Statistical analysis of high-dimensional biomedical data: issues and challenges in translation to medically useful results

Lisa Meier McShane, PhD
Biometric Research Program
Associate Director, Division of Cancer Treatment and Diagnosis
U.S. National Cancer Institute, National Institutes of Health

NIH NATIONAL CANCER INSTITUTE

67th Biometric Colloquium
Muenster, Germany (virtual)
March 15, 2021

Disclaimers

- The views expressed represent my own and do not necessarily represent views or policies of the U.S. National Cancer Institute.
- Examples I cite are all based on true stories or published articles, but I have made minor modifications in some cases to conceal identities.
- My examples focus on omics-based tests, but the principles apply more generally, particularly for high-dimensional data

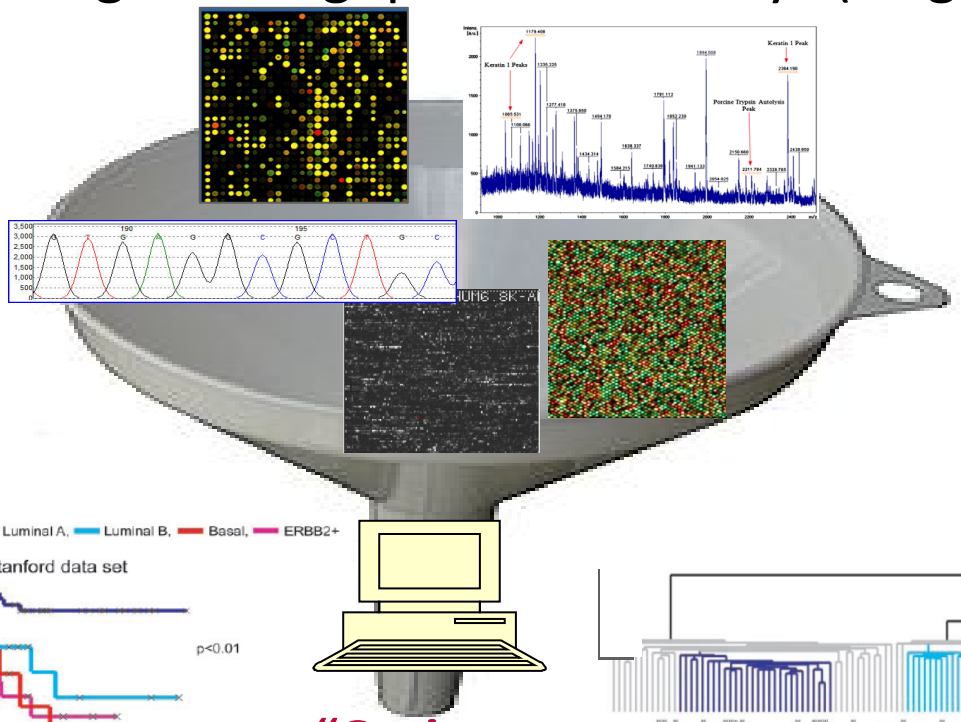
My perspective

- Statistical/scientific reviewer of NCI-sponsored clinical trials and studies for development and validation of biomarker- and omics-based tests, e.g., for precision medicine
- Journal editorial board member
- Statistical reviewer for numerous biomedical journals
- Statistical collaborator in research projects involving biomarkers and omics tests
- **Co-chair of STRATOS High-Dimensional Data (HDD) Topic Group (“TG9”)**

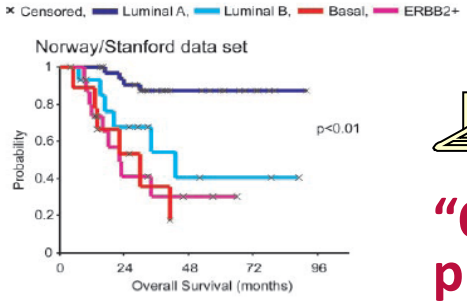
Translation from omics discoveries to clinically useful omics-based tests to guide clinical care

"Omics" is a term encompassing multiple molecular disciplines, which involve the characterization of global sets of biological molecules such as DNAs, RNAs, proteins, and metabolites."

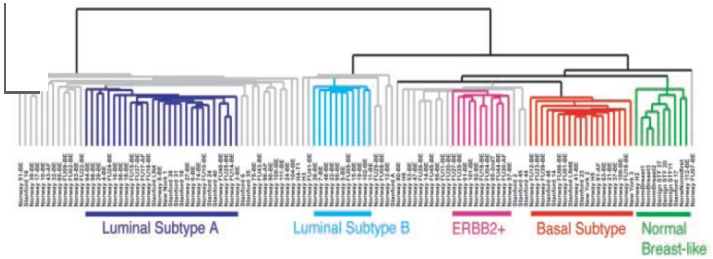
High-throughput omics assays ("high-dimensional")



Apply computational methods to derive models, risk scores, classifiers



"Omics predictor"



Skepticism, disappointment, and scandal

OvaSure diagnostic test for ovarian cancer

MISSING THE MARK

Why is it so hard to find a test to predict cancer?

BY LIZZIE BUCHEN

On 3 March, two studies appeared online that offered 19 pages of gloomy reading for anyone interested in cancer. They women — to ask whether these seemingly breakthrough biomarkers were better at identifying women with early ovarian cancer than contrast, detected 63%.) Mor's panel already had a tortured history. A primary research paper behind it had been criticized by other

Lizzie Buchen, *Nature*, v. 471, March 24, 2011

. . . AND MANY OTHER FAILURES THAT WE NEVER HEAR ABOUT

Genomic predictors (chemo-sensitivity & prognosis) developed by Anil Potti at Duke University

Duke University allowed a controversial set of clinical trials to continue despite serious concerns about the validity of the data on which they were based.

ETHICS

Cancer trial errors revealed

University officials admit data withheld from review panel before misconduct charges arose.

BY EUGENIE SAMUEL REICH

It was a weekend that Michael Cuffe, vice-president for medical affairs at Duke University in Durham, North Carolina, says he will never forget. It began on 16 July 2010, when Cuffe learned of a damning revelation in *The Cancer Letter*, a Washington DC-based publication with a reputation for probing controversial topics in cancer research. A story in that day's issue alleged that Anil Potti, a cancer geneticist whose data had been used to design three clinical trials then under way at Duke, had lied on multiple federal grant applications, Yet Cuffe and Kornbluth had decided to restart them when a review panel seemed to validate Potti's method. The allegations that Potti, who worked at Duke's Institute for Genome Sciences and Policy (ISGP), had padded his CV changed everything. "When it comes to light that someone may have been less than honest in one aspect of their professional life, one begins to wonder whether they have been less than honest in another aspect," says Kornbluth. That weekend, she and Cuffe suspended the trials once again, and initiated a misconduct investigation that is still ongoing. Potti, who could not be reached for comment, Freedom of Information Act, Kornbluth and Cuffe have offered their account of the mistakes that led the trials to be restarted even after they learned of potential flaws in the underlying data. The affair will have an impact beyond Duke, as the Institute of Medicine, part of the US National Academies in Washington DC, begins to examine research on genome-based patient testing. Originally commissioned to investigate Duke's controversial trials, the institute's US$687,000 study is now expected to focus on providing broader recommendations for the design of clinical trials that similarly use genomic data from individual patients to tailor therapy.

