Direct and Indirect Measures of Treatment Effects Including Surrogate Endpoints

John H Powers, MD
SAIC in support of Division of Clinical Research, NIH
Associate Clinical Professor of Medicine
George Washington University School of Medicine

Introduction

* Definitions and Regulatory Standards for Outcomes
* Strengths and Limitations of Surrogate Endpoints
* Uses of surrogates – prognosis vs treatment effect
* Regulatory Standards and Evaluation of Potential Surrogate Endpoints
* Challenges in infectious diseases trials with indirect outcomes/surrogates

Outcome Measures

*"The methods of assessment of subjects' response are **well-defined** and **reliable**. The protocol for the study and the report of results should explain the **variables measured**, the **methods of observation**, and **criteria used to assess response**."*

21 CFR 314.126(b)(6)

Types of Outcome Assessments

* **Direct Outcome Measure** = measure of how a patient feels, functions or survives ("clinical" endpoint)
  * Mortality
  * Ability to perform in daily life (function)
  * Symptoms of disease
    * 21 CFR 314.500; US v Rutherford 1979;Temple JAMA 1999;282(8):790-795, FR Notice April 1992

* **Indirect Outcome Measure** = physical signs, observations or laboratory test used as a substitute for direct outcome measure; surrogate endpoint by itself does not confer *direct* clinical benefit to the patient
  * FDA Clinical Outcome Assessment Workshop, October 2011

Categorization of Nomenclature

Outcome Assessments

Direct Measures of Patient "Functions, Feels, Survives"

- Patient (symptoms)
- Clinician (PANNS for schizophrenia syndrome)
- Observer (seizures, infant behavior, death)

Indirect Measures

Psychomodulated – (Dependence on patient motivation or clinician judgment to perform the test)

- Patient (rescue meds for pain, alcohol presentation test)
- Clinician (TM bulging, Limb Spasticity, 6MWT, PFTs, 9-hole peg test)
- Observer (rescue meds for pain)

Biomarkers – (e.g. H_bA_{1c}, CD-4, PSA, CEA, antibody levels, TIMI-III flow HDL, LDL, blood pressure, body temperature, urine GAG, urine KS cardiac rhythm, blood cultures, PCR, quantitative measures from radiology imaging.)

Presumes that relationship to a direct outcome has been demonstrated

Strengths and Limitations

* **Allow measurement of outcome earlier in time –**
  * **Most commonly used in chronic diseases**
  * **HIV viral load as a surrogate for death and opportunistic infections**

* **Decrease sample size**
  * **More frequent events with surrogate than direct outcome**
  * **Larger effect size on surrogate than direct outcome**

