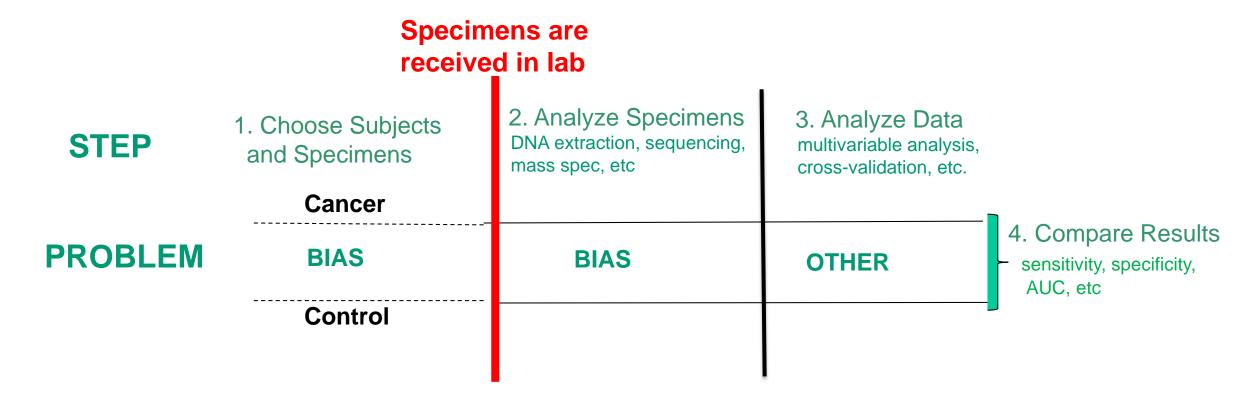
Definition:

•Bias can occur in a comparison when the groups compared differ systematically in some way other than 'cancer vs not'.

Importance:

- Discrimination caused by bias won't reproduce or 'validate'.
- Bias may be fatal.
- •Bias in non-experimental (observational) research is so difficult to avoid and so important that 'a study is guilty until proven innocent'.





MECHANISMS OF DISEASE

Mechanisms of disease

3 Use of proteomic patterns in serum to identify ovarian cancer

Emanuel F Petricoin III, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, Charles Simone, David A Fishman, Elise C Kohn, Lance A Liotta

Summary

Lancet 2002; 359: 572-577

Method: In Ca vs not: assess mass spec proteomics patterns

Result: ~100% sensitivity, specificity for ovarian cancer

MECHANISMS OF DISEASE

Mechanisms of disease

3 Use of proteomic patterns in serum to identify ovarian cancer

Emanuel F Petricoin III, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, Charles Simone, David A Fishman, Elise C Kohn, Lance A Liotta

Summary

Lancet 2002; 359: 572-577

Method: In Ca vs not: assess mass spec proteomics patterns

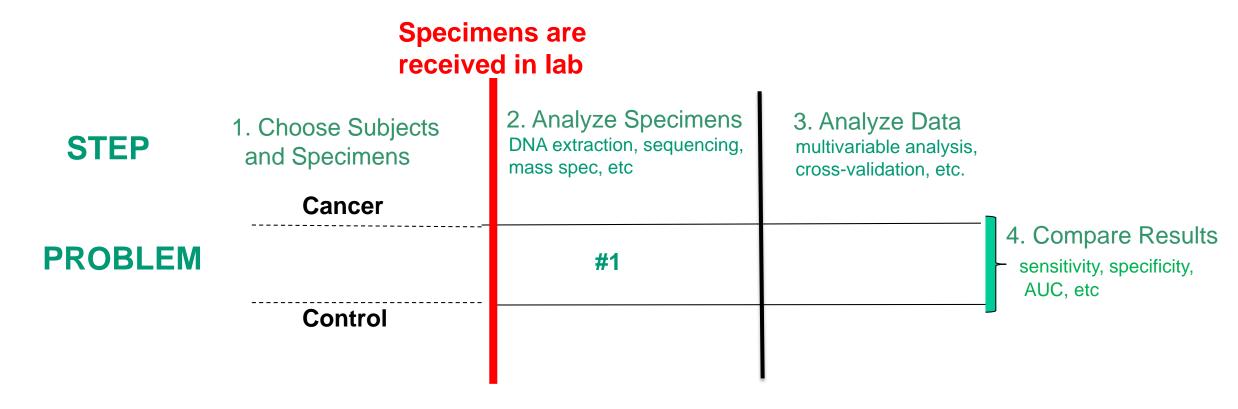
Result: ~100% sensitivity, specificity for ovarian cancer

