Design and Analysis of Crossover Trials

Chin-Shang Li, Ph.D.
Department of Public Health Sciences
Division of Biostatistics
University of California, Davis
Davis, CA

Outline

- Introduction to design of crossover trials
- Statistical analysis of crossover trials data
- Power analysis for crossover trials

Introduction to design of crossover trials

- A crossover design is a repeated/longitudinal measurements design.
  - Patients (experimental units) cross over from one treatment to another during the trial course.
  - In contrast to a parallel design where patients are randomized to a treatment and remain on that treatment throughout the trial duration
  - Two advantages over both a parallel study and a non-

crossover longitudinal study

♠ Advantage 1: The influence of confounding covariates is reduced because each patient serves as her/his own control.

- In a non-crossover study, even randomized, different treatment groups are often found to be unbalanced on some covariates.
- In a controlled, randomized crossover design, such imbalances are implausible unless covariates are changed systematically during the study.

♠ Advantage 2: Optimal crossover designs are sta-

tistically efficient and require fewer subjects than do non-crossover designs, even other repeated measures designs.

- Popular in medicine, agriculture, manufacturing, education, and many other disciplines
- Not preferred routinely because of the problems inherent with this design
  - In medical clinical trials the disease should be chronic and stable, and the treatments should not result in total cures but only alleviate the disease condition.
    - If treatment A cures the patient during the first

period, treatment B will not have the opportunity to demonstrate its effectiveness when the patient crosses over to treatment B in the second period.

- This type of design works only for those conditions that are chronic, e.g., asthma where there is no cure, and the treatments attempt to improve quality of life.

Introduction to design of crossover trials, cont'd

- A sequence: the order of treatment administration in a crossover experiment
  - ♣ Sequences should be determined a priori, and the experimental units are randomized to sequences.
- A period: the time of a treatment administration
- 2×2 crossover design: 2-sequence, 2-period, 2-treatment crossover design, with sequences AB and BA
  - ♣ Most popular crossover design

- Experimental units, randomized to sequence AB, receive treatments A and B in the first and second period, respectively; experimental units, randomized to sequence BA, receive treatments B and A in the first and second period, respectively.

Table 1: 2×2 crossover design (AB|BA)

	Period 1	Period 2
Sequence AB	A	B
Sequence BA	B	A

- Examples of some other crossover designs

Table 2: 3-period, 2-treatment crossover designs (ABB|BAA)

	Period 1	Period 2	Period 3
Sequence ABB	A	B	B
Sequence BAA	B	A	A

Table 3: 3-period, 2-treatment crossover designs (AAB|ABA|BAA)

	Period 1	Period 2	Period 3
Sequence AAB	A	A	B
Sequence ABA	A	B	A
Sequence BAA	B	A	A

Table 4: 3-period, 3-treatment crossover designs (ABC|BCA|CAB)

	Period 1	Period 2	Period 3
Sequence ABC	A	B	C
Sequence BCA	B	C	A
Sequence CAB	C	A	B

Table 5: 3-period, 3-treatment crossover designs (ABC|BCA|CAB|ACB|BAC|CBA)

	Period 1	Period 2	Period 3
Sequence ABC	A	B	C
Sequence BCA	B	C	A
Sequence CAB	C	A	B
Sequence ACB	A	C	B
Sequence BAC	B	A	C
Sequence CBA	C	B	A

Introduction to design of crossover trials, cont'd

- Designs incorporate non-crossover sequences, e.g.,
  - ♣ Balaam's design: unusual, with elements of both parallel and crossover design

Table 6: Example of Balaam's design

	Period 1	Period 2
Sequence AB	A	B
Sequence BA	B	A
Sequence AA	A	A
Sequence BB	B	B

Introduction to design of crossover trials, cont'd

- Disadvantages of crossover trials
  - Main disadvantage: Carryover effects may be aliased (confounded) with direct treatment effects, i.e., these effects cannot be estimated separately
    - A carryover effect: the effect of the treatment from the previous time period on the response at the current time period
    - Significant carryover effects can bias the interpre-

tation of data analysis, so it should be proceeded cautiously whenever considering the implementation of a crossover design.

♠ How to deal with this carryover effect?

- The incorporation of lengthy washout periods in the experimental design can diminish the impact of carryover effects.
- A washout period: the time between treatment periods

♠ How long of a washout period should there be?

- Based on an investigator's expertise

- For example, in a trial involving pharmaceutical products a washout period equivalent to 5 (or more) times the length of the half-life of the drug concentration in the blood

♠ Recommendation: avoid the problems caused by differential carryover effects at all costs by employing lengthy washout periods and/or designs where treatment and carryover are not aliased with each other.

Introduction to design of crossover trials, cont'd

- **Definitions with a crossover design**
- **First-order and higher-order carryover effects**
  - **Within time period j, $j = 2, \ldots, p$, it is possible to have carryover effects from treatments administered during periods $1, \ldots, j-1$. Usually in period j only consider first-order carryover effects (from period $j-1$) because**
    - **if first-order carryover effects are negligible, higher-**

order carryover effects usually are negligible;

- ♠ the designs needed for eliminating the aliasing between higher-order carryover effects and treatment effects are very cumbersome and not practical. Therefore, usually assume that these higher-order carryover effects are negligible.

- ♣ The length of the washout periods between treatment administrations may be the determining factor about whether higher-order carryover effects should be considered.

- ♣ Focus on designs for dealing with first-order carry-

over effects, but the development can be generalized if higher-order carryover effects need to be considered.

- Uniformity
  - A crossover design is uniform within sequences if each treatment appears the same number of times within each sequence.
  - A crossover design is uniform within periods if each treatment appears the same number of times within each period.
  - AB|BA is uniform within sequences and periods (each

sequence and each period has 1 A and 1 B).