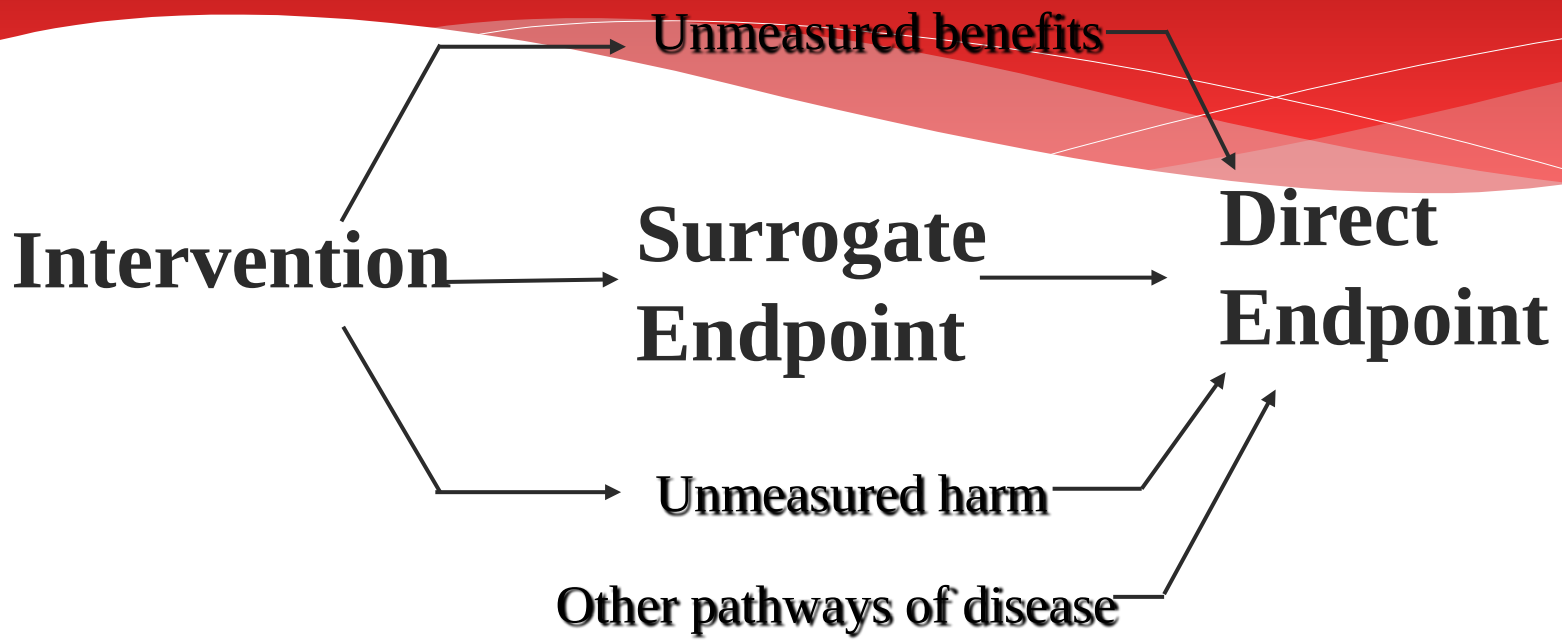
Strengths and Limitations

* **Potential surrogate endpoint/indirect measure may not reflect effects on direct outcome of interest**

* **Benefits are measured on different "scale" than harms which make benefit-harm analysis more challenging – nature of events and numerically**
  * International Conference on Harmonization Guidance E-9

  http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf page 7

Strengths and Limitations



Reasons why surrogate may not accurately predict direct outcomes:

- unmeasured harms or benefits caused by intervention
- other mechanisms of disease other than those affected by intervention
- issues with measuring surrogate
- issues with measuring clinical outcomes

Comparison of Effect Sizes Associated With Biomarkers Reported in Highly Cited Individual Articles and in Subsequent Meta-analyses

John P. A. Ioannidis, MD, DSc
Orestis A. Panagiotou, MD

Context Many biomarkers are proposed in highly cited studies as determinants of disease risk, prognosis, or response to treatment, but few eventually transform clinical practice.

Objective To examine whether the magnitude of the effect sizes of biomarkers proposed in highly cited studies is accurate or overestimated.

Data Sources We searched ISI Web of Science and MEDLINE until December 2010.

Study Selection We included biomarker studies that had a relative risk presented in their abstract. Eligible articles were those that had received more than 400 citations in the ISI Web of Science and that had been published in any of 24 highly cited biomedical journals. We also searched MEDLINE for subsequent meta-analyses on the same associations (same biomarker and same outcome).

Data Extraction In the highly cited studies, data extraction was focused on the disease/outcome, biomarker under study, and first reported relative risk in the abstract. From each meta-analysis, we extracted the overall relative risk and the relative risk in the largest study. Data extraction was performed independently by 2 investigators.

Results We evaluated 35 highly cited associations. For 30 of the 35 (86%), the highly cited studies had a stronger effect estimate than the largest study; for 3 the largest study was also the highly cited study; and only twice was the effect size estimate stronger in the largest than in the highly cited study. For 29 of the 35 (83%) highly cited studies, the corresponding meta-analysis found a smaller effect estimate. Only 15 of the associations were nominally statistically significant based on the largest studies, and of those only 7 had a relative risk point estimate greater than 1.37.

Conclusion Highly cited biomarker studies often report larger effect estimates for postulated associations than are reported in subsequent meta-analyses evaluating the same associations.

