Steps to Validation of Early Endpoints to Support Drug Development in Neuroblastoma: Key Concepts

Lisa M McShane, PhD
Associate Director, Division of Cancer Treatment & Diagnosis
Chief, Biometric Research Program
National Cancer Institute, NIH

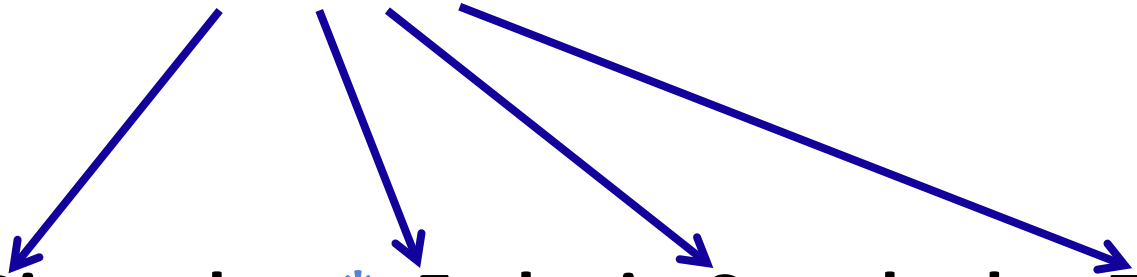
Presentation to FDA Pediatric Oncology Subcommittee of the Oncologic Drugs Advisory Committee (pedsODAC) Meeting

May 12, 2022

DISCLOSURES

- I have no financial relationships to disclose.
- I will not discuss off label use and/or investigational use in my presentation.
- The views expressed represent my own and do not necessarily represent the views or policies of the National Cancer Institute or National Institutes of Health (NIH).
- Although I will be referring to definitions from the BEST glossary, which is a joint FDA-NIH initiative, I am not an employee of the U.S. Food and Drug Administration (FDA); the views expressed represent my own and should not be interpreted as official views or policies of the FDA.

BEST Resource



(Biomarkers*, EndpointS, and other Tools)

Biomarker Working Group charged by **FDA-NIH Joint Leadership Council** to develop a glossary of harmonized terminology for biomarkers, endpoints, and other tools useful in medical product development or regulated product evaluation.

I will discuss considerations in use of end-of-induction response as an early endpoint "biomarker" in drug development for high-risk neuroblastoma using illustrative examples from pediatric cancers.

http://www.ncbi.nlm.nih.gov/books/NBK326791/

Response biomarker (https://www.ncbi.nlm.nih.gov/books/NBK402286/)

Biomarker used to show that a biological response, potentially beneficial or harmful, has occurred in an individual who has been exposed to a medical product or an environmental agent.

- **Pharmacodynamic biomarker:** Response biomarker that indicates biologic activity of a medical product or environmental agent without necessarily drawing conclusions about efficacy or disease outcome or necessarily linking this activity to an established mechanism of action.
- **Surrogate endpoint biomarker:** Response biomarker that is an endpoint used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives.
    - Does not measure the clinical benefit of primary interest in and of itself, but rather is expected to predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.

Potential uses for (early) response endpoints in clinical trials

- **Pharmacodynamic biomarker**
  - Early endpoint to **enrich** for patients for whom a modified treatment strategy may be evaluated
    - Measured post-treatment initiation as a **"prognostic" biomarker** ("correlate") of long-term clinical outcome
  - Early endpoint **to drop less active drugs (i.e., drug screening)**
    - Phase 2 $\rightarrow$ 3 in drug development program (e.g., tumor response)
    - Drop arms within a multi-arm phase 2 or 3 trial based on early endpoint
    - Interim endpoint in phase 2/3 trial design
- **Surrogate endpoint** (trial-level)

Different supporting evidence required to justify these distinct uses.

Enrichment

"Prospective use of any patient characteristic to select a study population in which detection of a drug effect (if one is in fact present) is more likely than it would be in an unselected population"

- Reduce inter-patient and intra-patient heterogeneity
- Prognostic enrichment strategies
  - Poor prognosis: More events ⇒ more statistical power to evaluate new therapies
  - Very favorable prognosis: Potentially reduce therapy
- Predictive enrichment strategies − choosing patients more likely to respond to the drug treatment (e.g., use of treatment selection biomarker)

http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf

EOI response (by INRC 1993) is prognostic for better event-free and overall survival in neuroblastoma

(Pinto et al. *Eur J Cancer* 2019;112:66-79)

EOI: End of induction

1280 patients (A3973: 470; ANBL02P1: 31; ANBL0532: 638; ANBL12P1: 141) met eligibility criteria. All four trials (**high risk**) used intensive induction chemotherapy for 5-6 cycles; response graded uniformly using the 1993 INRC criteria.

Event-free survival (EFS): time from initial diagnosis to first episode of disease relapse or progression, second malignancy or death, with patients without event censored at the last follow-up.

