Crossover and Self-Controlled Designs in Clinical Research

Louis T.A., Lavori P.W., Bailar J.C., and Polansky M.

Clinical Trials Reading Assignment (BSDS 6314)

Presented by: Oluwafunmibi Omotayo Fasanya

Louisiana State University Health Sciences Center, New Orleans

November 19, 2025

Abstract

Purpose:

- To evaluate the design and analysis of crossover and self-controlled studies published in medical journals during 1978–1979.
- To highlight statistical and clinical advantages of these designs in achieving valid results with fewer participants.

Methods and Findings:

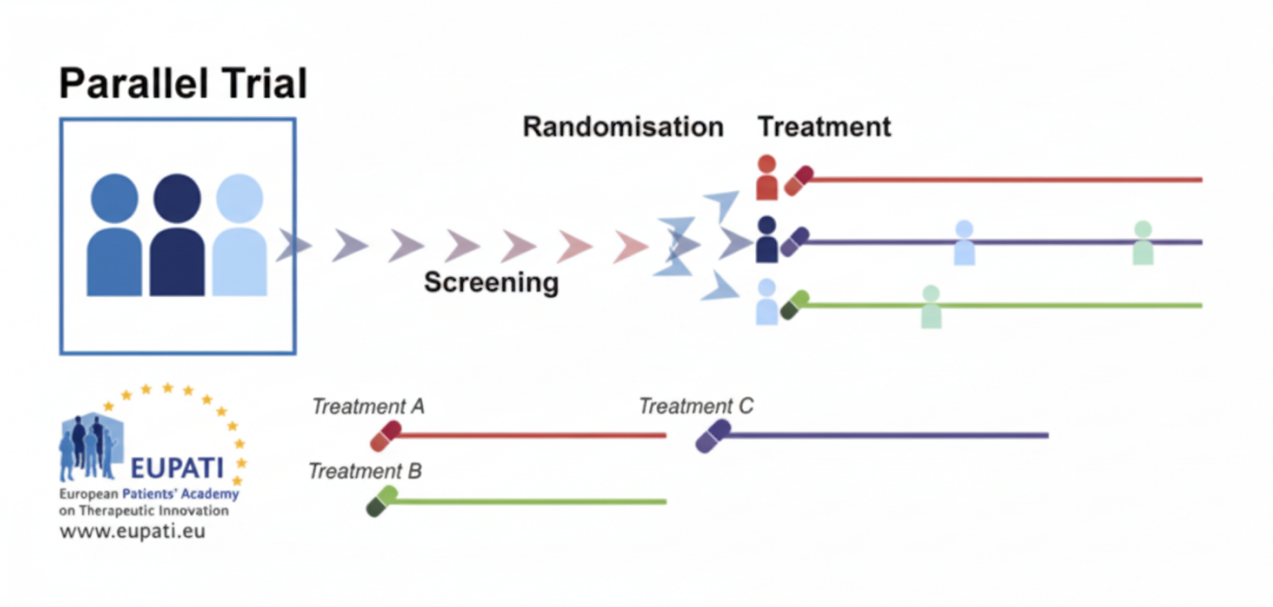
- Reviewed 13 crossover studies examining the following:
  - Method of randomization to initial treatment (7/13 studies)
  - Criteria for switching treatments (10 used time-dependent rule; 3 used disease-state dependent rule)
  - Blinding of crossover point (3 concealed; 4 found it infeasible)
  - Assessment of treatment order effects (1/13 studies)
  - Use of minimally acceptable statistical methods (11/13 studies)
- Additionally reviewed 28 self-controlled studies for comparative context.

Conclusion:

- Crossover and self-controlled designs offer efficiency but require rigorous attention to randomization, blinding, and statistical analysis to ensure validity.

