Statistical Evaluation of Low Cost Non-inferior Imaging Diagnostic Tests

Joong-Ho (Johann) Won, Xiao Wu, and Ying Lu

Seoul National University, Harvard University, and Stanford University

December 12, 2015

Outline

- Introduction
  - biomarkers
  - designs for clinical validation
  - imaging biomarkers - characteristics
- New design for prognostic values of imaging markers
  - illustration of the idea
  - simple 2x2-state model
  - estimation procedure
- Simulation studies
- Conclusion

Biomarkers

New biomarkers are discovered in an accelerated pace (ESR Executive Council, 03/2010)

Biomarkers have multiple applications (NCCN Task Force Report, 11/2011):

- Prognostic: associate some clinical outcomes independent of the treatment rendered
- Diagnostic: define a classification of the patient sample to establish the particular disease state or disease subtypes
- Predictive: predict the activity of a specific class or type of therapy and are used to help make more specific treatment decisions
- Companion diagnostic: diagnostic, prognostic, or predictive markers useful to identify a subgroup of patients for whom a therapy has shown benefits

Biomarker validation

Before their clinical uses, these biomarkers need to go through rigorous validation

- Analytic validation: how accurately and reliably the marker measures the truth
- Clinical validation: the strength of association between the marker measurement and the clinical outcome of interests
- Clinical utility: ability of the marker measures to improve clinical decision-making and patient outcome
- Report marker validation according to REMARK criteria

Designs for biomarker validation

- Various designs for clinical validation of a **prognostic** marker
  - prospective cohort study: expensive and long study duration for rare outcomes
  - nested case-control study: cost effective and shorter study duration
- **They do not work for imaging markers!**
  - techniques change rapidly and prospective cohort study may take too long to be relevant
  - technique is new and often *no historical samples*
- Worse for low cost non-inferior test – price cap

Imaging biomarkers

Anatomical, functional or molecular parameters detected with imaging

- Characteristic of imaging markers
  - non-invasive
  - may change over time
  - can be measured longitudinally
- Typical study designs for imaging research
  - cross sectional cohort study or case-control study: cost saving, good for diagnostic test, but **not for prognostic prediction** because of bias estimation
  - prospective longitudinal cohort study: good for both diagnosis and prediction but takes too long and cost too much for rare endpoints

Imaging markers: examples

Osteoporosis is one of the major public health challenges in modern society

- Bone mineral density (BMD) measured by dual X-ray absorptiometry (DXA) technique, both diagnostic and prognostic marker for osteoporotic hip fracture
- Speed of sound (SOS) measured by quantitative ultrasonometry (QUS) is an alternative prognostic predictor

Imaging markers: examples (cont'd)

- QUS is safe, simple, free of radiation, portable, and cost-effective, but a late comer: first publication in 1969, 3 in 1980s and only started in 1990s (later than DXA)
- Cost of DXA is already low ($50/scan)