JAMA. 2011;305(21):2200-2210 www.jama.com

MANY NEW BIOMARKERS ARE continuously proposed[1-3] as potential determinants of disease risk, prognosis, or response to treatment. The plethora of statistically significant associations[4,5] increases expectations for improvements in risk appraisal.[6] However, many markers get evaluated only in 1 or a few studies.[7] Among those evaluated more extensively, few reach clinical practice.[8]

This translational attrition requires better study. Are the effect sizes proposed in the literature accurate or overestimated?[9] It is interesting to address this question in particular for biomarker studies that are highly cited. Many of these risk factors are also evaluated in meta-analyses[10] that allow overviews of the evidence. However, some meta-analyses may suffer bias from selective reporting, especially among small data sets[11-13]; then large studies

Uses of Indirect Measures including Biomarkers

* **Diagnosis** – select those with and without disease
* **Staging of disease** – classification of natural history
* **Prognosis** – measure corresponds to outcome independent of treatment as a measure of natural history of disease
* **Replacement endpoint for measure of direct benefit** - measure *treatment effects* defined as difference between test group and control group in monitoring response to interventions

* Much confusion about difference between *prognosis* and *treatment effects*

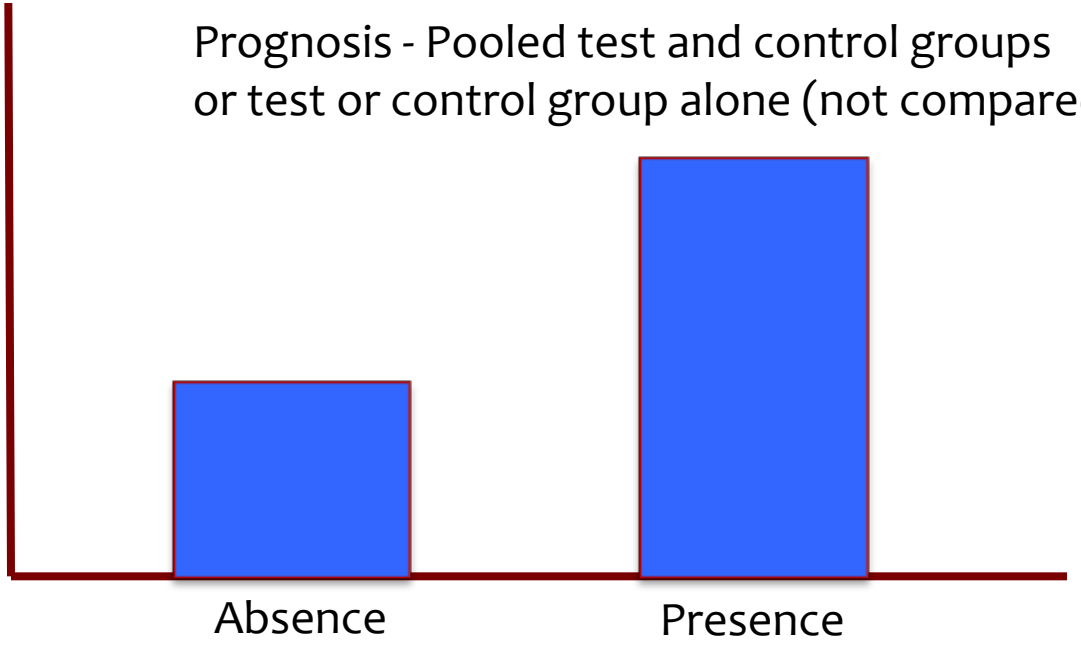
Prognosis vs Endpoint

* **Prognosis**- measures natural history (treated or untreated)
  * E*xposure* is amount or presence/absence of the indirect measure/biomarker and *outcome* is the direct outcome assessment
  * Done by pooling test and control groups or test or control group alone
  * Can be obtained from observational studies
  * Baseline comparability of groups not an issue

* **Outcome measure/endpoint** – measure treatment effects of interventions
  * *Exposure* is test or control intervention and outcome is indirect or direct outcome measure
  * Must evaluate test and control group separately (not pool)
  * Challenging to obtain from observational studies, usually randomized trials
  * Baseline comparability of groups is the issue