- ABA|BAB is uniform within periods but is not uniform within sequences because the sequences differ in the numbers of A and B (in sequence ABA, 2A and 1 B; in sequence BAB, 2B and 1A)
- A uniform design: a design is uniform within sequences and uniform within periods.
  - Example: ABC|BCA|CAB is a uniform crossover design (each sequence and each period has 1 A, 1 B, and 1 C).
- If the design is uniform across periods, we will be

able to remove the period effects.

♣ If the design is uniform across sequences, we will be also able to remove the sequence effects.

Introduction to design of crossover trials, cont'd

- Latin Squares
  - Provide the foundation for r-period, r-treatment crossover designs because they yield uniform crossover designs
    - Each treatment occurs only once within each sequence and once within each period.
  - Also serve as building blocks for other types of crossover designs.
  - Example: Latin squares for 4-period, 4-treatment

crossover designs

Table 7: 4-period, 4-treatment crossover designs (ABCD|BCDA|CDAB|DABC)

	Period 1	Period 2	Period 3	Period 4
Sequence ABCD	A	B	C	D
Sequence BCDA	B	C	D	A
Sequence CDAB	C	D	A	B
Sequence DABC	D	A	B	C

Table 8: 4-period, 4-treatment crossover designs (ABCD|BDAC|CADB|DCBA)

	Period 1	Period 2	Period 3	Period 4
Sequence ABCD	A	B	C	D
Sequence BDAC	B	D	A	C
Sequence CADB	C	A	D	B
Sequence DCBA	D	C	B	A

- Latin squares are uniform crossover designs (i.e., uniform both within periods and within sequences).
- Although with 4 periods and 4 treatments there are

$4! = 24$ possible sequences from which to choose, the Latin square only requires 4 sequences.

- Balanced Designs
  - The Latin square in Table 8 has an additional property that the Latin square in Table 7 does not have.
    - In Table 8 each treatment precedes every other treatment the same number of times (once).
    - This is an advantageous property for design in Table 8; the same property does not occur in Table 7.
    - The crossover design, as in Table 8, is balanced

with respect to (w.r.t.) first-order carryover effects.

- Strongly Balanced Designs
  - A crossover design is strongly balanced w.r.t. first-order carryover effects if each treatment precedes every other treatment, including itself, the same number of times.
  - A strongly balanced design can be constructed by repeating the last period in a balanced design.
  - Example: a 4-sequence, 5-period, 4-treatment crossover design strongly balanced w.r.t. first-order carryover

effects

Table 9: 4-sequence, 5-period, 4-treatment crossover design strongly balanced with respect to first-order carryover effects

	Period 1	Period 2	Period 3	Period 4	Period 5
Sequence ABCDD	A	B	C	D	D
Sequence BDACC	B	D	A	C	C
Sequence CADBB	C	A	D	B	B
Sequence DCBAA	D	C	B	A	A

♠ The uniformity of the Latin square design disappears because the design in Table 9 is no longer uniform within sequences.

▶ Uniform and Strongly Balanced Design

♣ Latin squares yield uniform crossover designs, but strongly balanced designs constructed by repeating

the last period of a balanced design are not uniform crossover designs.

♣ Example: 4-sequence, 4-period, 2-treatment crossover design is an example of a strongly balanced and uniform design.

Table 10: 4-sequence, 4-period, 2-treatment crossover design: a strongly balanced and uniform design

	Period 1	Period 2	Period 3	Period 4
Sequence ABBA	A	B	B	A
Sequence BAAB	B	A	A	B
Sequence AABB	A	A	B	B
Sequence BBAA	B	B	A	A

Introduction to design of crossover trials, cont'd

- Statistical Bias
  - Does a particular crossover design have any nuisance effects, such as sequence, period, or first-order carry-over effects, aliased with direct treatment effects?
    - Only consider first-order carryover effects.
    - If the design incorporates washout periods of inadequate length, treatment effects could be aliased with higher-order carryover effects as well.

♠ Assume the washout period is adequate for eliminating carryover beyond 1 treatment period.

♣ Example: in the 2×2 crossover design in Table 1 if including nuisance effects for sequence, period, and first-order carryover, model for this in Table 11 as follows:

Table 11: Expected values of responses of an AB|BA crossover design

	Period 1	Period 2
Sequence AB	$\mu_A + v + p$	$\mu_B + v - p + \lambda_A$
Sequence BA	$\mu_B - v + p$	$\mu_A - v - p + \lambda_B$

♠ μ_A and μ_B: population means for the direct effects of treatments A and B, respectively

♠ v: a sequence effect

- **p: a period effect**
- **λ_A and λ_B: carryover effects of treatments A and B, respectively**
- **An estimator of μ_A (or μ_B): the average over all cells**

$$\hat{\mu}_A = \frac{1}{2}\left(\bar{Y}_{AB,1} + \bar{Y}_{BA,2}\right) \text{ and } \hat{\mu}_B = \frac{1}{2}\left(\bar{Y}_{AB,2} + \bar{Y}_{BA,1}\right) \quad (12)$$

♠ **The expectations of these estimators:**

$$E\left(\hat{\mu}_A\right) = \frac{1}{2}(\mu_A + v + p + \mu_A - v - p + \lambda_B) = \mu_A + \frac{1}{2}\lambda_B$$

$$E\left(\hat{\mu}_B\right) = \frac{1}{2}(\mu_B + v - p + \lambda_A + \mu_B - v + p) = \mu_B + \frac{1}{2}\lambda_A$$

$$E\left(\hat{\mu}_A - \hat{\mu}_B\right) = \mu_A - \mu_B - \frac{1}{2}\left(\lambda_A - \lambda_B\right) \qquad (13)$$

- **From (13) the direct treatment effects and the treatment difference are not aliased with sequence or period effects, but are aliased with the carry-over effects.**
- **The treatment difference is not aliased with carryover effects when the carryover effects are equal,**

i.e., $\lambda_A = \lambda_B$.