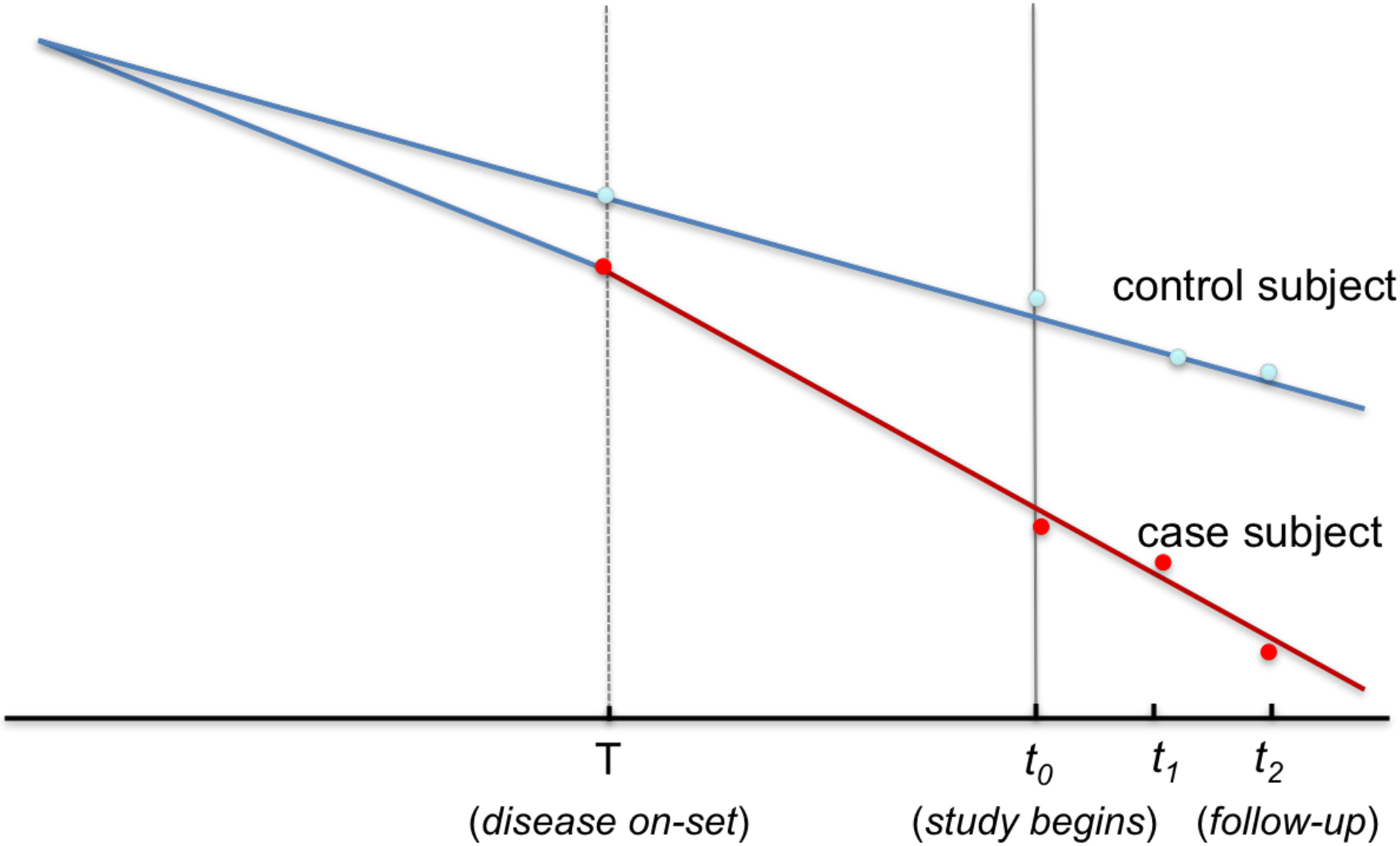
Prognostic prediction

- A population-based European multicenter prospectively designed observational study (OPUS, Gleur et al. 2004) conducted to estimate odds ratio of SOS with incident fracture by following participants – the best but expensive design.
- Low chance to recover the study cost for hip fracture.
- In epidemiology, a case-control study is also used for this purpose – key exposure prior to event.
- A true case-control study is often not possible for imaging techniques, e.g., for SOS, due to lack of measurement (exposure) history

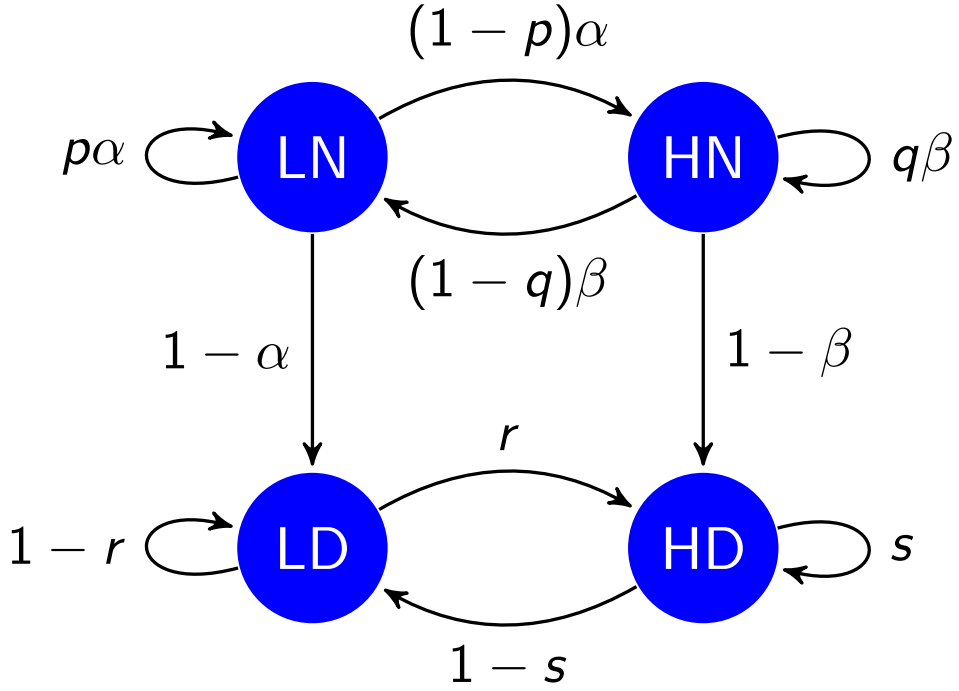
A new design

- Goal: to develop a cost-effective design to establish association of prognostic values of imaging markers
- Ideas:
  - take advantage of case-control design to enrich cases for *none-death* outcomes
  - take advantage of *prior event times* for the case
  - take advantage of repeated measures of imaging markers via *short-term follow-up*
  - use statistical modeling to impute the missing marker measures prior to the outcome

Illustration of the new design



Simplification: 2x2-state Markov chain model



- $X(t) = (Y(t), S(t))$
  - measurement $Y(t) \in \{H, L\}$; disease status $S(t) \in \{N, D\}$
- Assumptions
  - No recovery
  - Continuous transition: no change in measurement when diseased

2x2-state Markov chain model: algebra

- The transition probability matrix of this chain can be written

$$P = \begin{array}{c} \\ \text{LN} \\ \text{HN} \\ \text{LD} \\ \text{HD} \end{array} \begin{array}{c} \begin{array}{cccc} \text{LN} & \text{HN} & \text{LD} & \text{HD} \end{array} \\ \begin{bmatrix} p\alpha & (1-p)\alpha & 1-\alpha & 0 \\ (1-q)\beta & q\beta & 0 & 1-\beta \\ 0 & 0 & r & 1-r \\ 0 & 0 & 1-s & s \end{bmatrix} \end{array}.$$

- Assume that the chain starts from a normal popluation:

$$\pi^{(0)} = \begin{array}{c} \begin{array}{cccc} \text{LN} & \text{HN} & \text{LD} & \text{HD} \end{array} \\ \begin{bmatrix} \pi^{(0)}_{LN} & \pi^{(0)}_{HN} & 0 & 0 \end{bmatrix} \end{array}, \quad \pi^{0}_{LN} + \pi^{(0)}_{HN} = 1.$$

- Model parameters: p, q, r, s, α, β, and $\pi^{(0)}_{LN}$.

