UC DAVIS HEALTH | MIND INSTITUTE

Applied Statistics Seminar Series for Translational Researchers, April 14, 2021

Design and Analysis of Biomarker Studies

Kyoungmi Kim, Ph.D.

Professor of Biostatistics

Director, Biostatistics, Bioinformatics, and Research Design Core, the MIND IDDRC Center

UC Davis School of Medicine

Sponsored by Biostatistics Programs affiliated with the NIH-funded Centers

- Clinical and Translational Science Center
- MIND Institute Intellectual and Developmental Disability Center
- Comprehensive Cancer Center
- Environmental Health Sciences Center

UC DAVIS HEALTH | SCHOOL OF MEDICINE
Clinical and Translational Science Center

UC DAVIS HEALTH | MIND INSTITUTE

UC DAVIS HEALTH | COMPREHENSIVE CANCER CENTER

UC DAVIS
ENVIRONMENTAL HEALTH SCIENCES CENTER

Biomarker

- A characteristic that is objectively measured as an indicator of normal biological process, pathogenic processes, or pharmacological responses to a therapeutic intervention.
  - Single measurement
  - Signature based on multiple measurements (potentially biomolecular, imaging, clinical, etc.) – more common in recent biomarker research

- Types of Biomarkers
  - Diagnostic (to detect disease)
  - Prognostic (to predict disease progression)
  - Predictive (to predict clinical response to treatment)
  - Endpoint – intermediate and surrogate markers

Goals of Biomarker Studies

- To develop prescriptive/predictive biomarkers that indicate who has disease or will benefit from new treatment over standard care.

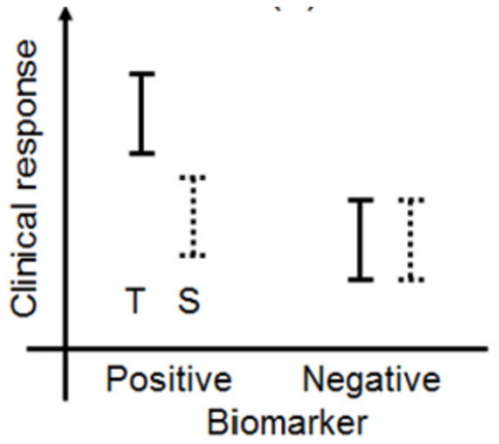
- To identify biomarkers that can be potential drug targets for therapeutic development.

Prognostic Biomarker

- Gives information on outcome independent of treatment
- Generally not informative for treatment decisions
- Measure of inherent aggressiveness of disease
- Intend to identify patients with differing (high vs. low) risks of a specific outcome such as disease progression or death (not known yet at baseline – so it is prospective)
- A prognostic biomarker should be associated with future trajectory of disease that eventually is instrumental in identifying high risk patients.
- Major problems of prognostic biomarker studies
  - Inadequate focus on intended use of the biomarker
  - Poor evaluation

Predictive Biomarker

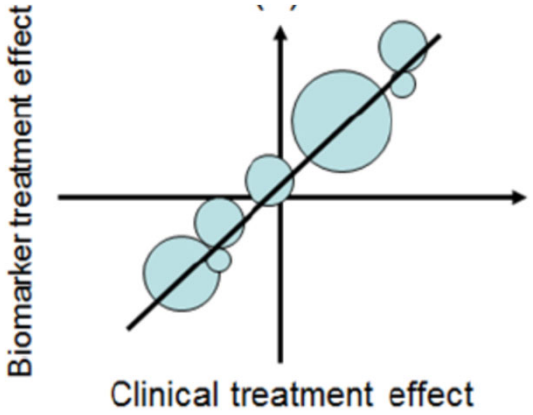
- Gives information on relative effectiveness of treatments
- Informative for treatment decisions
- Development and validation of a predictive biomarker requires data from an RCT
  - Need measurements from patients on the targeted treatment and patients on standard care
- Intend to identify patients who are more likely to respond to a specific treatment and thus have better outcomes (survival at 3 years, tumor shrinkage) compared to standard care.
- Example (figure): biomarker(+) patients responded to the new trt or to a greater degree than biomarker(-) patients