Bias: Compared groups differed, but not due to cancer.

- Cancer vs controls were analyzed on different days
- •Mass spec drifts over time; 'discrimination' is hardwired into results.

(Baggerly. Bioinformatics 2004; JNCI 2005)

Identifying this bias took 3 yrs, even with raw data.



Imaging, Diagnosis, Prognosis

Diagnostic Markers for Early Detection of Ovarian Cancer

kene Visintin, ¹Ziding Feng.² Gary Longton, ² David C. Ward, ³ Ayesha B. Alvero, ¹Yinglei Lai, ⁴
Jeannette Tenthorey, ¹Aliza Leiser, ¹Ruben Flores-Saaib, ⁵ Herbert Yu, ⁶ Masoud Azori, ¹
Thomas Rutherford, ¹Peter E. Schwartz, ¹ and Gil Mor ³ Clin Cancer Res 2008;14:1065

Method:

- Measure specific markers
- Pts from different clinics: Ca (high-risk, with mass) v nl (screening)

Result: ~100% sensitive/specific for OvCa

Bias: Compared groups may be different:

- •'Stress' proteins could differ because of stress, not cancer.
- •Specimen handling could differ in the 2 clinics.

McIntosh M. Ovarian cancer early detection claims are biased.Clin Cancer Res 2008;14(22):7574.

'What might be wrong': not discussed in article or press release.

Both OvCa blood test claims were in New York Times

Example #1: 2004

New York Times, 2.3.04

New Cancer Test Stirs Hope and Concern

By ANDREW POLLACK

Jill Doimer's mother died in 2002 from ovarian cancer, detected too late to be effectively treated.

So Ms. Doimer is eagerly awaiting the introduction of a new test that holds the promise of detecting early-stage ovarian cancer far more accurately than any test available now, using only blood from a fineer prick.

Not only does she plan to be tested, but an advocacy group she helped found, Ovarian Awareness of Kentucky, also intends to spread the word to women and doctors.

"H it's going to happen to me or anyone I know, I want it to be caught at an early stage," said Ms. Doimer, who lives in Louisville.

The new test, expected to be available in the next few months, could have a big effect on public health if it works as advertised. That is because when ovarian cancer is caught early, when it is treatable by surgery, more than 90 percent of women live five years or longer. But right now, about three-quarters of cases are detected after the cancer has advanced, and then only 35 percent of women survive five years.

The test is also the first to use a new technology that some believers say could revolutionize diagnostics. It looks not for a single telltale protein — like the prostate-specific antigen, or P.S.A., used to diagnose prostate cancer — but rather for a complex fingerprint formed by all the proteins in the blood. Similar tests are being developed for prostate, pancreatic, breast and other cancers. The technique may work for other diseases as well.

"I've been in cancer research for 40 years and I think it's the most important breakthrough in those years," said Dr. Continued on Page 6

Example #2: 2008

Cancer Test For Women Raises Hope, And Concern

By ANDREW POLLACK

A new blood test aimed at detecting ovarian cancer at an early, still treatable stage is stirring hopes among women and their physicians. But the Food and Drug Administration and some experts say the test has not been proved to work.

Differential exoprotease activities confer tumor-specific serum peptidome patterns

Josep Villanueva, David R. Shaffer, John Philip, Carlos A. Chaparro, Hediye Erdjument-Bromage, Adam B. Olshen, Martin Fleisher, Hans Lilja, Edi Brogi, Jeff Boyd, Marta Sanchez-Carbayo, Eric C. Holland, Carlos Cordon-Cardo, Howard I. Scher, and Paul Tempst

J Clin Invest 2006;116:271

Method: In persons with/without PrCa, measure peptide patterns

Results: ~100% sensitive, specific for PrCa.