Overall survival (OS): time from initial diagnosis to death, with surviving patients censored at the last follow-up.

Responder analysis: EFS or OS compared between patients with and without EOI response, irrespective of treatment assignment.

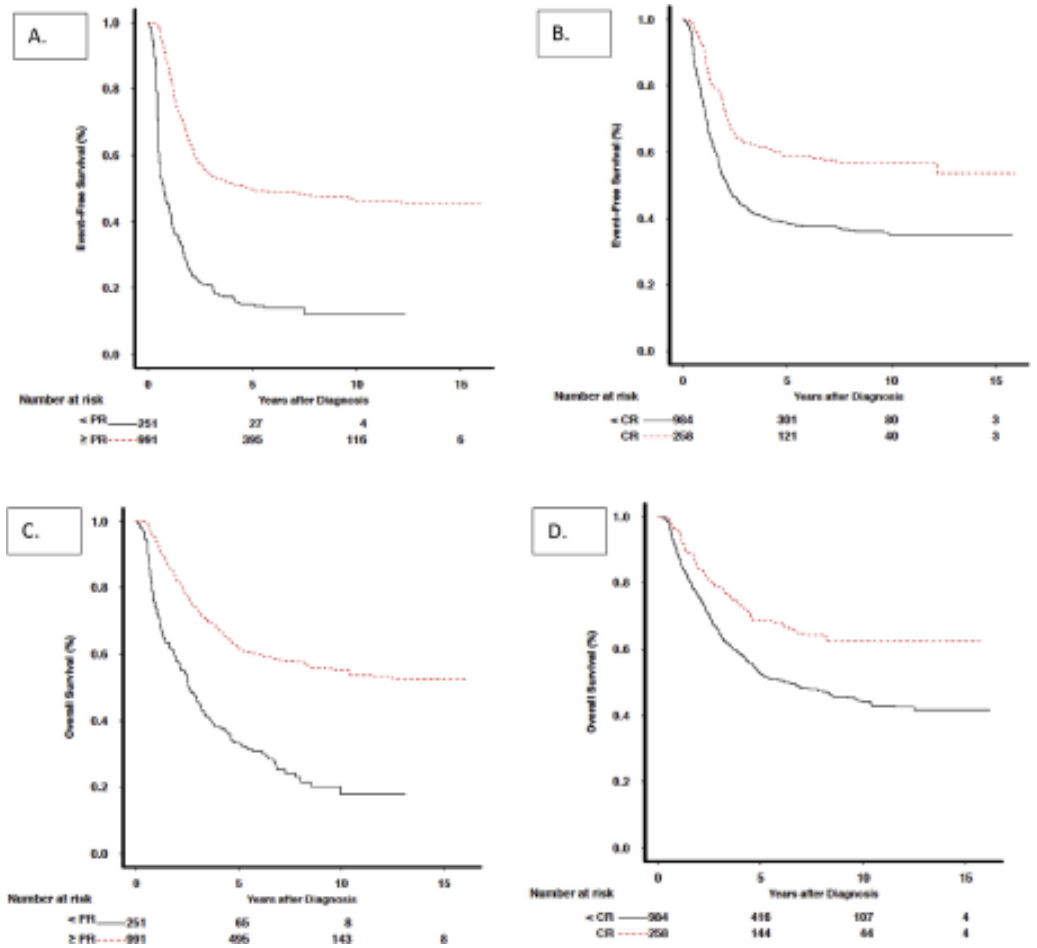


Figure 2. **A**. EFS according to EOI partial response (PR) or better vs. less than PR. **B**. EFS according to EOI complete response (CR) vs. less than CR. **C.** OS according to EOI PR or better vs. less than PR. **D**. OS according to EOI CR vs. less than CR. Log-rank test $p < 0.0001$ for all panels.

EOI response (by INRC 2017) is prognostic for better overall survival in high-risk neuroblastoma

(Barr et al. *Pediatr Blood Cancer* 2020;67:e28390)

2017 International Neuroblastoma Response Criteria (INRC) overall response. Components include individual assessments of response at primary site, metastatic sites, and bone marrow.

Response	Criterion
Complete response (CR)	CR in all components
Partial response (PR)	PR in at least one component and all other components are CR, MD*
Minor response (MR)	PR or CR in at least one component but at least one component with SD. No PD
Stable disease (SD)	No component better than SD or no involvement. No PD.
Progressive disease (PD)	Any component with PD.

*MD (minimal disease) is a unique response classification exclusive to the bone marrow assessment.