Prognosis

- Interested in the 1-year odds ratio of incidence given that the subject is normal at age t:

$$\begin{aligned}
&\frac{\Pr(T = t+1 | X_t = \text{HN}) / \Pr(T > t+1 | X_t = \text{HN})}{\Pr(T = t+1 | X_t = \text{LN}) / \Pr(T > t+1 | X_t = \text{LN})} \\
&= \frac{\Pr(X_{t+1} = \text{HD} | X_t = \text{HN}) / \Pr(X_{t+1} = \text{HN or LN} | X_t = \text{HN})}{\Pr(X_{t+1} = \text{LD} | X_t = \text{LN}) / \Pr(X_{t+1} = \text{LN or HN} | X_t = \text{LN})} \\
&= \frac{(1-\beta)/\beta}{(1-\alpha)/\alpha}.
\end{aligned}$$

- With logistic transforms

$$\begin{aligned}
\alpha &= 1/(1+e^{\gamma_0}), \\
\beta &= 1/(1+e^{\gamma_1}).
\end{aligned}$$

the log odds ratio is $\gamma_1 - \gamma_0$.

Estimation

- Write P as

$$P = \begin{bmatrix} \bar{\Lambda} P_N & \Lambda \\ & P_D \end{bmatrix},$$

where

$$P_N = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}, \quad P_D = \begin{bmatrix} r & 1-r \\ 1-s & s \end{bmatrix},$$

and

$$\Lambda = \text{diag}(1-\alpha, 1-\beta), \quad \bar{\Lambda} = \text{diag}(\alpha, \beta).$$

- P_N is the conditional transition submatrix *before* the disease on-set given that the subject remains normal after the transition, i.e., conditioned on $T > t+1$
- P_D is the conditional transition submatrix *after* the disease on-set, or $T \leq t$.

Conditional transition submatrices

- The conditional transition submatrix P_D after the disease on-set can be easily estimated, because the MLE of r and s is

$$\hat{r} = \frac{\#\{\text{LD} \to \text{LD}\}}{\#\{\text{LD} \to \text{LD}\} + \#\{\text{LD} \to \text{HD}\}}$$
$$\hat{s} = \frac{\#\{\text{HD} \to \text{HD}\}}{\#\{\text{HD} \to \text{HD}\} + \#\{\text{HD} \to \text{LD}\}}$$

  and these transtions are well observed by the **short-term follow-up** of the case group.

Conditional transition submatrices (cont'd)

- Short-term follow-up also allows us to estimate the conditional transition matrix P_N before the disease on-set efficiently:

$$\hat{p} = \frac{\#\{\text{LN} \to \text{LN}\}}{\#\{\text{LN} \to \text{LN}\} + \#\{\text{LN} \to \text{HN}\}}$$
$$\hat{q} = \frac{\#\{\text{HN} \to \text{HN}\}}{\#\{\text{HN} \to \text{HN}\} + \#\{\text{HN} \to \text{LN}\}},$$

Estimation of odds ratio

- Normal-to-disease transition is very rare, i.e., $\alpha, \beta \approx 1$
- Estimation of the odds ratio based on the observed number of transtions is unstable
  - it is not unlikely that no transition occurs during the short-term follow-up.
- Assuming that P_N and P_D are accurately estimated from the short-term follow-up, we can proceed with the estimation of α and β using the structure of the case-control design.

Data structure

		control (N)	case (D)			
		$\tau > t_0$	$\tau = 1$	$\tau = 2$	$\cdots$	$\tau = t_0$
risk	L	n_1	m_{11}	m_{12}	$\cdots$	m_{1t_0}
	H	n_2	m_{21}	m_{22}	$\cdots$	m_{2t_0}
	sum	n	m			

- $\tau = 1, \ldots, t_0$ are the observed disease on-set times for the case subjects
- $n = \#\{\text{ctrl}\}$, $m = \#\{\text{case}\}$ are fixed

Control group

- Probability of the observation in the control group (N by time t_0) is binomial:

$$\binom{n}{n_1} \Pr(X_{t_0} = \text{LN} | T > t_0)^{n_1} (1 - \Pr(X_{t_0} = \text{LN} | T > t_0))^{n_2}$$

where the observations (counts) are made at t_0, and

$$\begin{aligned} \Pr(X_{t_0} = \text{LN} | T > t_0) &= \Pr(X_{t_0} = \text{LN} | X_{t_0} = \text{LN or HN}) \\ &= \frac{\pi_{LN}^{(t_0)}}{\pi_{LN}^{(t_0)} + \pi_{HN}^{t_0}} = \frac{\pi_N^{(0)} (\bar{\Lambda} P_N)^{t_0} e_1}{\pi_N^{(0)} (\bar{\Lambda} P_N)^{t_0} \mathbf{1}} \end{aligned}$$