- The results in (13) are due to the fact that the AB|BA crossover design is uniform and balanced w.r.t. first-order carryover effects.

  ♦ Any crossover design, which is uniform and balanced w.r.t. first-order carryover effects, such as the designs in Tables 5 and 8, also exhibits these results.

♣ Example: Consider the ABB|BAA design, which is uniform within periods, not uniform within sequences, and is strongly balanced.

Table 12:

	Period 1	**Period 2**	**Period 3**
Sequence ABB	$\mu_A + v + p_1$	$\mu_B + v + p_2 + \lambda_A$	$\mu_B + v - p_1 - p_2 + \lambda_B$
Sequence BAA	$\mu_B - v + p_1$	$\mu_A - v + p_2 + \lambda_B$	$\mu_A - v - p_1 - p_2 + \lambda_A$

♠ **An estimator of μ_A (or μ_B): the average over all cells**

$$\hat{\mu}_A = \frac{1}{3}\left(\bar{Y}_{ABB,1} + \bar{Y}_{BAA,2} + \bar{Y}_{BAA,3}\right)$$
$$\hat{\mu}_B = \frac{1}{3}\left(\bar{Y}_{ABB,2} + \bar{Y}_{ABB,3} + \bar{Y}_{BAA,1}\right) \tag{15}$$

♠ **The expectations of these estimators**

$$
\begin{aligned}
E\left(\hat{\mu}_A\right) &= \frac{1}{3}(\mu_A + v + p_1 + \mu_A - v + p_2 + \lambda_B + \mu_A - v - p_1 - p_2 + \lambda_A) \\
&= \mu_A + \frac{1}{3}(\lambda_A + \lambda_B - v) \\
E\left(\hat{\mu}_B\right) &= \frac{1}{3}(\mu_B + v + p_2 + \lambda_A + \mu_B + v - p_1 - p_2 + \lambda_B + \mu_B - v + p_1) \\
&= \mu_B + \frac{1}{3}(\lambda_A + \lambda_B + v) \\
E\left(\hat{\mu}_A - \hat{\mu}_B\right) &= \mu_A - \mu_B - \frac{2}{3}v \qquad (16)
\end{aligned}
$$

- **From (16), the direct treatment effects are aliased with the sequence effect and the carryover effects.**
- **The treatment difference only is aliased with the**

sequence effect.

- The results in (16) are due to the ABB|BAA crossover design being uniform within periods and strongly balanced w.r.t. first-order carryover effects.

Introduction to design of crossover trials, cont'd

- **Higher-order Carryover Effects**
  - **The lack of aliasing between the treatment difference and the first-order carryover effects does not guarantee that the treatment difference and higher-order carryover effects also will not be aliased or confounded.**
    - **Example: let λ_{2A} and λ_{2B} denote the second-order carryover effects of treatments A and B, respec-**

tively, for the design in Table 2 (Second-order carryover effects look at the carryover effects of the treatment that took place previous to the prior treatment.)

Table 13:

	Period 1	**Period 2**	**Period 3**
Sequence ABB	$\mu_A + v + p_1$	$\mu_B + v + p_2 + \lambda_A$	$\mu_B + v - p_1 - p_2 + \lambda_B + \lambda_{2A}$
Sequence BAA	$\mu_B - v + p_1$	$\mu_A - v + p_2 + \lambda_B$	$\mu_A - v - p_1 - p_2 + \lambda_A + \lambda_{2B}$

- **An estimator of μ_A (or μ_B): the average over all cells**

$$\hat{\mu}_A = \frac{1}{3}\left(\bar{Y}_{ABB,1} + \bar{Y}_{BAA,2} + \bar{Y}_{BAA,3}\right)$$
$$\hat{\mu}_B = \frac{1}{3}\left(\bar{Y}_{ABB,2} + \bar{Y}_{ABB,3} + \bar{Y}_{BAA,1}\right)$$

- **The expectations of these estimators**

$$
\begin{aligned}
& E\left(\hat{\mu}_A\right) \\
=& \frac{1}{3}(\mu_A + v + p_1 + \mu_A - v + p_2 + \lambda_B + \mu_A - v - p_1 - p_2 + \lambda_A + \lambda_{2B}) \\
=& \mu_A + \frac{1}{3}(\lambda_A + \lambda_B + \lambda_{2B} - v) \\
& E\left(\hat{\mu}_B\right) \\
=& \frac{1}{3}(\mu_B + v + p_2 + \lambda_A + \mu_B + v - p_1 - p_2 + \lambda_B + \lambda_{2A} + \mu_B - v + p_1) \\
=& \mu_B + \frac{1}{3}(\lambda_A + \lambda_B + \lambda_{2A} + v) \\
& E\left(\hat{\mu}_A - \hat{\mu}_B\right) = \mu_A - \mu_B - \frac{1}{3}(\lambda_{2A} - \lambda_{2B}) - \frac{2}{3}v \qquad (18)
\end{aligned}
$$

- **From (18) the treatment mean difference is aliased**

with second-order carryover effects.

- Summary of Impacts of Design Types
  - If the crossover design is uniform within sequences, sequence effects are not aliased with treatment differences.
  - If the crossover design is uniform within periods, period effects are not aliased with treatment differences.
  - If the crossover design is balanced w.r.t. first-order carryover effects, carryover effects are aliased with treatment differences.
    - If the carryover effects are equal, carryover effects

are not aliased with treatment differences.

- If the crossover design is strongly balanced w.r.t. first-order carryover effects, carryover effects are not aliased with treatment differences.

Complex Carryover

- The type of carryover effects we modeled here is called "simple carryover" because it is assumed that the treatment in the current period does not interact with the carryover from the previous period.
- "Complex carryover" refers to the situation in which such an interaction is modeled.

