STRATIFICATION OF POPULATION IN CLINICAL TRIALS

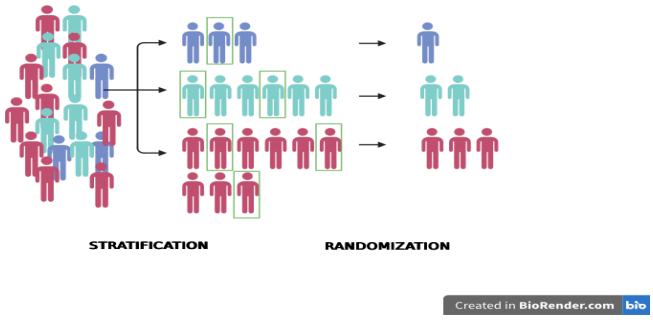
1. INTRODUCING THE CONCEPT OF RANDOMIZATION AND STRATIFICATION

When we are performing a clinical trial, we must have a population with a determinate condition over which we are conducting the experiments. Therefore, we must select a sample of this population to represent accurately the population of the study, in this case we will be talking about random sampling.

The term *randomization* in the context of clinical trials, refers to the assignment of treatments to patients using a chance procedure to ensure unbiased and fair allocation of participants in each group of treatment. Therefore, in a randomized study each participant has an equal chance of being assigned to any of the treatments being compared.

Considering all factors known to greatly influence the trial's results, to prevent biases in the analyzed data that might result in incorrect conclusions and could potentially mask any genuine differences between the treatments due to excessive data variability. When the factors influencing response are known, we can consider it in the initial randomization of the population.

Stratification refers to the process of dividing a population into distinct subgroups or strata based on specific characteristics or variables. Each stratum represents a subset of the population that shares similar or homogeneous characteristics regarding the chosen variables. This method aids in minimizing biases and ensuring that the groups under study are more comparable, enhancing the reliability of comparisons made between them.



2. TYPES AND METHODS OF STRATIFICATION

2.1 Different stratification variables

The factors used to stratify a population can be very diverse and, depending on the type of research being conducted, can be very different from one another.
To illustrate some factors used in stratification, a couple of variables are identified in the following table.

Example: Factors influencing handwriting

Table 8.3 Factors used for stratification in our sampling process

Reference in H&H	Stratification variable	Strata definition	Strata proportion in USA (confirmed except writing system) (%)
A	Writing systems	Location of third grade schooling in USA	80.0
		Location of third grade schooling NOT in USA	20.0
B	Gender	Male	49.0
		Female	51.0
C	Age	18–30	33.0
		> 30–50	36.0
		> 50	41.0
D	Handedness	R	90.0
		L	10.0
C	Temporal state	Night (after 8pm)	
		Day (before 8 pm)	
D	Education	HS or less	49.0
		> HS	51.0
N/A	Race	W	63.7
		B	12.6
		H	16.3
		A	4.8
N/A	US region (where samples are taken)	North West	
		North East	
		Middle West	
		South West	
		South East	

2.2 Methodology in stratification

To better explain the methodology to be followed to perform a stratification of the population, a simplified model of a clinical trial with two treatment groups, the active and the control, will be followed.

Four parameters must be defined at the start of the treatment allocation.

a) Deciding how to measure differences between treatment groups based on a patient's condition before they receive treatment.

b) Figuring out how to measure differences across all patient conditions when dividing them into active or control groups.

c) Deciding how much importance to give to different patient conditions when calculating overall differences.

d) Deciding the chance (probability) that a patient is placed in the treatment group to create the smallest overall difference between groups based on patient conditions.

The level of imbalance between the treatment groups for a particular level of a prognostic factor may be calculated by the standard deviation or variance of the number in each treatment group who occupy that level of the prognostic factor. So, if there are two treatment groups, the standard deviation of these numbers is equivalent to the magnitude of the difference between the 2 numbers.