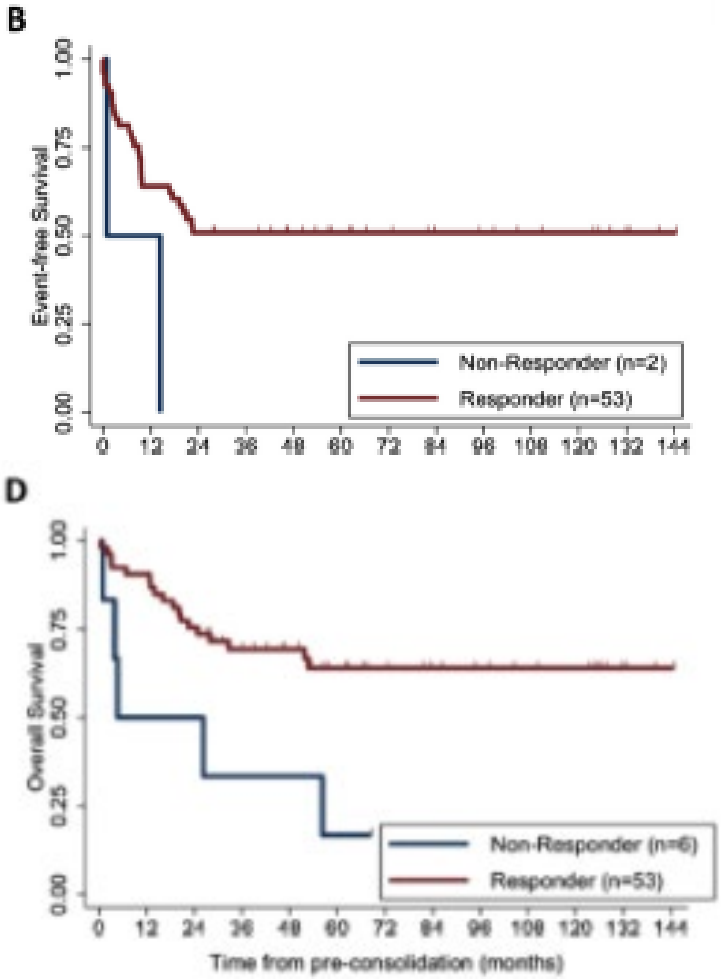
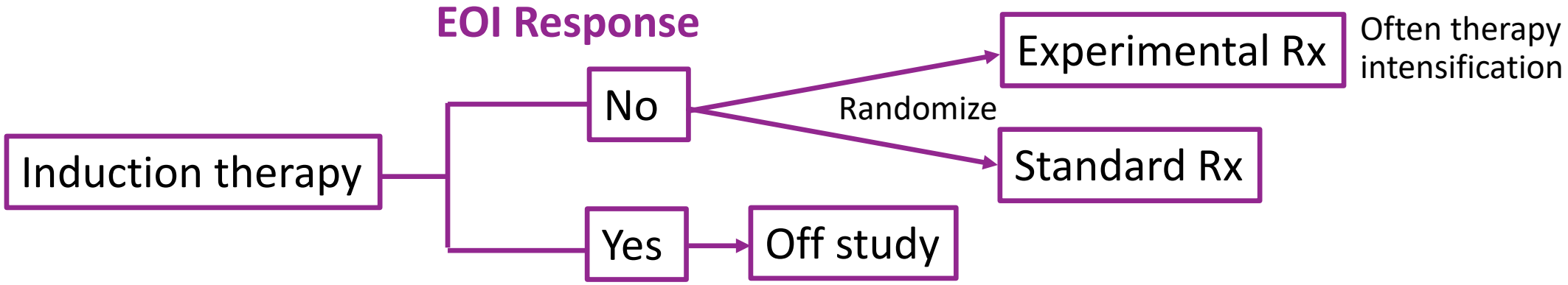


Figure 1. EFS (**B**) and OS (**D**) for responders (CR, PR, MR) vs nonresponders (SD, PD*) based on 2017 INRC (P = 0.08 and P = 0.01, respectively).

Use of EOI response for clinical trial enrichment

- **Prognostic enrichment strategy**
  - Poor prognosis: More events ⇒ more statistical power to evaluate new therapies

