

Surrogate Markers and Correlative Endpoints

- ❑ Surrogate outcomes: outcomes in the causal pathway of true outcome.
- ❑ How closely tied are early PET results and our clinical outcome?
- ❑ Issues with correlative endpoints

Surrogate Markers

- ❑ Replace a distal endpoint (survival) by proxy endpoint (metabolic tissue activity).
- ❑ Benefits of using surrogate markers
 - Reduction in sample size
 - Reduction in trial duration
 - Reduction in cost
 - Reduction in time to evaluate new therapies
 - Potential benefits to the patient: earlier intervention when initial therapy is not effective
- ❑ Their use is NOT AS EASY AS IT SOUNDS...
- ❑ Use of a marker as surrogate for outcome requires that you first identify one.

What is a surrogate marker?

□ ***Defining Characteristic:***

- a marker must predict clinical outcome, in addition to predicting the effect of treatment on clinical outcome

□ ***Operational Definition***

- establish an association between marker & clinical outcome
- establish an association between marker, treatment & clinical outcome, in which marker mediates relationship between clinical outcome and treatment

Surrogate Markers

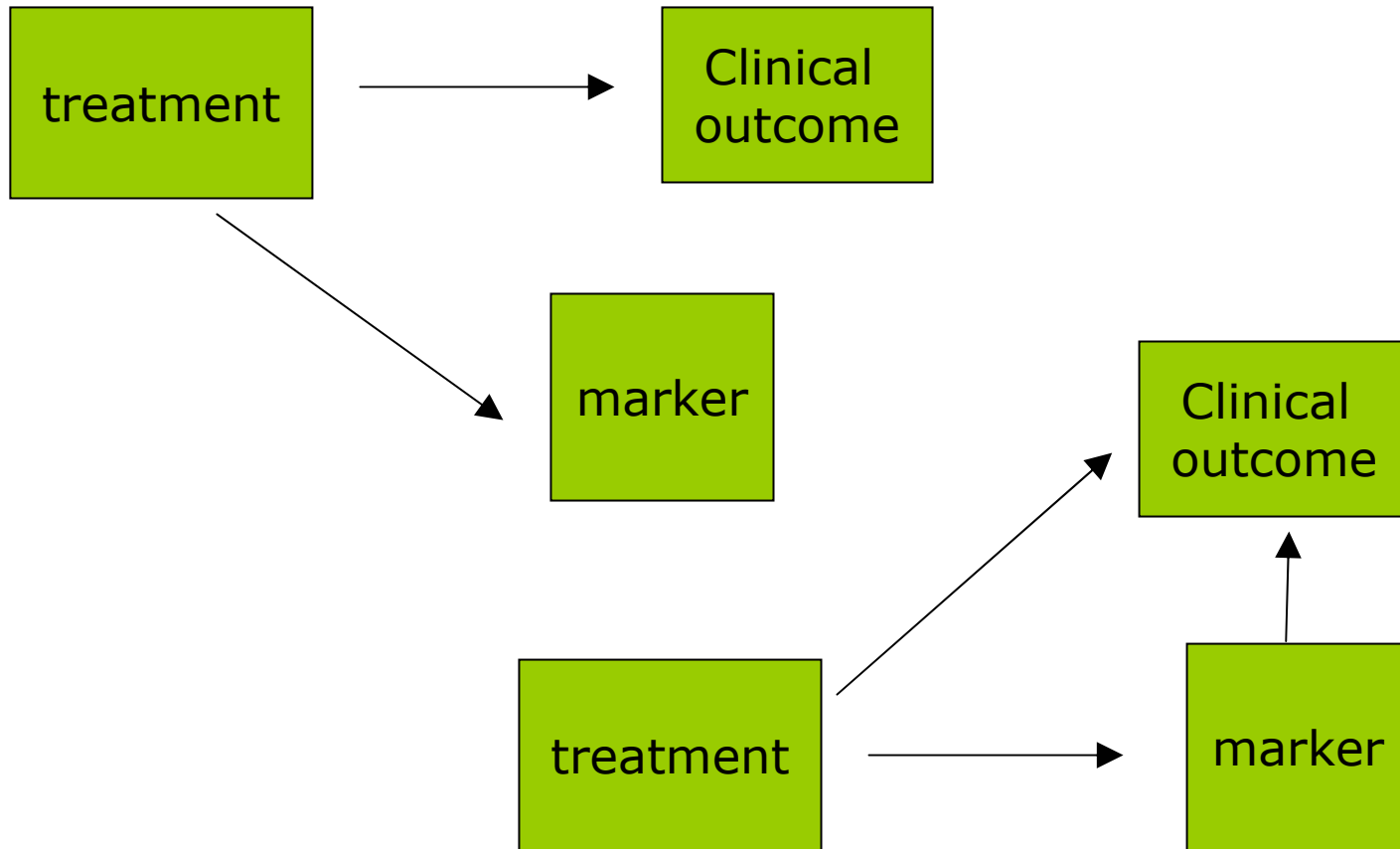
1) establish an association between marker & clinical outcome.