Control group (cont'd)

- Thus the log-likelihood of the control group is

$$l_{\text{ctrl}}(\alpha, \beta, \pi_{LN}^{(0)}) = n_1 \log(\pi_N^{(0)}(\bar{\Lambda} P_N)^{t_0} e_1) + n_2 \log(\pi_N^{(0)}(\bar{\Lambda} P_N)^{t_0} e_2) - n \log(\pi_N^{(0)}(\bar{\Lambda} P_N)^{t_0} \mathbf{1}) + \text{const.}$$

Case group

- Sampling distribution of the case group is multinomial with $(p_{\text{LD}}^{(1)}, p_{\text{HD}}^{(1)}, \ldots, p_{\text{LD}}^{(t_0)}, p_{\text{HD}}^{(t_0)})$, where

$$p_{\text{LD}}^{(\tau)} = \frac{\Pr(X_{t_0} = \text{LD}, T = \tau)}{\Pr(T \leq t_0)}, \quad \tau = 1, 2, \ldots, t_0.$$

- Numerator:

$$\begin{aligned}\Pr(X_{t_0} = \text{LD}, T = \tau) &= \Pr(X_{\tau-1} = \text{LN}, X_\tau = \text{LD}, X_{t_0} = \text{LD}) \\ &\quad + \Pr(X_{\tau-1} = \text{HN}, X_\tau = \text{HD}, X_{t_0} = \text{LD}) \\ &= \pi_N^{(\tau-1)} \Lambda P_D^{t_0-\tau} e_1 = \pi_N^{(0)} (\bar{\Lambda} P_N)^{\tau-1} \Lambda P_D^{t_0-\tau} e_1\end{aligned}$$

Case group (cont'd)

- Denominator:

$$\begin{aligned}\Pr(T \le t_0) &= \Pr(X_{t_0} = \text{LD or HD}) = \pi_{LD}^{(t_0)} + \pi_{HD}^{(t_0)} = \pi_D^{(t_0)}\mathbf{1} \\ &= \pi_N^{(0)} \left(\sum_{k=0}^{t_0-1} (\bar{\Lambda} P_N)^k \right) \Lambda \mathbf{1},\end{aligned}$$

since

$$\begin{aligned}\pi_D^{(t)} &= \pi_{(0)} \Lambda P_D^{t-1} + \pi_N^{(1)} \Lambda P_D^{t-2} + \cdots + \pi_N^{(t-1)} \Lambda \\ &= \pi_N^{(0)} \Lambda P_D^{t-1} + \pi_N^{(0)} (\bar{\Lambda} P_N) \Lambda P_D^{t-2} + \cdots + \pi_N^{(0)} (\bar{\Lambda} P_N)^{t-1} \Lambda\end{aligned}$$

and $P_D^k \mathbf{1} = \mathbf{1}$

Case group (cont'd)

- Thus the log likelihood of the case group is

$$l_{\text{case}}(\alpha, \beta, \pi_{LN}^{(0)}) = \sum_{t=1}^{t_0} \Big(m_{1t} \log(\pi_N^{(0)} (\bar{\Lambda} P_N)^{t-1} \Lambda P_D^{t_0 - t} e_1) + m_{2t} \log(\pi_N^{(0)} (\bar{\Lambda} P_N)^{t-1} \Lambda P_D^{t_0 - t} e_2) \Big) - m \log \left(\pi_N^{(0)} \left(\sum_{k=0}^{t_0 - 1} (\bar{\Lambda} P_N)^k \right) \Lambda \mathbf{1} \right) + \text{const}$$

- Estimation maximizes $l_{\text{ctrl}}(\alpha, \beta, \pi_{LN}^{(0)}) + l_{\text{case}}(\alpha, \beta, \pi_{LN}^{(0)})$
- Remember the goal is to estimate OR $((1-\beta)\alpha)/((1-\alpha)\beta)$

