

Statistical and operational efficiency of stratified clinical trials

Neal Thomas

Pfizer Inc,
Groton, CT 2018

Stratification

- Stratification
 - ▶ Small number (< 10) of pre-defined exhaustive non-overlapping subgroups (strata). Strata can be analyzed separately and then combined across the strata
 - ▶ Stratification in design means that a separate independent randomization list is utilized for each strata
- Goals of stratification
 - ▶ Improve estimate of an 'overall' treatment effect
 - ▶ Improve assessment of potential treatment differences (interactions) for subjects in different strata

Objectives

- Review theory and introduce new software to assess statistical and operational efficiency of stratified designs and analyses
- Distinguish between stratification in design and stratification in analysis
- Compute the financial costs of different designs/analyses
- Focus on continuous endpoints. Binary and survival endpoints have similar issues but involve additional complexity

Section 1

Stratified estimators

Notation

- Y_k is the outcome of subject $k = 1 \dots N$
- m_s is the number of strata, m_t is the number of treatments
- $s_{ik}, i = 1, \dots, m_s$, are the indicator variables ($k = 1 \dots N$) for strata membership
- $t_{jk}, j = 1, \dots, m_t$ are the indicator variables for treatment group membership
- $n_{ij} = \sum_k s_{ik} t_{jk}$ are the number of subjects in strata i receiving treatment j (count matrix is strata by treatment)
- The \bar{Y}_{ij} are the sample response means in stratum i for trt j

Stratified estimators

- Overlapping baseline characteristics are not modelled, only non-overlapping groups are represented
 - ▶ All combinations of classifications are included, e.g. male/age < 60, male/age \geq 60, female/age < 60, female/age \geq 60
 - ▶ This approach is recommended for confirmatory analyses. It does not assume classifications are additively related to outcome. The inclusion of all combinations typically uses only a few extra degrees of freedom.

Model-based estimator

$$Y_k = \sum_{j=1}^{m_t} \mu_j t_{jk} + \sum_{i=2}^{m_s} \beta_i s_{ik} + \epsilon_k$$

- Additive model. The μ_j are the treatment group means in the first strata. The β_j are the differences in strata means.
- Least squares estimation. Maximum likelihood when the ϵ_k are homogeneous and normally distributed.
- General form of the additive model for covariate adjustment:

$$Y_k = \sum_{j=1}^{m_t} \mu_j t_{jk} + \sum_{i=1}^m \beta_i x_{ik} + \epsilon_k$$

where the x_{ik} are m baseline regressor variables that can be represented by a $N \times m$ matrix $[X_1 | \dots | X_m]$.

Model-based estimator (cont)

- The general form for the least squares estimator of $\mu_j - \mu_{j'}$ is

$$\hat{\mu}_j - \hat{\mu}_{j'} = (\bar{Y}_j - \bar{Y}_{j'}) - \hat{\beta}' (\bar{X}_j - \bar{X}_{j'})$$

where \bar{Y}_j are the treatment group response means, and \bar{X}_j are the vectors of treatment-specific covariate means, and the $\hat{\beta}$ is the vector the least squares estimators of the β

- The covariate-adjusted and unadjusted means are equal when the covariate means are equal
- The standard error of the mean difference is minimized when the treatment-specific covariate means are equal (assuming homogeneous residuals)
- Equal treatment-specific covariate means translates into the same proportion of subjects in a stratum across all treatment groups. This condition is called 'balance'

Efficient Weighted estimator

- Consider only two treatments simultaneously, even when there are more than two treatments
- Compute the mean treatment difference in each strata and compute a weighted average:

$$\tilde{\mu}_j - \tilde{\mu}_{j'} = W^{-1} \sum_{i=1}^{m_s} w_i (\bar{Y}_{ij} - \bar{Y}_{ij'})$$

$$w_i = (1/n_{ij} + 1/n_{ij'})^{-1}, \quad W = \sum_{i=1}^{m_s} w_i$$

- The model-based estimator and weighted difference are equal when there are only two treatments, or the treatment counts in each strata are 'balanced'. Both estimators equal the simple unweighted difference in sample means when there is balance.

