# Section IV: Study design

- Ideal design: A traditional RCT
- Other RCT design options
- Early-stage study designs

# The ideal design

A traditional RCT is the ideal setting in which to discover and evaluate markers and treatment rules.

- Subjects randomized to treatment ($A = 1$) or standard of care ($A = 0$)
- Marker $X$ measured at baseline
- Subjects followed for clinical outcome ($D$)

$X$ may be measured prospectively or retrospectively, e.g. using stored baseline samples.

Randomization ensures comparability of treatment groups– overall and conditional on $X$.

# Variations on the traditional RCT

- Stratify the randomization on pre-specified binary marker "signature"– to ensure balance in treatment assignment conditional on the signature
- Retrospectively sub-sample RCT participants for $X$ measurement, e.g. using case-control sampling, to conserve resources

These variations have implications for estimation, but are not conceptually different from the traditional RCT.

Sargent et al. (*J Clin Onc* 2005); Simon et al. (*JNCI* 2009)

# Criteria for sizing the RCT

When marker discovery/evaluation is a primary or secondary study objective, the most common approach is to size the study to test for a marker-by-treatment interaction.

It is more compelling to size the study to quantify marker performance, e.g.

- Control the magnitude of errors in marker-based treatment recommendations
- Rule out $H_0 : \mathcal{I} = 0$ with high probability
- Ensure $\mathcal{I}$ can be estimated with sufficient precision

Simulation-based sample size calculations are flexible, and can accommodate evaluation of a data-derived rule.

Janes et al. (*Stat Med* 2015); Laber et al. (*Stat Med* 2015)

# When the marker is measured on a select subset of RCT participants

Unfortunately, marker measurements are sometimes only available for a subset of trial participants *who are not selected by design*.

If the outcome or marker distribution is different in this subset vs. in the population as a whole, results that are generated do not generalize to the population.

This situation should be avoided whenever possible.