Surrogate Endpoint

- is a substitute for a clinical endpoint.
- is an alternative measure of the treatment effect that correlates with the true endpoint (e.g., death) but does not necessarily have a guaranteed relationship with it directly.
- Use to prospectively predict clinical benefit or outcome

- Example (figure):
  - Imaging screen (CT, MRI, PET)
  - Pharmacodymanic biomarker that is known to be directly related with the endpoint outcome



Developing a Prognostic biomarker

- Generally need discovery (training) data
  - A dataset of patients with the disease of interest including
    - An outcome (related to progression of disease such as tumor growth)
    - Baseline measures (genetic mutation, serum measurements, gene expression values) – serve as predictors

- Building a prognostic model
  - Can estimate the probability of developing the outcome using baseline measures
  - Based on the estimated probabilities, stratify high-risk vs. low-risk patients
  - Then evaluate the accuracy of classification
  - Usually use regression-based approaches

Statistical approaches to model building

- Use existing data from a retrospective study
  - Data on patients:
    - X_i = serum measurement for patient i (can be more than one measurement)
    - T_i = Treatment received of patient i
    - Y_i= response/outcome of patient i
    - C_i= known important clinical variables (i.e., covariates) for patient i

- Then develop a statistical model such that a function of X_i is a prediction of Y_i.

  e.g., $Y_i = \beta_0 + \beta_1 f(X_i) + \beta_2 f(X_i) \times T_i + \beta_3 C_i$

Statistical approaches to model building (cont.)

- Need to address questions
  - Model construction
    - If there is more than one serum variables,
      - which serum variables to include in a model?
      - how many variables?
      - How to combine multiple variables (i.e., what kind of combination)?
  - Evaluation of model performance for intended use
    - Is the prediction close to the true response/outcome?
  - Clinical benefit
    - Do serum variables add any information over and above clinical variables?

Model building

- $f(X_i)$ is a statistical model
  - Well known models
    - Multiple linear regression
    - Logistic regression
    - Cox model
    - Linear discriminant analysis
  - More complex models for big (mostly high-dimensional) data
    - Random forests
    - Support vector machines
    - Gradient boosting
    - Neural networks
    - Other supervised machine learning models
    - Many of these models are "black boxes"- it may be ok if the goal is prediction not interpretation.
- Don't expect one model to be always better for every dataset and every context.
- Sometimes complex models are not better than simple models (overfitting issues).

Stages of Development

- Need to demonstrate that the prediction from the developed model $f(X_i)$ will be good for "future" patients.
- Major issue: Overfitting
    - The outcome data from discovery (training) dataset has been used to fine-tune the details of the model $f(X_i)$.
    - The model looks great on the data used to develop it and gives great predictions on this data.
        - look impressive, but
        - biased and meaningless to show prediction results
    - The predictions won't be good on future observations (generalization issues).
    - Hence need to balance between complexity and generalizability

Overfitting Reduction Strategies

- Use hold-out (training-test split) or cross-validation methods
  - Training set to train the model
  - Test set to estimate the prediction error rate of the model
- Refine the model so that it works well on data that was not used to build the model (i.e., portion of test set)
- Gives a final model ready to use on new data

- Validation of the model – how good is the model?
  - External validation - need to be evaluated on a dataset you have not seen or touched before (i.e., independent validation datasets)
  - External validation data is vital to demonstrate validity of the model for data collected in different settings (i.e., a measure of generalizability).