Interaction

- Primary analysis assumes the same mean difference between each pair of treatment groups across all strata
- The least squares/ML estimator of an 'interaction' is very simple. It is the difference in the differences of the sample means under the fully saturated interaction model (every treatment/strata mean combination has its own parameter)

Section 2

Randomization designs

Randomization designs

- Much of statistical theory assumes a 'simple restricted' or 'random allocation' of subjects (Lachin [1988]), e.g., randomly select 1/2 of the subjects and assign them to the treatment group.
- In clinical trials, unlike agriculture field trials, the subjects are collected sequentially and they are not known at the beginning of the trial. Exact implementation of 'simple' randomization can not be reliably achieved because we seldom exactly achieved the pre-specified sample size.
- Another simple randomization plan is the 'completely randomized (CR)' design. Each patient is assigned to treatment by an independent, identically distributed process (e.g. coin toss). Sample sizes are targetted in expectation only. This plan is used as a common theoretical benchmark, but is rare in practice.

Randomization designs (cont)

- To improve precision compared to completely randomized design, from 1970-1990, statisticians researched several methods to implement randomization in clinical trials (e.g., Matts and McHugh [1978], Wei and Lachin [1988], Pocock and Simon [1975])
- By the time the E9 guidance was completed (1998), sequences of randomly permuted blocks had been accepted as the default technique

Randomization in the E9 Guidance

"Although unrestricted randomisation is an acceptable approach, some advantages can generally be gained by randomising subjects in blocks. This helps to increase the comparability of the treatment groups, particularly when subject characteristics may change over time, as a result, for example, of changes in recruitment policy. It also provides a better guarantee that the treatment groups will be of nearly equal size."

- Completely randomized designs are an acceptable approach
- Note the acknowledgement that the clinical outcomes and treatment assignments depend on calendar time.

E9 on stratified randomization designs

"It is advisable to have a separate random scheme for each centre, i.e. to stratify by centre or to allocate several whole blocks to each centre. More generally, stratification by important prognostic factors measured at baseline (e.g. severity of disease, age, sex, etc.) may sometimes be valuable in order to promote balanced allocation within strata; this has greater potential benefit in small trials."

Implications for analysis

- Mantra: "Analyze as randomized"
 - ▶ Use the randomization distribution. Regard all data in the study as fixed. Probability calculations are based on the randomness in the treatment assignments, i.e., the randomization distribution
 - ▶ From likelihood theory, the analysis should condition on any baseline data used when treatment was assigned. Propensity scores incorporate this advice for both observational and experimental data (Rosenbaum and Rubin [1983])
- E9 guidance
 - ▶ There is no reference in the guidance to randomization distributions, permutation based inference, etc, so it is not likely they interpreted the mantra in this manner
 - ▶ "Factors on which randomisation has been stratified should be accounted for later in the analysis"

Implications for analysis (cont)

- Ignoring blocking in the analyses
 - ▶ Blocked designs are always nested within calendar time
 - ▶ Randomization blocks are correctly viewed as numerous small strata with restricted simple randomization within the 'strata'
 - ▶ Entry of subjects into clinical trials is often clustered at clinical centers, so block assignments are dependent on center even when blocking within center was not implemented
 - ▶ "Analyze as randomized" is rarely if ever applied as this would require terms (conditioning) at a minimum on block, calendar time, and center (even for designs not nesting blocks within centers)

Guidance from the statistics literature

- Ignoring blocks is conservative provided subjects recruited close in time and/or from the same center are not more heterogeneous than randomly selected subjects (Matts and Lachin [1988], Gansky and Koch [1994], Student [1937]). Mathematical counter-examples are possible

Advice from E9

- E9 guidance:
 - ▶ "In other trials it may be recognised from the start that the limited numbers of subjects per centre will make it impracticable to include the centre effects in the statistical model. In these cases it is not appropriate to include a term for centre in the model, and it is not necessary to stratify the randomisation by centre in this situation."
- Change in practice since E9 guidance
 - ▶ Designs with a large numbers of international centres are now common. Including centre effects is impracticable. Nesting randomization blocks within centre, however, can be operationally efficient.