♠ **For example, suppose we have a crossover design and want to model carryover effects.**

- **With simple carryover in a two-treatment design, there are two carryover parameters λ_A and λ_B.**
- **With complex carryover, however, there are four carryover parameters λ_{AB}, λ_{BA}, λ_{AA} and λ_{BB}.**
  - ♦ **λ_{AB}: the carryover effect of treatment A into a period in which treatment B is administered**
  - ♦ **λ_{BA}: the carryover effect of treatment B into a period in which treatment A is administered**
  - ♦ **λ_{AA}: the carryover effect of treatment A into a**

period in which treatment A is administered

- **λ_{BB}: the carryover effect of treatment B into a period in which treatment B is administered**
- **This will certainly complicate things!**

Introduction to design of crossover trials, cont'd

- Implementation Overview
  - Obviously, an "ideal" crossover design is uniform and strongly balanced.
  - There are situations, however, where it may be reasonable to assume that some of the nuisance parameters are null, so that resorting to a uniform and strongly balanced design is not necessary (although it provides a safety net if the assumptions do not hold).

- ♠ **Example 1: some researchers argue that sequence effects should be null or negligible because they represent randomization effects.**
- ♠ **Example 2: in bioequivalence trials some researchers argue that carryover effects should be null because blood concentration levels of the drug or active ingredient are monitored and any residual drug administered from an earlier period would be detected.**

♣ **Every proposed crossover trial should be examined to determine which, if any, nuisance effects may play**

a role. Once this determination is made, then an appropriate crossover design should be employed that avoids aliasing of those nuisance effects with treatment effects. This is a decision that the researchers should be prepared to address.

♠ Example:

- An investigator wants to conduct a two-period crossover design, but is concerned that he will have unequal carryover effects so he is reluctant to invoke the 2×2 crossover design.
- If the investigator is not as concerned about se-

quence effects, Balaam's design in Table 8 may be appropriate.

- Balaam's design is uniform within periods but not within sequences, and it is strongly balanced. Therefore, Balaam's design will not be adversely affected in the presence of unequal carryover effects.

♣ Some researchers consider randomization in a crossover design to be a minor issue because a patient eventually undergoes all of the treatments (this is true in most crossover designs).

- ♠ Obviously, randomization is very important if the crossover design is not uniform within sequences because the underlying assumption is that the sequence effect is negligible.
- ♠ Randomization is important in crossover trials even if the design is uniform within sequences because biases could result from investigators assigning patients to treatment sequences.

♣ At a *minimum*, it always is recommended to invoke a design that is uniform within periods because *period effects are common*. Period effects can be due to

- **increased patient comfort in later periods with trial processes;**
- **increased patient knowledge in later periods;**
- **improvement in skill and technique of those researchers taking the measurements.**

♣ A list of various crossover designs with some, all, or none of the properties

Uniform within Sequences	Uniform within Periods	Balanced	Strongly Balanced	Examples
no	no	no	no	AAB|ABB, ABCC|BCAA
yes	no	no	no	ABB|BAB, ABC|CBA
no	yes	no	no	ABCC|BCAA|CABB
no	no	yes	no	ABAA|BAAB
no	no	yes	yes	AABBA|BAABB
yes	yes	no	no	ABC|BCA|CAB
yes	no	yes	no	AABA|ABAA
no	yes	yes	no	ABA|BAB
yes	no	yes	yes	AABBA|ABBAA
no	yes	yes	yes	ABB|BAA, AB|BA|AA|BB
yes	yes	yes	no	AB|BA
yes	yes	yes	yes	ABBA|BAAB|AABB|BBAA

♣ It would be a good idea to go through each of these designs and diagram out what these would look like, the degree to which they are uniform and/or bal-

anced.

- **Make sure you see how these principles come into play!**

Introduction to design of crossover trials, cont'd

- Statistical Precision
  - Which crossover design provides the best precision?
    - Label one design as more precise than another if it yields a smaller variance for the estimated treatment mean difference.
    - Compare designs w.r.t. which designs are best for estimating and comparing variances.
    - Focus on differences in estimated treatment means

in two-period, two-treatment designs.

- 2×2 crossover design AB|BA in Table 1
- Balaam's design AB|BA|AA|BB in Table 6
- two-period parallel design AA|BB

♠ Assume that the total sample size, n, is a positive integer divisible by 4.

- $\frac{1}{2}n$ patients will be randomized to each sequence in the 2×2 crossover design: AB|BA design.
- $\frac{1}{2}n$ patients will be randomized to each sequence in the two-period parallel design: AA|BB design.
- $\frac{1}{4}n$ patients will be randomized to each sequence

in the Balaam's design: AB|BA|AA|BB design.

♠ Because the designs involve repeated measurements on patients, the statistical modeling must account for between-patient variability and within-patient variability.

- Between-patient variability accounts for the dispersion in measurements from one patient to another.
- Within-patient variability accounts for the dispersion in measurements from one time point to another within a patient.

- **Within-patient variability tends to be smaller than between-patient variability.**

♠ **The variance components are as follows:**

- W_{AA}**: between-patient variance for treatment A**
- W_{BB}**: between-patient variance for treatment B**
- W_{AB}**: between-patient covariance between treatments A and B**
- σ_{AA}**: within-patient variance for treatment A**
- σ_{BB}**: within-patient variance for treatment B**

♠ **Expressions for the variance of the estimated treatment mean difference for each of the two-period,**

two-treatment designs:

Table 14:

Design	Variance
Crossover	$\sigma^2/n = \{1.0(W_{AA} + W_{BB}) - 2.0W_{AB} + (\sigma_{AA} + \sigma_{BB})\}/n$
Balaam	$\sigma^2/n = \{1.5(W_{AA} + W_{BB}) - 1.0W_{AB} + (\sigma_{AA} + \sigma_{BB})\}/n$
Parallel	$\sigma^2/n = \{2.0(W_{AA} + W_{BB}) - 0.0W_{AB} + (\sigma_{AA} + \sigma_{BB})\}/n$