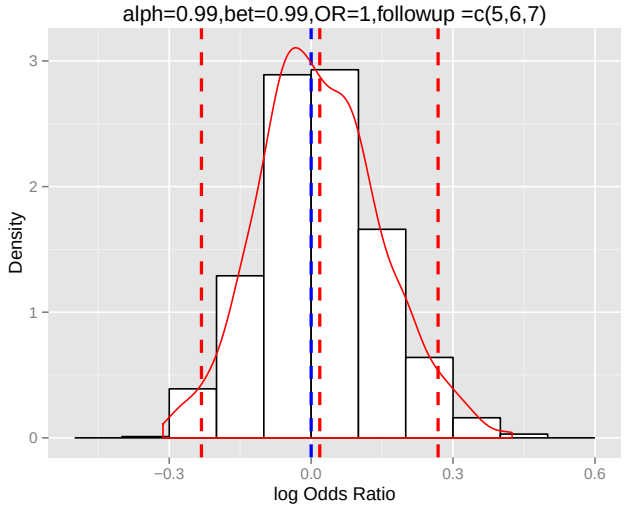
Simulation I: null case

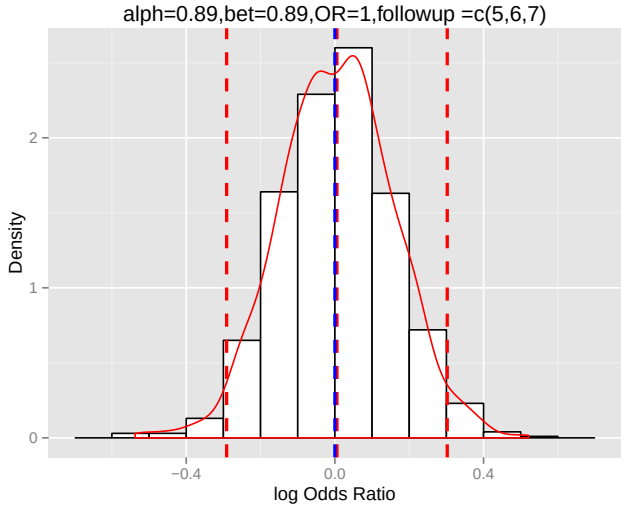
- Generated sample datasets with 500 cases and 500 controls, 500 times
- Study time point is fixed (t_0=5)
  - e.g., 65-yr old female fractured b/w ages 60–65
- Short-term follow-up visits are fixed to 2 times (t_1=6, t_2=7)
- Conditional transition:

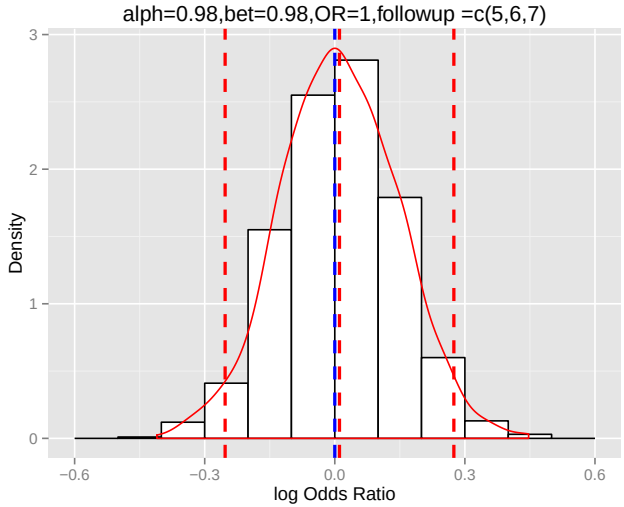
$$P_N = \begin{array}{c} \\ \text{LN} \\ \text{HN} \end{array} \begin{array}{c} \begin{array}{cc} \text{LN} & \text{HN} \end{array} \\ \begin{bmatrix} 0.95 & 0.05 \\ 0.15 & 0.85 \end{bmatrix} \end{array} \quad P_D = \begin{array}{c} \\ \text{LD} \\ \text{HD} \end{array} \begin{array}{c} \begin{array}{cc} \text{LD} & \text{HD} \end{array} \\ \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix} \end{array}$$

- Comparision between disease with different incidences but same OR (=1):
  - $(1-\alpha, 1-\beta)$=(0.001,0.001)
  - $(1-\alpha, 1-\beta)$=(0.020,0.020)
  - $(1-\alpha, 1-\beta)$=(0.110,0.110)

Simulation I (cont'd)





