What Constitutes Reasonable Evidence of Efficacy and Effectiveness?

Daniel Sargent, PhD

Professor of Biostatistics & Oncology

Mayo Clinic, Rochester, MN

February 9, 2009

Overview

- Definitions
- How to generate evidence
- How to evaluate strength of evidence
- What kind of evidence we will need in the future

Definitions: 'Reasonable'

- Definition: agreeable to reason or sound judgment; logical[1]

- Challenges in face of life threatening disease
  - Pressure to accept less evidence
  - Discount harms when alternative dire

[1]Dictionary.com

Definitions: 'Evidence'

- Definition: that which tends to prove or disprove something; ground for belief; proof[1]
- Challenges
  - Not mathematics: cannot prove A > B when based on imperfect observations on a sample
  - Not a laboratory: patients must consent to participation
  - Evidence is often unavailable/ inconclusive/ contradictory

[1]Dictionary.com

Definitions: 'Efficacy'

- Definition: capacity for producing a desired result or effect

- Challenges
  - What is the desired result?
  - Can we measure the desired result?
  - Is the result transferrable to other settings?

[1]Dictionary.com

‘Reasonable Evidence of Efficacy’

- Who decides whether ‘evidence’ of ‘efficacy’ is ‘reasonable’?
- Current (Oncology drugs):
  - FDA, often guided by ODAC
    - Primarily a scientific decision, some patient input
  - Community, guided by
    - Guidelines (ASCO, NCCN)
    - Scientific literature
    - Pharma/marketing

‘Reasonable Evidence of Efficacy’

- Is the current practice for evaluating efficacy evidence ‘reasonable’?

- Mostly yes
  - Input from many parties, in an organized manner
  - Therapies must have efficacy
  - Clear standards ($p < 0.05$)

Definitions: 'Effectiveness'

- Definition: how well a treatment works in practice, as opposed to efficacy, which measures how well it works in clinical trials or laboratory studies[2]

- Challenges
  - How can we predict effectiveness from efficacy measures?
  - Do we ever measure this?

[2]wikipedia.com

‘Reasonable Evidence of Effectiveness’

- Is the current practice for evaluating effectiveness evidence ‘reasonable’?

- Mostly no
  - Input from many parties, unorganized
  - Therapy effectiveness unclear
  - Clear standards lacking

How to generate evidence: hierarchy of research designs[3]

- I - Evidence from at least one properly randomized, controlled trial
- II-1 Evidence from well-designed controlled trials w/o randomization
- II-2 Evidence from well-designed cohort or case–control analytic studies
- II-3 Evidence from multiple time series with or w/o the intervention.
- III Opinions of authorities, clinical experience; descriptive studies and case reports; reports of expert committees

[3]U.S. Preventive Services Task Force

Evidence for efficacy: Level I or II-I

- I - Evidence from at least one properly randomized, controlled trial
  - Gold standard for FDA full approval

- II-1 Evidence from well-designed controlled trials w/o randomization
  - Acceptable for accelerated approval

Why are RCTs the gold standard?

- Randomization allows causal inference: A causes B

- All other forms of evidence potentially biased by selection effects & other hidden biases
  - Propensity scores, other modeling approaches try to adjust, but imperfect

Elements of Quality that Apply to Both Level I and II-1 studies

- Pre-specified hypothesis
  - Primary, secondary endpoints
  - Specified data cut-offs
- Defined sample set
  - Eligibility criteria
  - As inclusive as possible
- Power calculations to show have data to address primary aim, pre-specified analysis plans

Elements of Quality that Apply to Both Level I and II-1 studies

- Unbiased endpoint ascertainment
  - Blinding if possible
  - Protocol specified criteria
  - Independent review if possible
- Complete information
  - Standard follow-up per schedule
  - Full assessment of outcome on all patients (few lost to follow-up)

Evidence for Effectiveness: Current Paradigm

- RCT done to achieve initial approval (establish efficacy)
- Adoption by community
- Refined further study / community use
- Refinements rarely studied rigorously
- Ultimate pseudo-validation through meta-analysis or observational study (maybe)