Parallel Design

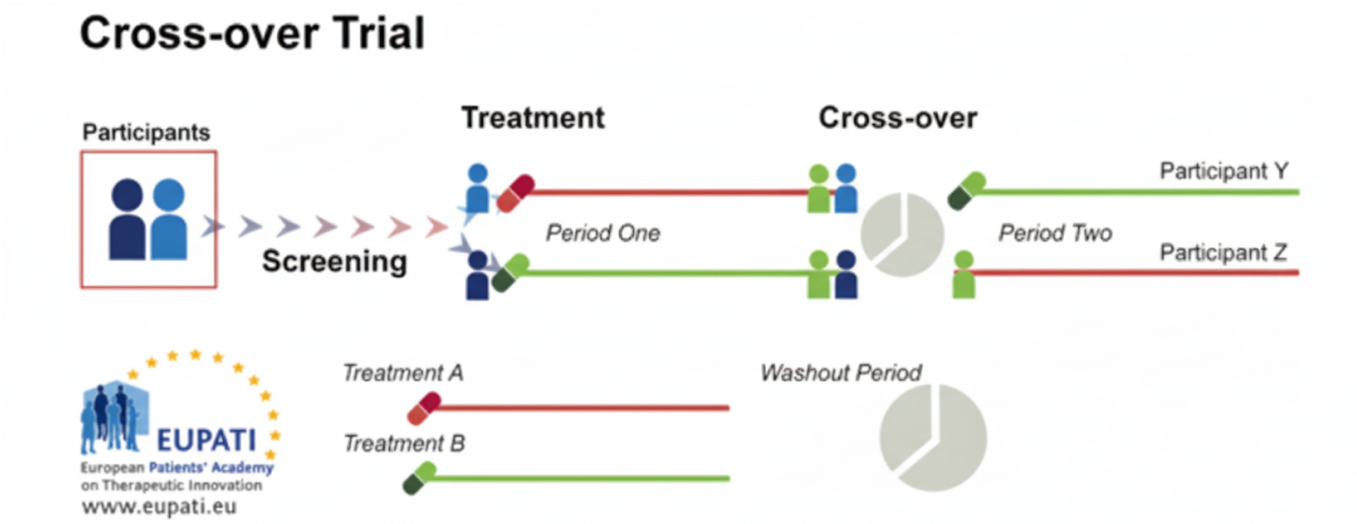
- Comparing two or more groups of patients treated separately but concurrently as part of the same study.
- Each patient receives only one treatment and response in one group are compared with those in another.



- If patients assigned to separate treatment groups differ by prognosis, e.g. age, results may be affected by this difference as well as by the treatment.

Crossover Design

- Each patient receives two or more treatments in sequence reducing individual differences.



- *Biologic or measurement variation within a patient is not removed, but if small, a crossover can achieve the same precision as a larger parallel trial.*
- A small crossover study can achieve same accuracy of larger parallel study.

Decision to use a crossover design should not be based solely on the potential saving in sample size, because powerful designs are also potential disasters.

Power of the Crossover Design

Study: Raskin & Unger: *Hyperglucagonemia and its suppression in diabetes control*
Goal: Compare the effects of insulin infusion regimens on blood and urine chemistry.

- **Treatment 1 (IS):** Insulin + somatostatin (administered first)
- **Treatment 2 (ISG):** Insulin + somatostatin + glucagon (administered second)
- **Response:** Rate of urea nitrogen excretion (g/24 hr)
- Washout period incorporated between treatments

	Parallel Design	**Crossover Design**
Sample size	8 patients	4 patients
Difference in means (g/24hr)	3	3
SE of difference	2.76	0.40
t-statistic	1.08 (ns)	7.5 ($p < 0.01$)
Degrees of freedom	6	3
Conclusion	Not significant	Statistically significant

Efficiency and Limitations of the Crossover Design

Efficiency of Crossover Design:

- No treatment order effect and variation in response is same as crossover, parallel design require **14 times more patients** (56 vs. 4) to achieve the same statistical significance.
- No variance reduction, crossover designs require only **half the sample size** of parallel designs for equivalent precision.