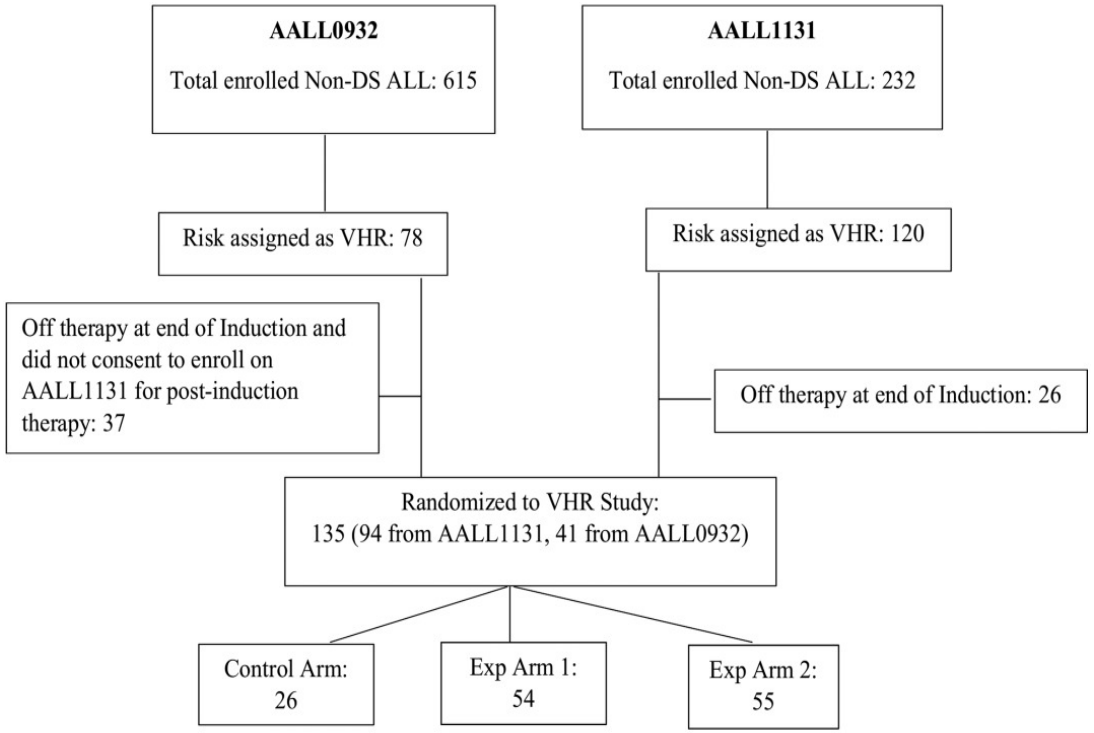


This has been a successful strategy in many pediatric cancer settings, but not always . . .

- **Predictive enrichment strategy**
  - Identify biological characteristics associated with EOI response/non-response
  - Evaluate new therapies that target biological characteristics of the responder or non-responder group (requires assessment of whether better EFS or OS results)

Example: Intensifying therapy for non-responders not successful in AALL1131 randomized cohort for VHR B-ALL

Randomized phase 3 trial to evaluate post-induction therapy using cyclophosphamide (CPM), etoposide (ETOP), and clofarabine (CLOF) for patients with VHR B-ALL



*"Consolidated Standards of Reporting Trials diagram of **pre-amendment** clofarabine at 30 mg/m/d 3-5 **between February 27, 2012, and September 13, 2012** (including only eligible patients).*

Saltzer et al., *Cancer* 2018; 124: 1150-1159, Figure 1

At end of induction on AALL1131, patients were eligible for randomization on the VHR stratum if any of:

- Age ≥ 13 yrs or CNS3 at diagnosis
- Day 29 bone marrow MRD ≥ 0.01%
- Induction failure (>25% blasts in the bone marrow [M3] on day 29)
- iAMP21
- KMT2A rearrangement
- Severe hypodiploidy

In addition, patients on AALL0932 eligible to be randomized after induction on the VHR stratum if certain conditions met

AALL1131 VHR B-ALL example (continued)

Patients enrolled in AALL1131 received a standard 4-drug induction. Patients identified as having VHR B-ALL were randomized 1:2:2 after induction to one of three arms:

Control Arm

- COG–modified augmented BFM, including cyclophosphamide, ARA-C, and 6-MP (during consolidation) or thioguanine (during delayed intensification)

Experimental Arm 1

- Control arm with addition of cyclophosphamide and etoposide during the second half of consolidation and delayed intensification

Experimental Arm 2

- Experimental arm 1 with cyclophosphamide, etoposide, and clofarabine during the second half of consolidation and delayed intensification.

The remainder of the therapy was identical for the patients in these three treatment arms.

Exp Arm 2 CLOSED EARLY FOR TOXICITY:

- **Rates of grade 4/5 infections and grade 3/4 pancreatitis significantly increased in Exp Arm 2**
- Dose of CLOF was reduced to 20 mg/m2/d 3-5, and myeloid growth factor was required after CLOF.
- Despite these changes, **4 of 39 patients (10.3%) developed grade 4 infections**, with 1 of these patients developing a grade 5 acute kidney injury attributed to CLOF, whereas **only 1 of 46 patients (2.2%) in Exp Arm 1 developed grade 4 infections, and there were no grade 4/5 infections in the control arm (n=20).**
- **Four patients in Exp Arm 2 had prolonged cytopenias** for >60 days, whereas none did in the control arm or Exp Arm 1. Counts failed to recover for 2 of these patients, one having a grade 5 acute kidney injury and the other removed from protocol therapy; both events occurred 92 days after the start of consolidation part 2.
- **Trial continued with control and Exp Arm 1 only, using 1:2 randomization.**