Metrics for Assessment of the Model as Biomarker

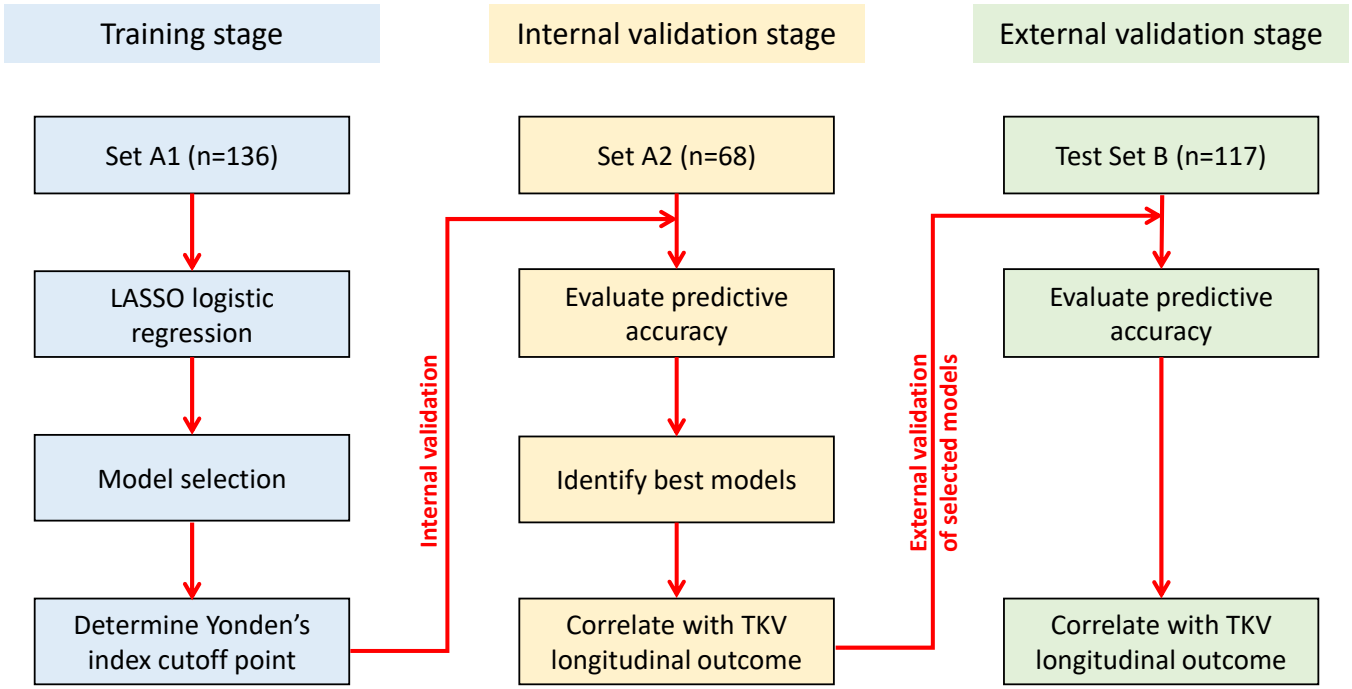
- ROC curves (Receiver Operating Characteristics)
- AUC (Area under the ROC curve)
  - Rough rule of thumb – if AUC >0.75, it is promising and needs further validation.
- Concordance index (C-index)
- Sensitivity and Specificity – require to define a cutoff threshold
- Positive and negative predictive values (PPV, NPV)
- Decision analysis metrics that incorporate costs and consequences of decisions
- Etc
- Keep in mind that statistical significance ($p<0.05$) is not sufficient to tell you about biomarker performance.

- Refer to the previous seminar "Logistic Regression – Application to Clinical Classification" (3/10/2021) for discussion on this topic.

