DISCUSSION
Randomization- *vs* Model-based Inference

KyungMann Kim
University of Wisconsin-Madison
kmkim@biostat.wisc.edu

Outline

- Model- *vs* Randomization-based Inference
- A lady tasting tea
- Simple *vs* Stratified Randomization
- Covariate adjustment using auxiliary covariates
- An example: A skin cancer prevention trial
- Conclusion

Statistical Properties of Randomization in Clinical Trials

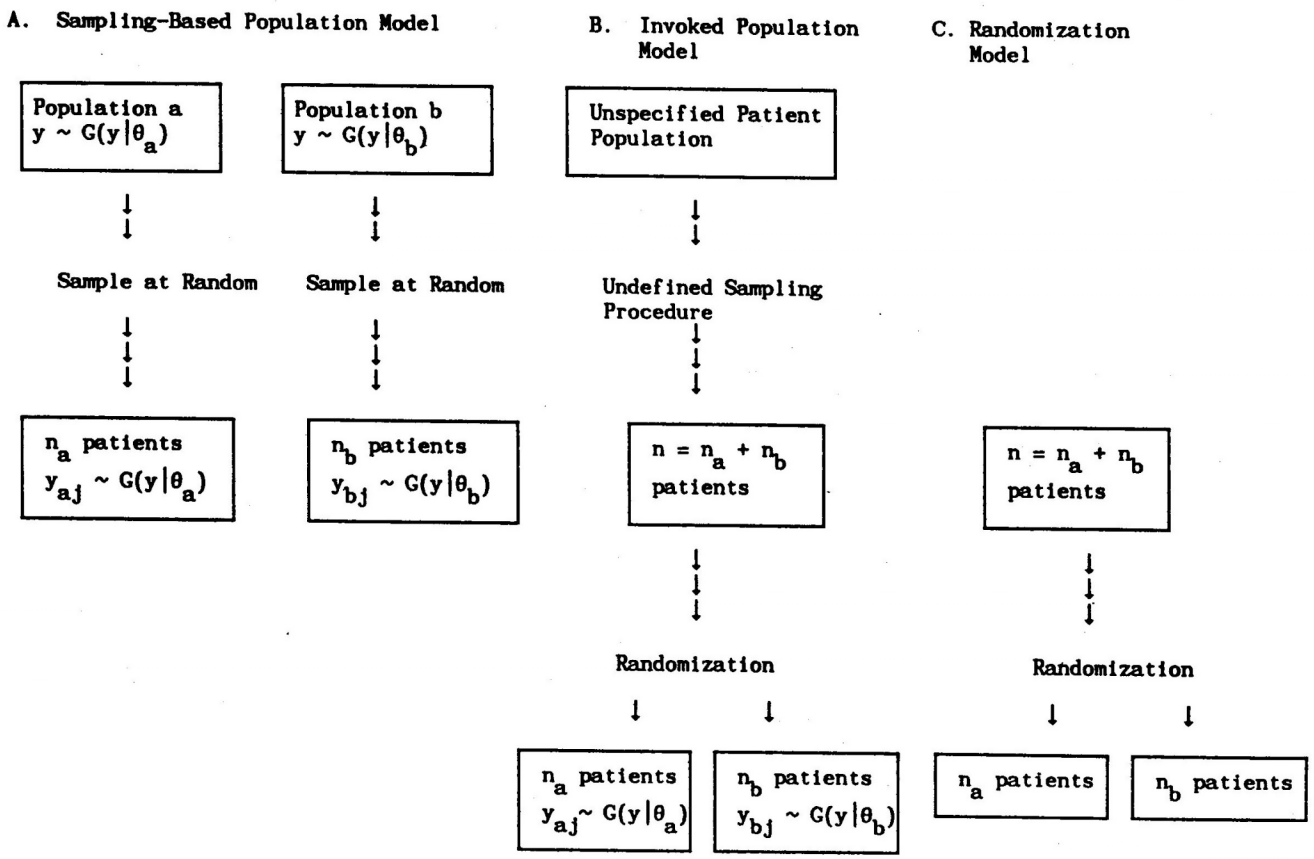


Figure 2 Population sampling models, the patient selection process, and the randomization model for a clinical trial.

Lachin: *Control Clin Trials* 1988;9:289-311.

