

SOME STATISTICAL ISSUES IN THE DESIGN AND ANALYSIS OF VACCINE CLINICAL TRIALS IN CANCER PATIENTS

iSBTc Workshop; 10/28/09

DOUGLAS M. POTTER

Biostatistics Department, Graduate School of Public
Health, University of Pittsburgh,
and Biostatistics Facility, University of Pittsburgh
Cancer Institute

OUTLINE

- Biomarker “Discovery”: a cautionary tale
- Definition of subject immunological response: ELISPOT
- Warning: adult content

- “Despite years of research and hundreds of reports on tumor markers in oncology, the number of markers that have emerged as clinically useful is pitifully small. Often, initially reported studies of a marker show great promise, but subsequent studies ...yield inconsistent conclusions or stand in direct contradiction....”

[Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics, *Reporting recommendations for Tumor Marker Prognostic Studies*, JNCI 97:1180, 2005]

The cautionary tale: a failed retrospective biomarker study

- 30 markers; 56 pts; banked samples from several different trials; no design
- Luminex multiplex bead assay
- Endpoint: survival
- Analysis: Cox regression; fdr controlled at 10%
- Results: 15 significant markers with p-values from 0.00011 to 0.048

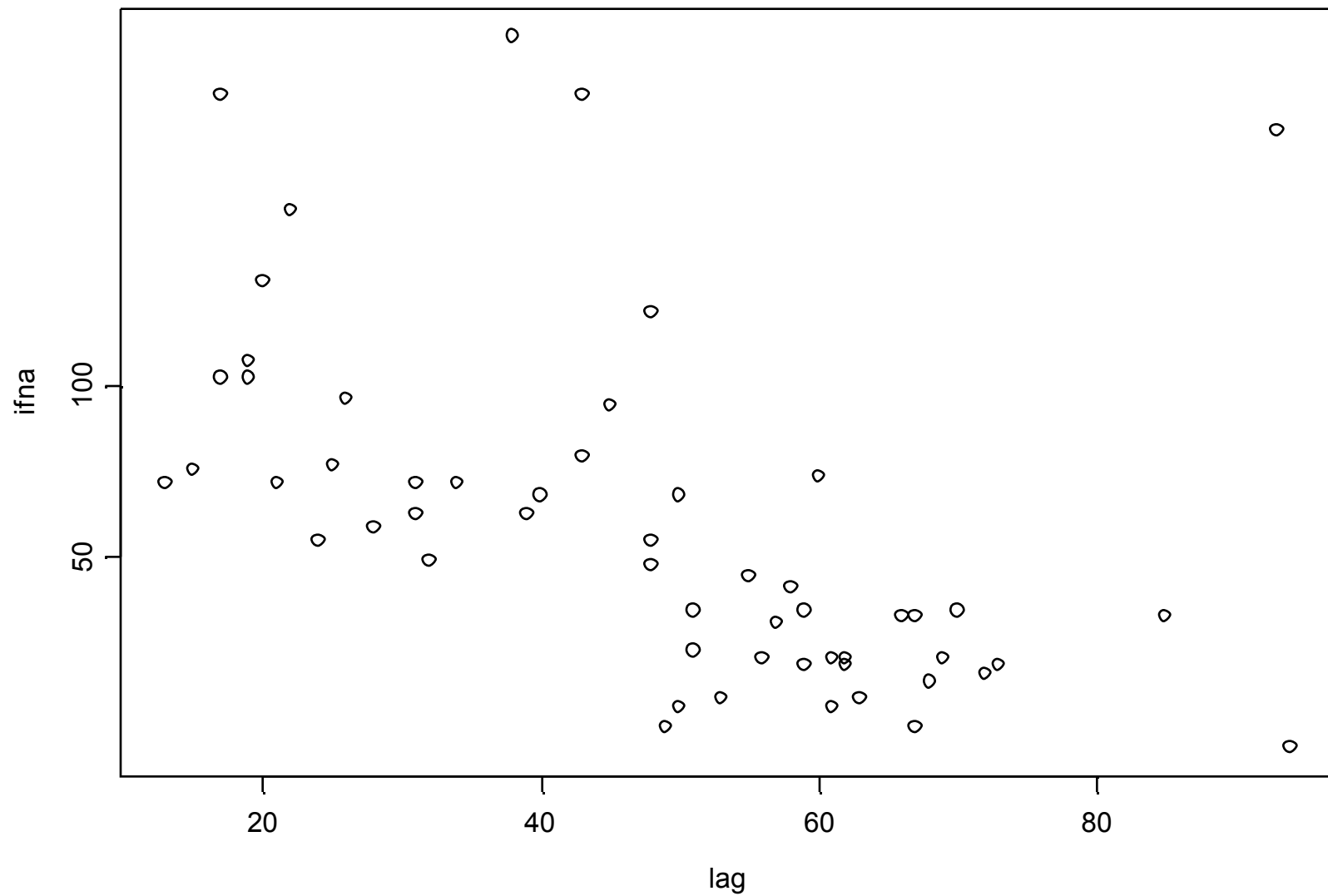
The red flags

- “Too many” significant markers; p-values “too small” (i.e., not expected due to small sample size & large inter-pt variability)
- Significant markers all positively correlated
- For all markers, high levels associated with increased survival
- So, results just not believable