Limitation of the Crossover Design:

- **No randomization:** All patients received IS first, then ISG – result could either reflect true treatment effect or treatment-order effect (nitrogen excretion may be due to receiving ISG treatment second)
- **Small sample size:** 4 patients limits clinical generalizability despite significance

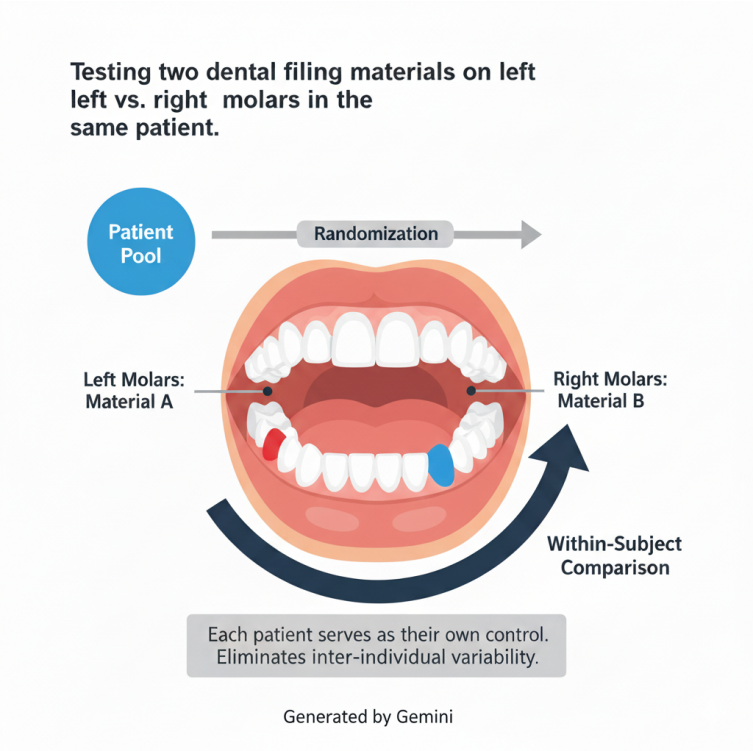
Limitations of the Crossover Design

- **Confounding possibilities:**
  - **Regression to the mean:** If patients were selected based on unusually low urea nitrogen values, their values would naturally increase over time regardless of treatment.
  - **Disease progression:** Diabetes condition could have changed between the first and second treatment periods.
  - **Carryover effects:** Despite the washout period, residual effects from IS treatment may have influenced ISG measurements.
  - **Time/period effects:** Natural physiological changes over the study duration could affect nitrogen excretion.
- **Stronger design:** Half should have started with IS, half with ISG (randomized treatment sequence).

While the crossover design achieved significance with far fewer patients, lack of randomization prevents us from distinguishing treatment effects from order effects.

Paired-Organs Design

- A special type of crossover design where paired organs within the same patient receive different treatments.
- One treatment is applied to one organ (e.g., left eye, one tooth) and another treatment is applied to the paired organ (e.g., right eye, another tooth)



- Shares similar strengths and weaknesses with standard crossover designs

Factors to Consider Before Designing a Crossover Study

General Guidelines on their Consideration

Application to the 13 Crossover Studies

Factors to Consider Before Designing a Crossover Study

The success of a crossover design over a parallel design is based on the following five scientific and statistical assumptions:

- Order Effect (Carry-over and period effects on treatments)
- Treatment sequencing and patient assignment.
- Crossover rules and timing of measurements.
- Dropouts, faulty data, and other data problems.
- Statistical analysis and sample size.

An investigator choosing between parallel and crossover designs should consider the above five factors as it determines the effectiveness of a crossover design.

Crossover Study used in the Article

- The above 5 assumptions were used in comparing the use of crossover design in 13 crossover studies that appeared in the journal of the New England Journal of Medicine from volume 298 to 301 in 1978-1979.
- These studies were identified using the classification system developed for clinical studies.

Condition used for classifying a study as a crossover design:

- Both treatments must be a realistic candidate for clinical use
  - Both treatments must be a viable option that could be used in clinical practice, they can't just be experimental or theoretical treatments.
- Sequential administration of treatment must be possible
  - Each treatment must be able to be given after the other treatment. Patients receives both treatment in sequence.