2) establish an association between marker, treatment & clinical outcome, in which marker completely mediates relationship between clinical outcome and treatment.



NOT Surrogate Markers



Marker Studies:

- How are marker studies different than other studies?
 - Primary outcomes are 'efficacy-related', but not clinical.
 - Outcomes are 'surrogate' outcomes or 'correlative' outcomes.
 - Measuring these outcomes is often more invasive and more costly than standard safety or efficacy trials.
 - Measurement of these outcomes can be complicated.

Early PET: surrogate marker?

		Early		total
		+	-	
Late	+	20	0	20
	-	13	47	60
total		33	47	80

$$\text{Sensitivity} = 20/20 = 100\%$$

$$\text{Specificity} = 47/60 = 78\%$$

$$\text{PPV} = 20/33 = 61\%$$

$$\text{NPV} = 47/47 = 100\%$$

Early PET: surrogate marker?

		Early		total
		+	-	
CR+PR	+	21	44	65
	-	15	8	23
total		36	52	88

$$\text{Sensitivity} = 15/23 = 65\%$$

$$\text{Specificity} = 44/65 = 67\%$$

$$\text{PPV} = P(\text{CR+PR negative} \mid \text{early PET positive}) = 15/36 = 42\%$$

$$\text{NPV} = P(\text{CR+PR positive} \mid \text{early PET negative}) = 44/52 = 85\%$$

Early PET: surrogate marker?

event = death, progression, or relapse.

		Early		total
		+	-	
event	+	16	6	22
	-	20	48	68
total		36	54	90

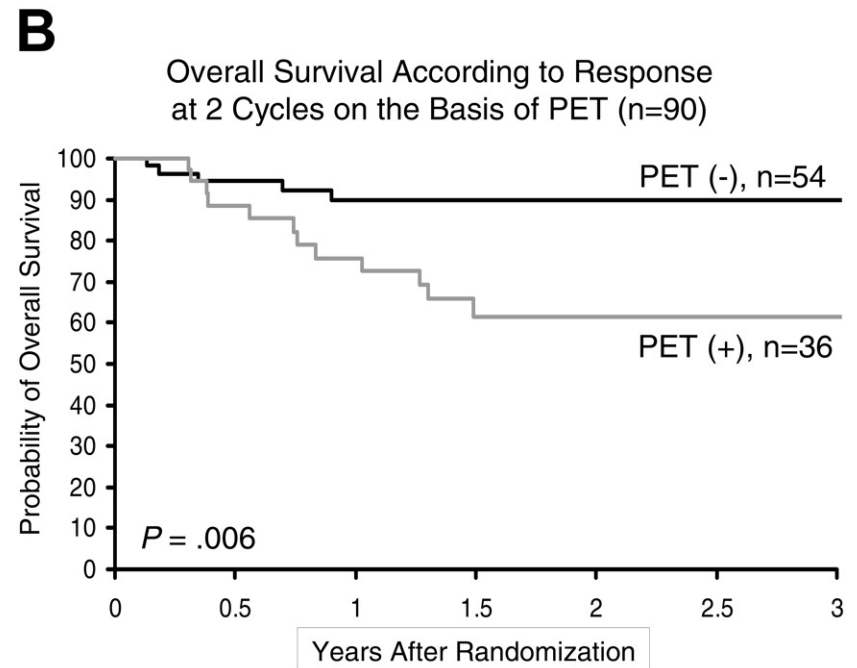
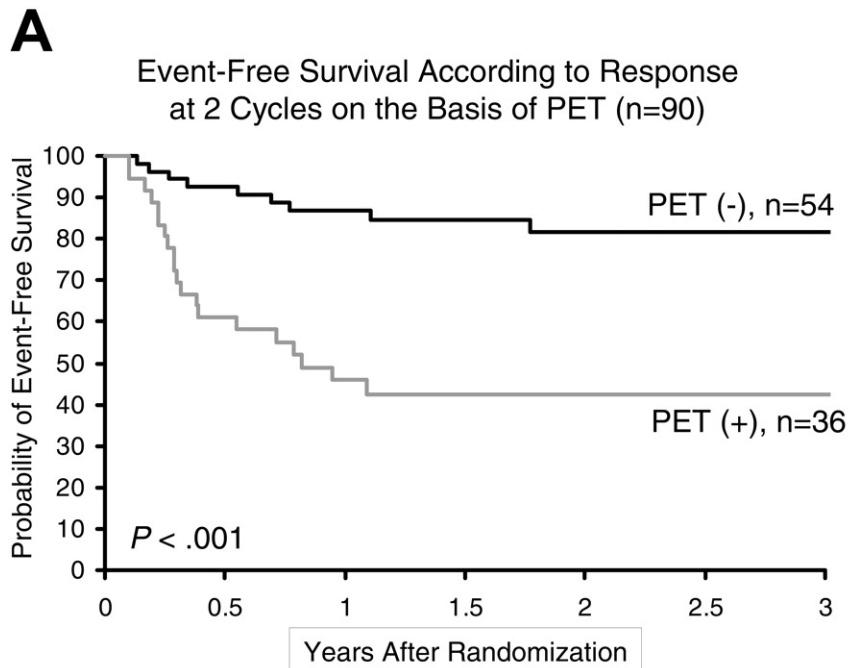
$$\text{Sensitivity} = 16/22 = 73\%$$