- **Under most circumstances, W_{AB} will be positive.**
  - **The 2×2 crossover design yields the smallest variance for the estimated treatment mean difference, followed by Balaam's design and then the parallel design.**
- **The investigator needs to consider other design issues, however, prior to selecting the 2×2 crossover design.**

- **In particular, if there is any concern over the possibility of differential first-order carryover effects, the 2×2 crossover design is not recommended.**
  - **In this situation the parallel design would be a better choice than the 2×2 crossover design.**
  - **Balaam's design is strongly balanced so that the treatment difference is not aliased with differential first-order carryover effects, so it also is a better choice than the 2×2 crossover design.**

A sample size calculation

- **The total sample size, n, required for a two-sided,**

α significance level test with $100(1-\beta)\%$ power and effect size $\mu_A - \mu_B$ is

$$n = (z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2 / (\mu_A - \mu_B)^2.$$

- **Suppose we want to conduct a two-period trial but is not sure whether to invoke a parallel design, a crossover design, or Balaam's design.**
- **Use a 0.05 significance level test with 90% power for detecting the effect size of $\mu_A - \mu_B = 10$.**

♠ **Assume**

Between-patient variances $W_{AA} = W_{BB} = W_{AB} = 400$

Within-patient variances $\sigma_{AA} = \sigma_{BB} = 100$

♠ **The sample sizes for the three different designs:**

Parallel $n = 190$, **Balaam** $n = 108$, **Crossover** $n = 21$

- **The crossover design yields a much smaller sample size because the within-patient variances are one-fourth that of the inter-patient variances (which is not unusual).**

- **Another issue in selecting a design is whether the experimenter wishes to compare the within-patient variances σ_{AA} and σ_{BB}.**
- **For the 2×2 crossover design, the within-patient variances can be estimated by imposing restrictions on the between-patient variances and covariances.**
- **The estimators of σ_{AA} and σ_{BB}, however, may lack precision and be unstable. Hence, the 2×2 crossover design is not recommended when comparing σ_{AA} and σ_{BB} is an objective.**

- **The parallel design provides optimal estimation of the within-unit variances because it has $\frac{1}{2}n$ patients who can provide data in estimating each of σ_{AA} and σ_{BB}, whereas Balaam's design has $\frac{1}{4}n$ patients who can provide data in estimating each of σ_{AA} and σ_{BB}.**
- **Balaam's design is a compromise between the 2×2 crossover design and the parallel design.**

Statistical analysis of crossover trial data

▶ Continuous Outcome

♣ **Assume carryover effects are equal** $(\lambda_A = \lambda_B = \lambda)$ **from a** 2×2 **crossover trial.**

♣ **The statistical model for continuous data from the** 2×2 **crossover trial:**

Table 15:

	Period 1	**Period 2**
Sequence AB	$\mu_A + v + p$	$\mu_B + v - p + \lambda_A$
Sequence BA	$\mu_B - v + p$	$\mu_A - v - p + \lambda_B$

- **In sequence AB, the Period 1 vs. Period 2 difference expectation**

$$\mu_{AB} = \mu_B + v - p + \lambda_A - (\mu_A + v + p) = \mu_A - \mu_B + 2p - \lambda.$$

- **In sequence BA, the Period 1 vs. Period 2 difference expectation**

$$\mu_{BA} = \mu_A - v - p + \lambda_B - (\mu_B - v + p) = \mu_B - \mu_A + 2p - \lambda.$$

- **To compare the two sequences w.r.t. these differences, test**

$$H_0 : \mu_{AB} - \mu_{BA} = 0 \Longleftrightarrow H_0 : \mu_A - \mu_B = 0$$

- Two-sample t test
- Wilcoxon rank-sum test (Mann-Whitney-Wilcoxon rank-sum test)

♣ Example: Example 3.1 from the book of Senn (Senn, S. *Cross-over Trials in Clinical Research*, 2nd ed. Chichester, England: John Wiley & Sons, 2002).

♠ 13 children enrolled in a trial to investigate the effects of two bronchodilators, formoterol and salbutamol, in the treatment of asthma

- Outcome variable: peak expiratory flow (PEF) rate (liters per minute) and was measured eight

hours after treatment

- One-day washout period between treatment periods
- The estimated treatment mean difference was 46.6 L/min in favor of formoterol ($p = 0.0012$) and the 95% confidence interval for the treatment mean difference is (22.9, 70.3).
- The Wilcoxon rank-sum test also indicated statistical significance between the treatment groups ($p = 0.0276$).

Statistical analysis of crossover trial data, cont'd

- **Binary Outcome**
  - **Suppose that the response from a crossover trial is binary and that there are no period effects.**
  - **The probabilities of response:**

	Failure (0) on B	Success (1) on B	marginal probabilities
Failure (0) on A	p_{00}	p_{01}	$p_{0\cdot} = p_{00} + p_{01}$
Success (1) on A	p_{10}	p_{11}	$p_{1\cdot} = p_{10} + p_{11}$
marginal probabilities	$p_{\cdot 0} = p_{00} + p_{10}$	$p_{\cdot 1} = p_{01} + p_{11}$	

  - **Test the null hypothesis of the probability of success on treatment A $p_{1\cdot}$ = the probability of success**

on treatment B $p_{\cdot 1}$

$$H_0 : p_{1\cdot} - p_{\cdot 1} = 0$$

$$\Longleftrightarrow H_0 : p_{1\cdot} - p_{\cdot 1} = (p_{10} + p_{11}) - (p_{01} + p_{11}) = p_{10} - p_{01} = 0,$$

which indicates that only the patients who display a (1,0) or (0,1) response contribute to the treatment comparison.

- **If they failed on both, or were successful on both, there is no way to determine which treatment is better.**
- **Denote the frequency of responses from the study**

data instead of the probabilities listed above.