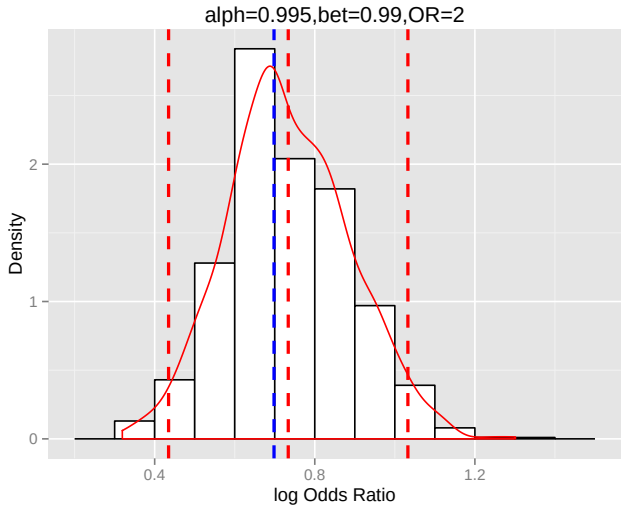


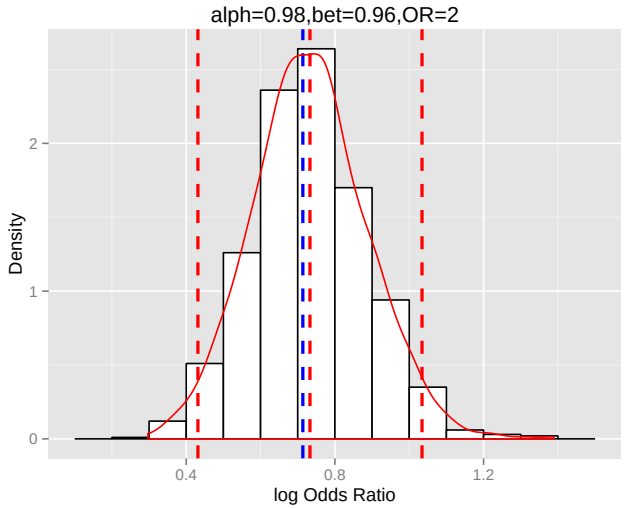
Simulation II: non-null case

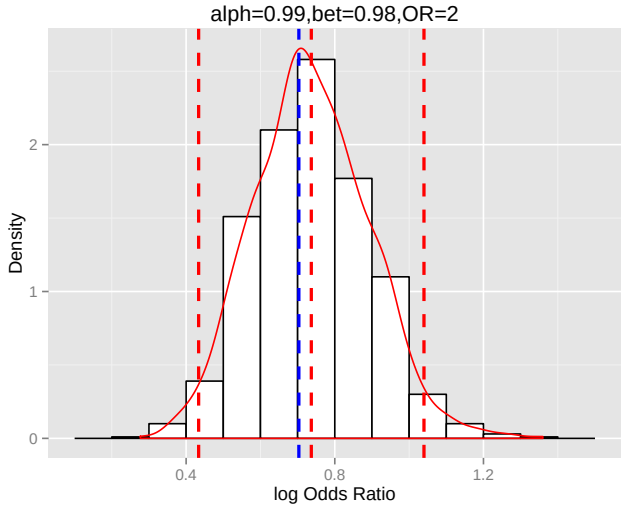
Same as Simulation I, except for

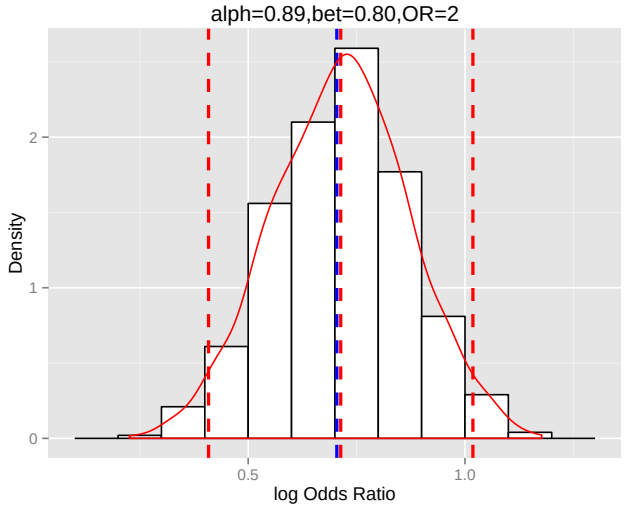
- comparision between disease with different incidences but about the same odds ratio (≈ 2):
  - $(1-\alpha, 1-\beta)$=(0.005,0.010)
  - $(1-\alpha, 1-\beta)$=(0.010,0.020)
  - $(1-\alpha, 1-\beta)$=(0.020,0.040)
  - $(1-\alpha, 1-\beta)$=(0.110,0.200)

Simulation II (cont'd)







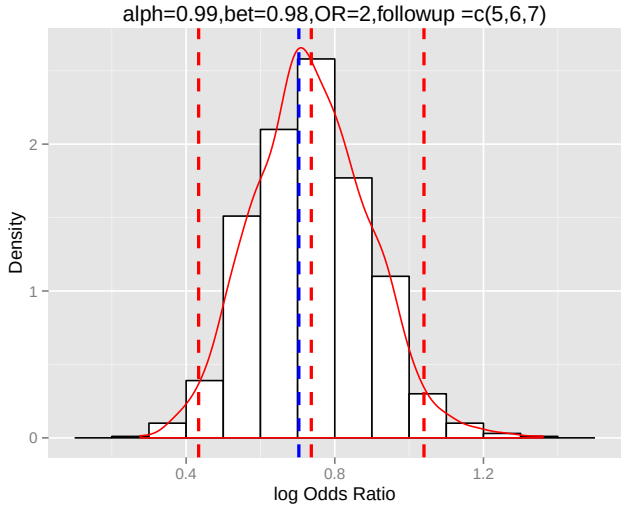


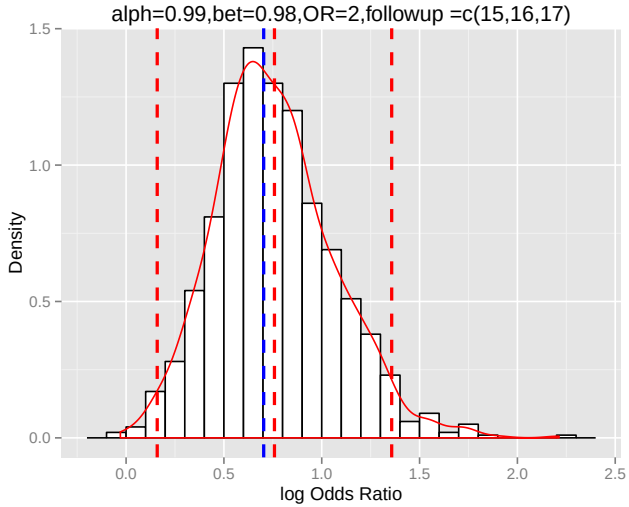
Simulation III: non-null with varying study times