Model-based Inference

- Not affected by treatment allocation scheme
  - Begg, Kalish: *Biometrics* 1984
- "Essentially, all models are wrong, but some are useful."
  - George E. P. Box
- Assumption: No unmeasured covariates

Randomization-based Inference

- The use of randomization provides a basis for an assumption-free statistical test of the equality of treatments
- Such tests were originally proposed by Fisher and are known as randomization or permutation tests
- With sample sizes n_a and n_b, the conditional reference set Ω_c consists of ${}_nC_{n_a}$ possible combinations of n_a patients on treatment *a* out of *n* patients

A Lady Tasting Tea (1)

- Fisher: *The Design of Experiments* 1935.
- A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup
- Test H_0: No faculty of discrimination
- The experiment
  - Mix 8 cups of tea, 4 in one way and 4 in the other, prepared otherwise the same
  - Present them to the lady for judgement in a random order
- 70 ways of choosing a group of 4 objects out of 8

A Lady Tasting Tea (2)

- A test of significance
  - With 4 correct, p=1/70=0.014 (<0.05)
  - With 3 correct and 1 wrong, p=(1+16)/70=0.243
- How would you conclude?
- The randomization formed the physical basis of the validity of the test
- The sensitivity of the experiment depends on the number of cups prepared

Simple *vs* Stratified Randomization (1)

- Basic Benefits of Randomization
  - Eliminate assignment bias
  - Tend to produce comparable groups
  - Produce valid statistical tests
- Grizzle: *Control Clin Trials* 1982;3:365-8.
  - The Lumpers
    - Simple randomization
    - Adjusted analysis by ANCOVA
  - The Splitters
    - Stratified randomization
    - Stratified analysis

Simple *vs* Stratified Randomization (2)

- Is the efficiency reduced by chance imbalance in simple randomization?
  - Yes for small sample trials
  - Covariate adjustment may improve the efficiency
    - How to do it properly?
- Is the administrative burden worth the added complexity in stratified randomization?
  - More importantly, stratification itself can lead to loss in efficiency

Covariate Adjustment Using Auxiliary Covariates (1)

- *Y, Z, X*: Outcome, treatment, auxiliary covariates
- Unbiased estimating function $m(Y,Z;\theta)$
- With Augmentation

$$m^*(Y,X,Z;\theta) = m(Y,Z;\theta) - (Z-\pi)\, a(X)$$

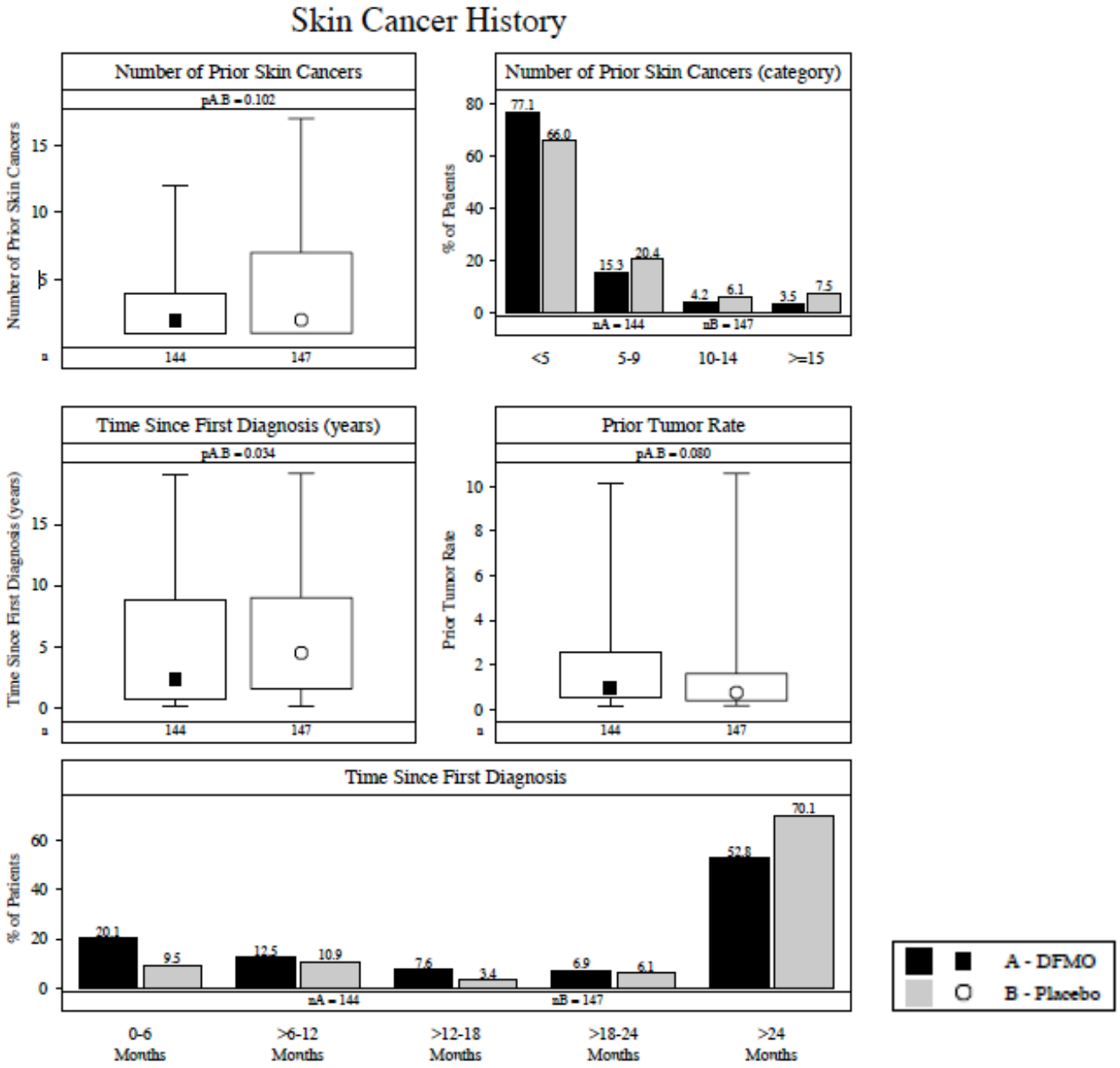
  where $\pi = E(Z)$
- Elegant semiparametric theory
- Key assumptions
  - $Z \perp X$, independent
    - Covariate-adaptive randomization does not work!
  - No unmeasured covariates?