	Failure (0) on B	**Success (1) on B**
Failure (0) on A	n_{00}	n_{01}
Success (1) on A	n_{10}	n_{11}

♠ McNemar's test

- **Given the number of patients who displayed a treatment preference,** $n_{10}+n_{01}$**,** n_{10} **follows a** ***binomial(***$n_{10}+n_{01}, p$***)*** **distribution and the null hypothesis reduces to testing**

$$H_0 : p = 0.5$$

i.e., we would expect a 50-50 split in the number

of patients that would be successful with either treatment in support of the null hypothesis, looking at only the cells where there was success with one treatment and failure with the other.

- The data in cells for both success or failure with both treatments would be ignored.

♣ Example: data from a 2×2 crossover trial with a binary outcome of failure/success.

♠ Fifty patients were randomized.

	Failure (0) on B	Success (1) on B
Failure (0) on A	21	15
Success (1) on A	7	7

- ♠ 22 patients displayed a treatment preference, of which 7 preferred A and 15 preferred B.
- ♠ McNemar's test indicated that this was not statistically significant (exact $p = 0.1338$).

♣ A problem that can arise from the application of McNemar's test to the binary outcome from a 2×2 crossover trial can occur if there is non-negligible period effects.

- ♠ If that is the case, the treatment comparison should account for this.
- ♠ This is possible via logistic regression analysis.

- **The probability of a 50-50 split between treatments A and B preferences under the null hypothesis is equivalent to the odds ratio (OR) for the treatment A preference to the treatment B preference being 1.0.**
- **Because logistic regression analysis models the natural logarithm of the odds, testing whether there is a 50-50 split between treatments A and B preferences is comparable to testing whether the intercept term is null in a logistic regression analysis.**

- To account for the possible period effect in the 2×2 crossover trial, a term for period can be included in the logistic regression analysis.

♠ Example: Use the same data set partitioned as to patients within the two sequences

Sequence AB	Failure (0) on B	Success (1) on B
Failure (0) on A	10	7
Success (1) on A	3	5

Sequence BA	Failure (0) on B	Success (1) on B
Failure (0) on A	11	8
Success (1) on A	4	2

- The logistic regression analysis yielded a non-significant result for the treatment comparison (exact $p = 0.2266$).

Statistical analysis of crossover trial data, cont'd

- Time-to-Event Outcome
  - Not often to see a cross-over design used in a time-to-event trial
    - If the event is death, the patient would not be able to cross-over to a second treatment.
    - Even when the event is treatment failure, this often implies that patients must be watched closely and perhaps rescued with other medicines when event

failure occurs.

- A time-to-event outcome within the context of a 2×2 crossover trial actually can reduce to a binary outcome score of preference.
  - Suppose that in a clinical trial, time to treatment failure is determined for each patient when receiving treatments A and B.
  - If the time to treatment failure on A equals that on B, the patient is assigned a (0,0) score and displays no preference.
  - If the time to treatment failure on A is less than

that on B, the patient is assigned a (0,1) score and prefers B.

- If the time to treatment failure on B is less than that on A, the patient is assigned a (1,0) score and prefers A.
- If the patient does not experience treatment failure on either treatment, the patient is assigned a (1,1) score and displays no preference.
- Hence, we can use the procedures which we implemented with binary outcomes.

Statistical analysis of crossover trial data, cont'd

- Outcome Data From More Complex Designs
  - When a design more complex than the 2×2 crossover, extensive modeling is required.
    - Continuous Outcome
      - A mixed-effects linear model (e.g., SAS PROC MIXED) to account for the repeated measurements that yield period, sequence, and carryover effects and to model the various sources of intra-

patient and inter-patient variability

♠ Binary Outcome

- Generalized estimating equations (GEE) (e.g., SAS PROC GENMOD) to account for the repeated measurements that yield period, sequence, and carryover effects and to model the various sources of intra-patient and inter-patient variability

Power analysis for crossover trials

- **Five components in power analyese**

  1. **Probability of type I error ($\alpha = 0.05$ or 0.01) (significance level)**

     - $P(\textbf{reject } H_0 | H_0) = \alpha$
       - H_0: **null hypothesis;** H_a: **alternative hypothesis**
     - **Type I error: concluding that a treatment effect or difference exists when, in reality, it does not**
     - **a false positive**

2. **Power = 1- probability of type II error (=$1 - \beta$ = 0.8 or 0.9)**

   ♠ $P(\textbf{accept } H_0 | H_a) = \beta$

   ♠ **Type II error: concluding that a treatment effect or difference does not exist when, in reality, it does.**

   • **a false negative**

   ♠ **Power =** $P(\textbf{reject } H_0 | H_a) = 1 - \beta$ **: the chance of detecting a specified (effect) size as being statistically significant**

3. **Treatment effect (size) or treatment difference** (δ) **based on clinical considerations**

4. Nuisance parameter (σ)

- ♠ usually unknown
- ♠ needed to be estimated from either literature or preliminary data collected from a pilot study
- ♠ If not available, e.g., for normally distributed data, consider a standardized treatment effect (difference) δ/σ while doing power analyses

5. Sample size (N)

- ♠ the number of patients or subjects on a study

Power analysis for crossover trials, cont'd

- Simple AB|BA crossover designs
  - Under certain simplifying assumptions, we can test the treatment difference in an AB|BA crossover trial by using either a paired or two-sample t test (or equivalence test, depending on the hypothesis).
    - Example: A study comparing difference between two new medications, "Xilodol" and "Brantium"
      - Half of the patients would be assigned to sequence AB, getting a dose of Xilodol in the first

treatment period, a washout period of one week, and then a dose of Brantium in the second treatment period.