Eugenie Samuel Reich, *Nature*, v. 469, January 13, 2011

Institute of Medicine initiates a study to examine field of translational omics

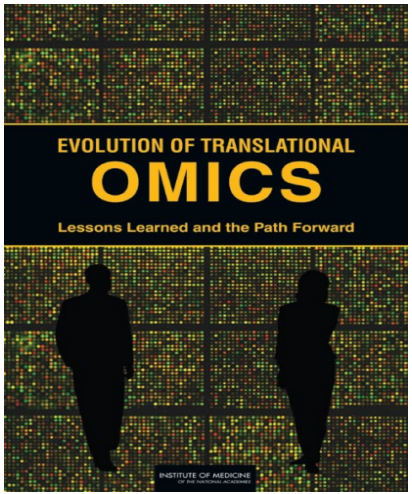
NEWS&ANALYSIS

CLINICAL MEDICINE

Jocelyn Kaiser

Science, v. 335, March 30, 2012

Biomarker Tests Need Closer Scrutiny, IOM Concludes



"There are a lot of lessons here that surely apply to other places."

—GILBERT S. OMENN, UNIVERSITY OF MICHIGAN, ANN ARBOR

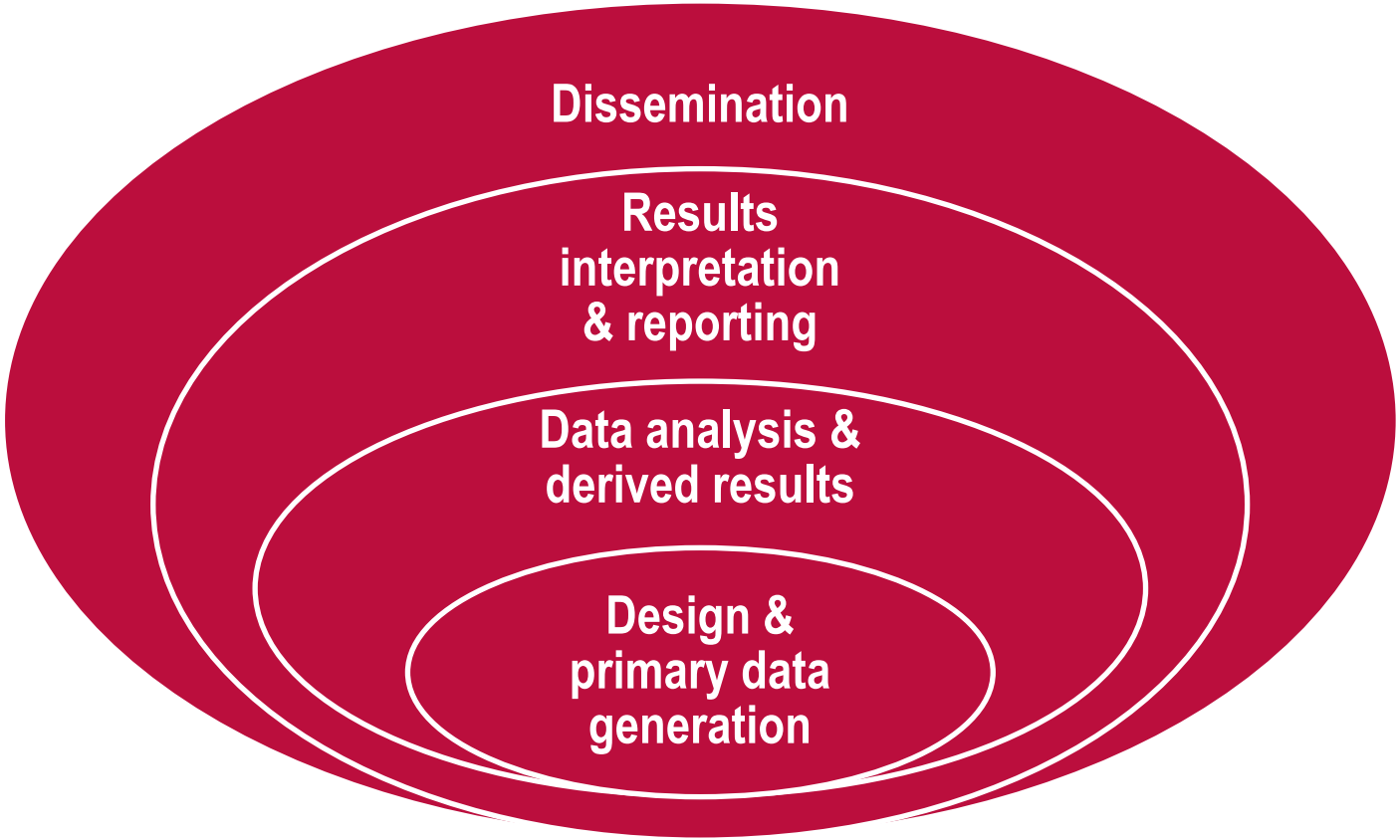
http://www.iom.edu/Reports/2012/Evolution-of-Translational-Omics.aspx

“Plane crash investigation” approach

To improve the integrity and quality of a system we must understand how it can fail

- Some examples are from a research misconduct scandal involving omics research conducted by Anil Potti at Duke University.
  - > 100 patients enrolled on trials using flawed chemosensitivity predictors
  - All information I cite for these examples is in the public domain.
- Other examples I have encountered over the last few years through collaborations, reviewing protocols and journal submissions, and as a reader of published papers reporting omics studies.
  - It is not my intent to imply that research misconduct was involved in these other examples.

Understanding sources of irreproducible research



Design considerations

Patient characteristics, potential confounding

- Published example: “Differential exoprotease activities confer tumor-specific serum peptidome patterns”
  - **100% sensitive and specific for prostate cancer**
  - Patient characteristics:
    - Cancer cases: mean age 67 yrs, 100% male
    - Controls: mean age 35 yrs, 58% women
  - Journal correspondence: Authors cite unpublished data that patterns not associated with age or sex.
    - Is that convincing?
    - Would pattern even be associated with prostate cancer in a new study?

Data generation

Specimen or data handling errors

Does anything look odd?

Pairwise correlations between gene expression profiles (arrays of thousands of gene expression measures per tumor)

ID1/ID2	1063	2094	3756	3781	4941	5725	6941	7461
1063	1	0.85	0.87	0.87	0.89	0.88	0.87	0.85
2094		1	0.88	0.86	0.85	0.87	0.88	0.87
3756			1	0.89	0.87	0.86	0.89	0.87
3781				1	0.88	0.89	0.88	0.86
4941					1	0.99	0.89	0.87
5725						1	0.87	0.89
6941							1	0.89
7461								1

(Disguised real example)

Data generation
Data handling errors

Data Set 1

ID	X1	X2	X3	X4	X5
2094	1	48	75	781	57
1063	1	34	36	686	42
5221	1	21	51	592	32
9089	1	79	66	328	52
3781	1	37	49	903	49

(Disguised real example)

Anything suspicious?