Bias: The compared groups are different:

•Cancer: mean age 67y.o.; 100% men

Differential exoprotease activities confer tumor-specific serum peptidome patterns

Josep Villanueva, David R. Shaffer, John Philip, Carlos A. Chaparro, Hediye Erdjument-Bromage, Adam B. Olshen, Martin Fleisher, Hans Lilja, Edi Brogi, Jeff Boyd, Marta Sanchez-Carbayo, Eric C. Holland, Carlos Cordon-Cardo, Howard I. Scher, and Paul Tempst

J Clin Invest 2006;116:271

Method: In persons with/without PrCa, measure peptide patterns

Results: ~100% sensitive, specific for PrCa.

Bias: The compared groups are different:

•Cancer: mean age 67y.o.; 100% men

•Control: mean age 35y.o.; 58% women

Differential exoprotease activities confer tumor-specific serum peptidome patterns

Josep Villanueva, David R. Shaffer, John Philip, Carlos A. Chaparro, Hediye Erdjument-Bromage,
Adam B. Olshen, Martin Fleisher, Hans Lilja, Edi Brogi, Jeff Boyd, Marta Sanchez-Carbayo,
Eric C. Holland, Carlos Cordon-Cardo, Howard I. Scher, and Paul Tempst J Clin Invest 2006;116:271

JCI report referenced article about 'bias as a threat to validity'.

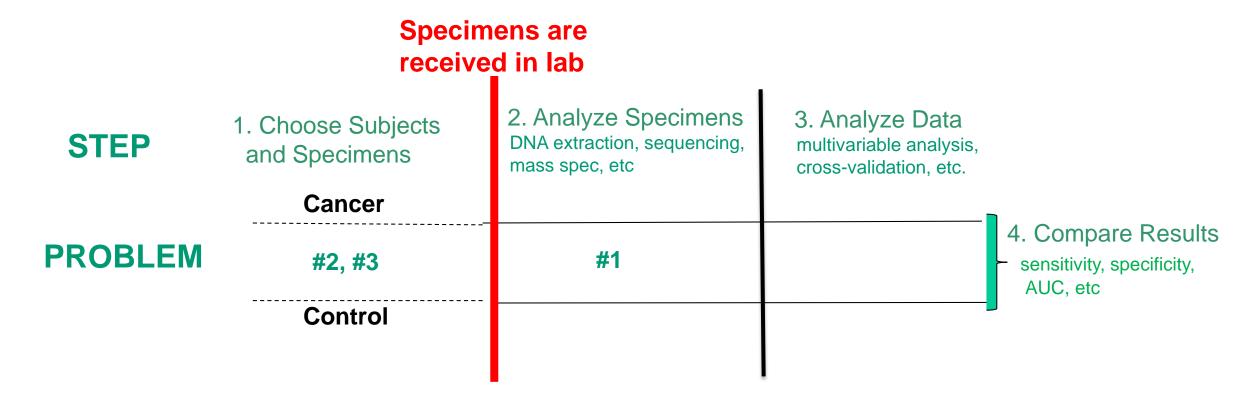
Bias as a threat to the validity of cancer molecular-marker research