Example of measurement of imbalance in stratification

Table I. Measurement of imbalance in adaptive stratification.

New patient allocated to	Stratification factor level	Number of patients in active treatment group	Number of patients in control group	Imbalance
Active treatment	age 61–70	10	8	2
	delay <6 hours	19	16	3
	ischaemic stroke	30	31	1
Control group	age 61–70	9	9	0
	delay <6 hours	18	17	1
	ischaemic stroke	29	32	3

Total imbalance if new patient allocated to active treatment is $2+3+1=6$.
Total imbalance if new patient allocated to control group is $0+1+3=4$.
Stroke type might be considered a more important prognostic factor than age or delay from stroke onset. If stroke type is given a weighting of 3 while the other factors are given a weighting of 1 in the calculation of imbalance:
Total imbalance if new patient allocated to active treatment is $2\times\mathbf{1}+3\times\mathbf{1}+1\times\mathbf{3}=8$
Total imbalance if new patient allocated to control group is $0\times\mathbf{1}+1\times\mathbf{1}+3\times\mathbf{3}=10$

In this situation, let's say there are already 80 patients in a clinical trial. A new patient, aged 65, with an ischemic stroke and admitted within 6 hours of the stroke onset, is about to join the trial. When looking at the overall differences between treatment groups, it seems the active treatment group has more imbalance than the control group. So, it might be better for the new patient to join the control group.

However, considering that the type of stroke could be more important than age or timing, we might decide to give stroke type a higher value in our calculation. If we give stroke type a weight of three, then the overall imbalance would show more for the control group, making the active treatment group a better choice. So, the decision about which treatment group is better can change based on how much importance we give to each factor.

3. IMPORTANCE OF STRATIFICATION

3.1 Assurance that compared groups are similar with respect to known prognostic factors.

The defining purpose of stratified randomization is to provide greater assurance that compared groups are similar with respect to known prognostic features other than treatment.
For trials with small sample population (<100 patients) stratification is critical to perform a valid comparison. Otherwise, when sample size increases the risk of the outcomes being compromised by bias diminishes.

3.2 Protection against type I error.

Type I error occurs when there appears to be a difference in outcome rates between two treatment groups, suggesting one treatment is more effective, even though both treatments have equal effectiveness (a false positive).

Therefore, a falsely positive trial can occur if the randomization process places the patients with a better prognosis in the active treatment group and those whose prognosis is worse in the comparison group.

Example of type I error

	Scenario without stratification		**Scenario with stratification**			
Participants	200		200			
Stratification			< 40 years		> 40 years	
Randomization	100 (treatment A)	100 (treatment B)	50 (treatment A)	50 (treatment B)	50 (treatment A)	50 (treatment B)
Overall results	p-value= 0,04 (treatment A over treatment B)		< 40 years p-value = 0,25 (no significance) > 40 years p-value = 0,03 (treatment A over treatment B)			

STRATIFICATION APPLICATION: In this case stratification of the sample population, it can be observed that the outcomes change in the stratified group, therefore having more accurate results.

3.3 Protection against type II error

Type II error, describes the chance of not identifying a genuine difference between two treatment groups, resulting in a false-negative outcome.

The statistical power refers to the probability that a study will correctly detect a true effect or difference when it exists, such as the effectiveness of a treatment or the presence of a relationship between variables.
Therefore, a highly powered study has larger probability of detecting a specific treatment effect at *any level of statistical significance.*

Power goes down when there's more variation between two averages or rates being compared. Stratification helps cut down that variation, which should, theoretically, boost the study's ability to detect effects.

Example of type II error

Imagine a drug that has a known success rate of 75% in treating a condition, and we want to try the effectiveness of a new drug.

- Null Hypothesis (Ho): There is no difference between the new drug and the existing drug in treating the condition.
- Alternative Hypothesis (H1): The new drug is better than the existing drug in treating the condition.