Data Set 2 = Data Set 1 + additional assay runs

ID	X1	X2	X3	X4	X5
1063	1	48	781	75	57
2094	1	34	686	36	42
3781	1	21	592	51	32
5221	1	79	328	66	52
9089	1	37	903	49	49
3756	2	291	54	569	48
4941	2	428	61	747	58
5725	2	644	42	581	63
5894	2	503	83	470	50
8503	2	743	36	655	44

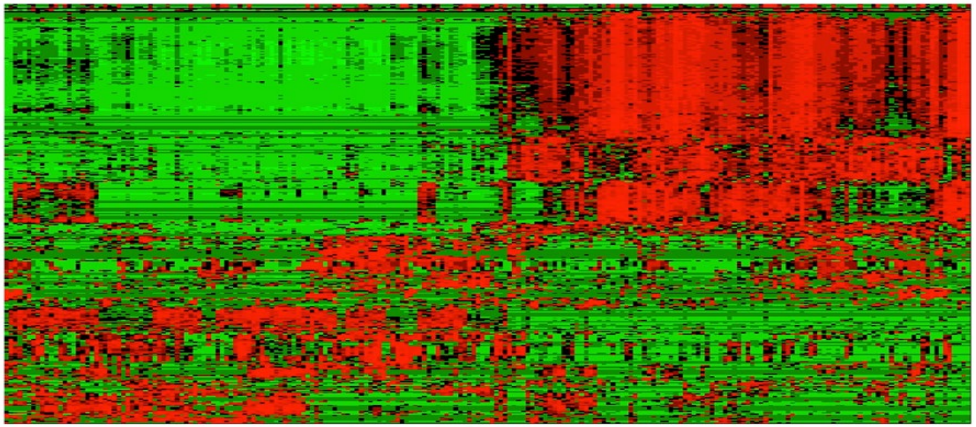
Data generation

Assay artifacts & batch effects

- Impact of changes in assay procedures, reagents, equipment, or technician during predictor development

Dramatic effect of change in RNA extraction procedure & reagents on tumor gene expression microarray profiles

Extraction method 1 ¦ Extraction method 2



215 tumor samples

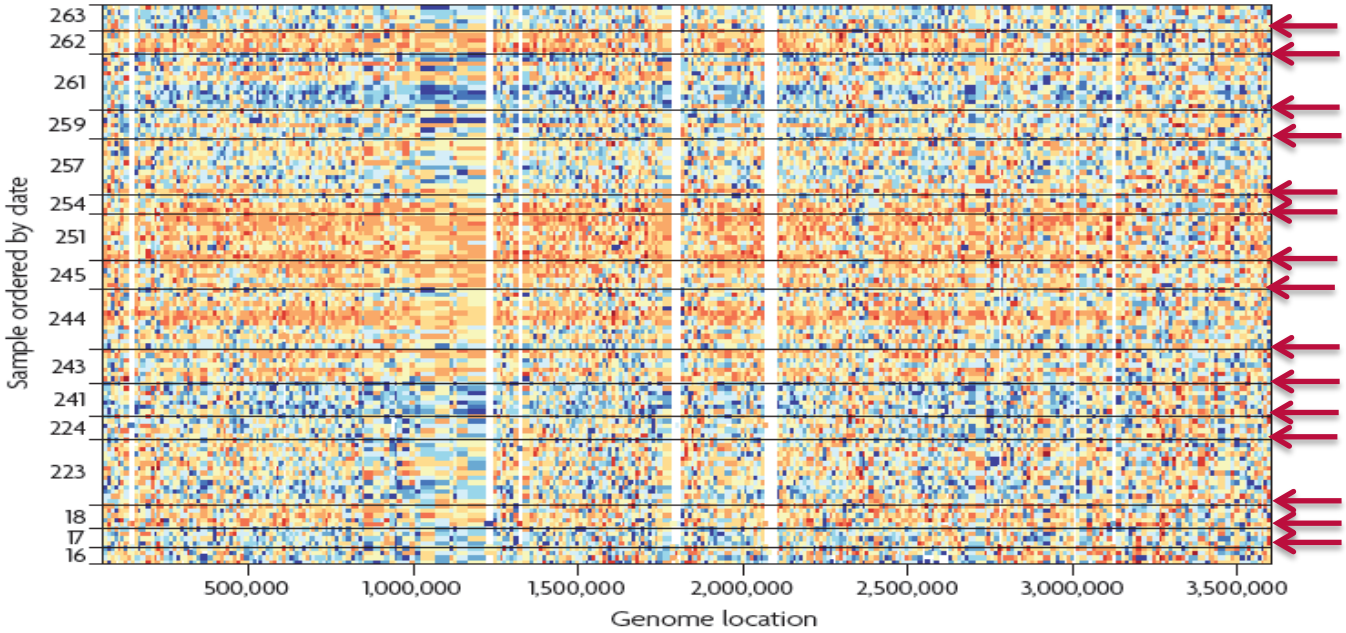
116 genes included in a genomic predictor of treatment response

(Shown with permission from an NIH grantee)

Data generation

Assay artifacts & batch effects

Example: 2nd generation sequence data from the 1000 Genomes Project. Standardized coverage data represented. Same facility, same platform.



Horizontal lines divide by date.

Figure 2 from Leek et al 2010, *Nature Rev Genet*

Data generation
Impact of changes in assay

MINDACT Trial

Cardoso F et al., *N Engl J Med* 2016;375:717-729

"A **change in the RNA-extraction solution** that was used in the calculation of the 70-gene signature (a change that was not communicated by the manufacturer) **caused a temporary shift in the risk calculation** from May 24, 2009, to January 30, 2010, at which time the issue was discovered and rectified . . .

Because of this shift, **162 patients who had been identified as being at high genomic risk were subsequently identified as being at low genomic risk** with the use of the correct . . .

The clinical effect of this risk revision was that an **additional 28 patients received chemotherapy before the results were corrected, although no patient was undertreated**."