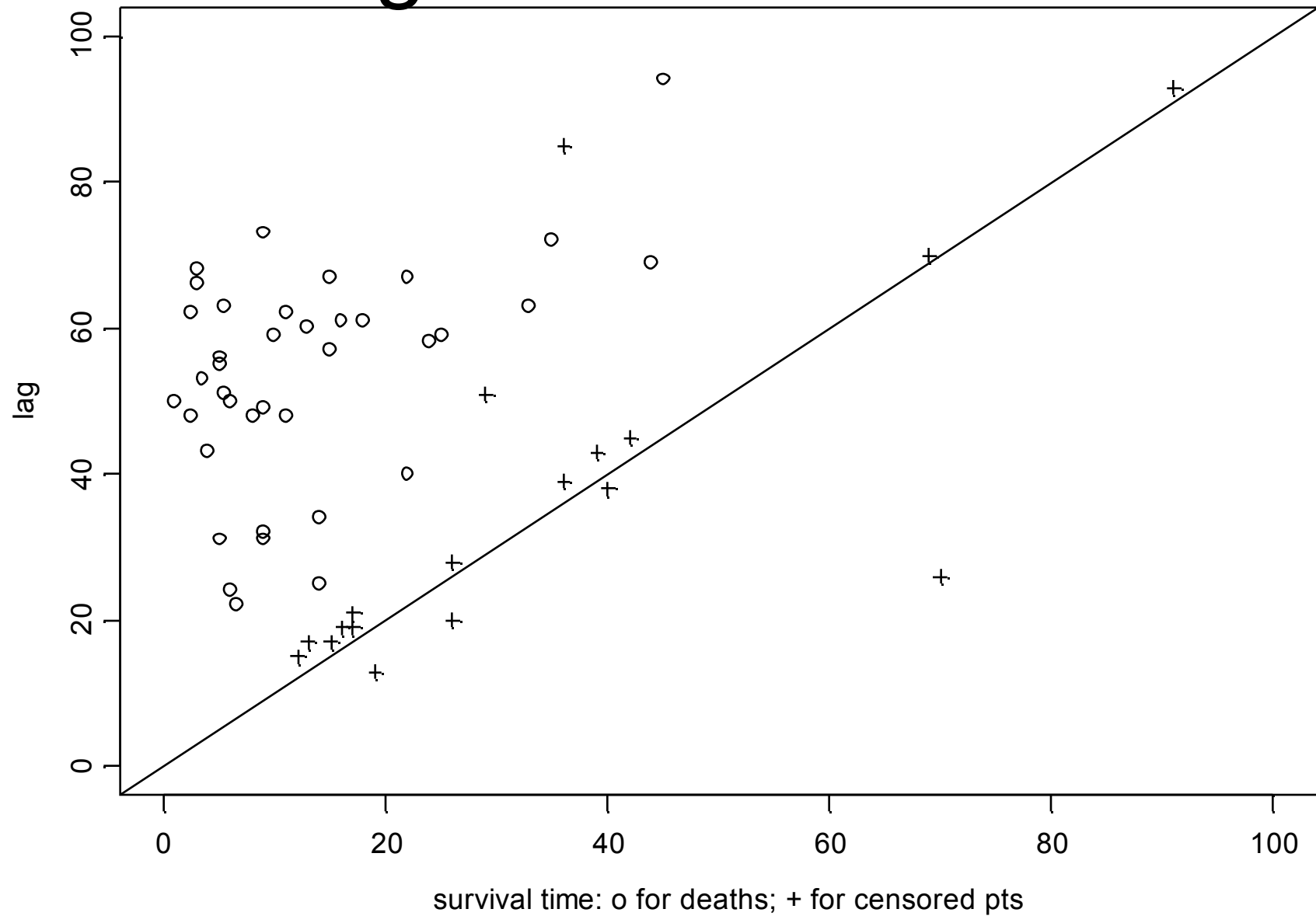
The source of the problem

- “lag”: the age of the sample (period of time between collection of sample and assay)
- lag correlated with survival & with biomarker level

IFN γ level vs. lag



Lag vs. survival time



Recommendations

- Don't do retrospective biomarker studies
- If you must, then...
 - Design study with statistician/cancer epidem.
 - Worry about sample handling & selection, pt prognostic variables, other potential sources of bias.... Demand complete documentation & tracking records for your samples.
 - Don't immediately & enthusiastically accept your newly "discovered" biomarkers. Good science demands skepticism: investigate alternative explanations for your results. Don't expect the statistician to know all potential sources of bias.

Remember....

- “Often, initially reported studies of a marker show great promise, but subsequent studies ...yield inconsistent conclusions or stand in direct contradiction....”
- It's not a small p-value that makes a clinically useful marker; it's high specificity & sensitivity—good separation between cases & controls.

IFN γ ELISPOT RESPONSE

- e.g.: response criterion used by UPCI IMCPL; developed by Bill Gooding
- 3 tst & 3 ctrl wells: $y = \text{mean}(\text{tst}) - \text{mean}(\text{ctrl})$
 - tst counts scaled by ratio of #tst cells/#ctrl cells
 - y set to 0 if any tst count/well < max(ctrl counts)
- Response to tx:
 - $y(\text{post-tx})/y(\text{pre-tx}) > 2$
 - $y(\text{post-tx}) > 10$
 - (a factor of 2 in background-corrected counts with protection against false positives due to small counts)