- The other half would be assigned to sequence BA, following the same schedule but with the drugs reversed.
- In each treatment period we would administer the drugs in the morning and then measure peak expiratory flow (PEF) at the end of the day, with higher PEF representing better lung function.
- We conjecture that the mean and standard devi-

ation of PEF are about $\mu_{\mathrm{A}} = 330$ and $\sigma_{\mathrm{A}} = 40$ for Xilodol and $\mu_{\mathrm{B}} = 310$ and $\sigma_{\mathrm{B}} = 55$ for Brantium, and that each pair of measurements on the same subject will have a correlation of about 0.3.

- The powers of both one-sided and two-sided paired t-tests of mean difference, with a significance level of $\alpha = 0.01$, for a sample size of 100 patients are 0.865 and 0.801, respectively.
  - Note that the allocation ratio of patients to the two sequences is irrelevant in this analysis.
- The choice of statistical test depends on which

assumptions are reasonable.

- One possibility is a t test.

- A paired or two-sample t test is valid when there is no carryover effect and no interactions between patients, treatments, and periods.

  - See, e.g., Senn (2002, Chapter 3)[1] for more details.

- The choice between a paired or a two-sample test depends on what we assume about the period effect.

  - If no period effect is assumed, a paired t test

[1] Senn, S. *Cross-over Trials in Clinical Research*, 2nd ed. Chichester, England: John Wiley & Sons, 2002

is the appropriate analysis for the design, with the first member of each pair being the Xilodol measurement (regardless of which sequence the patient belongs to).

♦ Otherwise, the two-sample t test approach is called for because this analysis adjusts for the period effect by using an extra degree of freedom.

- See Figure 1 for plot of power versus sample size for paired t analysis of crossover design for a range of 50 to 200 patients.

♠ Example: (Equivalence trial) A study establishing

similarity between two new medications, "Xilodol" and "Brantium"

- For example, the absolute mean PEF difference is at most 35.
- Consider this goal if, e.g., one of the drugs has fewer side effects and if a difference of no more than 35 is considered clinically small.
- Instead of a standard t test, conduct an equivalence test of the treatment mean difference for the two drugs.
- Test the hypothesis that the true difference is less

than -35 or more than 35

$$H_0 : \mu_A - \mu_B < -35 \text{ or } \mu_A - \mu_B > 35, \text{ i.e., } |\mu_A - \mu_B| > 35$$

against the alternative that the mean difference is between -35 and 35

$$H_a : -35 \leq \mu_A - \mu_B \leq 35, \text{ i.e., } |\mu_A - \mu_B| \leq 35$$

by using an additive model and a two one-sided tests ("TOST")[2] analysis.

- **Assuming no period effect and using the same pa-**

[2] Schuirmann, D.J. (1987). A Comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability, *Journal of Pharmacokinetics and Biopharmaceutics*, 15, 657-680.

rameter values in the above example, the power of the paired equivalence test, with a significance level of $\alpha = 0.01$, for a sample size of 100 patients is 0.598.

- See Figure 2 for plot of power versus sample size for paired equivalence test for crossover design for a range of 50 to 200 patients.