Prognosis vs Endpoint

Outcome =
Direct measure
e.g. mortality



Demonstrates correlation which is not necessarily causal

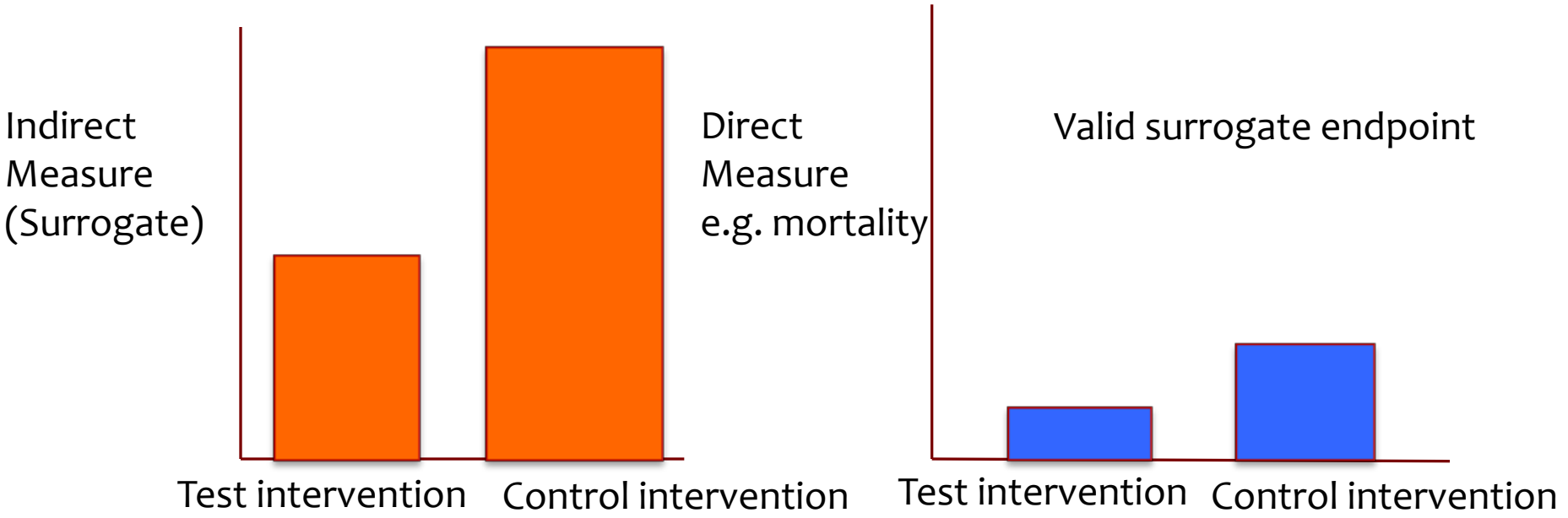
Natural History vs Drug Effects

* 21 CFR 314.126(a)

* **"The purpose of conducting clinical investigations of a drug is to distinguish *the effect of a drug* from other influences, such as spontaneous change in the course of the disease, placebo effect, or biased observation."**