Evidence for Effectiveness

- Current paradigm mostly prohibits generation of level 1 evidence of effectiveness

- Is level I evidence possible?
  - Large, simple trials
  - Cluster randomization

Generating effectiveness evidence: Cluster randomization

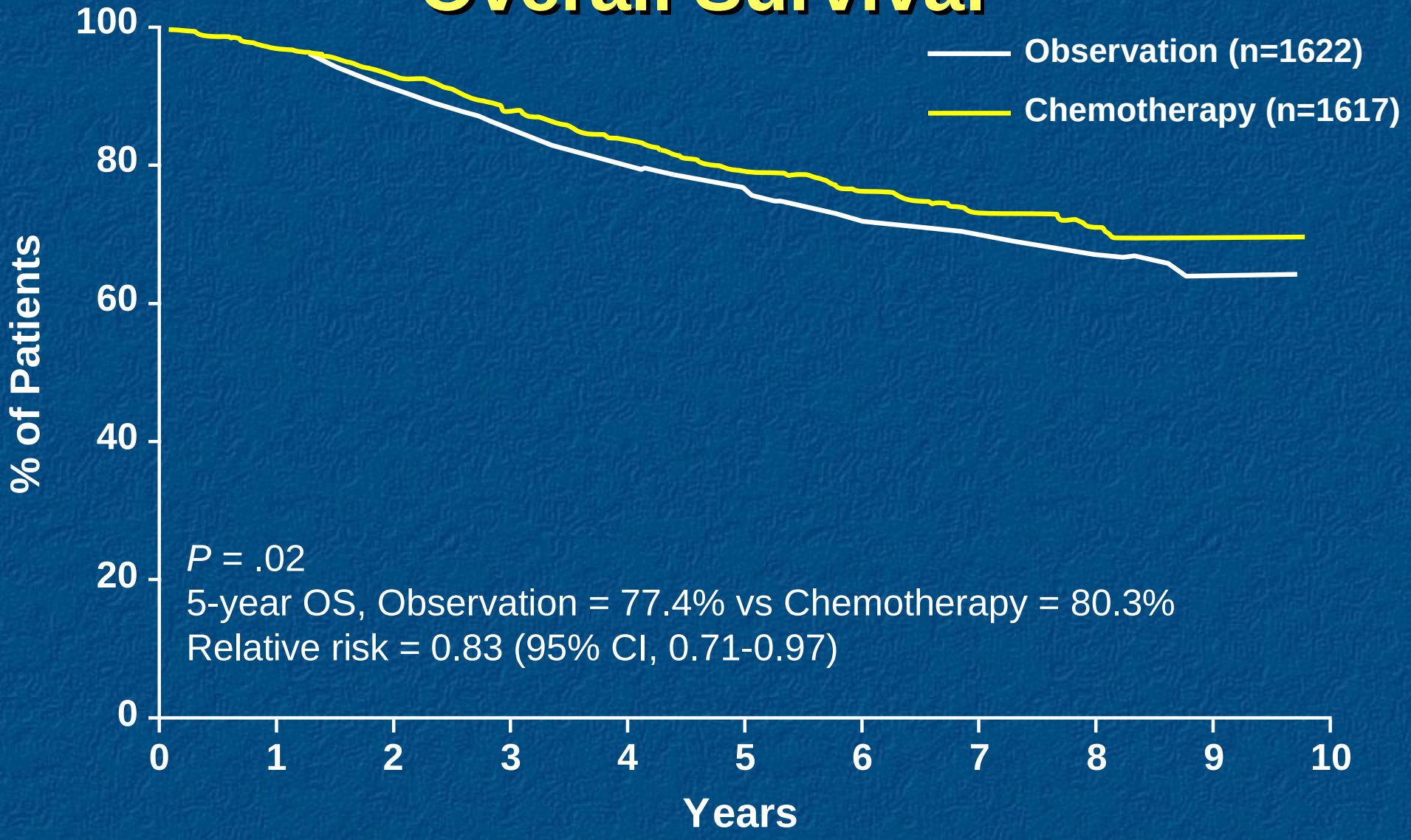
- If it is impossible to randomize individuals, can we randomize groups?
  - Physicians, Institutions, States, etc.
- Less powerful than randomizing patients, but still randomized
- Special analyses required, but feasible