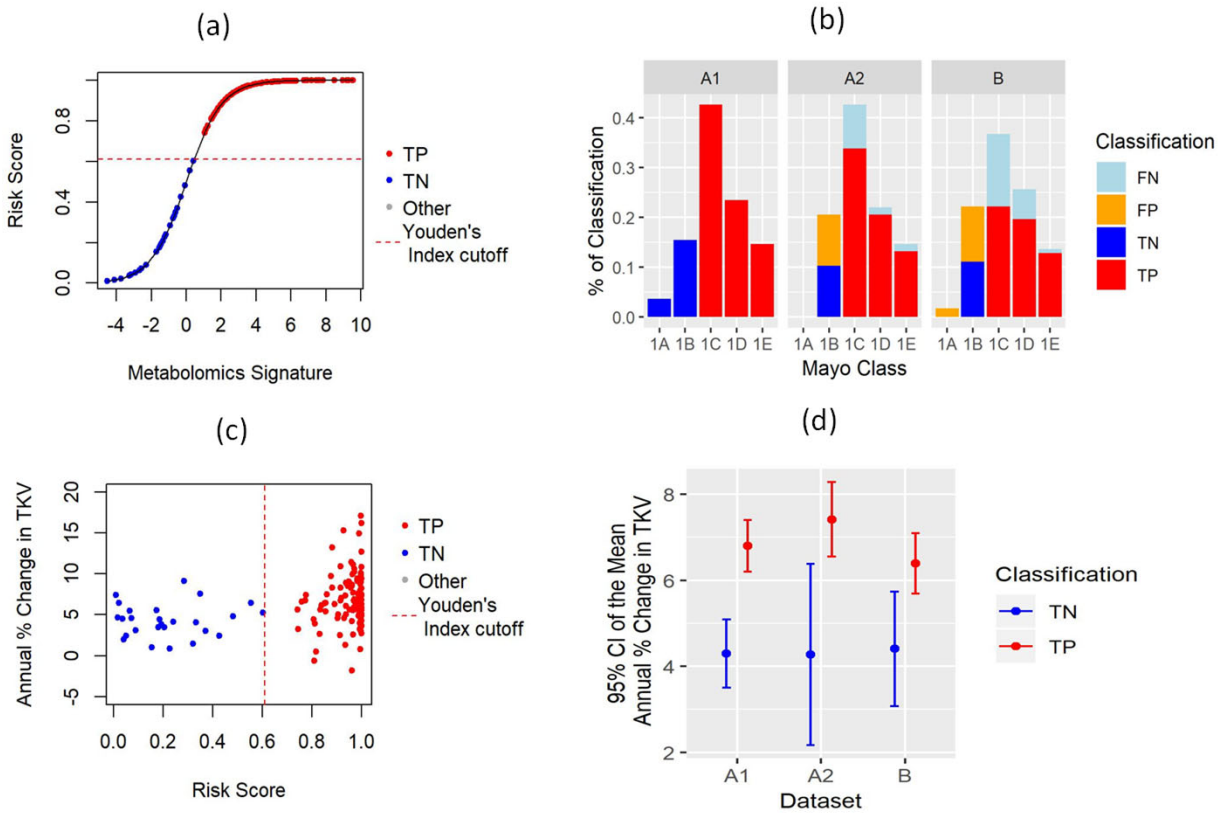
Application. Prognostic Biomarker for ADPKD

- Ongoing collaboration with Drs. Robert Weiss (UCDMC) and Arlene Chapman (U of Chicago) - unpublished
- Autosomal dominant polycystic kidney disease (ADPKD) is the most common hereditary kidney disease and is characterized by increased total kidney volume (TKV) decades prior to loss of kidney function and end stage kidney disease. There is no prognostic biomarker for ADPKD yet.
- Objective: to develop a metabolomic signature as a prognostic biomarker for ADPKD.
- Data from the HALT cohort:
  - Outcome: increased total kidney volume (TKV)
  - Baseline measures: 140 plasma metabolites and Mayo's class (A-E) based on baseline eGFR (estimated glomerular filtration rate)
  - Discovery set: 204 HALT participants with
  - External validation set: 117 HALT participants
- Statistical method: Least Absolute Shrinkage and Selection Operator (LASSO) regression

Flow chart of modeling, statistical analysis, and validation steps undertaken



Measures of accuracy of metabolite-signature biomarker



Design for Biomarker Discovery Study

Phase I Discovery

- Discovery studies probing associations
- Identify “differentiating measures” between groups (e.g., Cancer vs. Control, New Trt vs. Standard)

Phase II Quantification

- Large confirmatory studies with pre-stated hypotheses and precise quantification of the magnitude of the effect.
- Move “possible” to “potential” biomarker through the rigorous validation process.