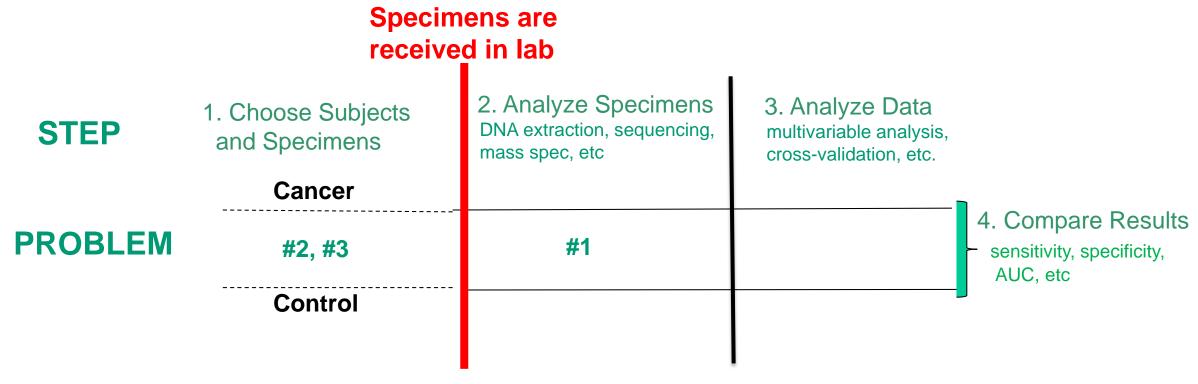
Nat Rev Cancer 2006 WW

Differential exoprotease activities confer tumor-specific serum peptidome patterns

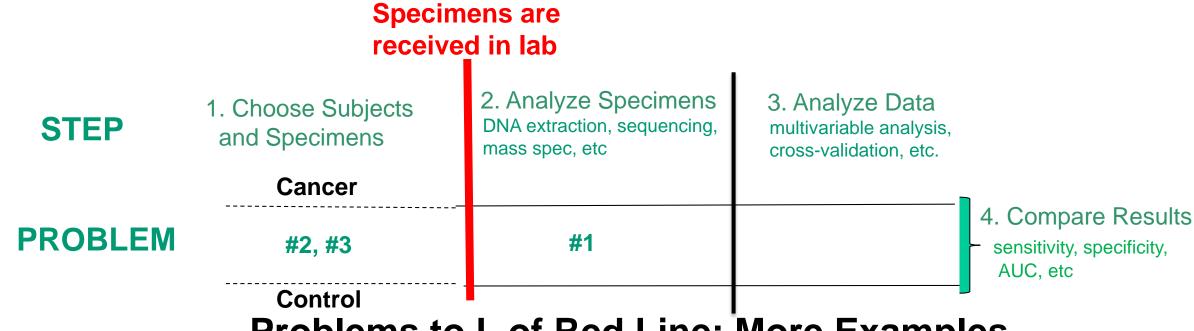
Josep Villanueva, David R. Shaffer, John Philip, Carlos A. Chaparro, Hediye Erdjument-Bromage, Adam B. Olshen, Martin Fleisher, Hans Lilja, Edi Brogi, Jeff Boyd, Marta Sanchez-Carbayo, Eric C. Holland, Carlos Cordon-Cardo, Howard I. Scher, and Paul Tempst

An obvious bias ('what might be wrong') -comparing men vs womenwasn't noted by investigators, editors, reviewers, or editorialists. What are lessons, then, for bias that is more subtle but may be fatal?





If comparison is fatally biased on the L or R of the Red Line, then GIGO, the problem, cannot be fixed by any amount of 'better analysis' of specimens or data.



Problems to L of Red Line: More Examples

•Did happen:

- -cases: serum; controls: plasma
- -cases from S.America; controls from USA
- -cases collected over 10 yrs, controls over last year

•Could happen:

- -thaw-freeze cycles more frequent in cancers than controls
- -knowledge of outcome affects interpretation of assay (and vice versa)

oto oto

3/4/2020 27

Do OvCa blood tests discriminate?

NCI designed a study to help answer; may have lessons for future.

By 2008, there were 5 claims of high sens/spec, but none 'validated'.

- In 2008, NCI designed a validation study: nested case-control design within NCI/PLCO bank of *serial* bloods, to compare 5 new assays with CA125.
- •Concept: Blood specimens from asymptomatic persons may be available 'near' the time of diagnosis; can be assessed blinded.
- •This design helps 'make other things equal', un-biased, re handling of bloods, assessment of outcome, etc.