Large Simple Trial Example: QUASAR[4]

- Streamlined trial design, with no extra investigations & minimal extra workload
- Notify trial office of serious unexpected adverse experiences
- Yearly follow-up for brief details of serious toxicity, recurrence, and death
- Health economic, compliance, toxicity, quality of life measured in a sub-study

[4]QUASAR Group, LANCET 2007

QUASAR (stage II colon cancer): Overall Survival



MAYO CLINIC

[4]QUASAR Group, LANCET 2007

Bridging the efficacy vs effectiveness gap in RCTs

- Multi-center recruitment
- Minimize eligibility criteria
- Intention to treat analysis
- Minimize accrual disincentives
  - Financial
  - Regulatory
  - Data Collection

Evaluating Strength of Evidence: The Endpoint Hierarchy

- True Clinical Efficacy Measure
- Validated Surrogate Endpoint (*Rare*)
- Surrogate Endpoint that is "reasonable likely to predict clinical benefit"
- None of the Above: A correlate that is solely a measure of Biological Activity

Evidentiary Requirements for Drug Approval

- Regular approval
  - Clinical benefit, or
  - Established surrogate for clinical benefit
- Accelerated approval
  - Surrogate (reasonably likely to predict clinical benefit)

Evidence: Surrogate Endpoints

- An endpoint obtained sooner, at less cost, or less invasively than the true endpoint of interest

- When using a potential surrogate endpoint, one would like to make the same inference as if one had observed a true endpoint (i.e. a health outcome)