Section 3

Statistical and operational efficiency

Statistical efficiency

- Comparing two estimators: Conditional Relative Efficiency (RE)
 - ▶ Example: compare the least squares and weighted estimators of a treatment difference

$$RE = \text{var}(\tilde{\mu}_j - \tilde{\mu}_{j'}) / \text{var}(\hat{\mu}_j - \hat{\mu}_{j'})$$

- ▶ The usual standard errors (squared) are computed 'conditional' on the treatment/strata sample sizes
- ▶ Variances are compared because their ratio is the proportional change in total sample size required to produce the same standard errors from the two estimators
- ▶ The smaller variance is usually reported in the denominator. The ratio is then the increase in total sample size for the less efficient estimator to yield the same *SE*.

Statistical efficiency (cont)

- Compare a stratified and unstratified design with both analyses using a stratified estimator.
 - ▶ Fix a common total sample size for each design
 - ▶ A ratio of variances is still compared, but the variances must be 'unconditional' to include differences that can arise due to different combinations of treatment/strata sizes yielding the same total sample size
 - ▶ The stratified estimator is conditionally unbiased, so the design RE is the ratio of the average of the conditional variances across the possible sample size configurations arising from the designs

Conditions that determine the statistical efficiency of stratified designs

- A stratified analysis will be pre-specified and performed. The benefit of corresponding stratification in the design increases when
 - ▶ There is a small total sample size (typically < 100)
 - ▶ Number of treatments/strata increase
 - ▶ There are unequal treatment allocations (e.g., 2:1) and/or unequal strata frequencies
- Improvement due to design stratification is typically small compared to stratification in the analysis
- Software to easily compute design efficiency will be described

The impact of randomization designs on operational efficiency

- Central designs— One randomization list for the study, or one list for each stratum in the study
- Center-based designs—a separate randomization list for each clinical center
 - ▶ Numerous randomization blocks can be generated in advance and pre-allocated to each center
 - ▶ The unblinded personnel that package and distribute experimental drugs to the clinical centers then know the order of treatment assignments for the center allowing them to ship drug more efficiently
 - ▶ Supply utilization varies between studies. For center-based designs, typically 25% of the shipped drug is not used. For central designs, the typical overage is 50%
 - ▶ The cost of goods and shipping costs vary between studies. These costs are often high in studies with active comparators.
 - ▶ When experimental drug supply is tight, trials may need to be delayed if a central randomization design is used.

Operational efficiency (cont)

- Combining center-based design with stratification in the design is not feasible
 - ▶ The number of partially filled blocks will be very high so stratification will not be achieved
 - ▶ The supply personnel will not know the strata classifications of future subjects, so drug supply predictability is lost
- Central designs are required for open-label studies and studies with high potential for functional unblinding
- The use of predictive algorithms for shipping drug in longer-term trials requiring re-supply of drug to subjects is common practice

Financial efficiency

- Stratified designs can reduce the number of subjects and save the clinical costs to treat and measure subjects
- Center-based unstratified designs can reduce the costs to make/purchase/distribute drug
- To evaluate the competing interests, we need 1) the cost to add a patient to the trial, and 2) the drug supply costs for center-based and central randomization designs.
- Estimates of these costs exist at the time of detailed protocol development. Two recent examples will illustrate the process.