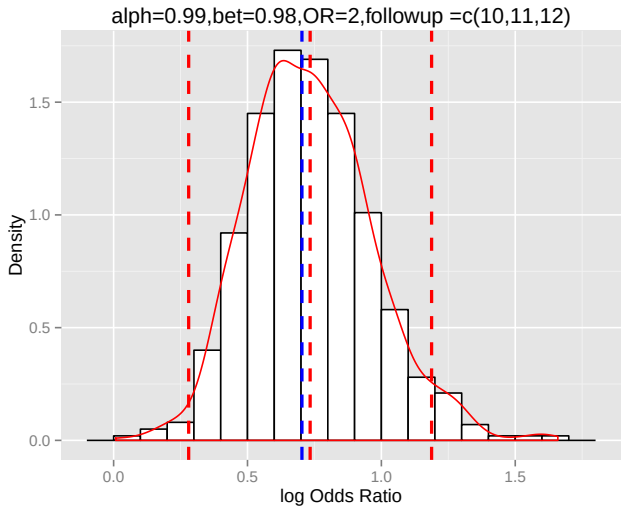
Same as Simulation I, except for

- incidences are fixed ($1-\alpha$=0.01, $1-\beta$=0.02; OR ≈ 2), and
- comparision between *different study time points* (t_0=5, 10, 15)

Simulation III (cont'd)





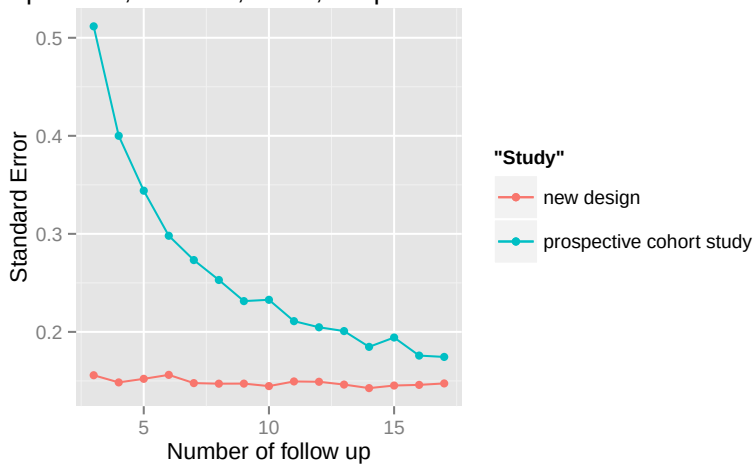


Simulation IV: benchmark 1

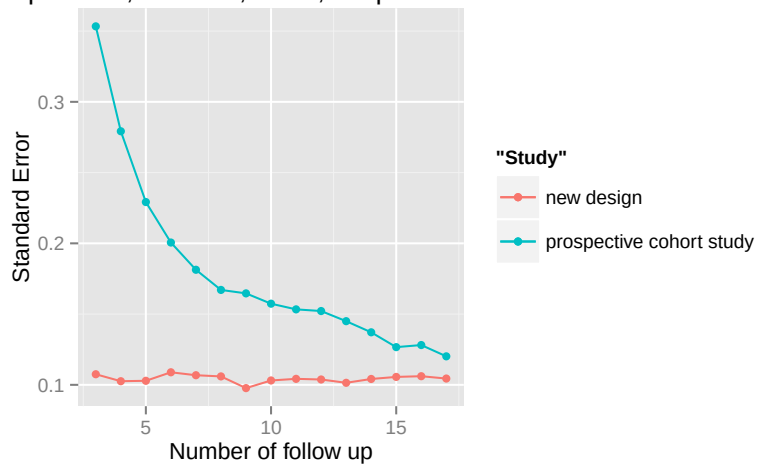
- Effectiveness of the new design compared to the prospective cohort study
- Similar setting to Simulation I:
  - Study time point is fixed (t_0=5)
  - Same conditional transition matrices as before
  - Incidences with OR $\approx$ 2: $(1-\alpha, 1-\beta)$=(0.005,0.010) and $(1-\alpha, 1-\beta)$=(0.010,0.020)
- Sample sizes of 500 cases and 500 controls; 1000 cases and 1000 controls
- Comparision of estimation accuracy of OR with varying number of visits for the follow-up

Simulation IV (cont'd)