Validation of Surrogate Endpoints

Property of a Valid Surrogate

Effect of the Intervention
on the Clinical Endpoint

is reliably predicted by the

Effect of the Intervention
on the Surrogate Endpoint

Validation of Surrogate Endpoints

Statistical

- Meta-analyses of clinical trials data

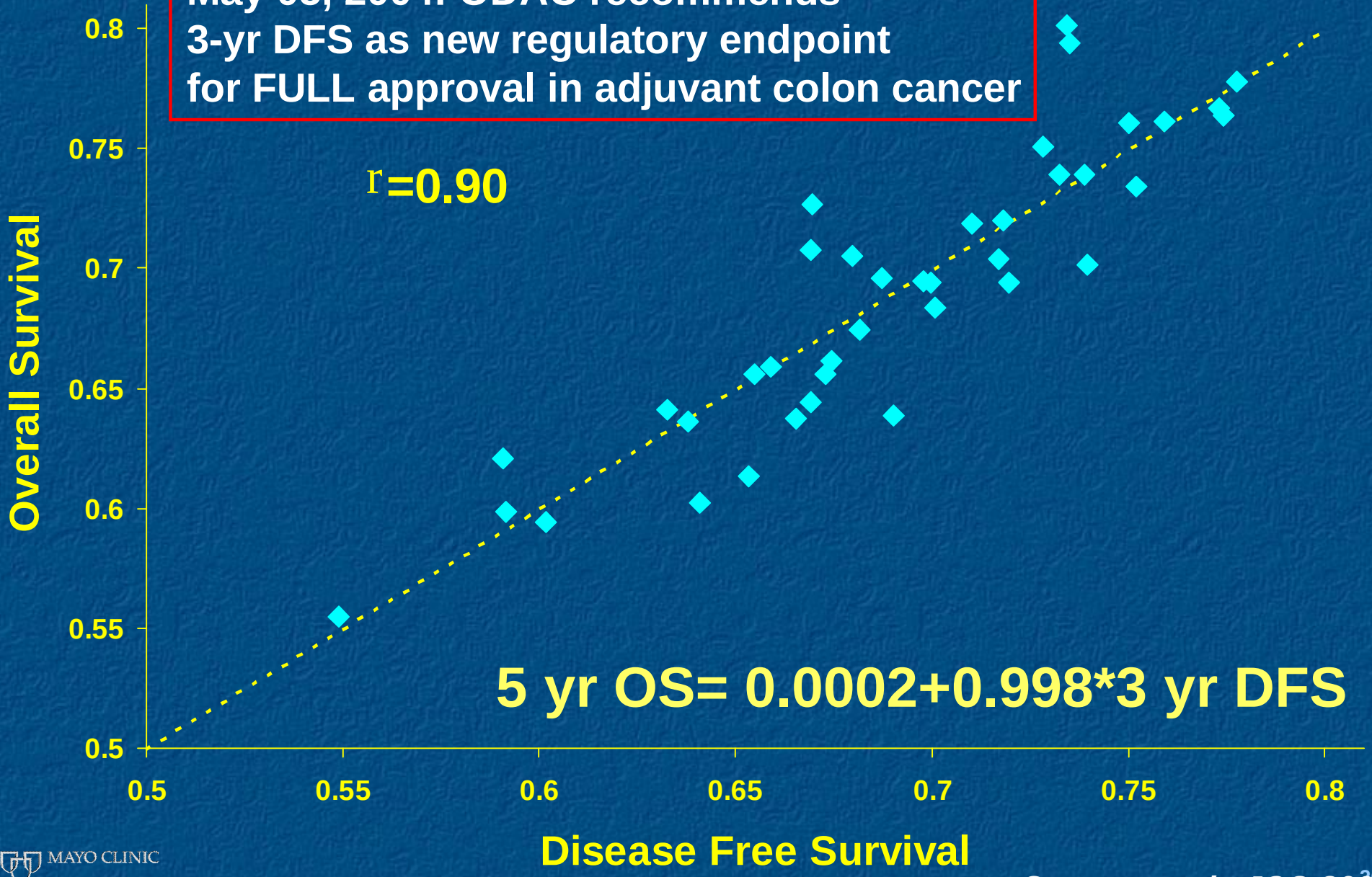
Clinical

- Comprehensive understanding of the
  - ~ Causal pathways of the disease process
  - ~ Intervention's intended and unintended mechanisms of action

No single gold standard approach

ACCENT Analysis: 3 Yr DFS vs 5 Yr OS

May 05, 2004: ODAC recommends
3-yr DFS as new regulatory endpoint
for FULL approval in adjuvant colon cancer



Sargent et al., JCO 2005

What type of evidence will we need in the future?