R package stratRCT for evaluating stratified designs

- *stratRCT is an R package that supports planning of stratified clinical trials*
- *The primary function is RECalculator*
 - ▶ *Statistical efficiency and related quantities computed with a combination of theoretical calculations and simulation*
 - ▶ *Several supporting functions: print, plot, vcov*
- *Function strataCosts that computes costs using output of RECalculator along with clinical and drug costs*

Example: protocol with an active comparator

- 700 patient randomized trial
- Large number of international clinical centers
- 1:2:2:2 with placebo, 2 doses, active comparator
- Primary endpoint is a responder variable
- Strata is baseline value for the primary endpoint: moderate/severe (2 levels). Prevalence and placebo response estimates based on medical literature

R code

```
library ( stratRCT )  
set.seed(12357)  
sampn<-700  
### rand ratio 1:2:2:2  
tratio<-c(1,2,2,2)  
# proportion in strata (moderate, severe)  
stratap<-c(0.60,0.40)  
trtcomp<-c(1,2)  
presp<-c(0.1,0.12)  
  
REobj<-RECalculator( treatRatio=tratio , sampSize=sampn ,  
                    strataP=stratap ,  
                    binaryParms = list (pResp=presp) ,  
                    trtComp=trtcomp , strataComp=1:2 ,  
                    nsim=100000)  
  
print ( REobj )
```

Stratification efficiency I

Relative efficiencies for the stratified estimator of $\text{trt } 2 \text{ less trt } 1$
 $RE > 1$ favor the stratified design

Each design versus the stratified permuted block design:

	CR Design	Perm Block	No Strata
Ave RE	1.008		1.001
Prop Zero-count designs	0.000		0.000

Note: The proportion of stratified designs with a zero count (completers) treatment/strata combination is 0

Relative efficiency of the interaction estimator for strata levels 1 and 2
 $RE > 1$ favor the stratified design

Each design versus the stratified permuted block design:

	CR Design	Perm Block	No Strata
Ave RE	1.015		1.008

Stratification efficiency II

**Relative efficiency of stratification in design/analysis
versus no stratification of either**

RE>1 favor stratification

Efficiency for risk difference estimators for the binary outcome:

	Stratified Analysis	Stratified Design/Analysis
Ave RE	0.999	1.001

Summary of statistical efficiency

- There is exceedingly little gain in statistical efficiency using stratification in the design. There is also very little improvement from stratification in the analysis in this setting.
- The distribution of patient counts across clinical centers was not specified. Results for the center-based block design are bounded by the CR design, and the unstratified permuted block design.
- Following standard analysis practice, calendar time and between center variation is ignored in the variance calculations.
- The gain in efficiency from the stratified design is slightly larger for estimates of interactions. This is common.

Operational/Financial efficiency

- The cost for labs and investigator time per patient is \$50,000
- The difference in cost for drug (raw materials, labeling, distribution) between center-based and central randomization is \$600,000.

```
strataCosts (REobj, iCost=50000, dCostDif=600000)
```

```
stratified central cost – unstratified center based cost  
306442.4
```

- The cost difference returned by *strataCosts* is the central stratified design minus the center-based unstratified design
- The study is over-powered. The sample size was determined by exposure needed to meet regulatory requirements for safety data. There will be no increase in sample size to account of the unstratified permuted block design nested within clinical centers, so the realized savings is \$600,000

Example: phase 2 protocol

- 98 patient randomized trial
- Approximately (< 30 US centers
- 1:2:2:2 with placebo, 3 treatments
- Primary endpoint is continuous
- Strata are baseline measurement of primary endpoint (low/high), and a comorbidity (no/yes) (4 combinations/levels)
- Strata prevalence and response within strata estimated from an internal study with similar design

R code

```
library(stratRCT)
set.seed(12357)
propS<-c(0.340,0.170,0.362,0.128)
withinSD<-0.306
r2strat<-0.695  ## reduction in residual variances
cparms<-list(withinSD=withinSD,r2=r2strat)

trtrat<-c(1,2,2,2)
n<-98

rel<-RECalculator(treatRatio = trtrat,
                  sampSize = n,strataP=propS,contParms = cparms,
                  nsim=100000)

print(rel)
```