$$\text{Specificity} = 48/68 = 71\%$$

$$\text{PPV} = P(\text{event} \mid \text{early PET positive}) = 16/36 = 44\%$$

$$\text{NPV} = P(\text{no event} \mid \text{early PET negative}) = 48/54 = 89\%$$

Dramatic, but true surrogate?

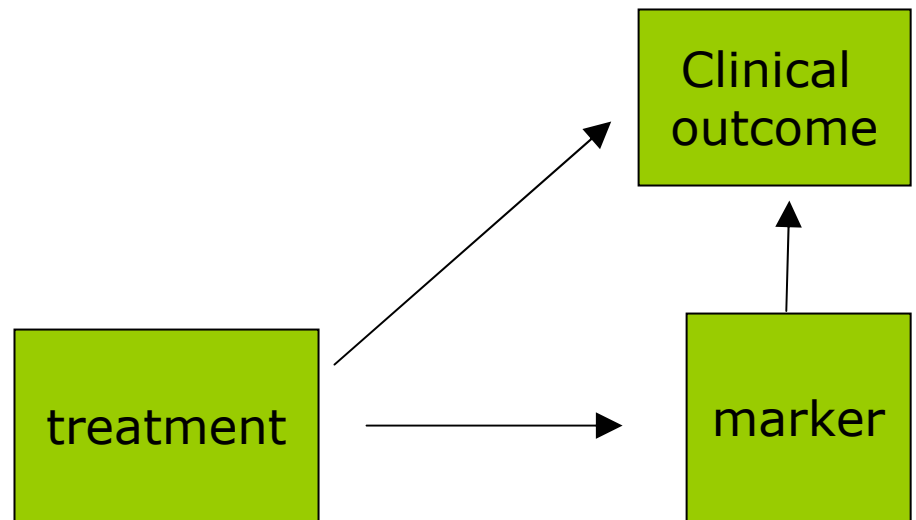


Surrogate Markers

- ❑ So, what would you do?
- ❑ If you perform an early PET and you get a **NEGATIVE** result, how would the result affect your course of action?
- ❑ If you perform an early PET and you get a **POSITIVE** result, how would the result affect your course of action?
- ❑ Put another way (more statistically): if you had two patients who were the same in every way (IPI, performance status, age, etc.), but one had a positive PET at 4 weeks and the other had a negative PET at 4 weeks, would these results be convincing enough for you to act differently?

Correlative Studies

- ❑ **It is still valid to look at markers that you would expect to be correlated with the clinical outcome.**
- ❑ But, we do not want to be overconfident by saying that they are true 'surrogates.'
- ❑ Correlative studies might include:
 - Pharmacokinetics
 - Pharmacodynamics
 - Imaging
 - Other biologic markers that can be measured in serum, biopsy samples, etc.