Prognosis vs Treatment Effect

Treatment effect – compare test and control groups in difference showing trials

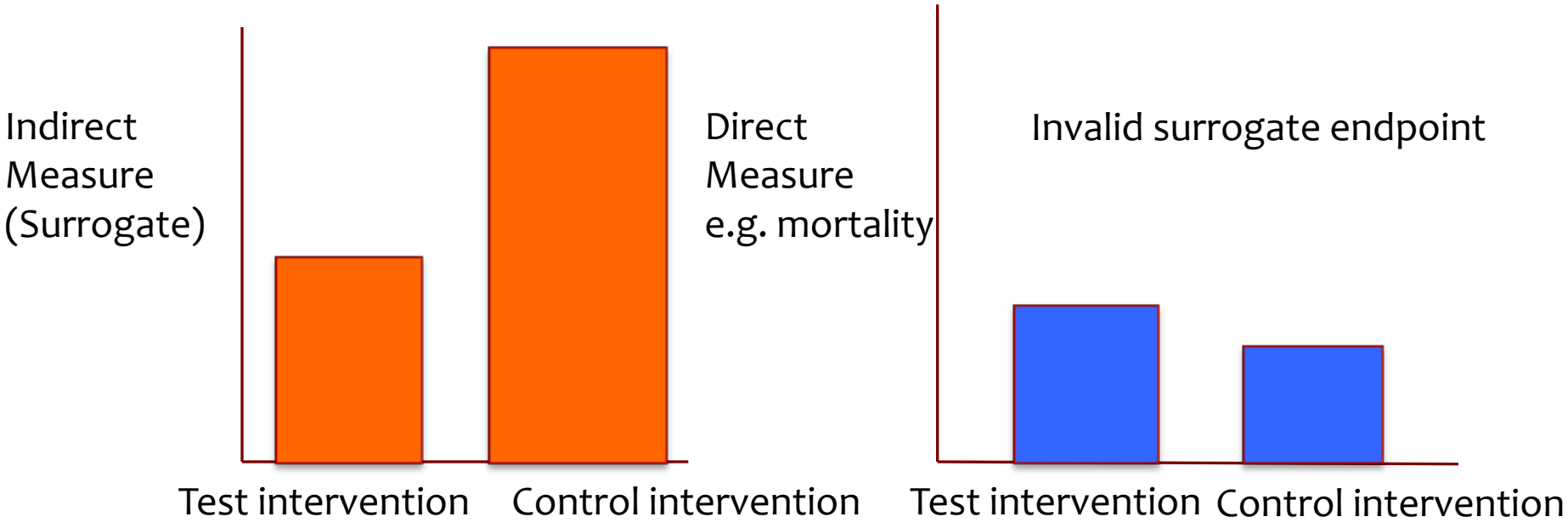


Exposure = test or control intervention

Prentice RL. Stat Med 8 19898:431-40.