Stratification efficiency I

Relative efficiencies for the stratified estimator of $\text{trt } 2 \text{ less trt } 1$
 $RE > 1$ favor the stratified design

Each design versus the stratified permuted block design:

	CR Design	Perm Block	No Strata
Ave RE	1.086		1.027
Prop Zero-count designs	0.363		0.312

Note: The proportion of stratified designs with a zero count (completers) treatment/strata combination is 0.05003

Relative efficiency of the interaction estimator for strata levels 1 and 2
 $RE > 1$ favor the stratified design

Each design versus the stratified permuted block design:

	CR Design	Perm Block	No Strata
Ave RE	1.137		1.111

Stratification efficiency II

**Relative efficiency of stratification in design/analysis
versus no stratification of either**

RE>1 favor stratification

	Stratified Analysis	Stratified Design/Analysis
Ave RE	3.16	3.246

Operational/Financial efficiency

- Cost per patient \$55,000
- Center vs central drug savings, \$10,000 (minimal)

strataCosts (re1 , iCost=55000, dCostDif=10000)

stratified central cost – unstratified center based cost
–452866.7

- There is a large savings due to stratification in the analysis:
 $3.2 * 98 * 55500 = \$17,404,800$ (statistical efficiency of 3.2 obtained from RECalculator)

Conclusions

- Kernan et al. [1999]
 - ▶ "In practice, investigators rarely take account of stratification in calculating sample size. Rather, they regard stratification as providing a margin of security in sample size estimates. The trade off, of course, is security that the trial will be adequately sized for cost savings that could be realized with smaller samples."
- The potential cost savings from seemingly small changes in our designs and analyses are very large. Consideration of costs would lead to much more careful statistical evaluation of study designs.
- Financial cost evaluations also highlight the importance of statistical excellence that is often under-appreciated

Section 4

Supplemental slides

Reductions in sample size using stratification

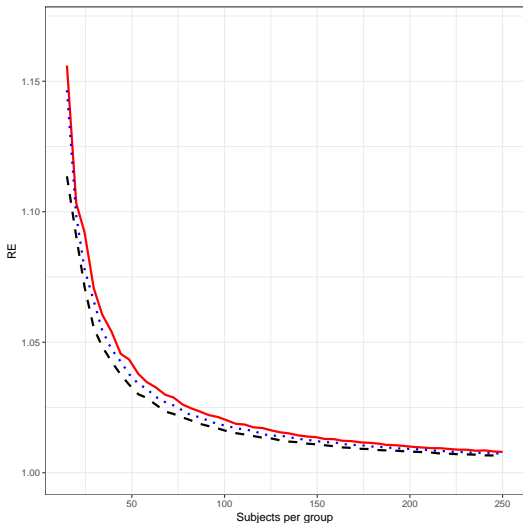
- Kernan et al. [1999]
 - ▶ "In practice, investigators rarely take account of stratification in calculating sample size. Rather, they regard stratification as providing a margin of security in sample size estimates. The trade off, of course, is security that the trial will be adequately sized for cost savings that could be realized with smaller samples."
- Large range in potential cost savings
 - ▶ First example there is no gain in statistical efficiency or cost savings from stratification in design or analysis
 - ▶ Second example has very large savings:
 $3.2 * 98 * 55500 = \$17,404,800$ (statistical efficiency of 3.2 obtained from RECalculator)
 - ▶ Excludes center startup costs

Reductions in sample size using stratification (cont)

- Quantify cost of conservative analyses
 - ▶ *RECalculator*, *stratCosts* make cost/efficiency calculations easy
 - ▶ Main effort is obtaining cost data and within-strata variance estimates
 - ▶ Decisions may result in compromise between strata-based sample size and sample size ignoring strata
- Contrast reductions in sample size due to stratification with the use of baseline in repeated-measures studies
 - ▶ It is routine to account for baseline when computing sample size for repeated-measures study. This is often arbitrarily 'conservative' based the calculations on an 'unadjusted' change from baseline variable.
 - ▶ In the second example, one stratifying factor was baseline. Sample size was computed using the variance for change-from-baseline ignoring strata.
- The potential costs of these sloppy 'conservative' practices is surprisingly large