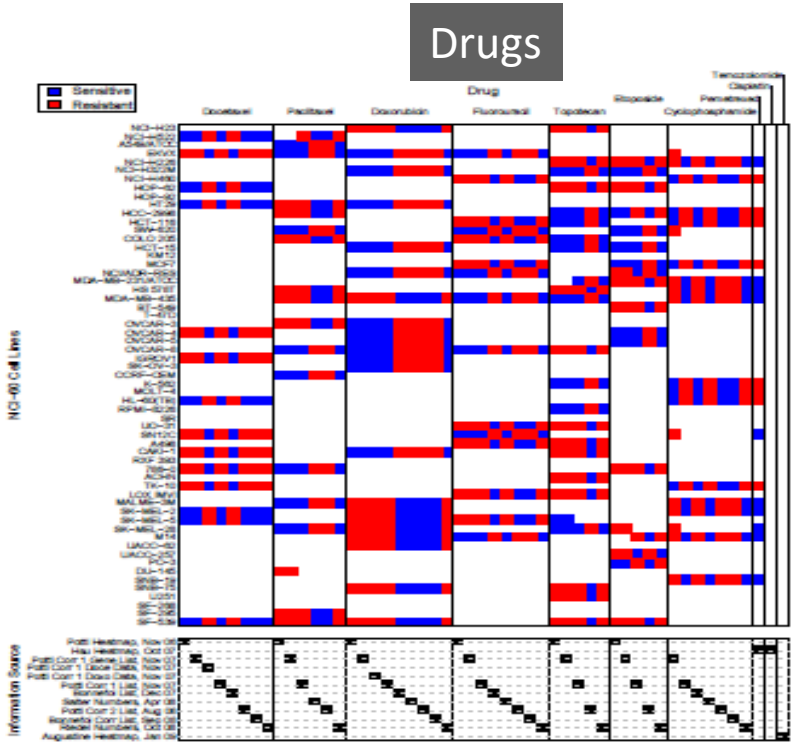
Massive data corruption

Baggerly & Coombes 2009 *Annals of Applied Statistics*

Drugs

Cell lines

Information source



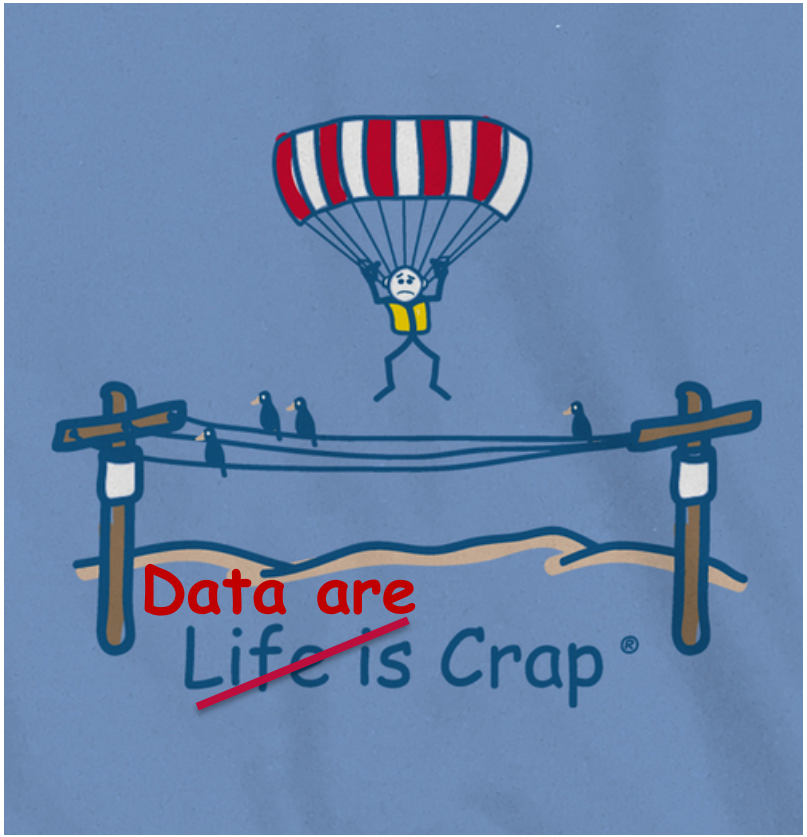
Each *block* of columns = one drug (reported across different publications)

Each row = one cell line

Within each block of columns, a given row should be either all blue (sensitive) or all red (resistant)

Several clinical trials using the Potti genomic predictors based on these data to select patient therapy were launched at Duke.

By the time of data analysis it might already be too late. . .



Data are worthless and potentially dangerous if there are major errors or “hardwired” biases (e.g., due to confounding factors).

Recommendations: Design and data generation

- ***Engagement of statisticians in the scientific process*.**
- Better training of scientists and statisticians in basic study design principles.
- Early involvement of statisticians and others with experimental design expertise early in research projects
- Ask basic questions about potential for bias and confounding (including batch effects).
- Full description and more consideration of patient characteristics and specimen sources and handling.

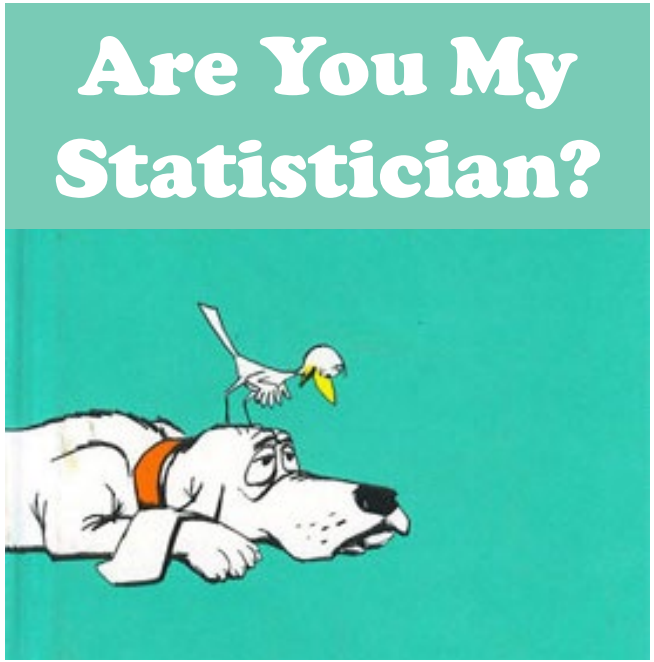
(cont →)

Recommendations: Design and data generation (cont.)

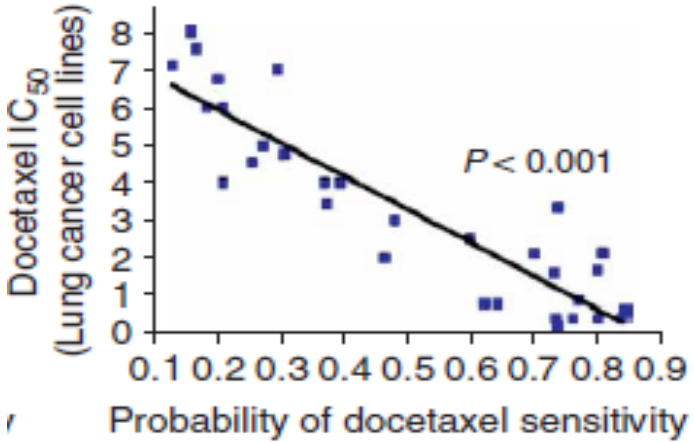
- Education on proper data management practices, including locked databases for prospective clinical studies.
- Reliable systems for data management and documentation of data provenance.
- Specifically designate qualified individuals responsible for data management.
- Better documentation of data – meta-data as well as data dictionaries defining individual variables

Data analysis

Ability to run statistical software
≠ statistical expertise



Data analysis
Appropriateness of cited statistical methods



Huh?

"The docetaxel sensitivity model developed from the NCI-60 panel again predicted sensitivity in this independent dataset, also with an accuracy exceeding 80% (P<0.001, **log-rank test**; Fig. 1c, right)." **???**
(Potti et al, *Nature Medicine* 2006)

Data analysis

Many varieties of multiple testing

- Multiple explanatory variables
- Multiple endpoints
- Multiple subgroups
- Multiple cutpoints applied to continuous variables
- Multiple models with multiple variables

Number of *independent* tests ($\alpha = 0.05$ per test)	Probability observe ≥ 1 statistically significant ($p<0.05$) result
1	0.05
2	0.10
3	0.14
4	0.19
5	0.23
6	0.26
7	0.30
8	0.34
9	0.37
10	0.40

Data analysis
Key considerations in predictor or model development

- Quality of data (clinical & omics) used to develop and validate predictor models (might not be “clinical trials grade” data)
- Appropriate statistical approaches for model/predictor development *and performance assessment*
- Meaningful “validation”
  - Define clinical context and use
    - Patient population
    - Clinical use - prognostic, predictive (treatment selection), etc.
  - “Locked down” test
  - Pre-specified performance criteria (not just a significant p-value)

