

Challenges of Using High-Throughput Assay Data to Improve Personalized Medicine

BST 226

Statistical Methods for Bioinformatics

David M. Rocke

Personalized Medicine

- Most medical research is focused on groups of patients with (supposedly) similar conditions.
- An example would be a clinical trial to help decide if chemotherapy is helpful after surgery for stage I and stage II gastric cancer patients.
- Personalized medicine would mean making this decision for each patient separately, based on (usually) genetic, genomic, or biomarker data.
- Although this sounds like a good idea, implementation is very difficult and has been achieved only in a few cases.

Breast Cancer

- Mutations in the tumor suppressor genes BRACA₁ and BRACA₂ are associated with elevated risk of breast cancer.
- Breast tumors can be evaluated by several markers that affect treatment and prognosis.
- Estrogen/progesterone receptor (ER/PR) tumors can be treated with tamoxifen or aromatase inhibitors.
- Her2/neu positive can be treated with Herceptin and related drugs.
- Triple negative tumors can be treated only with standard, less targeted chemotherapy.

Personalized Medicine Successes

- Treatment of breast cancer depends not only on the stage, but also on genetic characteristics of the tumor.
- This type of personalized medicine has been successful in cases in which the molecular mechanisms are well understood.
- Other successes come from sequencing the tumor genome and looking for mutations in genes that relate to specific therapies.

The Search for Criteria

- Much of the research in personalized medicine depends on gathering large amounts of information about patients and correlating the information with outcome, with response to therapy, or with diagnosis.
- For example, one may test serum or tissue samples using a gene expression array, which can have as many as 50,000 measurements.
- Sequencing transcripts or the genome of the patient or of a tumor also can yield tens of thousands of measurements.

The Problem of Scale

- Oncotype DX is a proprietary predictive signature based on the levels of 21 transcripts which supposedly predicts the risk of recurrence in breast cancer.
- It was derived from 250 candidate genes in a trial of 400 patients. The number of sets of 21 genes from 250 is 2×10^{30} , which is a large number (larger than a trillion trillion, which is only 2×10^{24})
- The number of gene sets of 21 or fewer is larger by a factor of 2 million
- If the genes had been selected from the full array of 25,000 to 50,000, the number is 10^{72} to 10^{79} , nearly the number of particles in the universe.

Biomarkers for Personalized Medicine

- Without exercising very great care, when one starts with 50,000 measurements on each patient, there is a large chance of accidental correlations that will not extend to a new group of patients, and is thus useless.
- We understand much about methods for preventing false discoveries, but there is still much to learn.
- It seems that many to most studies that purport to make such discoveries are not ever validated, and perhaps cannot be.

Rules for Diagnostic Markers

- Many studies use a comparison group that is essentially arbitrary and may differ from the patients in many ways other than the disease.
- Only effects that are so large that they could not be due to measured or unmeasured differences in the groups can be trusted.
- These almost never occur, since effects that large likely would already have been discovered.
- These and other biomarker studies are only convincing when repeatedly replicated in different settings by different investigators.

Design Standards

- The most believable studies start with a cohort containing the groups who will eventually develop, as the disease is diagnosed, or as the treatment is successful or unsuccessful.
- Ideally, the measurements are taken before the outcomes are known.
- Any candidate predictor is later validated blind.

What is Needed for Valid Studies

- Besides good design, an important contributor to valid studies is good statistical/bioinformatic methods.
- We know a lot about what not to do (clustering, data mining without consideration of the size of the problem).
- We need to put much effort into developing better methods of analysis, so that we can retain the sensitivity of broader methods while avoiding false discoveries as much as possible.

A Bad Example from Duke

- In 2006, Potti et al. introduced a method to combine array studies on the NCI160 cell lines with drug sensitivity assays to derive “signatures” of sensitivity to specific drugs.
- Coombes, Wang, and Baggerly (2007) at MD Anderson uncovered significant errors in the work, but Potti et al. did not acknowledge them, and stuck to their conclusions.
- They designed a clinical trial which “assigned subjects to either pemetrexed/gemcitabine or cisplatin/gemcitabine therapy using a genomic based platinum predictor to determine chemotherapy sensitivity and predict response to chemotherapy for first-line therapy in advanced non-small cell lung cancer.”

- All of these claims were examined by Baggerly, Coombes and colleagues (2009) and found to be corrupted by data errors, methodological errors, and other problems to the extent that all of the apparent conclusions were actually unsupported by the data.
- These errors lay somewhere on the continuum from sloppiness to fraud.
- Nonetheless, it was difficult even with all this for Baggerly and Coombes to attract enough attention to stop the trial.
- Duke carried out a secret internal review and then restarted the trial.

Outcome

- It was only when Dr. Potti was found to have claimed fraudulently on his CV that he was a Rhodes scholar that the trial was stopped.
- Four major papers were retracted and Potti lost his position.
- Baggerly and Coombes were proved correct in every particular.
- The systems of review for publication and internal review at Duke were both subject to question.

Conclusions

- Complex and/or large data require great care and often sophisticated analysis, and trained statisticians specialized in the area should always be involved.
- Reproducible research is vital. Ideally, the raw data, and all the code required should be made public on publication so that the data fed into the program yields the results in the paper.
- Journals need to take seriously disputes over serious errors, and not treat them as “cat fights”, aka dueling statisticians.

Conclusions

- Ionnidis et al. (2009) showed that of 18 microarray articles examined from Nature Genetics, a top journal, only 2 could be reproduced in full, 6 could be reproduced in part with some discrepancies, and 10 could not be reproduced at all.
- Steen (2011, J. Medical Ethics) showed that between 2000 and 2010, at least 80,000 patients had been in clinical trials based on research that was incorrect and later retracted, so the Duke case is not an isolated instance.
- Personalized medicine has great potential, but only if the research is conducted with great care.
- Most things that are tried (in any field) don't work, so persistence is required.