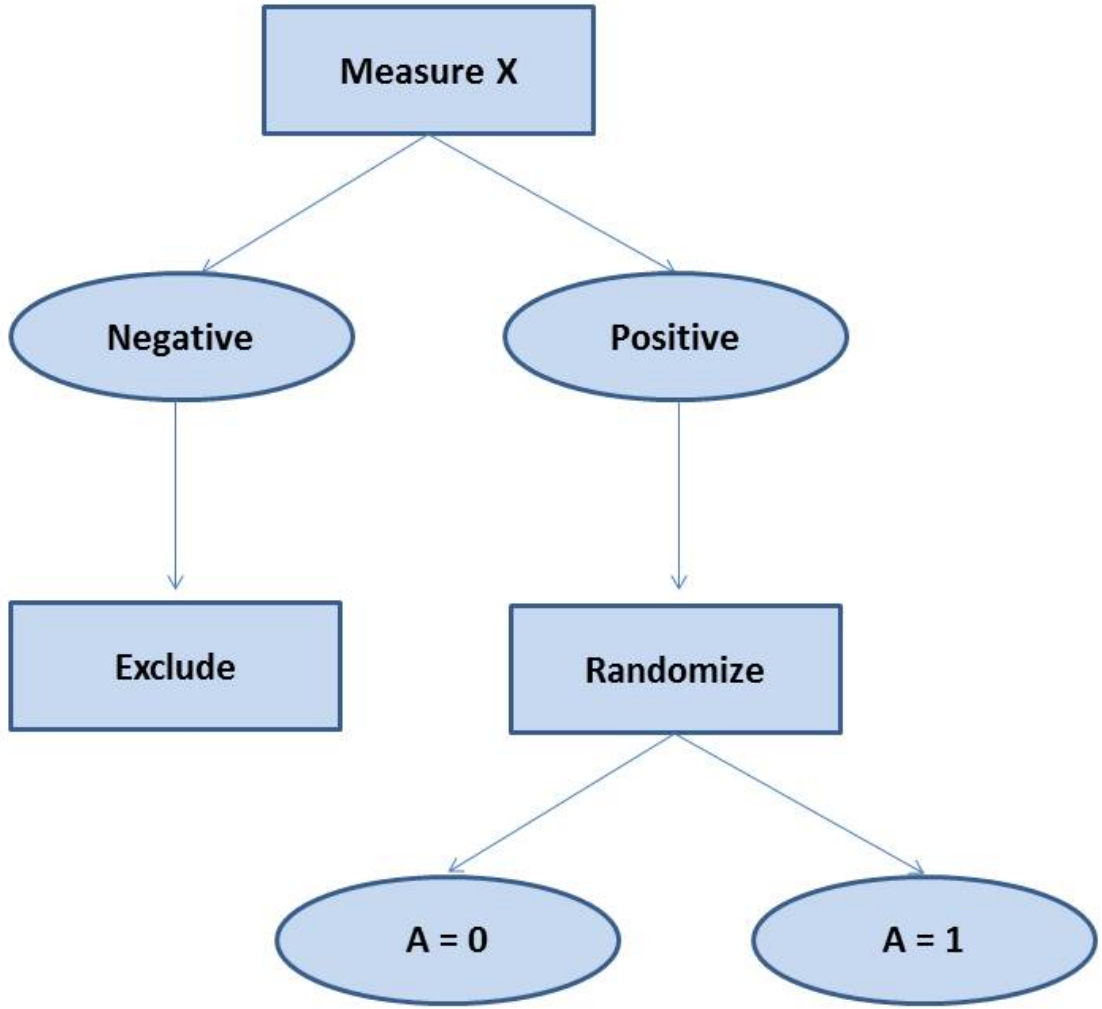
# Other RCT design options

- Targeted design, a.k.a. enrichment design
- Hybrid design
- Marker-strategy design
- Modified marker-strategy design

We comment on the merits of these designs for discovery and/or validation of markers and treatment rules.

In describing the designs, we focus on the context where standard of care ($A = 0$) is the default approach absent $X$.

# Targeted design




Enroll only those with a marker signature thought to predict benefit from treatment

Simon and Maitournam (*CCR* 2004); Mandrekar and Sargent (*J Clin Onc* 2009)

# Classic example of targeted design

NSABP B-31 and NCCTG N9831 trials of trastuzumab plus adjuvant chemotherapy for treating HER2-positive breast cancer (Romond et al. 2005)