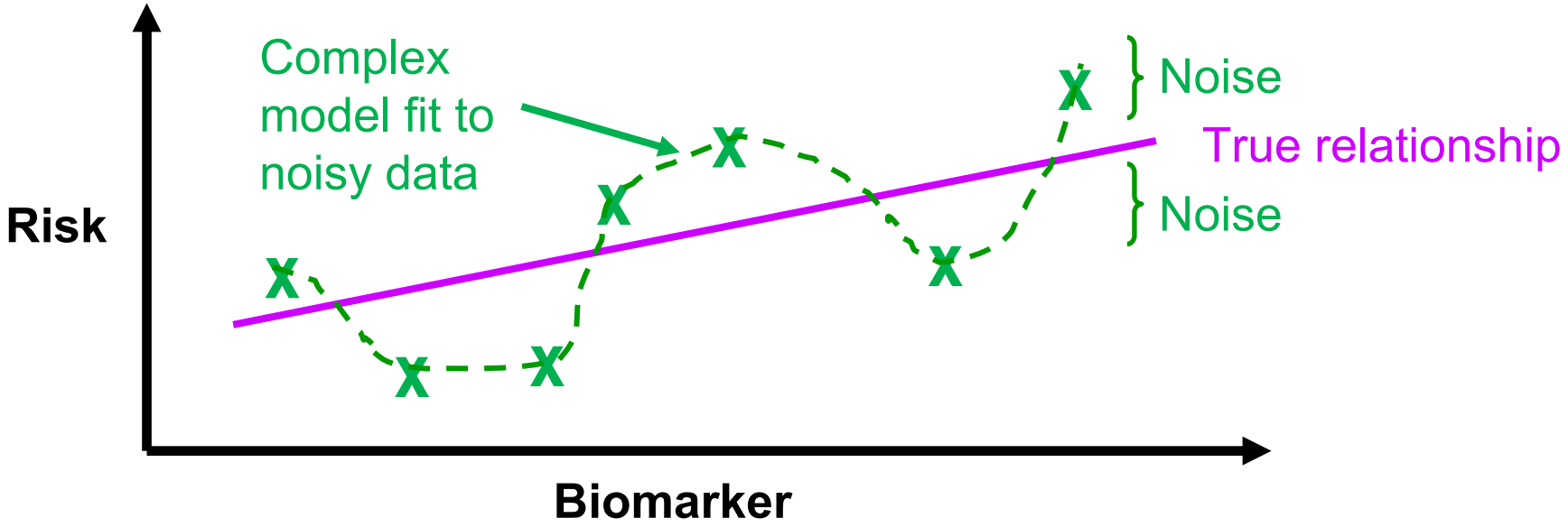
Data analysis

Common pitfalls in omics model development

- A statistical model is **OVER-FIT** when it describes random error or noise instead of the true underlying relationship
  - Excessively complex (too many parameters or predictor variables)
  - Will have poor predictive performance on independent data set
  - Naively fit omics predictors will always be overfit
- **RE-SUBSTITUTION** is the naïve evaluation of model performance by "plugging in" same data used to build it
  - Other more subtle forms of re-substitution (combining training & test, with covariates, comparative, partial)

(J Biopharm Statistics 2016;26(6):1098-1110)

Data analysis
Avoid pitfall of model over-fitting



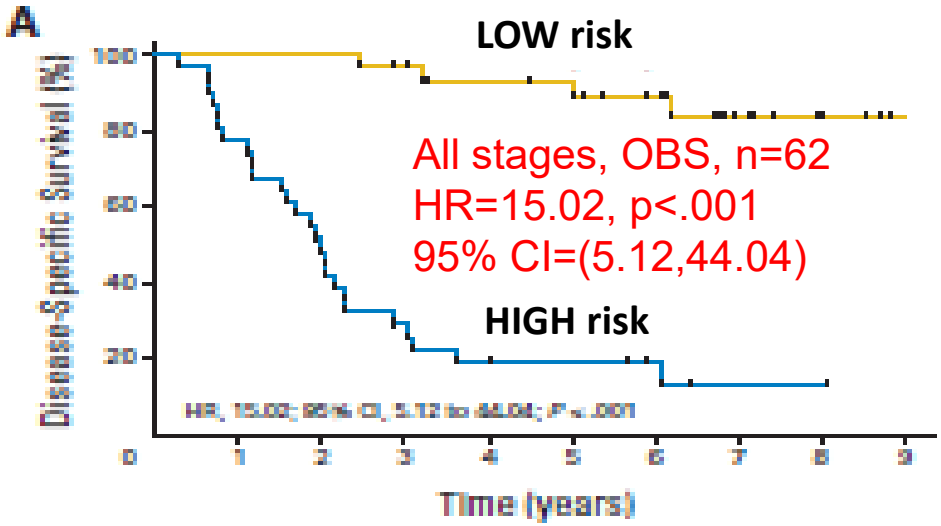
- Evaluation of a model's fit by data re-substitution will suggest fit is perfect
- In high dimensions (e.g., omics data), naively fit models are almost always over-fit and such models will rarely validate on an independent data set

Data analysis
Avoid various forms of resubstitution

- Full re-substitution (plug in exactly same data used to build predictor)
- Comparing to a re-substitution estimate
- Partial re-substitution (selection of informative variables on *full* data set with cross-validation post-selection)
- Combining training and test sets
- Resubstitution with covariate adjustment

Simon et al, *J Natl Cancer Inst* 2003;95:14-18
Subramanian & Simon, *J Natl Cancer Inst* 2010;102:464-474
Simon & Freidlin, [Correspondence] *J Natl Cancer Inst* 2012;103(5):445
Subramanian & Simon, *Contemporary Clinical Trials* 2013;36:636–641
Sachs & McShane, *J Biopharm Statistics* 2016;26(6):1098-1110