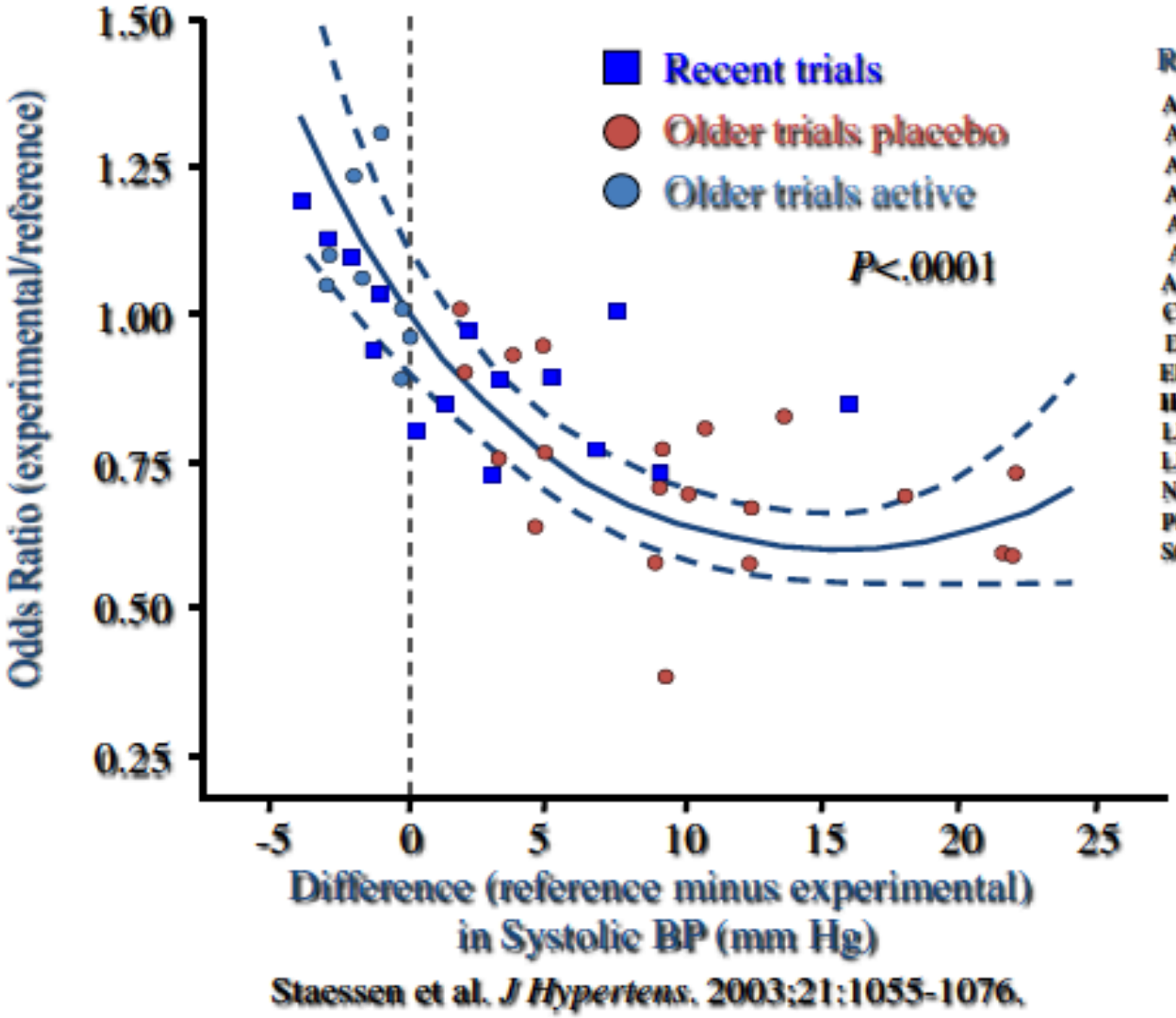
Prognosis vs Treatment Effect

Treatment effect – compare test and control groups in difference showing trials



Exposure = test or control intervention

Prentice RL. Stat Med 8 19898:431-40.



Recent	Older
AASK L vs. H	ALLHAT/Dox
ABCD/NT L vs. H	ATMH
ALLHAT/Aml	EWPHE
ALLHAT/Lis	HEP
ALLHAT/Lis ≥65	HOPE
ALLHAT/Lis Blacks	HOT
ANBP2	HOT M vs. H
CONVINCE	INSIGHT
DIABHYCAR	MIDAS/NICS/VHAS
ELSA	L vs. H
IDNT2	MRC
LIFE/ALL	MRC2
LIFE/DM	PART2/SCAT
NICOLE	PATS
PREVENT	PROGRESS/Per
SCOPE	PROGRESSION/Com
	RCT70-80
	RENAAL
	SHEP
	STONE
	STOP 1
	STOP2/CCBs
	STOP2/ACEIs
	Syst-China
	Syst-Eur
	UKPDS C vs. A
	UKPDS L vs. H

Staessen et al. *J Hypertens*. 2003;21:1055-1076.

Regulatory Standard for Surrogates

- **Subpart H--Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses**

- **21 CFR Sec. 314.500 Scope.** ***This subpart applies to certain new drug products that have been studied for their safety and effectiveness in treating serious or life-threatening illnesses and that provide meaningful therapeutic benefit to patients over existing treatments (e.g., ability to treat patients unresponsive to, or intolerant of, available therapy, or improved patient response over available therapy).***

[57 FR 58958, Dec. 11, 1992, as amended at 64 FR 402, Jan. 5, 1999]

Regulatory Standard for Surrogates

* Sec. 314.510 **Approval based on a surrogate endpoint or on an effect on a clinical endpoint other than survival or irreversible morbidity. FDA may grant marketing approval for a new drug product on the basis of adequate and well-controlled clinical trials establishing that the drug product has an effect on a *surrogate endpoint that is reasonably likely*, based on epidemiologic, therapeutic, pathophysiologic, or other evidence*, to predict clinical benefit or on the basis of an effect on a clinical endpoint other than survival or irreversible morbidity*.**