Phase III Application

- Test and establish the utility of biomarker in routine clinical practice settings.
- Evaluate clinical effort on patients’ outcomes and cost-effectiveness.

Validation of Biomarker

- Analytical validation
  - The assay is reproducible and robust
  - It measures the analytes accurately
  - Samples collected and handled as for the intended use
  - Batch effects – non-biological factors such as inter-machine or inter-operator variability, time of data, and day of week

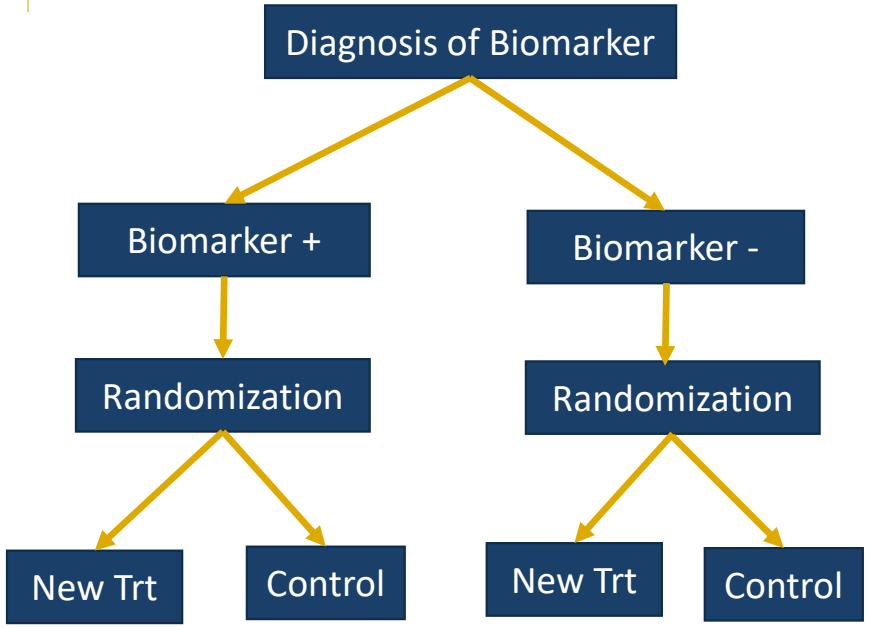
- Clinical validation
  - Correlation with clinical outcome
  - Sensitivity, specificity, PPV, NPV
  - Usually use retrospective samples

- Medical utility
  - The biomarker is actionable and improves outcome
  - It should be translatable in clinical settings.
  - Require prospective study