Correlative Studies

- Design can be tricky
- Often the standard Phase I and Phase II studies (based on clinical outcomes of toxicity and response) are not perfectly suited to looking at correlates
- We end up doing the best we can

Choosing the timing

□ Standard Phase I

- Patients are under surveillance for short term toxicities.
- Patients contact study team when a toxicity occurs after discharge.
- Not an issue of 'when to look.'

□ Add in correlative outcomes:

- Usually a pre-post setting: how does the measure compare after treatment to before treatment.
- Baseline (before treatment) measure is needed.
- Post-treatment measures:
 - When is treatment at its most potent?
 - How often should response be measured?
 - Expensive? Invasive?
 - Is it sufficient to look at clinic visit times?

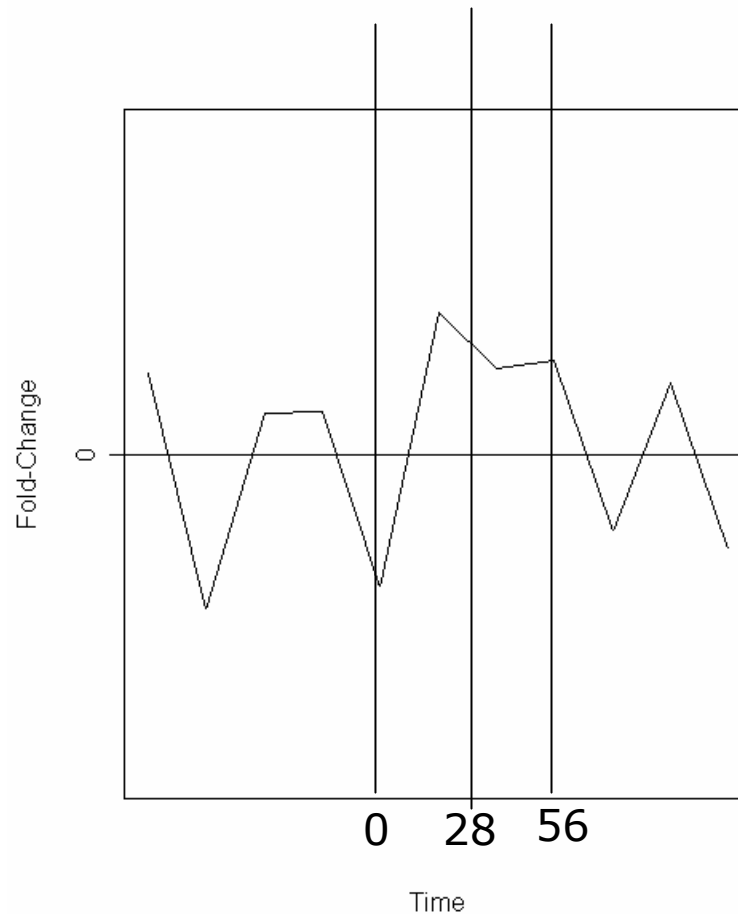
Phase II Designs

- Phase II designs: usually efficacy
- Standard efficacy outcomes
 - Complete response
 - Partial response
 - Overall survival
- For cytostatic agents
 - Progression
 - Progression-free survival
 - Laboratory outcomes...

Defining outcomes in laboratory studies

- ❑ They are often messy (e.g. cell counts, gene expression)
- ❑ More common binary outcomes have nice properties:
 - “Looking for 40% response vs. 20% response”
- ❑ Laboratory outcomes are not so nice:
 - Often skewed.
 - Often have ‘undetectable’ range.
 - Often do not know what to expect.
 - ➔ This makes it hard to plan (i.e. sample size, power).
- ❑ Novel assays: Not obvious what expected changes would be without treatment.
 - How much fluctuation would we expect to see?