Covariate Adjustment Using Auxiliary Covariates (2)

- PURSUIT trial as reported in Zhang *et al.* Biometrics 2008
- 5,710 pts and 35 baseline auxiliary covariates
- Relative efficiency of $1.06=(0.073/0.071)^2$
  - Unadjusted estimate of the log odds ratio
    - β_2=-0.174 (0.073)
  - Adjusted estimate
    - β_2=-0.163 (0.071)
- How about the bias?
  - Unadjusted estimate of the log odds ratio
    - Z_{unadj}=-2.38 and p_{unadj}=0.0171
  - Adjusted estimate
    - Z_{adj}=-2.30 and p_{adj}=0.0217>p_{unadj}!

A Skin Cancer Prevention Trial UWCCC CO9737

- Patients with prior history of nonmelanomatous skin cancer
- Simple randomization after a run-in period
- DFMO *vs* placebo daily for an average of four years
- 90% power to detect a 35% reduction in new skin cancer from 0.65 new cancer/yr at a two-sided 0.05 test
- 291 pts randomized
- Unadjusted *vs* adjusted analysis


Skin Cancer History
Number of Prior Skin Cancers
pA.B = 0.102
Number of Prior Skin Cancers
144
147
Number of Prior Skin Cancers (category)
% of Patients
77.1
66.0
15.3
20.4
4.2
6.1
3.5
7.5
nA = 144
nB = 147
<5
5-9
10-14
>=15
Time Since First Diagnosis (years)
pA.B = 0.034
Time Since First Diagnosis (years)
144
147
Prior Tumor Rate
pA.B = 0.080
Prior Tumor Rate
144
147
Time Since First Diagnosis
% of Patients
20.1
9.5
12.5
10.9
7.6
3.4
6.9
6.1
52.8
70.1
nA = 144
nB = 147
0-6 Months
>6-12 Months
>12-18 Months
>18-24 Months
>24 Months
A - DFMO
B - Placebo

Poisson Regression Models for Adjusted Analysis (1)

Model P1:

log(mean)=intercept + log(years on study) + β[DFMO indicator]

Model P2:

log(mean)=intercept + log(years on study) + β[DFMO indicator] + γ[log(# of prior cancers)]

Model P3:

log(mean)=intercept + log(years on study) + β[DFMO indicator] + γ[categorized # of prior cancers]

Model P4:

log(mean)=intercept + log(years on study) + β[DFMO indicator] + γ[log(prior tumor rate)]

Model P5:

log(mean)=intercept + log(years on study) + β[DFMO indicator] + γ[log(prior tumor rate)] + δ[age]

Poisson Regression Models for Adjusted Analysis (2)

Table 7: Poisson Regression Results

Poisson Model	Statistic	Parameter β	γ	δ	Overdispersion Parameter
P1	Value	-0.325			4.58
	SE	0.174			
	Exp	0.722			
	P-value	0.062			
P2	Value	-0.092	0.806		2.76
	SE	0.137	0.067		
	Exp	0.912	2.238		
	P-value	0.501	<0.001		
P3	Value	-0.178	0.776		3.30
	SE	0.149	0.087		
	Exp	0.837	2.173		
	P-value	0.231	<0.001		
P4	Value	-0.353	0.178		4.12
	SE	0.165	0.066		
	Exp	0.703	1.195		
	P-value	0.033	0.007		
P5	Value	-0.373	0.201	0.019	4.00
	SE	0.163	0.066	0.008	
	Exp	0.689	1.223	1.019	
	P-value	0.022	0.002	0.016	

Dilemma

- Different models lead to different results
- Specific models oftentimes not decided upon at the design stage
  - If you know the true model, you wouldn't be doing experiment in the first place
  - In this case we don't know which way of summarizing the prior cancer burden is appropriate
  - Leading to dilemma as to what to report
- Potential for data dredging is real!!!

Conclusion

- Some imbalance is a fact of life, especially for small studies
- Covariate adjustment
  - Damned if you do and damned if you don't?
  - Potential for misuse
  - Present both unadjusted and adjusted analyses
- Covariate adjustment using auxiliary covariates
  - Elegant theoretically
  - Very promising practically
  - Avoid problems of data dredging by decoupling evaluation of treatment from regression modeling