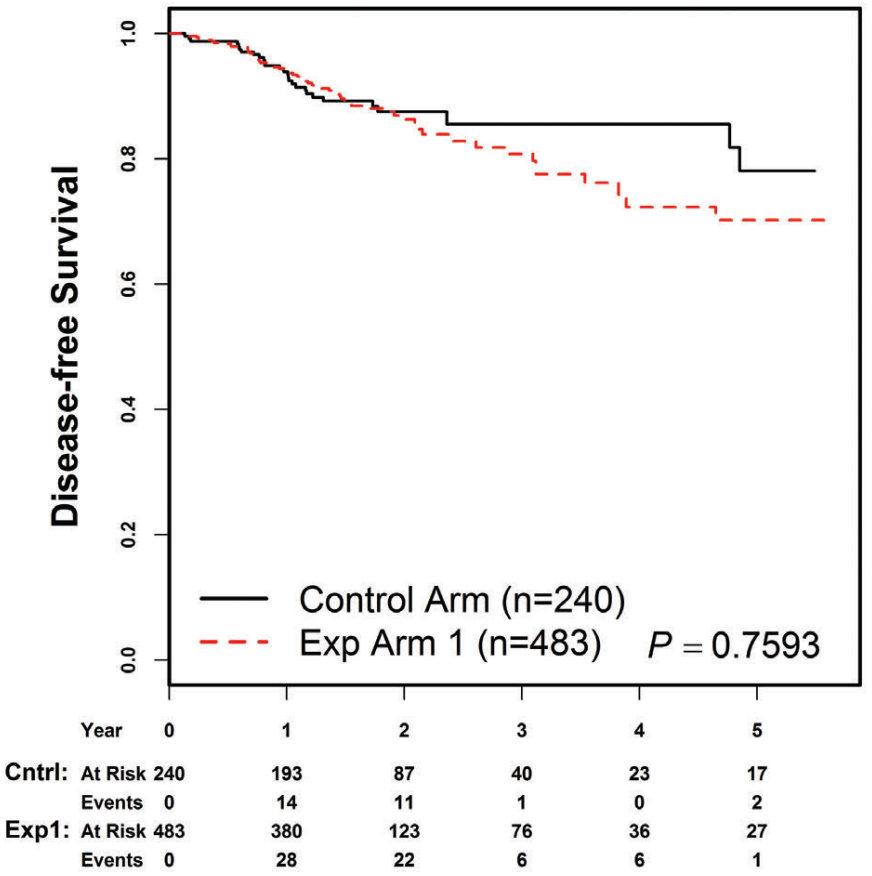
Saltzer et al., *Cancer* 2018; 124: 1150-1159

AALL1131 VHR B-ALL example (continued)

Exp Arm 1 CLOSED EARLY FOR FUTILITY:

- Feb 2017: Interim monitoring boundary crossed [hazard ratio 0.606 (95% confidence interval: 0.297 - 1.237) favoring control arm] and all experimental arms in the very high-risk cohort were closed.
- No significant differences in grade 3/4 adverse events between the two arms.
- **Of note, 4-year DFS of 85.5±6.8% reported for the control arm was higher than the 70% originally predicted based on data available for patients with VHR features treated in the preceding B-ALL studies for standard-risk (AALL0331) and high-risk (AALL0232) patients.**

Burke et al., *Haematologica 2019; 104(5): 986-992*



Dec 2017: With additional follow-up, ***evidence was even stronger*** that Exp Arm 1 was not superior to the control arm (**4-yr DFS 85.5±6.8% for control arm vs. 72.3±6.3% for Exp Arm 1, P=0.76).**

EOI response for drug screening

- Measure of **drug activity** (not necessarily efficacy)
  - Eliminate drugs showing no activity from further development
  - Usefulness may depend on disease subtype and drug class
  - Relies on ability to ***reliably screen out bad drugs*** and ***infrequently screen out good drugs*** (may depend on drug mechanism of action, e.g., cytotoxics, cytostatics, immunotherapies)
- Early **selection among** candidate drugs
  - Select the "winner" in a run-in phase of a trial
  - Screening in phase II trials to bring forward into later phase trials
  - Relies on early endpoint having ability to distinguish drugs with reasonably large efficacy differences, but can be risky (issues similar to surrogates)

Surrogate endpoint validation in a clinical trial

Prentice criteria (Prentice. *Stat Med* 1989;8:431-440)

- The treatment has an effect on the "true" (definitive for assessment of clinical benefit) endpoint (e.g., survival).
- The treatment has an effect on the surrogate.
- The surrogate is associated with (or prognostic for) the true (definitive) clinical outcome.
- Surrogate must fully capture the net effect of treatment on the true clinical outcome
  - Rarely holds even for one treatment much less for multiple treatments one might wish to compare

