Towards a More Rigorous Approach to Personalized Medicine

Frank E Harrell Jr

Department of Biostatistics
Vanderbilt University School of Medicine

Division of Biostatistics and Health Services Research
Department of Quantitative Health Sciences
UMass Medical School  N Worcester MA  15 Oct 12
1. Types of Personalization
2. Doing Simple Things First
3. Clinical Trial Designs
4. $n$-of-one Trials
5. What’s Gone Wrong with Omics & Biomarkers?
6. Challenges of Molecular Signatures
7. Reproducible Research
8. Summary
Types of Personalization

- Choice and sequencing of diagnostic procedures
- Choice of medical vs. surgical therapy
- Timing of interventions
- More rapid treatment in critical situations (e.g., t-PA for ischemic stroke; better initial diagnosis)
- Choice of drugs, dose and dose interval
- Optimizing compliance
- Concomitant therapies
Types of Personalization, continued

- Real-time diagnostic driven (post-CABG cardiac cath) vs. use of only initial diagnostic measurements
- Choice of variables to monitor in follow-up
- Follow-up interval
- Treatment duration
- Home vs. clinic vs. hospital-based treatment
- Access to care
- Listening
Modes of Differentiation

- Low– vs. high–dimensional
- Phenotype
- Genotype
- Moneytype
Information Used to Differentiate

- Patient preferences
- Static baseline vs. dynamic monitoring
- Demographics
- Vital signs
- Symptoms and physical exam
- Risk factors
- Disease stage/severity or post-surgical residual
- Validated biomarkers
- Unspecified molecular signatures
Carriers for loss-of-function CYP2C19 alleles: reduced conversion of clopidogrel to active metabolite

Suggested that clop. less effective in reducing CV death, MI, stroke

12,562 (clop. HR 0.8); 5059 genotyped (clop. HR 0.7)

<table>
<thead>
<tr>
<th></th>
<th>Carrier</th>
<th>Non-Carrier</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>0.69 (0.49, 0.98)</td>
<td>0.72 (0.59, 0.87)</td>
</tr>
<tr>
<td>Ratio of HRs</td>
<td>0.96 (0.64, 1.43)</td>
<td><em>(P = 0.8)</em></td>
</tr>
</tbody>
</table>

Paré et al. [2010]
A Pharmacogenomics Conversation

Dr. X: a nephrologist
Dr. Y: a clinical pharmacologist
Dr. Z: a biostatistician
Do Simple Things First

- Dose by weight
- Check for differential drug benefit (interaction between weight and treatment)
  - This checks for proper dosing
- Harvest clinical trial data better to learn about interactions between patient characteristics and treatments
Differential Treatment Effects

- Most common: expansion of absolute benefit with severity of disease
- Value of absolute effectiveness estimation
  - mathematical necessity
  - relation to cost-effectiveness
- Quantitative interactions
- Qualitative interactions
Absolute Benefit vs. Baseline Risk

Model for Probability of Death Within 30 Days in GUSTO-I

After Califf et al. [1997]
Absolute Benefit vs. Baseline Risk, continued

![Graph showing Absolute Risk Reduction with t-PA vs. Baseline Expected Risk with SK. The graph includes a mean and median line.]
Individual Response to Therapy: Assumptions To Check

- Do patients agree with themselves?
  - role of crossover studies and serial data
- What single-period/patient studies cannot tell us:
  - Does a drug give a small benefit to everyone or does it give a large benefit to a few?
### Sources of Variation in Clinical Trials

<table>
<thead>
<tr>
<th>Label</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Between treatments</td>
<td>Mean treatment effect</td>
</tr>
<tr>
<td>B</td>
<td>Between patients</td>
<td>Variation in Tx effect across Pts</td>
</tr>
<tr>
<td>C</td>
<td>Pt $\times$ Tx</td>
<td>Variation in within-Pt Tx effect over periods</td>
</tr>
<tr>
<td>D</td>
<td>Within-Pt error</td>
<td></td>
</tr>
</tbody>
</table>

Senn [2001]
### Estimable Effects in Clinical Trials

<table>
<thead>
<tr>
<th>Type of Trial</th>
<th>Description</th>
<th>Identifiable Effects</th>
<th>Error Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel</td>
<td>1 Tx/Pt</td>
<td>A</td>
<td>B + C + D</td>
</tr>
<tr>
<td>2 Per X-over</td>
<td>1 Per/Tx/Pt</td>
<td>A, B</td>
<td>C + D</td>
</tr>
<tr>
<td>≥ 4 Per X-over</td>
<td>≥ 2 Per/Tx/Pt</td>
<td>A, B, C</td>
<td>E</td>
</tr>
</tbody>
</table>

Senn [2001, 2009]