- We assume that; the existing drug success rate is 75% and that the desired power is 80%.
- The researchers want to detect a difference in success rates of at least 10% between new and existing drug.
- Desires power is 80%, meaning a type II error of 20%.

Now, assuming that the findings include a success rate of 80% (a 5% absolute improvement over the existing drug).
If we have a sample size calculated to detect a 10% difference between treatments, the study might not have enough statistical power to detect this smaller effect size (5% increase) due to sample size insufficient.

STRATIFICATION APPLICATION: If we apply stratification to the population, we'll be able to detect differences in analyzing subgroups separately, potentially decreasing the likehood of Type II error by revealing undetected differences that could be in overall analysis.

3.4 Decreasing number of recruited patients

The number of patients that are required to detect a difference in two treatments at a prespecified power and level of statistical significance. Therefore, for a specific treatment difference and significance level, a more highly powered trial will require more patients and a larger trial will have more power.

If we are assuming this direct relationship, when stratification improves power, it also reduces the required sample size.

3.5 Facilitation of subgroup analysis

The analysis and interpretation of data within subgroups have sparked considerable debate. Subgroup analyses often yield misleading results.

- They might miss significant treatment effects due to insufficient statistical power (false negatives).
- They can wrongly identify treatment effects that don't exist (type I errors, leading to false positives). The possibility of type I error is high in subgroup analyses because authors often simultaneously examine treatment effects within many subgroups.

To take into consideration the limitations of power, investigators may report the power of the subgroup analysis (given the obtained sample size) to detect the observed treatment effect.
STRATIFICATION APLICATION: Stratified randomization forces investigators to identify subgroups before the start of a study. Stratification helps to assure that treatment assignments within subgroups are balanced. Assuming that patients in each subgroup are similar among them in all regards except the treatment. Therefore, each subgroup becomes a small trial.

4. LIMITATIONS OF STRATIFICATION

4.1 Interaction with a small stratum

If we are measuring the effect of treatment differing by race, for example in EEUU there is a 12% of the population that is African American, so if we recruited randomly from the target population only 12 out of 100 would be African American, but to measure effect of the treatment in this stratum the sample will be too small and the data few reliable.
Therefore, is not a solution that the trial population is representative of the target population because in small strata won't be guaranteed a reliable information about the effect of the treatment. An overrepresented minority stratum would be required to known about the effects in this groups.

4.2 Overstratification

In permuted block randomization scheme if there are many strata, the patients will be distributed among this stratum and since there will be few patients in each stratum, some strata will have no blocks completed. Imbalances for prognostic variables can occur between treatment groups as a result of this incomplete filling.
Consequently, reducing the number of strata avoids the problem of multiple comparisons that may occur if the outcome results are shown separately by strata.

5. MAIN CONCLUSIONS

- Assigning treatments randomly minimizes bias in clinical trial participant allocation.
- Dividing populations by characteristics enhances group similarities for accurate comparisons.
- Guards against type I and type II errors by minimizing variability between treatment groups.
- Interaction with small strata or overstratification may compromise trial reliability.
- Balancing subgroup analyses' benefits against potential complexities and limitations is crucial.

6. REFERENCES

1. Stratified Randomization for Clinical Trials Walter N. Kernan, 2 LM, editores. Brass,2,3,5 and Ralph I. Horwitz1,2 PII.
2. Zelen M. The randomization and stratification of patients to clinical trials. J Chronic Dis [Internet]. 1974;27(7–8):365–75. Disponible en: http://dx.doi.org/10.1016/0021-9681(74)90015-0
3. Weir CJ, Lees KR. Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. Stat Med [Internet]. 2003;22(5):705–26. Disponible en: http://dx.doi.org/10.1002/sim.1366
4. Permutt T. A note on stratification in clinical trials. Drug Inf J [Internet]. 2007;41(6):719–22. Disponible en: http://dx.doi.org/10.1177/009286150704100604