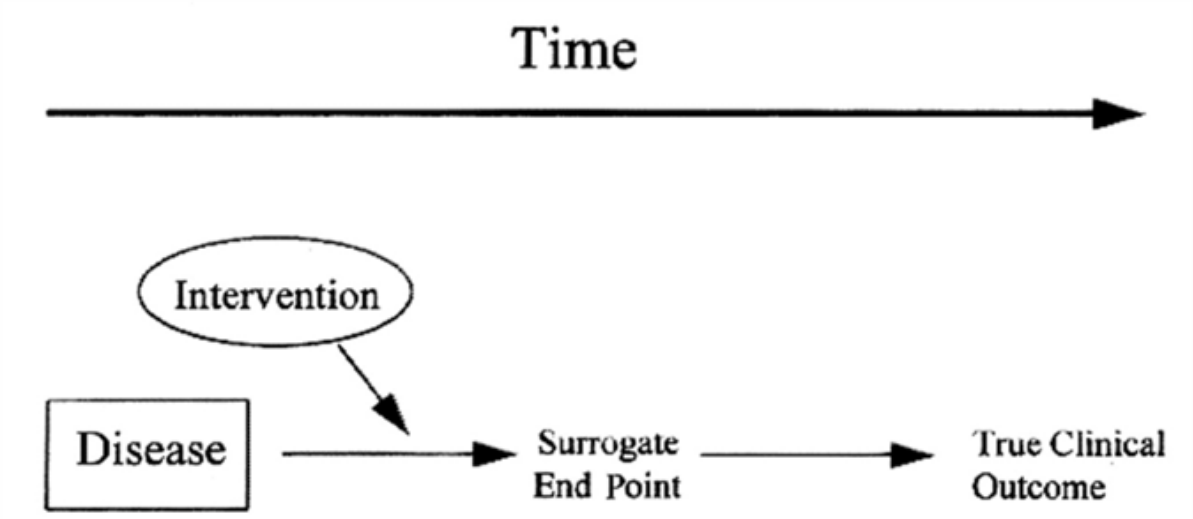
Conceptually appealing, but generally impractical because it is so stringent that it is almost never satisfied.

Surrogate endpoint validation

When Prentice criteria might hold

(Fleming & DeMets. *Ann Intern Med* 1996;125:605-613, Figure 2)

Setting that provides the greatest potential for the surrogate endpoint to be valid: intervention works exclusively through the casual pathway linking the surrogate to true clinical outcome



Surrogate depicted will be both prognostic and a good trial-level surrogate if all treatments under consideration work entirely through this same pathway

Individual-level versus trial-level surrogate endpoint

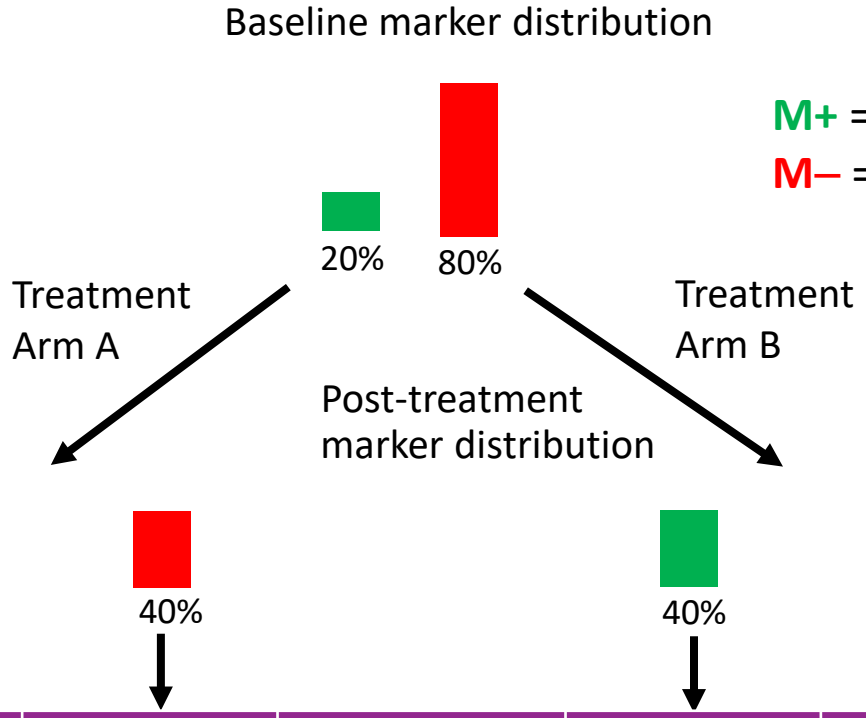
(Buyse et al. *Nat Rev Clin Oncol* 2010;7:309-317)

- **Individual-level "surrogacy"**
  - A variable that is a ***correlate or prognostic*** for the true endpoint within the context of specified treatment(s) and patient population
  - May be demonstrated in context of a ***single*** cohort or clinical trial
  - **Fleming and DeMets: "A correlate does not a surrogate make!"**
    (*Ann Intern Med* 1996;125:605-613, Figure 1)
- **Trial-level surrogacy**
  - A variable or endpoint that can ***replace*** the "true" (definitive) clinical trial endpoint
  - Requires **meta-analysis of clinical trials** to show that a **conclusion about treatment effect** based on the surrogate **reliably agrees with conclusion** obtained **using the true endpoint**
  - **Evidence needed to use a surrogate as a *replacement endpoint* in a NEW trial**