Zhu, Pinsky, Cramer, Ransohoff et al. Cancer Prev Res 2011:4:375 Ransohoff. J Clin Epi 2007;60:1205 Pepe. JNCI 2008;100:1432. [PRoBE design]

Result: 5 new assays no better than Ca125

Model	Sensitivity ^a ≤12 mo % (95% CI)	Sensitivity ^a 13–24 mo % (95% CI)	Specificity ^a % (95% CI)	ROC** Area (95% CI)	Sensitivity at 98% Specificity ^b % (95% CI)
A1	34.3 (23-46)	7.7 (1-25)	96.8 (95.2-98.4)	0.721 (0.64-0.80)	32.8 (22-44)
B1	69.2 (58-80)	12.5 (3-31)	96.6 (94.9-98.3)	0.892 (0.84-0.95)	64.6 (53-76)
C1	34.3 (23-46)	11.5 (2-30)	95.1 (93.1-97.1)	0.712 (0.63-0.79)	25.4 (15-36)
D1	95.4 (90-99)	76.0 (59-93)	32.2 (27.4-36.5)	0.858 (0.80-0.92)	52.3 (40-64)
E1	37.9 (26-50)	3.9 (0-20)	89.8 (87.0-92.6)	N/A°	N/A ^c
CA125 ^d	63.1 (51-75)	0.0 (0-13)	98.5 (97.4-99.6)	0.890 (0.84-0.94)	64.6 (53-76)
Step 2°	n = 30	n = 15	n = 237		n = 30
A2	53.3 (35-71)	6.7 (0-32)	96.6 (94.3-98.8)	0.852 (0.77-0.94)	36.7 (20-54)
B2	80.0 (66-94)	21.4 (5-50)	92.2 (88.7-95.7)	N/A ^c	N/A ^c
C2	70.0 (54-86)	6.7 (0-32)	91.9 (88.4-95.4)	0.848 (0.76-0.94)	46.7 (29-64)
D2	55.2 (37-73)	0.0 (0-22)	86.9 (82.5-91.3)	0.810 (0.72-0.90)	51.7 (34-69)
E2	30.0 (14-46)	13.3 (2-40)	96.2 (93.7-98.7)	0.590 (0.46-0.72)	23.3 (8-38)
CA125 ^d	72.4 (56-89)	0.0 (0-22)	97.9 (96.0-99.8)	0.898 (0.82-0.98)	72.4 (56-89)
Step 3 (Pan-site)	N/A [†]	N/A ^f	N/A ^f	0.911 (0.86-0.96)	68.2 (57–80)

Could appropriate cohort infrastructure be developed?

For example, using large population-based health systems, like HMOs, VA, Scandinavian countries.



JNCI J Natl Cancer Inst (2015) 107(4): djv012

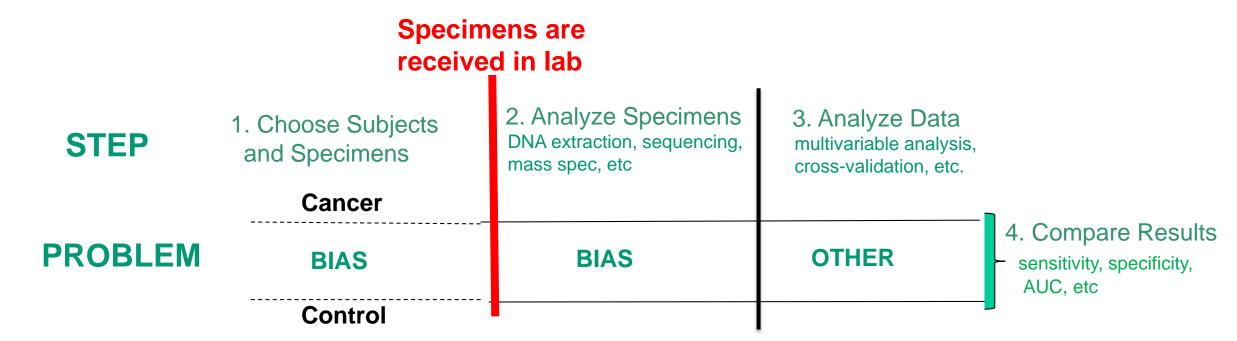
doi:10.1093/jnci/djv012 First published online February 16, 2015 Commentary