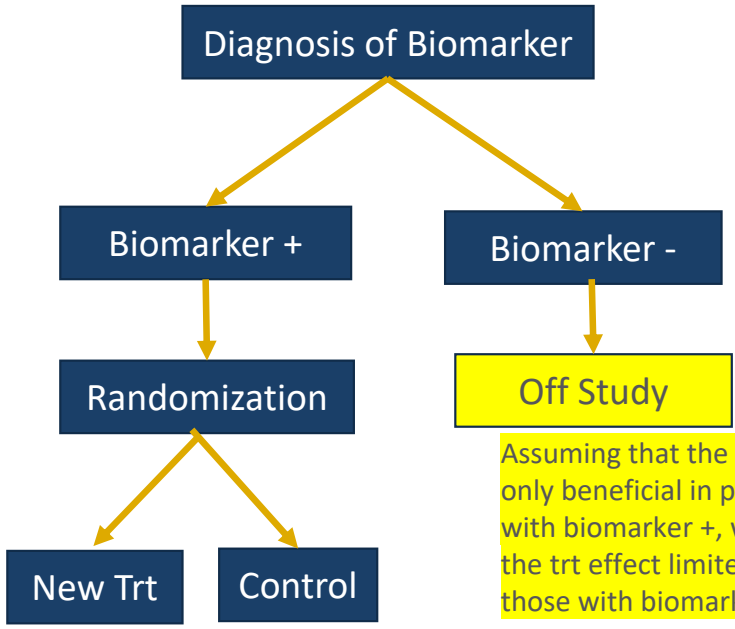
Biomarker-by-Treatment Interaction design

- Apply to test whether a treatment effect depends on biomarker status.
- Size each trial to have power 90% at significance level 0.05.

Design A



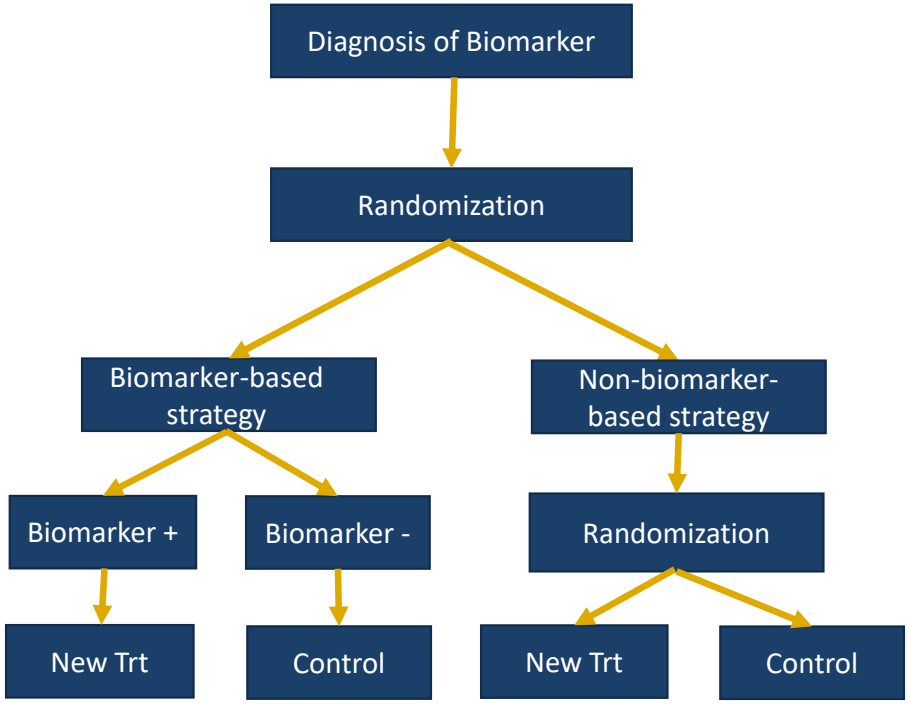
Design B (Enrichment Design)



Assuming that the test trt is only beneficial in patients with biomarker +, we test the trt effect limited to those with biomarker +.

Biomarker-Strategy Design

- Use to test the clinical utility of a biomarker



Questions?

Biostatistics Support is available

- CTSC and Cancer Center Biostatistics Office Hours
  - Every Tuesday from 12-2 p.m. currently via WebEx
  - Sign-up through the CTSC biostatistics website
- MIND IDDRC (full support but available only to IDDRC membership)
- EHSC Biostatistics Drop-in Office hours
  - Every Monday 2-4 p.m. or Upon request
- Request Biostatistics Consultations through the center websites
  - CTSC
  - MIND IDDRC
  - Cancer Center Shared Resources
  - EHSC