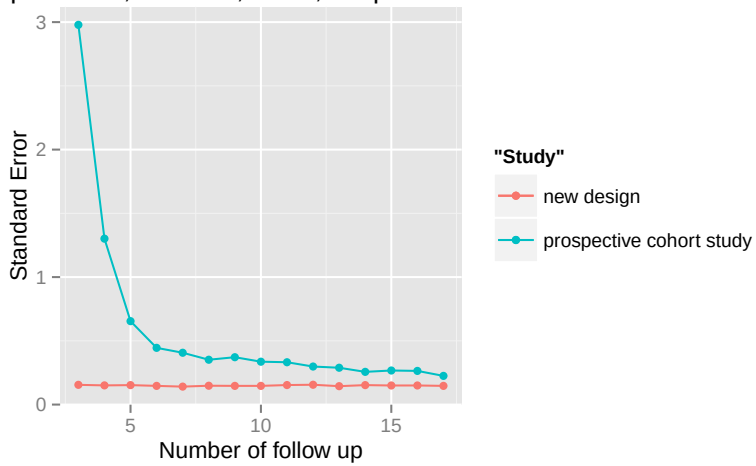
alph=0.99,bet=0.98,OR=2,samplesize=1000



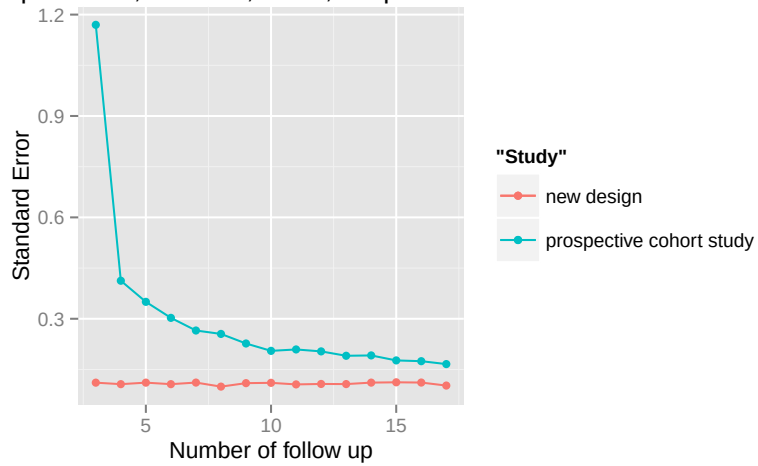
alph=0.99,bet=0.98,OR=2,samplesize=2000



alph=0.995,bet=0.99,OR=2,samplesize=1000



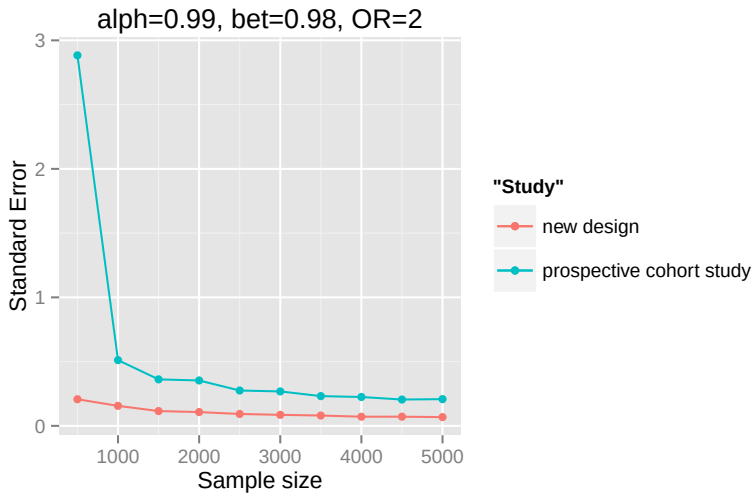
alph=0.995,bet=0.99,OR=2,samplesize=2000

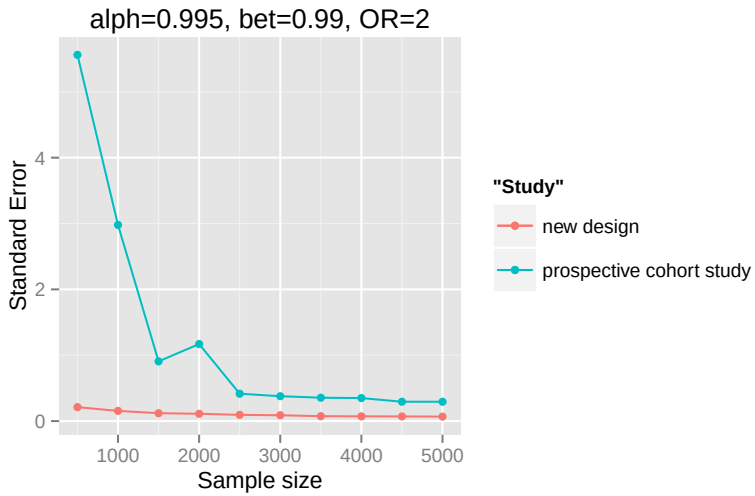


Simulation V: benchmark 2

- Effectiveness of the new design compared to the prospective cohort study
- Similar setting to Simulation I:
  - Study time point is fixed (t_0=5)
  - Same conditional transition matrices as before
  - Incidences with OR $\approx$ 2: $(1-\alpha, 1-\beta)$=(0.005,0.010) and $(1-\alpha, 1-\beta)$=(0.010,0.020)
- Total number of visits fixed to 3 (t_1=6, t_2=7)
- Comparision of estimation accuracy of OR with varying sample size

Simulation V (cont'd)





Conclusion

- New design gives an unbiased estimation for odds ratio, which is similar to the case-control design.
- In most common case follow-up with three visits is sufficient to obtain a reasonable estimate.
- Distribution of the estimate appears normal (esp. when study time is not far from the disease on-set); needs formal analysis.
- New design is superior to prospective cohort study in a relative shorter study length, especially for rare diseases.
- Under the same sample size and study length, the new design seems always superior to prospective cohort study in power.
- New design is recommended for rare diseases when the historical exposure to biomarkers are not available.

Limitations

- Assumption of homogeneous transitions – need to model aging, etc
- Discrete-time model – need to extend to continuous-time setting
- Markov model – need to incorporate long-term memory