![Graph showing Difference to placebo in DBP mmHg](image)
Strategy for Rapidly-Assessed Endpoints

- 4- or 6-period 2-treatment crossover study
- Variability = treatment, patient, patient \( \times \) treatment interaction
- patient \( \times \) treatment = individualized response to therapy
- People are more than their genes \( \rightarrow \) patient \( \times \) treatment interaction is upper bound on genome \( \times \) treatment interaction, summing over the entire genome
- Patient effect huge or patient \( \times \) treatment interaction is tiny \( \rightarrow \) search for a genomic signature is futile
- Otherwise, do genetic analysis to predict the within-patient treatment response

Senn [2001, 2004]
For chronic diseases with quickly assessed outcomes, can run a series of trials for the patient at hand
- Select the Tx that seems to work best
- But optimal efficacy estimates still require borrowing information across heterogeneous pts
Forty-five Trials Per Patient May Be Insufficient

Efron and Morris [1977]
Hackam and Redelmeier [2006]: Translation of research evidence from animals to humans

Screened articles having preventive or therapeutic intervention in in vivo animal model, > 500 citations

76 “positive” studies identified

Median 14 years for potential translation

37 judged to have good methodological quality (flat over time)

28 of 76 replicated in human randomized trials; 34 remain untested

↑ 10% methodology score ↑ odds of replication × 1.28 (0.95 CL 0.97–1.69)

Dose-response demonstrations: ↑ odds × 3.3 (1.1–10.1)

Note: The article misinterpreted $P$-values
Scientists at Amgen tried to confirm published findings related to a line of research, before launching development.

- Identified 53 ‘landmark’ studies
- Scientific findings confirmed in only 6 studies
- Non-reproduced articles cited far more frequently than reproduced articles

A Sharp Rise in Retractions Prompts Calls for Reform

By CARL ZIMMER

In the fall of 2010, Dr. Ferric C. Fang made an unsettling discovery. Dr. Fang, who is editor in chief of the journal Infection and Immunity, found that one of his authors had doctored several papers.

It was a new experience for him. “Prior to that time,” he said in an interview, “Infection and Immunity had only retracted nine articles over a 40-year period.”

The journal wound up retracting six of the papers from the author, Naoki Mori of the University of the Ryukyus in Japan. And it soon became clear that Infection and Immunity was hardly the only victim of Dr. Mori’s misconduct. Since then, other scientific journals have retracted two dozen of his papers, according to the watchdog blog Retraction Watch.

April 16, 2012
BMJ 1994; 308: 283 (Published 29 January 1994)

Editorial

The scandal of poor medical research

D G Altman

We need less research, better research, and research done for the right reasons.

What should we think about a doctor who uses the wrong treatment, either wilfully or through ignorance, or who uses the right treatment wrongly (such as by giving the wrong dose of a drug)? Most people would agree that such behaviour was unprofessional, arguably unethical, and certainly unacceptable.

What, then, should we think about researchers who use the wrong techniques (either wilfully or in ignorance), use the right techniques wrongly, misinterpret their results, report their results selectively, cite the literature selectively, and draw unjustified conclusions? We should be appalled. Yet numerous studies of the medical literature, in both general and specialist journals, have shown that all of the above phenomena are common. This is surely a scandal.
What’s Gone Wrong with Omics & Biomarkers?

- Subramanian and Simon [2010]: Gene expression-based prognostic signatures in lung cancer: Ready for clinical use?
- NSCLC gene expression studies 2002–2009, \( n \geq 50 \)
- 16 studies found
- Scored on appropriateness of protocol, stat validation, medical utility
- Average quality score: 3.1 of 7 points
- No study showed prediction improvement over known risk factors; many failed to validate
- Most studies did not even consider factors in guidelines
  - Completeness of resection only considered in 7
  - Similar for tumor size
  - Some only adjusted for age and sex
Biases might pose a special challenge for laboratory researchers who are used to biological reasoning and the tightly controlled conditions of experimental research. Such researchers unwittingly become non-experimental observational epidemiologists when they apply molecular assays in studies of diagnosis and prognosis, for which the experimental method is not available and for which biological reasoning might have limited usefulness.

Ransohoff [2005]
Authors of 108 studies who were laboratory scientists were 19-fold more likely to over-interpret the clinical utility of molecular diagnostic tests compared with clinic-based authors.

Lumbreras et al. [2009]
Each new field has a rapid exponential growth of its literature over 5–8 years (“new field phase”), followed by an “established field” phase when growth rates are more modest, and then an “over-maturity” phase, where the rates of growth are similar to the growth of the scientific literature at large or even smaller. There is a parallel in the spread of an infectious epidemic that emerges rapidly and gets established when a large number of scientists (and articles) are infected with these concepts. Then momentum decreases, although many scientists remain infected and continue to work on this field. New omics infections continuously arise in the scientific community.