Prognostic is not sufficient for surrogacy

Hypothetical example

*Assume long term event-free survival (EFS) is the outcome used to assess clinical benefit



M+ = favorable prognosis marker for EFS
M– = unfavorable prognosis marker for EFS

Difference (A-B) in **M+ rate** is +20% (60-40%) in all scenarios but **EFS difference can be + or –**

Scenario	Arm A, EFS for M+	Arm A, EFS for M–	Arm A, EFS* overall	Arm B, EFS for M+	Arm B, EFS for M–	Arm B, Overall EFS*	EFS* Diff (A-B)
1	80%	20%	56%	80%	20%	44%	12%
2	90%	20%	62%	60%	40%	48%	14%
3	60%	20%	44%	70%	30%	46%	–2%
4	80%	5%	50%	90%	40%	60%	–10%

Meta-analytic approach to trial-level surrogate endpoint validation

- Collection of data from a sufficient number of relevant trials
- Specify design
  - Clinical benefit measure of interest
    - Clinical outcome measure (e.g., OS, EFS) and quantification of benefit (e.g., hazard ratio)
    - Time-to-event endpoints generally require randomized treatment to assess treatment benefit
  - Method of measuring the surrogate
  - Class of drug (or intervention)
  - Patient population, including possibly biologically defined tumor types
- Extrapolation to new class of drugs or patient population not covered by the meta-analysis can be risky, so think carefully about scope

Example: Surrogacy analysis for minimal residual disease (MRD) in pediatric B-lineage acute lymphoblastic leukemia

Galimberti et al. *JNCI Cancer Spectrum* 2018;2(4):pky069

Initial prognostic evaluation: Is the early endpoint MRD associated with definitive endpoint (EFS)?

- 4830 patients from two large phase III trials that asked a randomized question on the effect of different corticosteroids (dexamethasone vs prednisone) during induction chemotherapy on EFS.
- Association between MRD [negative = 0, low positive = (>0 and <$5x10^{-4}$), and positive = (≥$5x10^{-4}$)] and EFS at the individual and trial levels was evaluated.
- Patients received either Capizzi or high-dose methotrexate regimens in COG trial

Prognostic value of MRD for EFS

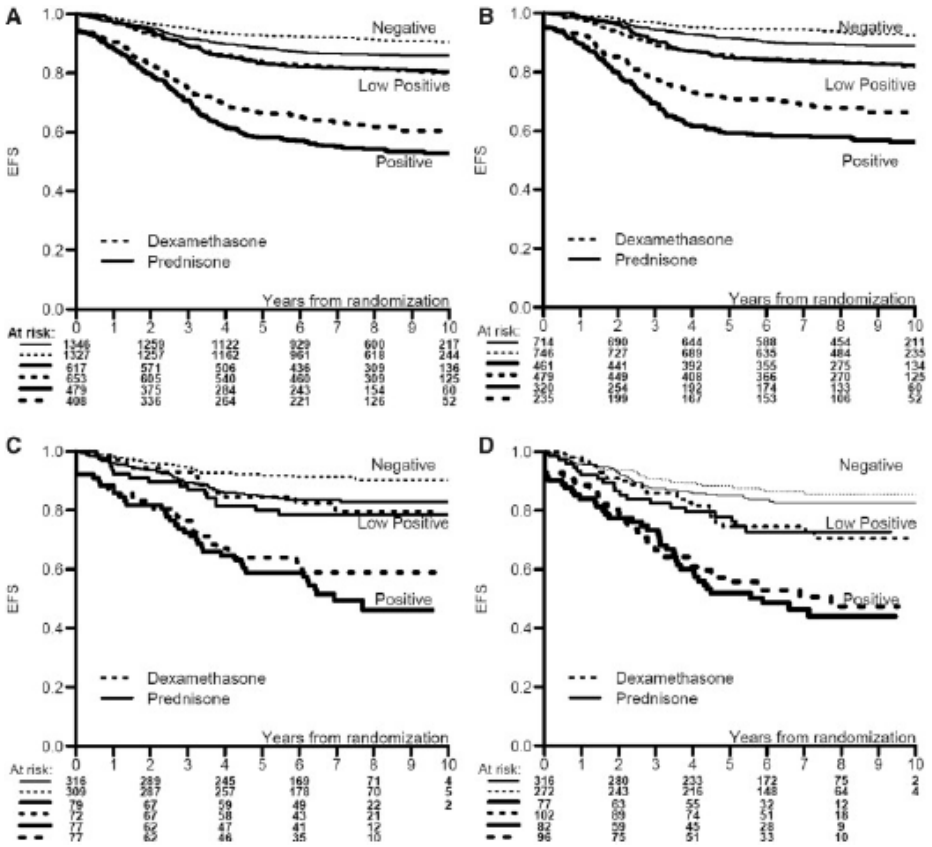


Figure 1 from Galimberti et al. *JNCI Cancer Spectrum* 2018;2(4):pky069

Trial-level surrogacy analysis for minimal residual disease (MRD) in pediatric B-lineage acute lymphoblastic leukemia

Galimberti et al. *JNCI Cancer Spectrum* 2018;2(4):pky069

Trial-level surrogacy validation: Do conclusions about treatment effect based on the early endpoint MRD reliably agree with conclusions obtained using "true" definitive endpoint (EFS)?