Order Effect (Carry-over and period effects on treatments)

Carry-Over Effects

- Persisting effects of the first treatment into the next treatment i.e. current treatment's activity depends on the previous treatment.
- Can bias treatment comparisons.
- Minimized via washout periods.

Period Effects

- Disease or patient condition changes naturally over time (progresses, regresses, or fluctuates).

Order Effects (Only 1 study tested whether treatment order affected the outcome)

- Reduces statistical power and complicates analysis.
- Assess using appropriate statistical models.
- Order effects large as treatment effects, crossover designs lose their advantage over parallel designs.
- Effects minimized using balanced sequencing and well-timed measurements.

Treatment Sequencing and Patient Assignment

Treatment Sequencing

- How investigator assigns and orders treatments for each patient in a crossover study.
- If treatment A is followed by treatment B (sequence AB) for all patients, then this comparison is based on a strong assumption that there is no carry-over or order effects i.e. effect of B after A = effect of A after B
- If disease or treatment characteristics make B after A different from A after B (e.g., chemotherapy after radiation vs. radiation after chemotherapy), then the assumption is violated, and a valid treatment comparison cannot be made.
- Careful sequencing and random assignment of patients to different treatment orders (e.g., AB and BA) are essential to reduce bias and maintain study validity.

Treatment Sequencing and Patient Assignment (cont'd)

Patient Assignment Strategies

- Fixed Sequence (4 papers):
  - All patients receive the same order.
  - Validity depends on external evidence.
- Random Assignment (7 papers):
  - Patients randomly assigned to sequences.
  - Recommended
- Deterministically Balanced (1 paper):
  - Balanced assignment (e.g., alternating AB, BA, AB, BA...)
  - May be valid but more prone to selection bias
- Uncontrolled/Haphazard (1 paper):
  - No defined procedure.
  - High risk of bias.

Crossover rules and timing of measurements.

Time-Dependent Crossover (Best Practice) (10 papers)

- Switch in treatment occurs after a specified length of time, e.g. 4 weeks on Treatment A, then switch to B.
- Measurements based on elapsed time only is most scientifically acceptable

Disease-State-Dependent Crossover (1 paper)

- Switch occurs based on patient's clinical characteristics
- Example: Switch when symptoms worsen or improve
- Difficult to interpret treatment effect when timing of measurements is based on response to symptom appearance or disappearance

Note: 2 papers had unclear crossover rules

Whenever possible, crossover points should be concealed from patients and observers (blinded) as this can influence patient's treatment response, observer's assessment of outcomes, reporting of symptoms.

Study: Only 3 blinded the crossover point (blinding impossible in 4 others).

Dropouts, faulty data, and other data problems

Dropouts and implausible data points are problematic in crossover studies because

- Each patient contributes a large proportion of total information
- Design is sensitive to departures from the ideal plan. single dropout could have altered conclusions
- Patient dropout increases the standard error.
- In the 13 Reviewed Studies, only 1 study reported having a subject drop out after initial assignment

Statistical Analysis and Sample Size

Concept:

- Repeated observations within a patient are **correlated**
- Analyses are more complex but powerful than parallel designs.
- The **patient**, not the individual measurement, is the unit of analysis.

Proper Analysis Approach

- Compare data **within patients** over time.
- Combine these comparisons **across patients**.

Statistical Techniques

- **Two observations per patient:** Paired t-test
- **Multiple observations per patient:** Multivariate regression or ANOVA for correlated data.