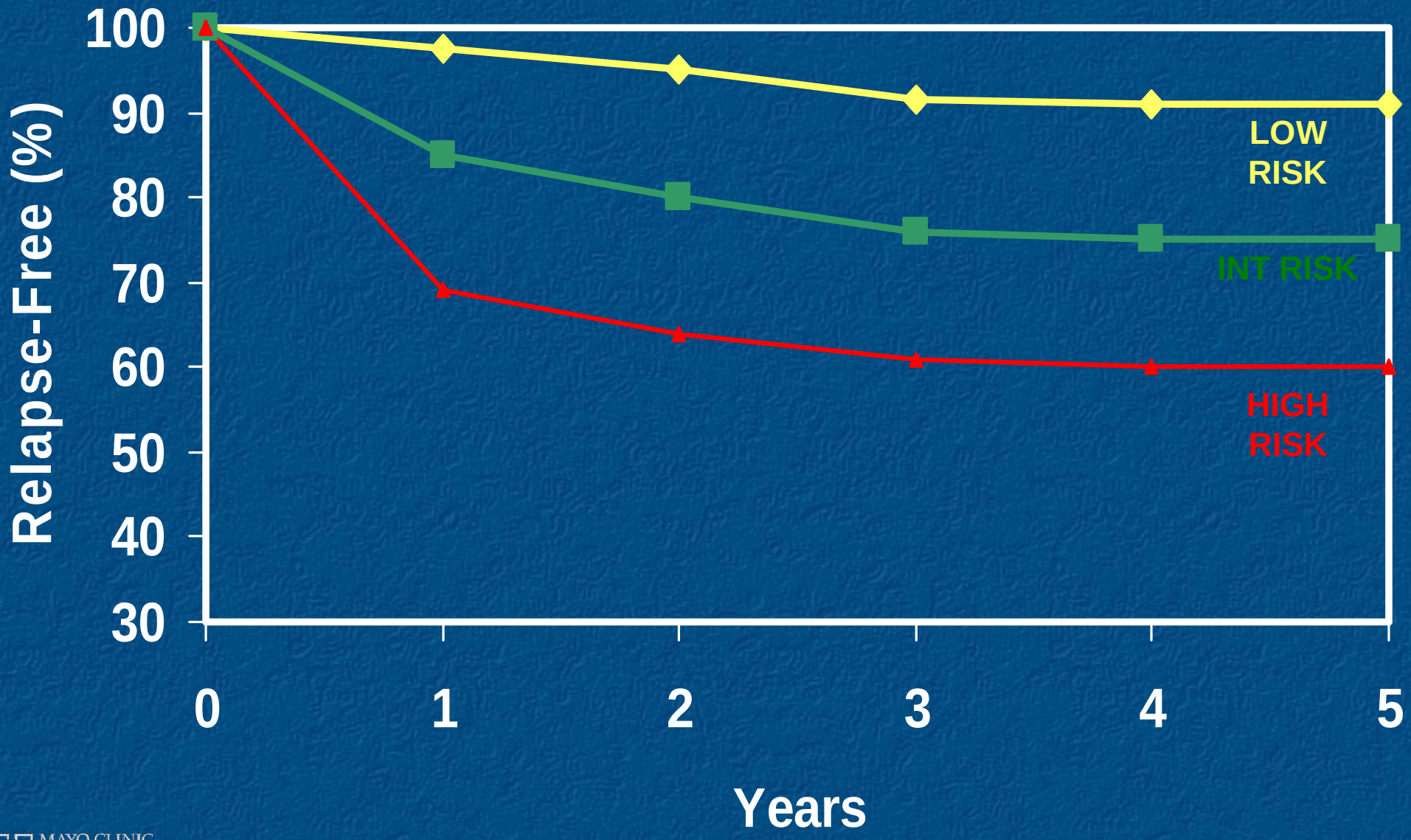
- Premise: Biomarkers will define patient populations based on
  - Risk
  - Potential to benefit

- Premise: Biomarkers will allow early assessment of treatment efficacy
  - As trial endpoints
  - As patient management tools