Data analysis
Avoid re-substitution



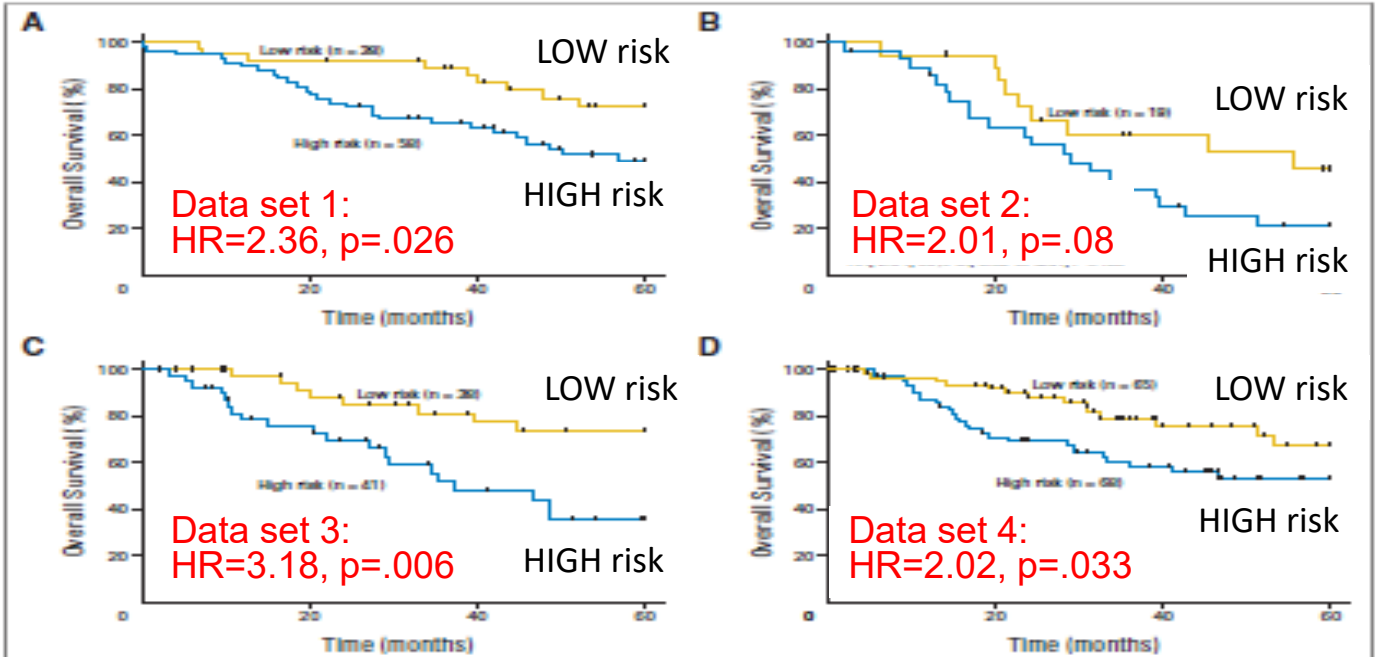
(J Clin Oncol 2010;28:4417-4424)

"A 15-gene signature *[for lung cancer]* separated OBS patients *[no chemotherapy after surgery]* into high-risk and low-risk subgroups with significantly different survival (hazard ratio [HR], 15.02; 95% CI, 5.12 to 44.04; *P <.001.*"

RE-SUBSTITUTION!

If this large separation in survival curves was real, the signature would have clinical utility. Patients designated as low risk could confidently avoid toxic chemotherapy.

Data analysis: What was validated?



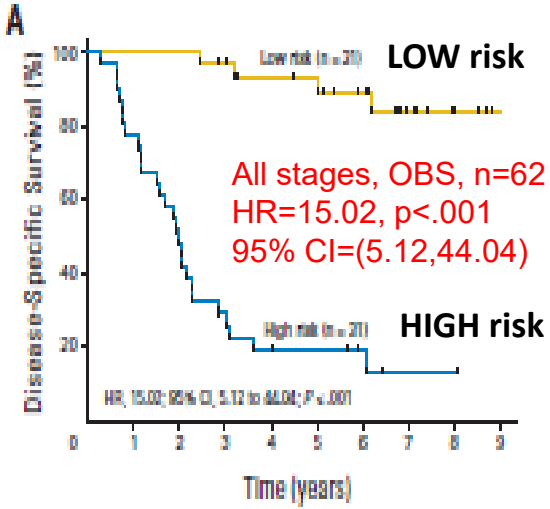
(J Clin Oncol 2010;28:4417-4424)

"... prognostic effect *[of 15-gene signature]* was **VALIDATED** consistently in four separate microarray data sets (total 356 stage IB to II patients without adjuvant treatment)."

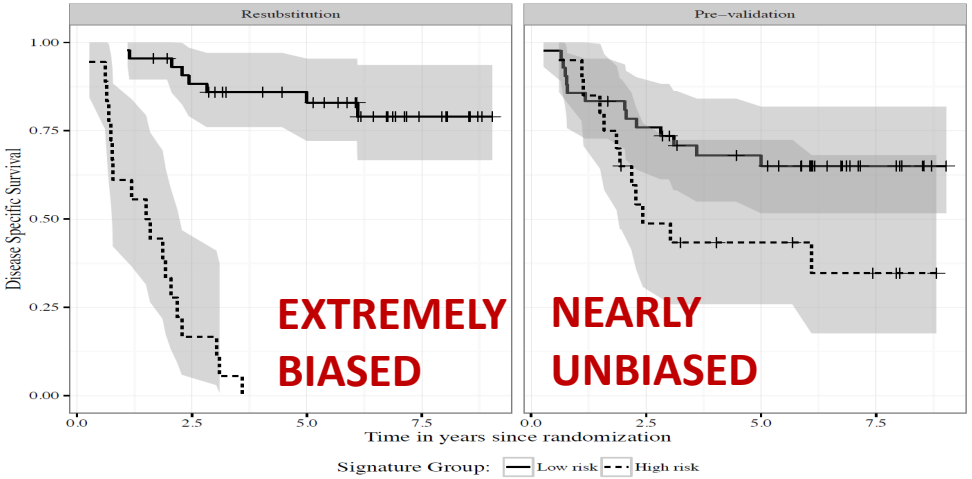
Endpoint: Disease-specific survival (DSS) → Overall survival (OS)
Timescale: 0 to 9 years → 0 to 60 months (5 years)
HR: 15.02→ ≈ 2-3 5-yr DSS ≈ 90% → 5-yr OS < 80%
Mixture of disease stages? Adjustment for standard covariates?

Data analysis

Use internal validation during model development



Original Kaplan-Meier curves (DSS) showing prognostic ability of 15-gene signature in OBS arm, using re-substitution
(J Clin Oncol 2010;28:4417-4424)

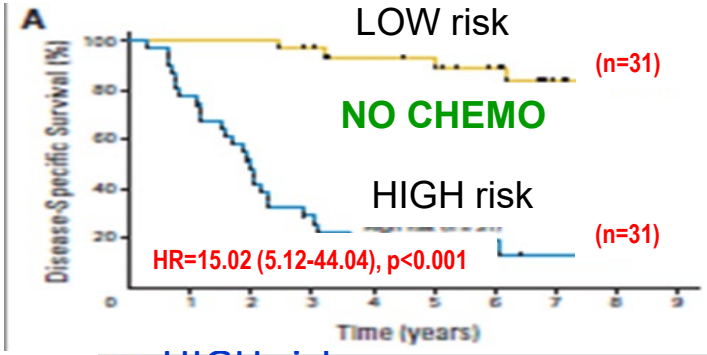


RE-SUBSTITUTION CROSS-VALIDATION

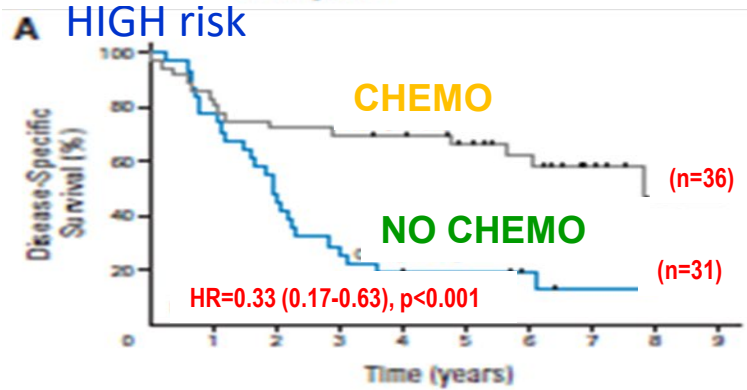
Reproduced (approx.) Kaplan-Meier curves (DSS) showing prognostic ability of 15-gene signature in OBS arm, using re-substitution (A) and cross-validation (B)