Expected Fluctuations

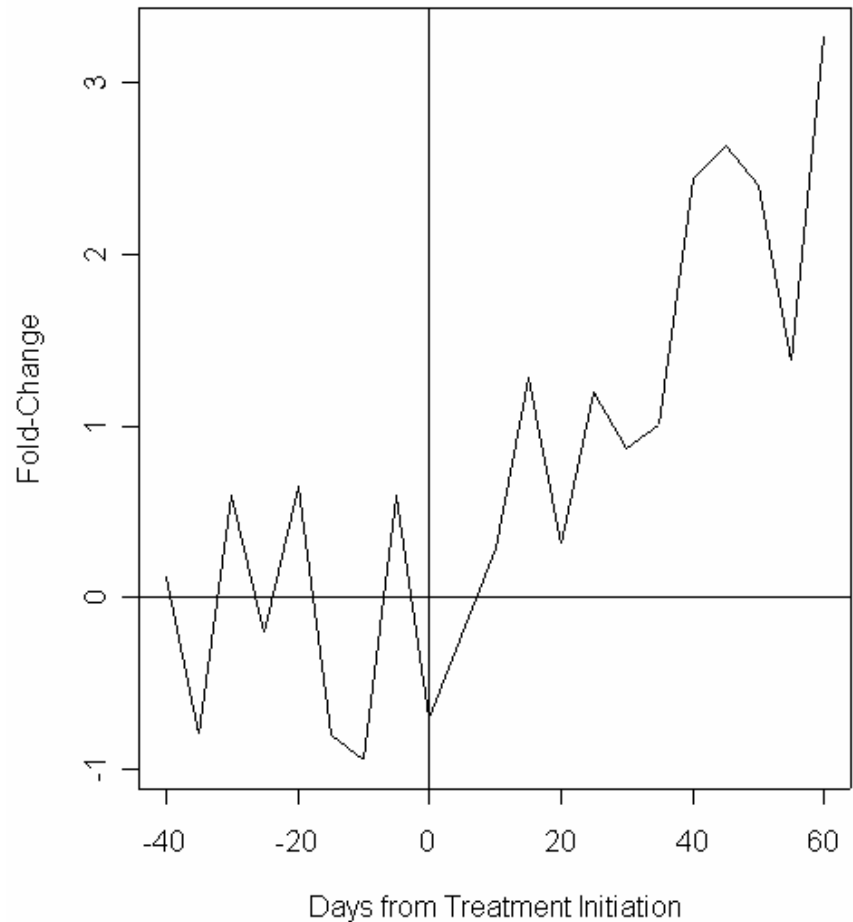


But, with multiple patients, these changes would tend to average to zero.

However, if half of patients have increase and half have decrease, we might conclude that treatment is 50% effective!

Expected Fluctuations

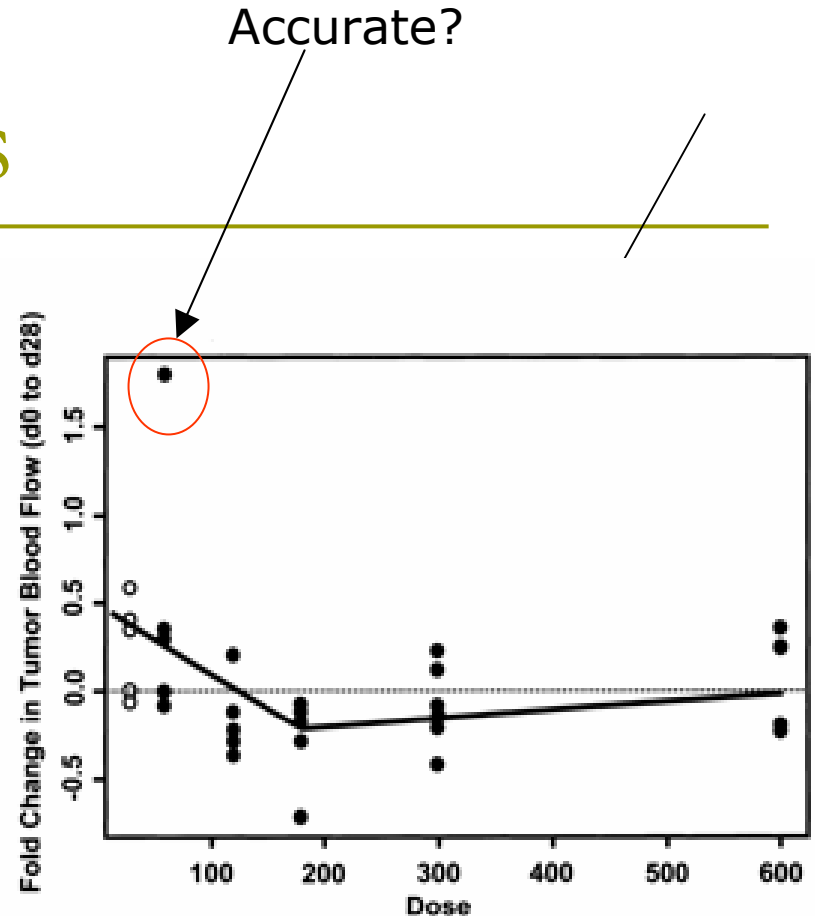
- Follow-up times are compared to baseline
 - That puts a lot of stock in baseline measure
 - Why not consider a 'burn-in' period?
- If baseline is inaccurately measured, all comparisons will be incorrect.