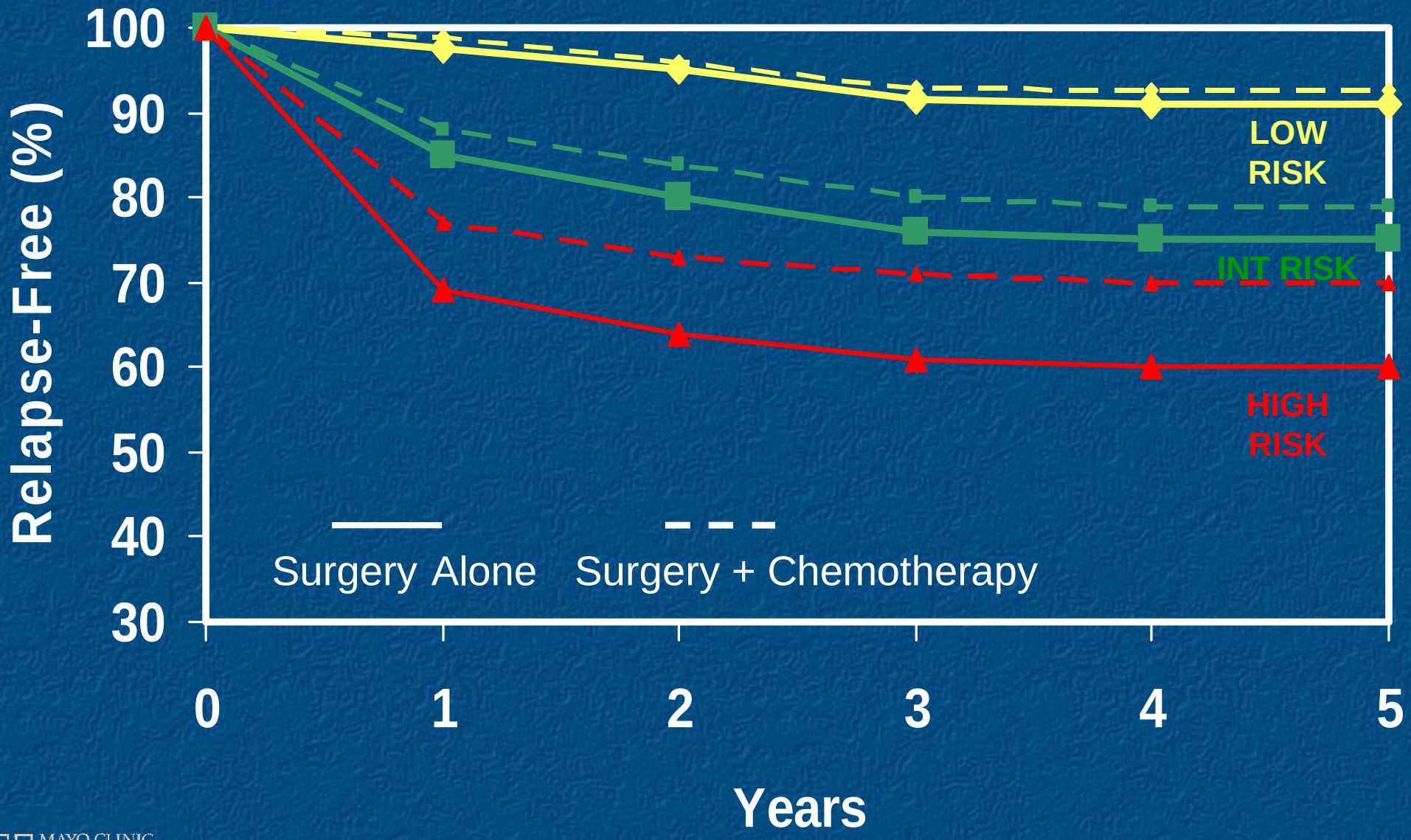
Prognostic & Predictive Biomarkers

- New technologies allowing large-scale measurement of genomic & other factors
- New (targeted) therapeutics emerging
- Individualizing therapy becoming increasingly desirable & theoretically feasible – clear value implications
- Very few potential biomarkers developed to the point of allowing reliable use in clinical practice

Prognostic Goal: Early Stage Cancer



Predictive Goal: Early Stage Cancer



Predictive marker validation: RCTs required

- Goal: Determine with treatment will work for which patient
- Vital: Patients treated with rx choices in question must be comparable
- Only true assurance: Patients randomized between treatments

Biomarker Classifier Development: Prospective Specification

- Inclusion/exclusion criteria
- Primary, secondary endpoints
- Precise definition of biomarker outcome (pre-specified cutpoints)
- Statistical analysis methodology
- Just like a prospective RCT clinical trial

Requirements for Retrospective Validation

- Samples available on large majority of patients to avoid selection bias
- Hypotheses, analyses techniques, patient population, and precise algorithm for assay techniques stated prospectively
- All marker subgroup analyses stated upfront, with appropriate sample size justification

Prospective clinical trial designs to validate biomarkers

- Targeted (selection) trial: enroll only those thought likely to respond to new therapy
- Unselected trial: Enroll all, but prospectively include biomarker in analysis plan
  - Prospective subgroup analysis by marker status

Designs for Targeted Trials

- Design can use standard approaches

Possible Issues

- Negative trials when agent has benefit since precise mechanism of action unknown; missed efficacy in other pts
- Inability to test association of the biologic endpoints with clinical outcomes
- Need to screen all patients anyway
- Need real time method for assessing patients who are / are not likely to respond