Regulatory Standard for Surrogates

* Sec 314.510: **Approval under this section will be subject to the *requirement that the applicant study the drug further*, to verify and describe its clinical benefit, where there is uncertainty as to the relation of the surrogate endpoint to clinical benefit, or of the observed clinical benefit to ultimate outcome. Post-marketing studies *would usually be studies already underway*. When required to be conducted, such studies *must also be adequate and well-controlled*. The applicant shall carry out any such studies with due diligence.**
*

Evaluation of Surrogates (IOM Paper – 2010)

* 1) For what direct measure is the indirect measure substituting?
* 2) **Analytical val**idation: analyses of available evidence on the analytical performance of indirect measure (e.g. CPIS and lack of reliability)
* 3) **Qualification**: assessment of available evidence on associations between indirect and direct outcomes including data showing *effects of interventions* on both the biomarker and clinical outcomes
* 4) **Utilization**: contextual analysis based on specific use proposed and applicability of available evidence to this use; includes a determination of whether evaluation provides sufficient support for the use proposed
* 5) **Reevaluate** analytical validation, qualification, and utilization on a continual and a case-by-case basis

Current Outcomes in HAP Trials

* "Clinical response" = "the clinical evaluation of pneumonia based on the improvement and resolution of clinical signs and symptoms of infection, including fever, leukocytosis, purulent sputum production, and radiographic lung infiltrates" based on clinician judgment and decision to administer another antibiotic
  * Rubenstein Clin Infect Dis 2001; 32:402–12
* Composite of indirect measures of clinician reported outcomes and biomarkers, actual variables measured or how they are combined not well-defined
* Biomarkers (body temperature, leukocytosis, CXR) are not on causal pathway of disease

Correlation and Causation

* “There is usually at least a theoretical possibility that the marker and the disease are not causally related but instead are associated with a common underlying factor. Fever and respiratory impairment occur with pneumonia but the fever does not cause the disease, and treating it will not improve the infection.”

* 57 Federal Register no 73 p 13235, 1992

Challenges for Surrogates in HAP Trials

* **Hypothesis in non-inferiority trials not to demonstrate added benefits over available therapies**
* **Most HAP/VAP trials are not difference showing trials**
* **Many proposed indirect measures/potential surrogates are not on causal pathway of disease – fever, leukocytosis, CXR**
* **Challenge in doing a single study so how will follow-up studies be done to evaluate indirect measure/surrogate? Other indications?**
* **Use of surrogate in NI trials includes need to know *how much* of a difference on surrogate reflects *how much* of a difference on direct outcome measure**
* **Lack of diagnostics makes it challenging to select appropriate populations**

Proposal

* Could use a Direct Outcome of Patient Reported symptoms in HAP patients who are mentally competent
  * Could include in current FNIH work on CABP outcomes
  * Interim outcome could use same as proposed for CABP (4 symptoms of cough, chest pain, dyspnea, sputum)
* Recent evidence in hospitalized CABP in patients with co-morbidities shows substantial burden of symptoms
* Recent registrational trials in VAP have mortality in range of 15-20% - mortality or potential surrogate for mortality – candidates?
* Potential for surrogate endpoints for mortality based on biomarkers needs further evaluation

Hospitalized CABP

- Retrospectively developed symptom questionnaire for CABP symptoms in older patients – 16 symptoms
- Interviewed 500 subjects over age 50 years (stratified above and below age 65) – 201 hospitalized
- 83.9% overall (90.8% of hospitalized) with co-morbid conditions
- Incidence of symptoms in hospitalized patients
  - Cough = 93.2%
  - Shortness of breath = 90.3%
  - Chest pain = 68.6%
  - Mucous or phlegm bothering them = 72.7%
  - Body aches = 93.9%
  - Weakness = 94.3%
  - Tiredness = 98.2%
  - Fever/warmth = 61.1%
    - Wywrich K et al, 49th Annual Meeting of IDSA October 2011

Back-Up Slides

Surrogate Endpoint

* "A surrogate endpoint of 'marker' is a laboratory measurement of physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions or survives and that is expected to predict the effect of the therapy"

57 Federal Register no 73 p 13235, 1992