- Response correlates with survival in melanoma trial(s):
Kirkwood, et al. Clin Cancer Res. 2009 Feb 15;15(4):1443-51.

IFN γ ELISPOT RESPONSE

- Many other definitions of response; some utilize sd or CV.
- Problem: CV not constant (because sd of counts $\geq \sqrt{\text{counts}}$, so $CV \geq 1/\sqrt{\text{counts}}$)
- Problem: the sd of interest is the within-pt sd over the same time period as response assessed; never available. sd will also depend on pt and peptide.

IFN γ ELISPOT RESPONSE

- All response definitions are effectively “seat of pants”; statistical properties unknown & in practice, unknowable—i.e., false positive & false negative rates unknown.
- Use of a single standardized definition could allow the results of different studies to be compared, but will not solve this problem.

IFN γ ELISPOT RESPONSE-- SUGGESTIONS

- Get as many pre-tx samples (at different times) for analysis as possible; useful for limiting pre-tx variability
- Tighten response criteria: require positive responses at 2 consecutive post-tx time points; useful for limiting post-tx variability
- Consider using clinical response to refine definition of immune response
- True immune response is continuous, not binary. Different definitions of binary response are arbitrary & will correlate differently with clinical outcome.
- Consider investigating the relationship of degree of immune response with (degree of) clinical outcome.

IFN γ ELISPOT RESPONSE-- SUGGESTIONS

- When immune response is primary outcome of interest in a trial, use non-parametric techniques to assess response of the entire sample of pts as a group. Don't attempt to define individual response.