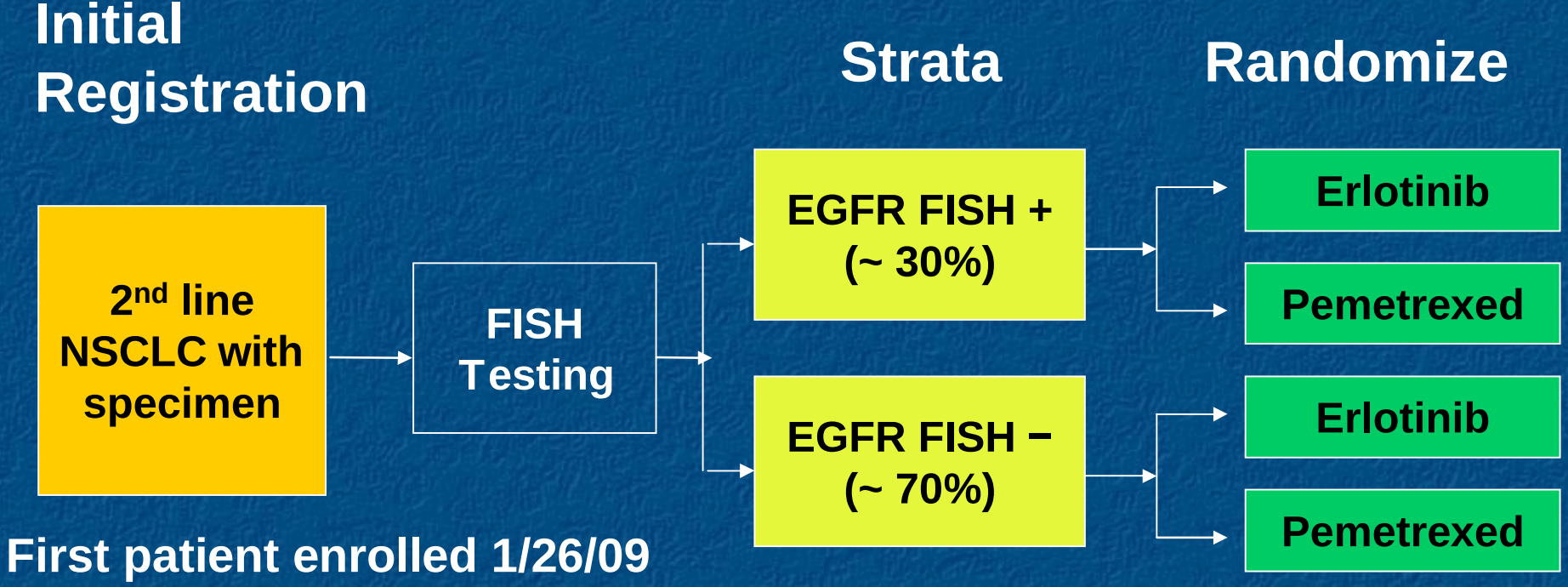
Unselected Biomarker Validation Design

- Randomize all pts between treatments, Biospecimens on all – prospectively analyze by marker

- Advantages:
  - Answers more questions
  - Allows retrospective analyses for even better markers

Sargent, JCO 2005

MARVEL - Marker Validation for Erlotinib in Lung Cancer



First patient enrolled 1/26/09

Primary Goal:

To evaluate whether there are differences in PFS between erlotinib and pemetrexed within the FISH positive and FISH negative subgroups (N = 957)

Conclusions

- As a medical community, we do a reasonable job of efficacy determination
  - Still very costly & burdensome
  - Need to reduce data collection, develop reliable early endpoints

Conclusions

- We rarely collect data to allow reliable determination of effectiveness
  - Careful experimental design critical to generate reliable evidence
  - Large simple trials, cluster randomization are possibilities
- Future medicine will be more complex, not less, and both efficacy and effectiveness determinations require prospective planning in RCTs