COMMENTARY

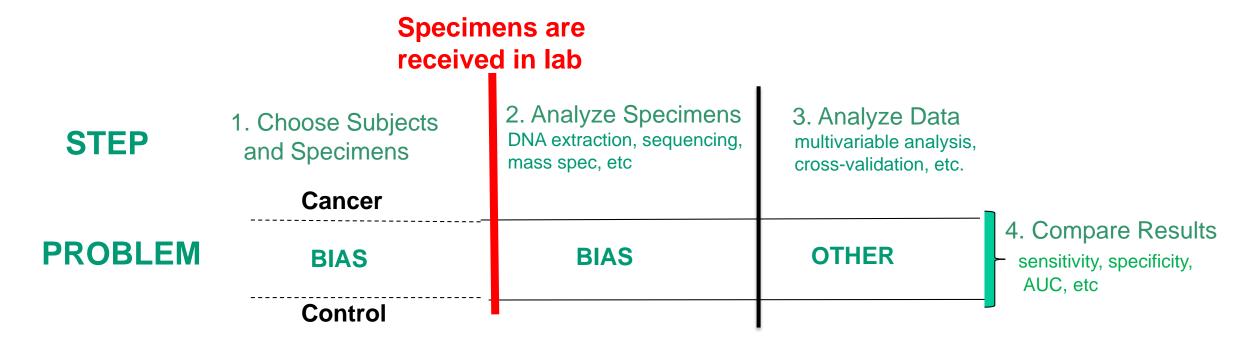
Leveraging Biospecimen Resources for Discovery or Validation of Markers for Early Cancer Detection

Sheri D. Schully, Danielle M. Carrick, Leah E. Mechanic, Sudhir Srivastava, Garnet L. Anderson, John A. Baron, Christine D. Berg, Jennifer Cullen, Eleftherios P. Diamandis, V. Paul Doria-Rose, Katrina A. B. Goddard, Susan E. Hankinson, Lawrence H. Kushi, Eric B. Larson, Lisa M. McShane, Richard L. Schilsky, Steven Shak, Steven J. Skates, Nicole Urban, Barnett S. Kramer, Muin J. Khoury, David F. Ransohoff

How to address challenges at each step



How to address challenges at each step



Challenges in asking 'What might be wrong?'

First, it's hard to know what to ask and where to look for answers. Obvious problems can be missed.

Second, each step requires different expertise, raising questions about communication, responsibility, leadership.

Third challenge: There may be incentive to not ask 'What might be wrong?'

Downloaded from http://rsos.royalsocietypublishing.org/ on September 28, 2016

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research





Cite this article: Smaldino PE, McElreath R. 2016 The natural selection of bad science. *R. Soc. open sci.* 3: 160384. http://dx.doi.org/10.1098/rsos.160384

Received: 1 June 2016 Accepted: 17 August 2016

The natural selection of bad science

Paul E. Smaldino¹ and Richard McElreath²

¹Cognitive and Information Sciences, University of California, Merced, CA 95343, USA ²Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

PES, 0000-0002-7133-5620; RME, 0000-0002-0387-5377

Poor research design and data analysis encourage false-positive findings. Such poor methods persist despite perennial calls for improvement, suggesting that they result from something more than just misunderstanding. The persistence of poor methods results partly from incentives that favour them, leading to the natural selection of bad science. This dynamic requires no conscious strategizing—no deliberate cheating nor loafing—by scientists, only that publication is a principal factor for career advancement. Some normative methods of analysis have

Third challenge: There may be incentive to not ask 'What might be wrong?'