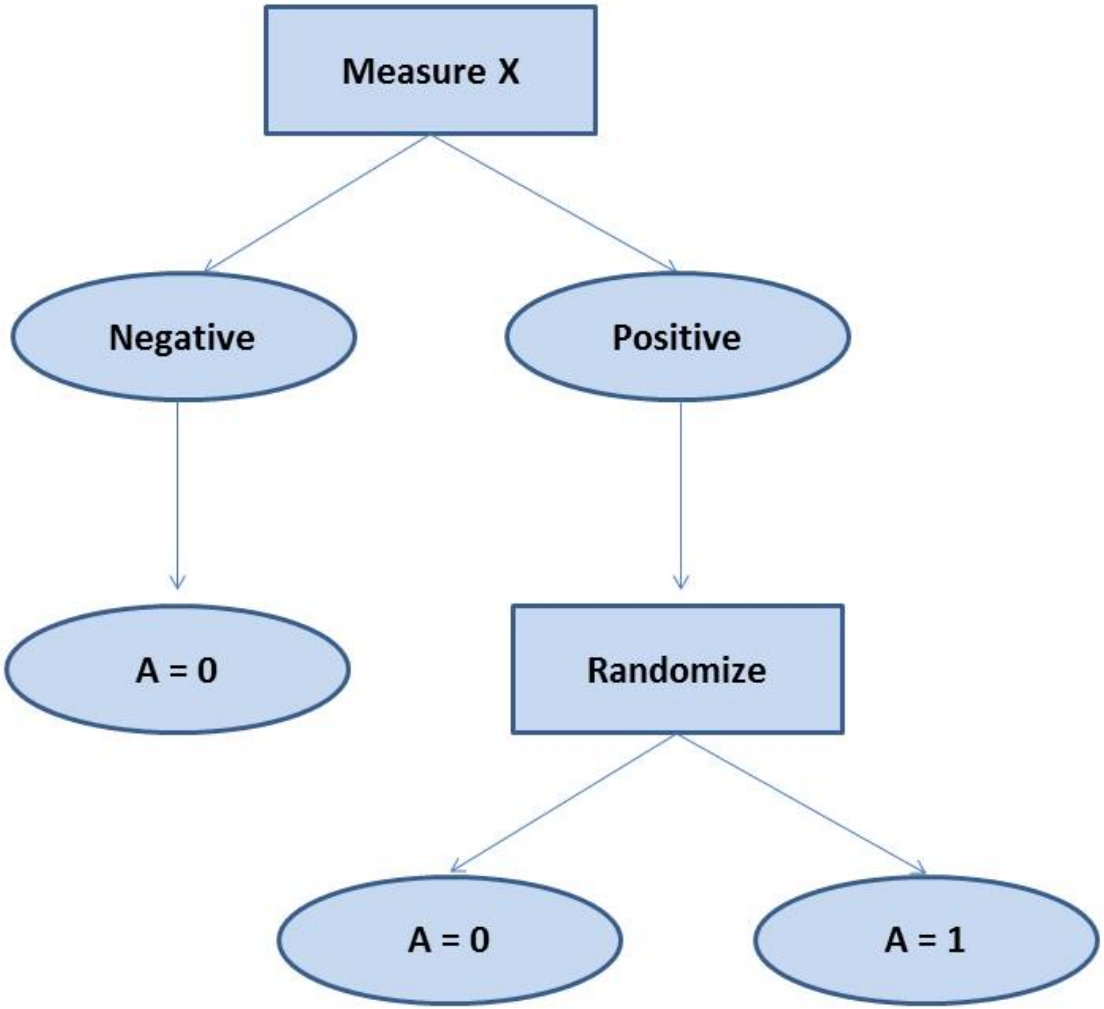
# Attributes of the targeted design

The design is commonly employed because, if the signature is thought to predict a large treatment effect, it requires a small sample size.

However, the design has fundamental limitations for marker discovery/validation:

- Only allows assessment of treatment efficacy among those with marker signature. Efficacy among other subjects cannot be assessed.
- Clinical impact of the marker signature cannot be assessed.
- Design also precludes evaluation of different markers and/or treatment rules

# Hybrid design




Targeted design, plus default treatment for those without signature

Mandrekar and Sargent (*J Clin Onc* 2009)