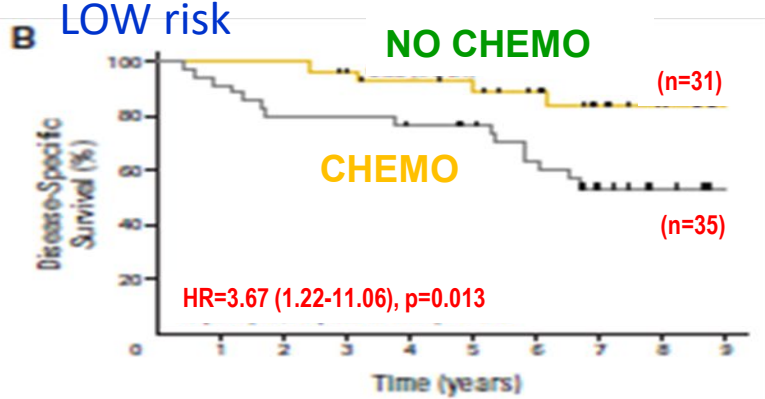
(J Biopharm Statistics 2016;26(6):1098-1110)

Data analysis:
Avoid *comparisons* with resubstitution estimates



Prognostic classifier fit using gene expression microarray data from clinical trial arm on which patients received **no adjuvant chemotherapy** (resubstitution)





Does the genomic predictor identify groups of patients who benefit differently from adjuvant chemotherapy? **Can't conclude anything.**

Data analysis
Avoid partial re-substitution

- Published example: "Metabolomic detection of early-stage ovarian cancer"
- **100% accuracy in cohort** (46 early stage (I/II) serous epithelial ovarian cancer (EOC) patients and 49 age-matched normal healthy controls)
- "Using all 255 metabolic features, a first SVM model was generated . . . **(accuracy 62%; specificity 57%; sensitivity 67%)**. Since SVM models built upon large datasets typically contain uninformative features . . . We employed . . . . . .RFE method to **select features that distinguished the early-staged EOC samples from controls with optimal accuracy.** . . . **100% accuracy . . .** obtained with . . . **16 features. . . accuracy of these 16 metabolites was independently validated** by orthogonal partial least squares-discriminant analysis (oPLS-DA) **using a variety of cross-validation approaches**" (details in Supplement)

Data analysis

Avoid combining training & test sets

Multivariable Model for OS **(Training and Test sets *combined*)**

Variable	HR	95% CI	P
Genomic score	2.43	1.94 – 3.06	< 0.001
Stand. molec. factor 1	1.77	1.41 – 2.22	< 0.001
Stand. molec. factor 2	0.66	0.48 – 0.93	0.02
Age group, ≥ 60 yrs vs < 60 yrs	2.22	1.76 – 2.79	< 0.001

- Combining Training data (*used to develop genomic score*) with Test data destroys the validation and interpretability of the adjusted effects
- Resubstitution with covariate adjustment: Nowhere in the paper was a multivariable analysis based solely on the Test set presented.

Requirements for rigorous predictor validation

- Must be able to **COMPLETELY SPECIFY** a **LOCKED-DOWN** predictor or algorithm
  - The lockdown includes *all steps* in the data pre-processing and prediction algorithm (including computer code).
  - A gene list alone does not specify a predictor
- Must be able to apply the predictor to **ONE INDIVIDUAL/PATIENT AT A TIME**

(cont →)

Requirements for rigorous predictor validation (cont.)

- Need **PRE-SPECIFIED PERFORMANCE CRITERIA**.
  - Not just a significant p-value!
- Ideally, **INDEPENDENT VALIDATION DATA** generated from specimens collected at a different time, or in a different place, and according to the pre-specified collection protocol.
- Assays for the validation specimen set should be run at a different time or in a different laboratory according to the **PRE-SPECIFIED ASSAY** protocol (including quality rejection criteria).

(cont →)

Requirements for rigorous predictor validation (cont.)

- Individuals who developed/have interest in the predictor must remain completely **BLINDED** to the validation data.
- The validation **DATA SHOULD NOT BE CHANGED** and **DATA VALUES SHOULD NOT BE SELECTIVELY ELIMINATED** after observing the performance of the predictor.
- **PREDICTOR SHOULD NOT BE ADJUSTED** (including cut-points) after its performance has been observed on any part of the validation data. Otherwise, the validation is compromised and a new validation may be required.

Recommendations: Data analysis

- Individuals with adequate statistical and bioinformatics expertise should be engaged in the research
- Understand where the data came from and how they might have been preprocessed or filtered
- Run quality checks on data prior to analysis
  - Inconsistent or illogical values
    - Examples: Disease relapse after death, male with ovarian cancer
  - Values out of expected range
    - Often indicate mix-up of measurement units or missing data codes
- Define clinically meaningful goals and pre-specify appropriate performance criteria

Results reporting and interpretation

Example: "Diagnostic markers for early detection of ovarian cancer"

- Six proteins are used to compare the plasma from ovarian cancer cases and healthy controls.
- Claim: 95.3% sensitivity and 99.4% specificity
- Patient characteristics:
  - Cancer cases: High risk of cancer, with masses
  - Controls: Healthy, seen in screening clinic
- Markers include "stress" proteins that could differ in compared groups.
  - Explanation for discrimination ability?

Results reporting and interpretation

"Diagnostic markers for early detection of ovarian cancer" (continued)

To the Editors: <Authors> "claim the ability to detect ovarian cancer early and with **95.3% sensitivity and 99.4% specificity**. Several **serious methodologic issues** lead us to conclude that these figures are **greatly exaggerated**. The **training set** specimens, derived from one cohort, were used to **fit several classifiers**, and the test set specimens. . .were used to estimate their performance. The accuracy reported in their conclusion, however, was determined from the **combined data and from the classifier that did best in the test set**. This violates fundamental principles of statistical analysis, . . . Had they properly . . . they would have had to report a lower sensitivity of either 84% or 88% at lower 95% specificity. The analysis they chose to highlight is **inappropriate and misleading**."

Results reporting and interpretation

"Diagnostic markers for early detection of ovarian cancer" (continued)

To the Editors: "The published report is noteworthy for the performance characteristics . . . based on the combined training and test sets, all ovarian cancers combined: **sensitivity, 95.3%; specificity, 99.4%; positive predictive value (PPV), 99.3%; and negative predictive value, 99.2%.** However, the PPV estimate of 99.3% . . . based on a prevalence of ovarian cancer near 50%. The **prevalence of ovarian cancer in any screened population will be much smaller than 50%. . . .** correction **. . .** assumed that the prevalence of ovarian cancer in the screened population was 1 of 2,500 (0.04%) and **recalculated the PPV to be only 6.5%**. . . . **only 1 in 15 women with a positive test result will, in fact, have ovarian cancer**

Given that this assay is currently **being marketed** to health care providers and consumers as a validated ovarian cancer screening test, this **difference is not academic**."

Recommendations: Results reporting and interpretation

- Term “validation” should not be used unless accompanied by appropriate explanation of what is being validated
- Approaches ensuring rigor of the validation should be described (e.g., blinding, honest broker)
- Journals and funding agencies need to ensure that there are qualified statistical reviewers for manuscripts and grant applications

Results dissemination

Once a “desirable” (but wrong) result has been obtained by a flawed analysis approach or data dredging, it’s hard to pull it back.