Downloaded from http://rsos.royalsocietypublishing.org/ on September 28, 2016

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research





Cite this article: Smaldino PE, McElreath R. 2016 The natural selection of bad science. *R. Soc. open sci.* 3: 160384. http://dx.doi.org/10.1098/rsos.160384

Received: 1 June 2016 Accepted: 17 August 2016

The natural selection of bad science

Paul E. Smaldino¹ and Richard McElreath²

¹Cognitive and Information Sciences, University of California, Merced, CA 95343, USA ²Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

(i) PES, 0000-0002-7133-5620; RME, 0000-0002-0387-5377

Poor research design and data analysis encourage false-positive findings. Such poor methods persist despite perennial calls for improvement, suggesting that they result from something more than just misunderstanding. The persistence of poor methods results partly from incentives that favour them, leading to the natural selection of bad science. This dynamic requires no conscious strategizing—no deliberate cheating nor loafing—by scientists, only that publication is a principal factor for career advancement. Some normative methods of analysis have



Also called: "self-serving statistical sloppiness."

Third challenge: There may be incentive to not ask 'What might be wrong?'

Downloaded from http://rsos.royalsocietypublishing.org/ on September 28, 2016

ROYAL SOCIETY **OPEN SCIENCE**

rsos.royalsocietypublishing.org

Research





Cite this article: Smaldino PE, McElreath R. 2016 The natural selection of bad science. R. Soc. open sci. 3: 160384. http://dx.doi.org/10.1098/rsos.160384

Received: 1 June 2016 Accepted: 17 August 2016

The natural selection of bad science

Paul E. Smaldino¹ and Richard McElreath²

¹Cognitive and Information Sciences, University of California, Merced, CA 95343, USA ²Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

PES, 0000-0002-7133-5620; RME, 0000-0002-0387-5377

Poor research design and data analysis encourage false-positive findings. Such poor methods persist despite perennial calls for improvement, suggesting that they result from something more than just misunderstanding. The persistence of poor methods results partly from incentives that favour them, leading to the natural selection of bad science. This dynamic requires no conscious strategizing—no deliberate cheating nor loafing by scientists, only that publication is a principal factor for career advancement. Some normative methods of analysis have



- Also called: "self-serving statistical sloppiness."
- Incentives may reflect a 'systems problem'. Deming says, "Every system is perfectly designed to get the results it gets."

Assessing the Evidence Base for Cancer Screening as New Technologies are Developed

Selected topics for further consideration (not a list of shovel-ready projects)

Resources

-Cultivate cohorts, when feasible, to create banks of subjects/specimens for 'nested case control studies' to be used in discovery and validation research; will be feasible someday.

Organization

-To effectively ask 'what might be wrong' in *planning* research, arrange for appropriate expertise/communication/leadership among involved scientists.

Motivation

-Consider how to motivate people to improve strength of science; what 'systems problems' need to be addressed.

Acknowledgements

Support

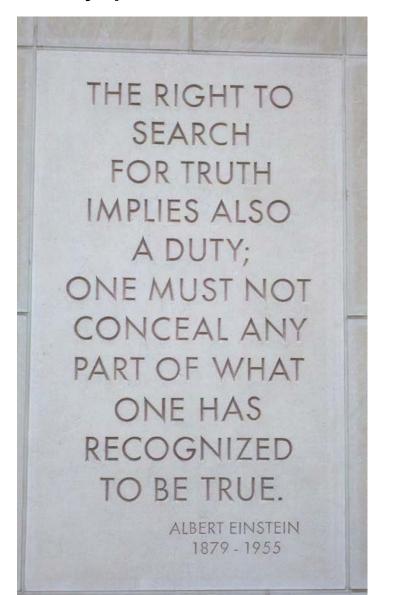
National Cancer Institute

- Division of Cancer Prevention: Early Detection Research Network
- Division of Cancer Control and Population Sciences
- Clinical Proteomic Tumor Analysis Consortium
- Division of Cancer Epidemiology and Genetics

Colleagues

- Steven Skates, Massachusetts General Hospital
- •Lisa McShane, National Cancer Institute
- Keith Baggerly, MD Anderson Cancer Center
- •Ziding Feng, Fred Hutchinson Cancer Research Center
- John Baron, University of North Carolina

Einstein's comment, related to 'Ask what might be wrong': "One must not conceal any part of what one has recognized to be true."



at entrance to Keck Center