# Examples of hybrid designs

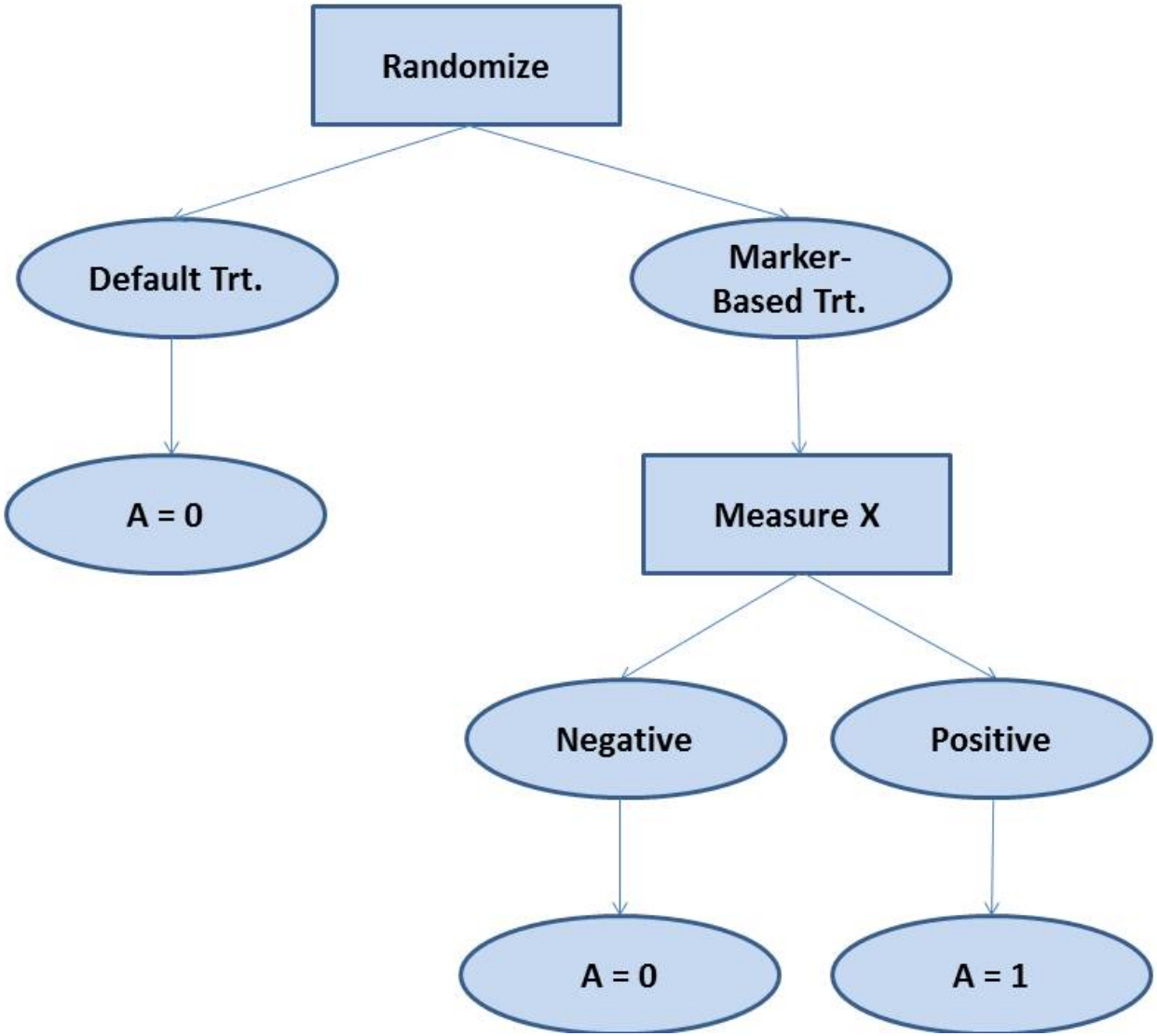
1. TAILORx trial evaluating adjuvant chemo. for ER+ or PR+/node-negative breast cancer, and the Oncotype DX RS (Sparano 2006):
   - Women with low RS provided hormonal therapy
   - Women with intermediate RS randomized to hormone therapy with or without chemo.
   - Women with high RS provided hormone therapy plus chemo.
2. MINDACT trial evaluating adjuvant chemo. for node-negative breast cancer, and the MammaPrint signature
3. ECOG 5202 evaluating oxaliplatin, leucovorin calcium, and fluororacil with vs. without bevacizumab in patients with resected stage II colon cancer at high risk for recurrence based on molecular markers

# Attributes of the hybrid design

It is said that the design provides "additional value" over and above the targeted design because patients without marker signature are followed under standard of care.

However, the key disadvantages of the targeted design for marker discovery and validation still apply.

# Marker strategy design




Randomize participants to marker-based or default treatment

Sargent et al. (*J Clin Onc* 2005)