Ionnidis [2010]
Biomarker Discoveries

Izvestia  (News)

Pravda  (Truth)

Big Effects  Validated Effects
Challenges of Molecular Signatures

- Signatures should explain a large amount of variation in outcomes
- Highest potential for pharmacogenomics when treatment has major effect on a substantial portion of the target population
- No general advantage of one type of information over another
  Most important: speed and low cost
Biology is Complex

- Goldstein’s estimate of the number of SNPs needed to explain 0.8 of variation in height: 93,000
- Yang et al: 295,000 SNPs considered simultaneously explain 0.45 of variation
- Heisenberg’s uncertainty principle: it is impossible to determine simultaneously both the position and momentum of a particle with great accuracy
- Molecular signature uncertainty principle: a signature can be parsimonious or very predictive but seldom both

Goldstein [2009] and Yang et al. [2010]
Disease incidence data from huge number of identical twins
24 diseases with high public health impact
Estimated absolute upper bound on ability of whole genome sequencing to identify individuals with substantially lower or higher risk of one of the diseases
“Genetic testing, at best, will not be a substitute for preventative medicine strategies incorporating routine checkups and risk management based on the history, physical status and lifestyle of the patient.”

“nature” or the experimental outcome chooses—to go to the right branch or the left; at the next fork, to go left or right; and so on. There are similar branch points in a “conditional computer program,” where the next move depends on the result of the last calculation. And there is a “conditional inductive tree” or “logical tree” of this kind written out in detail in many first-year chemistry books, in the table of steps for qualitative analysis of an unknown sample, where the student
Strong (Inductive) Inference, *continued*

- Devise alternative hypotheses
- Devise an experiment with alternative possible outcomes each of which will exclude a hypothesis
- Carry out the experiment
- Repeat
- Regular, explicit use of alternative hypotheses & sharp exclusions → rapid & powerful progress
- “Our conclusions . . . might be invalid if . . . (i) . . . (ii) . . . (iii) . . . We shall describe experiments which eliminate these alternatives.”
- Rushton: “A theory which cannot be mortally endangered cannot be alive.”

Platt [1964]
Does Anyone Still Care About Experimental Design?

...hypotheses are often detached from data collection, experimental design, and causal theories. ... We advocate for reinterpretation of the scientific method ... and for renewed appreciation of falsifiable hypotheses, so that we can learn more from our best mistakes.

... the GWAS research community has too often accommodated bad experimental design with automated post-experiment cleanup.

...major batch effects in all but 2 of 30 public and private GWAS ... the dependent variable ... was correlated with the order of sample collection and/or the order in which samples were batch-processed ...

Lambert and Black [2012]
Religion is a culture of faith; science is a culture of doubt.

Science is the belief in the ignorance of experts.  
 Richard Feynman

Fiction is about the suspension of disbelief; science is about the suspension of belief.

A true scientist is bored by knowledge; it is the assault on ignorance that motivates him.  
 Matt Ridley
Need for Validation & Sensitivity Analysis

- Unbiased validation of predictions
- Demonstrate dose-response
- Alternate explanations of molecular signatures by other “losing” molecular markers
- Alternate explanations of molecular signatures by combinations of clinical variables
- Is a new marker summative?
- Is it unique?
Non-Reproducible Research

- Tweaking instrumentation
- Pre-statistician “normalization” of data and background subtraction
- Poorly studied high-dimensional feature selection
- Programming errors
- Lack of documentation
- Failing to script multiple-step procedures
- Copying and pasting results into manuscripts
- Insufficient detail in scientific articles
- No audit trail
Non-Reproducible Research, continued
Reproducible Research

- **Imperative** to be able to reproduce all results
- **All steps** must be scripted
- Sweave or knitr = R + \texttt{\LaTeX} → pdf is one excellent tool
Some Random Thoughts

Kelvin’s curse: The unthinking and inappropriate worship of quantifiable information in medicine

Feinstein [1977]

...monetization of intellectual property appears to be a powerful force favoring methodological limitations and an excessive reductionism and fragmentation of biologic knowledge

Porta et al. [2007]

There is nothing wrong with cancer research that a little less money wouldn’t cure.

Nathan Mantel, NCI
Summary

- Medicine can be personalized in many ways including use of traditional information
- Personalized medicine is easy to do poorly, especially in exploratory mode
- There are great risks for
  - non-replicable results & non-unique signatures
  - customized treatment effect estimates that are less accurate than averages from RCTs
- Biostatisticians and clinical epidemiologists play important roles in
  - assessing the needed information content for a given problem complexity
  - minimizing bias
  - maximizing reproducibility
There are Many Faces to Personalized Medicine

This work used only free software

\LaTeX\,\ P gorgeous GNU/Linux\,\ R