- Only two trials were available, but large multicenter
- Centers within each trial grouped according to geographical area to define many trial units for the purpose of meta-analysis regression
- Groupings also accounted for whether patients received Capizzi or high-dose methotrexate regimen within COG trial

Treatment effects (dexamethasone vs. prednisone):

MRD: Proportional odds ratio (OR for high MRD)

EFS: Hazard ratio (HR)

Meta-analysis regression to evaluate MRD as a trial-level surrogate for EFS

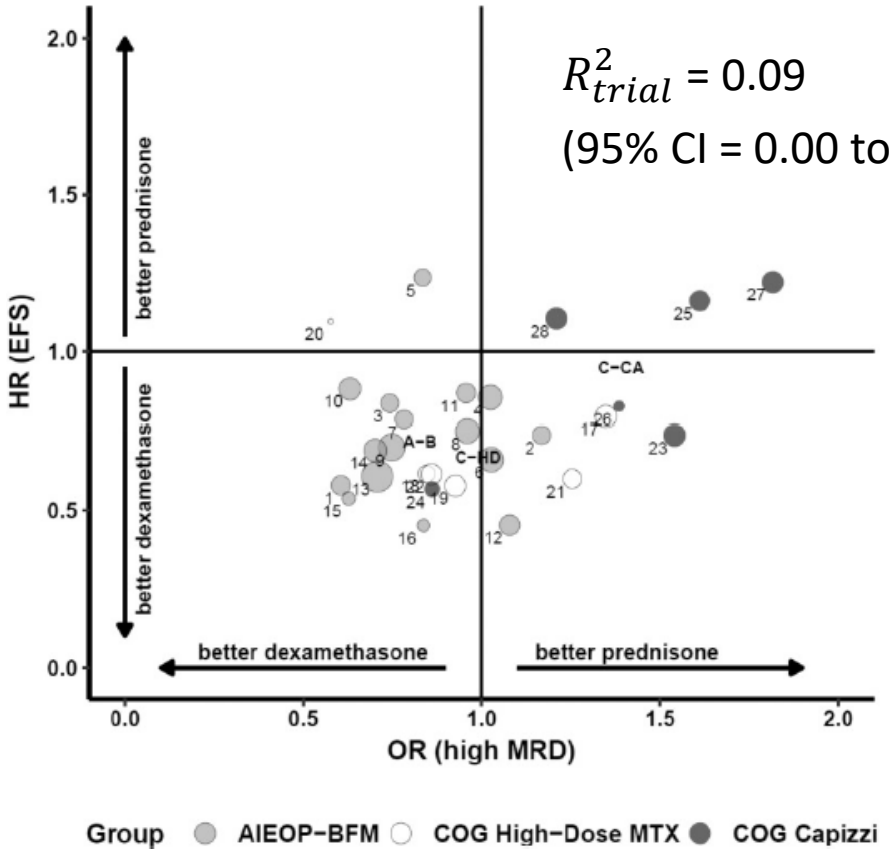


Figure 2 from Galimberti et al. *JNCI Cancer Spectrum* 2018;2(4):pky069

Potential explanations for why an early endpoint might fail to validate as a trial-level surrogate

- Early endpoint not capturing the relevant biology (e.g., disseminated tumor cells vs. cells in primary tumor bed)
- Not measuring early endpoint in best way or at right time (measurement closer to definitive endpoint has greater chance to be good surrogate)
- Effects of therapies delivered after measurement of early endpoint, and possibly influenced by observation of early endpoint
- Value of early endpoint may depend on biological subtypes of tumors
- Need to restrict to a particular class of therapeutic interventions (e.g., targeted vs. not targeted drugs; loco-regional vs. systemic therapy)
- Not enough trials
- Trials too small
- Insufficient range of treatment effect

Concluding remarks

- Clearly define ***intended role*** for early endpoint (e.g., enrichment, screening/activity signal, trial-level surrogate)
- Plan ahead to collect the ***right evidence to support intended role***
  - Harmonize measurement of early endpoint (assays, timing) across trials
  - Identify sufficient number of trials in the relevant patient population with drug/therapy class of interest, appropriate endpoints, etc.
  - Surrogacy analyses typically need ***randomized trials*** with early endpoint measured ***after delivery*** of treatments of interest
- Premature adoption of a “reasonably likely” surrogate may thwart efforts to complete ongoing phase III trials designed to assess a “true” definitive endpoint

THANK YOU