Impact of unplanned missing data on stratification

- Primary impact of unplanned missing data is the potential for bias. Here we only consider increase in standard errors
- Even with imputation, unplanned missing data creates imbalance in information collected. Optimal designs are thus not achieved.
- Assess the impact of unplanned missingness assuming MCAR and complete analysis
- The small efficiency gains are further reduced by unplanned missing data. 20% missing data appreciably attenuates the small efficiency gains (roughly $1/3$).

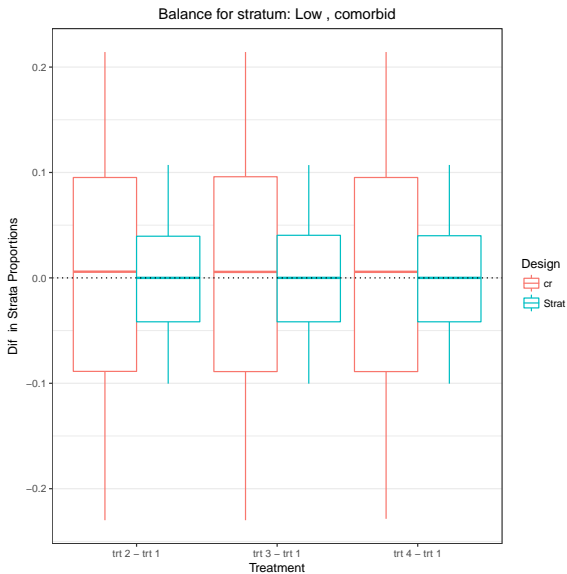


- 2 treatments, 4 strata. Solid-red 0 missing, black-dot 10% missing, blue-dash 10% missing. RE for Stratified vs CR design

More on design efficiency

- The improvement from design stratification decreases when there is unplanned missing data and mis-classified strata
- Improvement due to design stratification is typically small compared to stratification in the analysis
- Efficiency in design is NOT determined by how predictive (R^2) the strata are (Grizzle [1982]). This is surprising because R^2 is very important for assessing the efficiency of a stratified versus unstratified estimator

```
plot(re1,strata=2,dif=1, strataName = 'Low baseline, comorbidity')
```



Bibliography I

- S. Gansky and G. Koch. Evaluation of alternative strategies for managing stratification factors in the analyses of randomized clinical trials. *American Statistical Association Proceedings of the Biopharmaceutical Section*, pages 405–410, 1994.
- J. Grizzle. A note on stratifying versus complete random assignment in clinical trials. *Controlled Clinical Trials*, 3:365–368, 1982.
- W. Kernan, C. Viscoli, R. Makuch, L. Brass, and R. Horwitz. Stratified randomization for clinical trials. *Journal of Clinical Epidemiology*, 52: 19–26, 1999.
- J. Lachin. Properties of simple randomization in clinical trials. *Controlled Clinical Trials*, 9:312–326, 1988.
- J. Matts and J. Lachin. Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials*, 9:327–344, 1988.

Bibliography II

- J. Matts and R. McHugh. Analysis of accrual randomized clinical trials with balanced groups in strata. *Journal of Chronic Disease*, 31: 725–740, 1978.
- S. Pocock and R. Simon. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 31:103–105, 1975.
- P. Rosenbaum and D. Rubin. The central role of the propensity score in observational studies for causal effects. 70:41–55, 1983.
- Student. Comparison between balanced and random arrangements of field plots. *Biometrika*, 29:363–379, 1937.
- L. Wei and J. Lachin. Properties of urn randomization in clinical trials. *Controlled Clinical Trials*, 9:345–364, 1988.