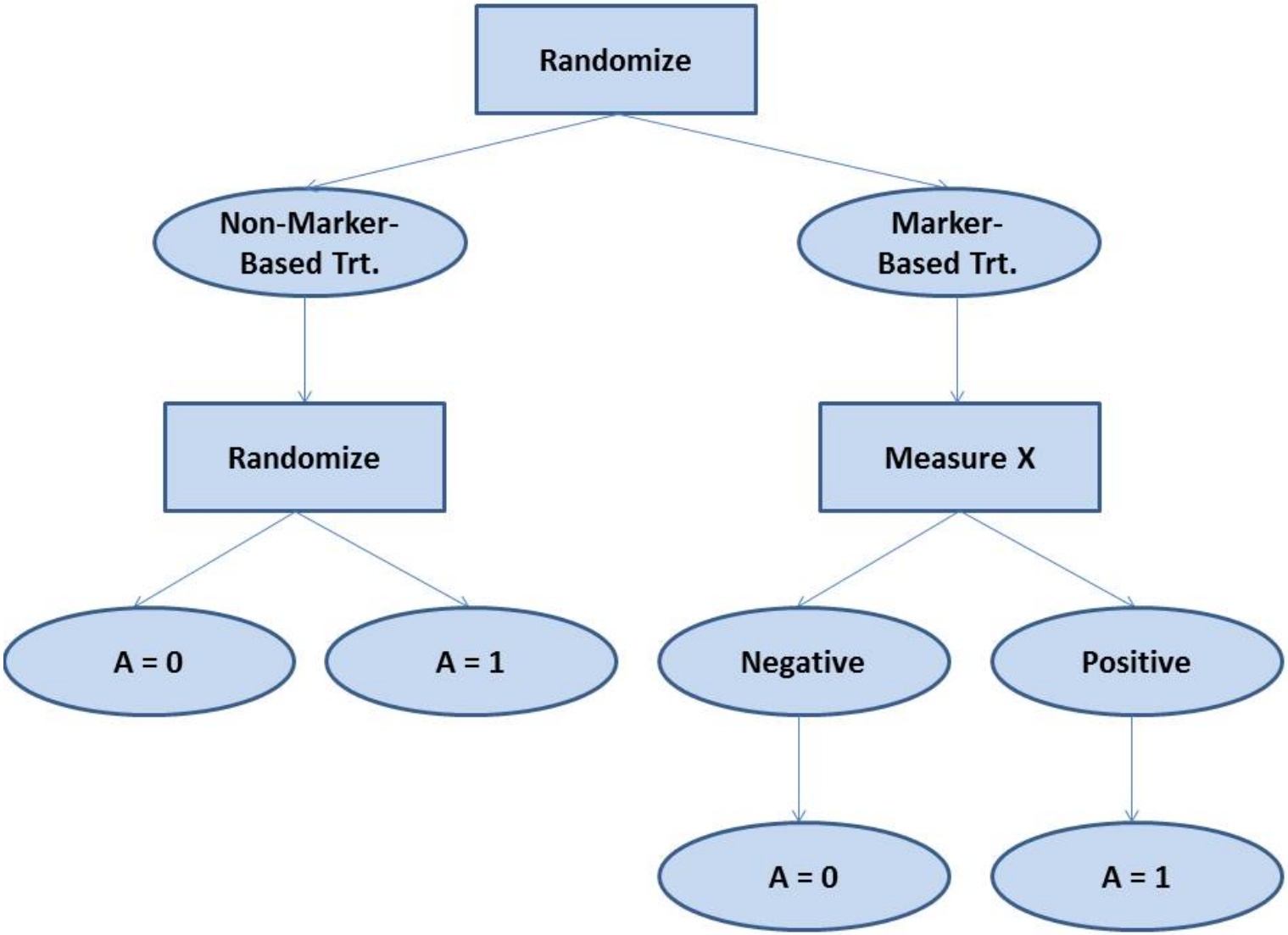
# Examples of marker strategy designs

1. Phase III trial of mRNA expression-based cisplatin for treating non small cell lung cancer (Cobo et al. 2007)
2. Phase III trial of genotype-guided warfarin dosing (Kimmel et al. 2013)
3. Phase III trial of HLA-based abacavir treatment for HIV infection (Mallal et al. 2008)

# Attributes of the marker strategy design

- The difference in expected outcomes between arms is the clinical impact of the marker signature– including the effect of potential non-compliance with marker-recommended treatment
- Fixing sample size, less efficient evaluation of clinical impact than traditional RCT if marker is only measured on marker-based treatment arm. Efficiency is recovered if marker measured for all participants
- Marginal treatment effect cannot be estimated
- Alternative markers/rules cannot be assessed $\implies$ perhaps best used as late-phase validation study of marker signature

# Modified marker strategy design




Sargent and Allegra (*Sem in Onc* 2002); Sargent et al. (*J Clin Onc* 2005)

# Attributes of the modified marker strategy design

- Retains the ability to estimate the clinical impact of the marker signature
- Allows marginal treatment effect estimation in the non-marker-based treatment arm
- Allows alternative markers/rules to be discovered in the non-marker-based treatment arm, albeit with low power
- Fixing sample size, less efficient evaluation of the marker signature than traditional RCT or marker-strategy design, even if $X$ measured for all– since subjects in non-marker-based treatment arm receiving $A = 1$ are not used for analysis

We know of no trials implementing this design.