Results dissemination

"...we can now predict with high accuracy which patients may benefit most from chemotherapy."

Genomic Clinical Trials

Studies focus on the personalized care of patients

Physician-researchers from the Duke Comprehensive Cancer Center and Duke's Institute for Genome Sciences & Policy are leading two new national clinical trials that use a patient's cancer genome to determine what treatment will likely be the most successful for that patient. One study will test whether chemotherapies already FDA approved for late-stage lung cancer are effective for certain early-stage lung cancer patients, too. The other study will use a genomic test developed at Duke to compare the effectiveness of an FDA-approved therapy with an investigational therapy for men with advanced prostate cancer.

"It's difficult for physicians to know which patients will benefit from which treatments," explains Duke surgeon David Harpole, MD, lead investigator of the trial for early-stage lung cancer patients. "Using a genomic signature from a patient's lung cancer, we can now predict with high accuracy which patients may benefit most from chemotherapy." Genome is defined as the collection of a person's genetic information from a person or his or her tumor.

"...we can now predict with high accuracy which patients may benefit most from chemotherapy."
David Harpole, MD

Harpole and his colleagues across the country will accrue 1500 early-stage lung cancer patients who are being treated at major medical centers in the United States. The trial is one of the largest trials ever conducted for early-stage lung cancer patients and the first to use genomics to determine which patients will likely benefit from receiving chemotherapy.

"Forty percent of patients with early-stage lung cancer die within five years," says Harpole. "It's unacceptable that so many people die, even when the disease is caught early."

continued on page 3

What accuracy measure?
Sensitivity? PPV???
75%? 80%?

Some predictors were developed on cell lines or different tumor types.

Genomic predictors generated advertising, patent applications, diagnostics company start-ups, consulting arrangements

Results dissemination

Dozens of papers from Potti group published in top journals

Nature Medicine 2006

nature medicine

Genomic signatures to guide the use of chemotherapeutics

Anil Potti[1,2], Holly K Dressman[1,3], Andrea Bild[1,3], Richard F Riedel[1,2], Gina Chan[4], Robyn Sayer[4], Janiel Cragun[4], Hope Cottrill[4], Michael J Kelley[2], Rebecca Petersen[5], David Harpole[5], Jeffrey Marks[5], Andrew Berchuck[1,6], Geoffrey S Ginsburg[1,2], Phillip Febbo[1–3], Johnathan Lancaster[4] & Joseph R Nevins[1–3]

NEJM 2006

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Genomic Strategy to Refine Prognosis in Early-Stage Non–Small-Cell Lung Cancer

Anil Potti, M.D., Sayan Mukherjee, Ph.D., Rebecca Petersen, M.D., Holly K. Dressman, Ph.D., Andrea Bild, Ph.D., Jason Koontz, M.D., Robert Kratzke, M.D., Mark A. Watson, M.D., Ph.D., Michael Kelley, M.D., Geoffrey S. Ginsburg, M.D., Ph.D., Mike West, Ph.D., David H. Harpole, Jr., M.D., and Joseph R. Nevins, Ph.D.

VOLUME 25 · NUMBER 28 · OCTOBER 1 2007

JOURNAL OF CLINICAL ONCOLOGY ORIGINAL REPORT

J Clin Oncol 2007

Pharmacogenomic Strategies Provi
to the Treatment of Cisplatin-Resist
Advanced Cancer

*David S. Hsu, Bala S. Balakumaran, Chaitanya R. Acharya, Van
Katherine Garman, Carey Anders, Richard F. Riedel, Johnathan
Joseph R. Nevins, Phillip G. Febbo, and Anil Potti*

VOLUME 25 · NUMBER 5 · FEBRUARY 10 2007

JOURNAL OF CLINICAL ONCOLOGY ORIGINAL REPORT

J Clin Oncol 2007

An Integrated Genomic-Based Approach to Individualized Treatment of Patients With Advanced-Stage Ovarian Cancer

Holly K. Dressman, Andrew Berchuck, Gina Chan, Jun Zhai, Andrea Bild, Robyn Sayer, Janiel Cragun, Jennifer Clarke, Regina S. Whitaker, LiHua Li, Jonathan Gray, Jeffrey Marks, Geoffrey S. Ginsburg, Anil Potti, Mike West, Joseph R. Nevins, and Johnathan M. Lancaster

Recommendations: Dissemination

- As a requirement for access to resources such as specimens and funding for research, there should be a commitment to report results completely and transparently, *regardless of findings*
  - Avoid publication bias and selective reporting
- Data and computer code should be made available for others to examine

Accountability

Jan. 27, 2011: After initiation of a research misconduct investigation, the Duke vice-chancellor for clinical research and the head of the Institute for Genome Science and Policy jointly send a letter to all co-authors of Potti:

"In keeping with our institutional commitment and mandate to maintain public trust, . . . **assure that you can identify the person or persons responsible for the data management, statistical analysis, and interpretation of the results**.

Based on the requirements for authorship, we ask you to **attest that you are confident that these elements of the manuscript are appropriate, accurate, and free of improper manipulation**.

If you cannot do so, we will work with you to reach the point of either assuring that the paper and its results are reasonable or retracting the article. . . In order to ensure that we as an institution as well as others in the scientific community can have confidence in the integrity of these papers, **we will select a small number at random for a detailed review**."

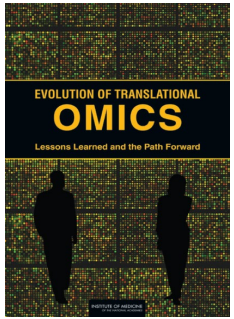
Consequences of poor practices in omics research

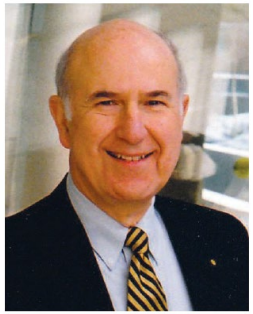
- Possible harm to patients if flawed omics-based test is used in a clinical trial or in clinical practice
- Waste of research effort and resources
- Damage to professional reputations, including for “innocent” members of lab or research team
- Lawsuits – patients or others misled about value of the research
- Prosecution for violation of regulatory requirements for use of an investigational medical device

Recommendations: Accountability & responsibility

- In any study, identify qualified individuals who will be accountable for laboratory work (specimens and assays), clinical data collection, data management, statistical and bioinformatic analyses, and interpretation of the results.
- Identify in manuscripts submitted for publication the specific contributions made by each author.
- Ensure that omics predictors to be used in clinical trials or to guide patient care undergo sufficient independent review and trials receive proper oversight (institutional and regulatory) as routinely expected for drug trials.

Recommended reading





"There are a lot of lessons here that surely apply to other places."
—GILBERT S. OMENN, UNIVERSITY OF MICHIGAN, ANN ARBOR

http://www.iom.edu/Reports/2012/Evolution-of-Translational-Omics.aspx

NCI criteria for the use of omics-based predictors in clinical trials:

McShane et al. *Nature* 2013;502:317-320 (checklist)

McShane et al. *BMC Medicine* 2013;11:220 (explanation & elaboration)

Nearing completion: TG9 overview paper "Statistical analysis of high-dimensional biomedical data: A gentle introduction to analytical goals, common approaches and challenges"

THANK YOU!

www.cancer.gov

www.cancer.gov/espanol