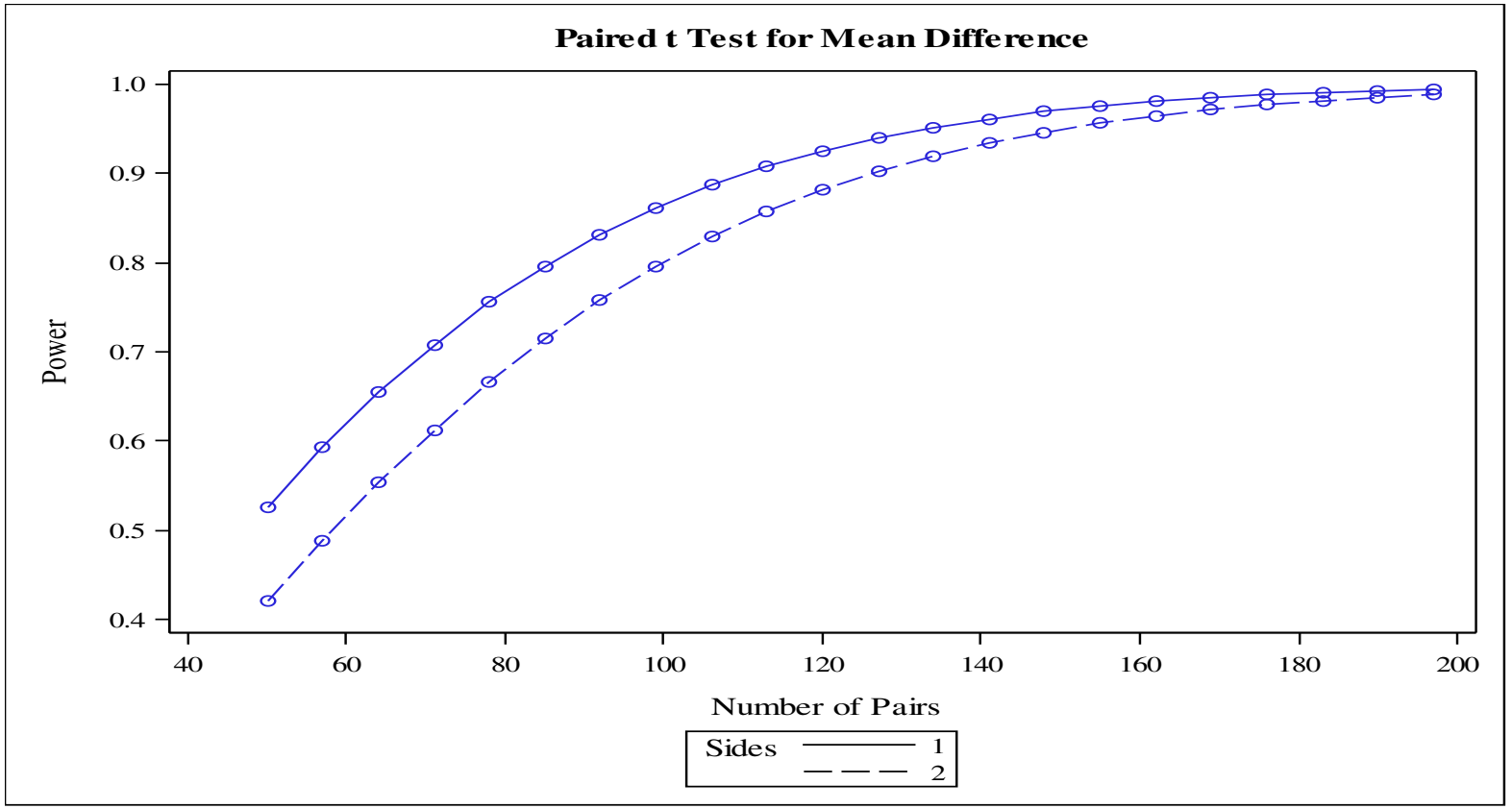

Figure 1: Plot of Power versus Sample Size for Paired t Analysis of Crossover Design for a Range of 50 to 200 Patients

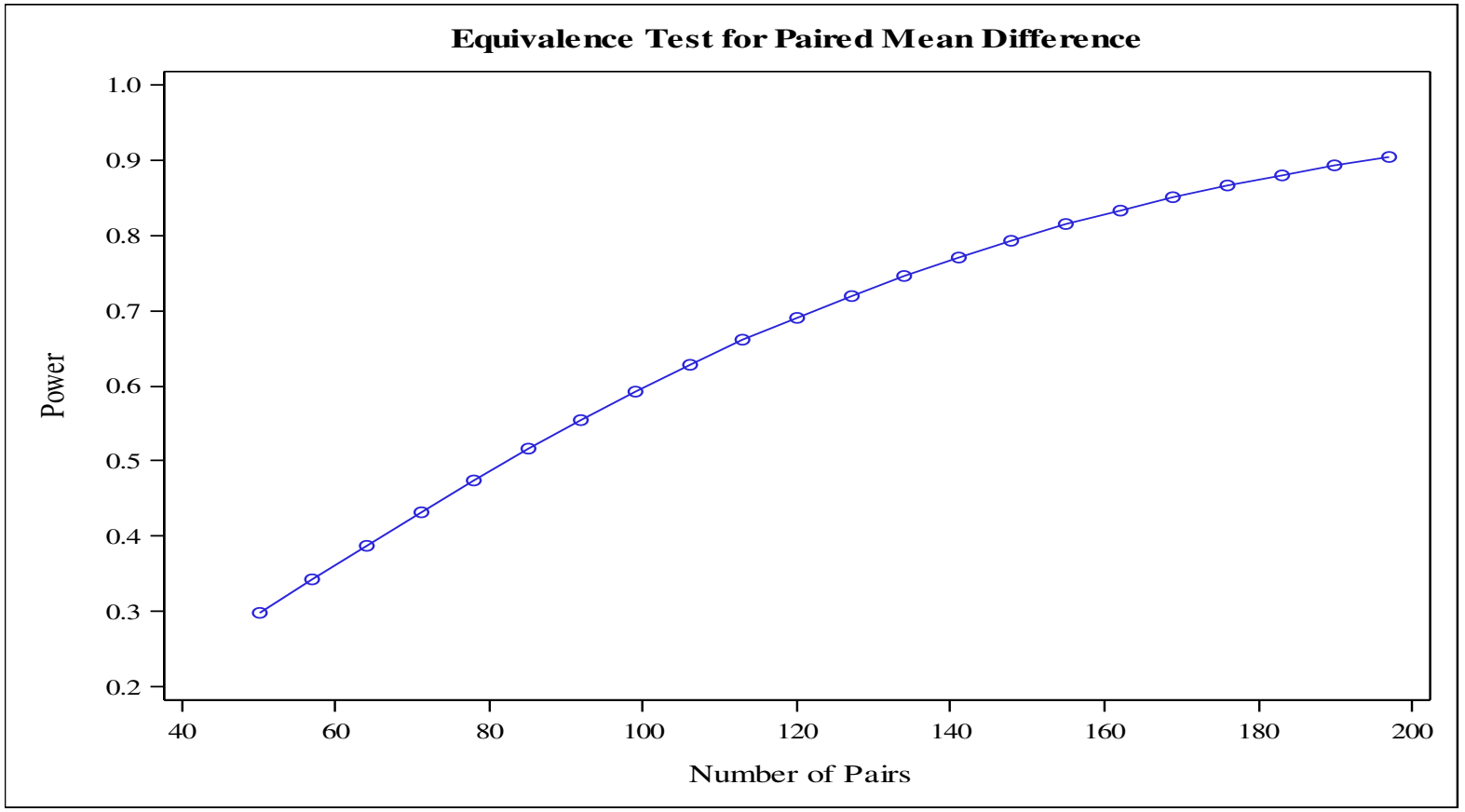

Figure 2: Plot of Power versus Sample Size for Paired Equivalence Test for Crossover Design for a Range of 50 to 200 Patients

References

1. https://onlinecourses.science.psu.edu/stat509/node/123

2. Jones, B. and Kenward, M.G. (2015). *Design and analysis of cross-over trials*, 3rd ed. New York: Chapman & Hall/CRC.

3. R packages for crossover designs

   - crossdes
   - Crossover

4. **Schuirmann, D.J. (1987). A Comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability, *Journal of Pharmacokinetics and Biopharmaceutics*, 15, 657-680.**

5. **Senn, S. (2002). *Cross-over Trials in Clinical Research*, 2nd ed. Chichester, England: John Wiley & Sons.**

6. **SAS PROC POWER, V9.4, SAS Institute Inc., Cary, NC, USA.**