| | Marker objectives | Can marker performance be evaluated? | Must pre-specify marker? signature? | Suitable for marker discovery? |
|---|---|---|---|---|
| Traditional RCT | 1. Evaluate $X$<br>2. Discover $X_2$ | Yes | No | Yes |
| Targeted design | 1. Evaluate trt. eff. given $X$-signature | No | Yes | No |
| Hybrid design | 1. Evaluate trt. eff. given $X$-signature | No | Yes | No |
| Marker strategy design | 1. Evaluate $X$-signature | Yes, $X$-signature | Yes | No |
| Modified marker strategy design | 1. Evaluate $X$-signature<br>2. Discover $X_2$ | Yes | Yes | Yes* |

* With lower power than traditional RCT

# Adaptive signature design

Divide data from a traditional RCT design into 2 "stages"

- Stage 1 data used to develop a biomarker-based treatment rule
- Stage 1+2 data used to evaluate the overall treatment effect
- Stage 2 data used to evaluate the treatment effect in the marker-defined subset *if the overall treatment effect is not statistically significant*
- Study is powered to detect *any treatment efficacy*, defined as significantly positive treatment efficacy overall or in the marker-defined subset

Design is not powered to evaluate the marker, but to address: Is there a subgroup for whom treatment is beneficial? This objective may be compelling to a drug developer.

Freidlin and Simon (*CCR* 2005); Freidlin et al. (*CCR* 2010)

# Early-stage study designs

# Early stage marker studies

Biomarkers are often discovered and first evaluated in settings other than RCTs.

We comment on the merits of (3) early-stage designs:

- A cohort study of subjects treated with standard of care
- A single-arm treatment trial– a cohort study of subjects treated with the experimental intervention
- An observational study where some subjects are treated ($A = 1$) and some are not ($A = 0$)

# An untreated cohort study

A cohort of subjects treated with standard of care and followed for the clinical outcome. $X$ is measured at baseline, prospectively or retrospectively.

Markers first developed using untreated cohort data:

- Gail breast cancer risk model
- Oncotype DX recurrence score

The underlying concept is that individuals likely to have poor outcomes absent treatment may have more potential to benefit from treatment.

# Analysis of untreated cohort data

Evaluate how well the marker predicts outcomes absent treatment:

- Model $E(D|A = 0, X)$.
- Evaluate the model's predictive capacity using metrics such as AUC , $R^2$ (PEV), or mean squared prediction error

Importantly, good prediction does not necessarily imply the marker is useful for treatment selection– the marker may be equally predictive of outcomes under treatment, and therefore not predictive of treatment efficacy.

An untreated cohort study may also miss markers predicting treatment efficacy that do not predict outcomes without treatment.

# A treated cohort study

A single-arm treatment trial, i.e. a cohort of subjects experimentally treated and followed for the clinical outcome. $X$ is measured at baseline, prospectively or retrospectively.

Estrogen receptor expression, as a predictor of response to endocrine therapy for treatment of breast cancer, was first evaluated using treated cohort data.

The underlying concept is that individuals likely to have good outcomes under treatment may have greater treatment efficacy.

# Analysis of treated cohort data

Evaluate how well the marker predicts outcomes under treatment:

- Model $E(D|A = 1, X)$.
- Evaluate the model's predictive capacity using metrics such as AUC , $R^2$ (PEV), or mean squared prediction error

Again, however, good prediction does not necessarily imply the marker is useful for treatment selection.

And a marker predicting treatment efficacy that does not predict outcomes under treatment may be missed.

# An observational study

Data for a set of subjects followed for the clinical outcome. Some subjects are treated and some are not, based on subject/physician choice. $X$ is measured at baseline, prospectively or retrospectively.

The well-known limitation of this design is that there is potential confounding: treatments may be chosen based on factors (measured and unmeasured) related to the outcome.

Two approaches to estimation:

1. Include potential confounders $Z$ in the outcome model– and develop and evaluate treatment rules that depend on $(X, Z)$
2. Use a propensity score model for $P(A = 1|X, Z)$ to adjust for confounding in developing and evaluating treatment rules based on $X$

# Summary

- The ideal study design is a traditional RCT
- Other RCT design options should be evaluated in terms of the research objectives they address
- Early-stage study designs we discussed have their respective merits, but their limitations must also be recognized.