Measurement Issues

- ❑ Regardless of natural fluctuations...
- ❑ How accurately are we measuring our outcome?
- ❑ How accurate can we measure blood flow?
- ❑ Are we on target?

A



Measurement Issues

- Why would it be measured inaccurately?
 - Often LONG protocol to get 'final' measure.
 - Lots of room for errors!
- Although the classic efficacy outcome of 'response' is relatively soft, it has been objectively defined.
- Laboratory endpoints require assays and other measurements, and also assumptions.
- Often, assay is being developed along with the trial.

Assay Properties

- ❑ IS THE ASSAY SENSITIVE?
- ❑ Sensitivity: does the assay detect abnormalities in cases where abnormalities exist?
- ❑ Specificity: does the assay find normal levels in normal cases?
- ❑ “It has been shown that at high flow rates, measuring blood flow by PET underestimates blood flow.”
 - ➡ bias in results.

More Measurement Issues

□ Reliability of assay

- How reproducible are the results?
 - Two samples taken from the same patient on the same day from different lesions?
 - Two samples taken from the same patient on the same day from the same lesion?
 - One sample analyzed twice using the same method?
- Subjectivity
- Inter-rater agreement
- Intra-rater agreement
- In what ways can 'error' come into the procedure?

Great study + bad assay = bad study

Measurement Issues

- ❑ How can we know the reliability of the measures?
- ❑ Preliminary studies (pre-clinical).
- ❑ Build it into the study design!
 - Reliability substudy:
 - ❑ Inter-rater agreement
 - ❑ Intra-rater agreement
 - Incorporate burn-in period
 - Take multiple measures
 - Run the assay (test) more than once

Study Design Issues

- ❑ Trials looking at surrogate and correlative outcomes often need creative/novel study designs
- ❑ Understanding the properties (i.e. reliability, sensitivity) of your assay/measurement technique are crucial.
- ❑ Think carefully about:
 - Are your markers are truly surrogates?
 - Timing
 - The potential for missing data, especially in biopsy studies (Jackie Walling).....

Practical Issue: Biopsies in clinical trials

- Including biopsies is great for research
 - Allows investigation of correlative endpoints
 - Helps understand mechanism of action
- But, practically:
 - Often hurts accrual
 - Mandatory versus optional?
 - Potentially large proportion are 'unevaluable'
 - Might only be useful if **both** of paired samples are evaluable
 - Is it worth the effort (and is it ethical) if you only end up with useable information on a subset of the patients?

Assessability of ex vivo samples used in chemosensitivity assays (from Shrag et al 2004)

Ref	Assay	N	% pts with assessable assays
1	HTCA	470	64
2	HTCA	168	45
3	CCS	68	71
4	DiSC	79	33
5	DiSC	45	78
6	Mod DiSC	SCLC:115 NSCLC:165	29 36
7	MTT	83	88
8	DiSC	18	56
9	ATP	31	93
10	EDR	50	100

Opinion Leader Survey: Assessability in Paired Samples

- 17 OLs at Major Cancer Centers in the US and UK
- 2 phase I studies, 1 phase I/II, 12 phase II studies, largely ongoing
- 11/15 studies mandatory bx

Number of patients recruited (range)	9-58
% Range of paired mandatory biopsies	43 -100
% of Patients with paired evaluable biopsies in mandatory studies	27 - 61
% of Patients with paired evaluable biopsies in optional studies	10 - 58

OL Survey: Take Home Messages

- Paired biopsy studies are hard!
 - No center has this cracked
- Multiple issues:
 - Clinical Trial Design
 - Institutional
 - Technical
 - QC
 - Unknown false positives/negatives
 - Assays
 - Few / no correlations with xenograft / pk data/clinical data
 - Standardization/ Quantitation
 - Ethics

References from Jackie Walling

1. Von Hoff et al 1983
2. Von Hoff et al 1991
3. Von Hoff et al 1990
4. Gazdar et al 1990
5. Wilbur et al 1992
6. Shaw et al 1993,1996
7. Xu et al 1999
8. Cortazar et al 1997
9. Kurbacher et al 1998
10. Loizzi et al 2003