Statistical Analysis and Sample Size

- The above methods (Paired and Multivariate Regression/ANOVA)
  - Model correlation among measurements from single individual
  - Use this association in computing standard errors
  - Allow adjustment of P values for multiple tests on same series of measurements
  - Operate by linking together results of several paired t-tests

State of Practice in Reviewed Studies:

- 2 of 13 studies used no multivariate methods
- 11 employed multivariate methods with various degrees of sophistication
- ALL authors failed to exploit the data structure fully
- Billewicz reported similar failure in 9 of 20 medical literature examples from the medical literature
- Of the 13 studies, 12 studies used small number of patients (4 to 22)

Comparing Parallel Design Adjusted for Sex with a Crossover Design

Crossover vs. Parallel Design Adjusted for Sex

Table 2. Urinary Excretion of Urea Nitrogen in Four Diabetic Patients.*

Patient No.	Treatment IS	Treatment ISG	Difference †
	g of urea nitrogen/24 hr		
1	14	17	3
2	6	8	2
3	7	11	4
4	6	9	3
Mean	8.25	11.25	3.00
S.E.M.	1.90	2.00	0.40

*Data are adapted from Raskin and Unger.[6] IS denotes intravenous insulin and somatostatin, and ISG intravenous insulin, somatostatin, and glucagon.

†The S.E. of the difference between the means of ISG and IS: $2.76 = \sqrt{(1.90)^2 + (2.00)^2}$, if the groups were unpaired.

Table 3. Variation on the Same Study Shown in Table 2, with a Hypothetical Sex Variable.*

	Treatment IS	Treatment ISG	Mean	Sex Effect
		g of urea nitrogen/24 hr		
Men	14 7	11 17	12.25	2.5
Women	6 6	9 8	7.25	−2.5
Mean	8.25	11.25	9.75	
Effect	−1.5	1.5		

*Response = base-line value + (treatment effect) + (sex effect) + residual; residuals (not shown above) are the deviations between observed responses and those predicted by the model. They make the right-hand side equal the left-hand side in the general formula and numerical examples.

Examples:	Base Line		Treatment		Sex		Residual
14 =	9.75	−	1.5	+	2.5	+	3.25
8 =	9.75	+	1.5	−	2.5	−	0.75

Table 2: Urea Nitrogen Excretion (Crossover)

Table 3: Parallel Design Adjusted for Sex

- In Table 3, men excrete 5.0g more urea nitrogen per 24 hours than women.
- Mean treatment difference remains 3.0g/24hr (same as Table 2).
- Including "sex" explains some patient-to-patient variation in the parallel design.

Analytical Insights and Covariance Adjustment

ANOVA Results (Parallel Design):

- Urea nitrogen explained by:
  - Sum of baseline values
  - Effect of treatment (Insulin + Somatostatin vs. + Glucagon)
  - Effect of sex (male vs. female)
  - Residual (deviation between the observed and predicted.)
- SE = 2.09 (24% reduction from 2.76 in the previous parallel design)
- $t = 1.44$, df $= 5$, $p = 0.20$ (vs. $t = 1.08$, df $= 6$, $p = 0.32$)
- Paying one degree of freedom for sex makes the experiment more precise despite non-significance.

Covariance Adjustment:

- Adjusting for sex = covariance adjustment.
- Parallel design with sex adjustment needs $\approx 11\times$ crossover sample size (vs. $14\times$ unadjusted).
- Must be interpreted carefully expecially if characteristics was discovered through an exhaustive search.

Authors' Position and Recommendations

Authors' Position:

> *"In choosing between a parallel and a crossover design, the burden of proof should rest on those favoring the crossover design to demonstrate clear improvement over the parallel design."*

Evidence Supporting Crossover Success:

- Prior studies show absence of carry-over effects.
- Low dropout rate.
- Stable disease process.

Authors' Recommendation: Design choice should be guided by:

- Research goals, and the Disease Process.
- Trade-off between statistical power and study fragility.
- These considerations requires collaborative effort among Clinical, Laboratory, and Statistical Scientist

Other Studies Not Classified as Crossover Designs

Self-Controlled Studies Without True Crossover

- Self-controlled studies: A single treatment is evaluated by comparing a patient's status **before and after treatment**.
- A study where patients take a treatment and then stop (measuring "on treatment" vs "off treatment" readings) would not be classified as a crossover unless "no treatment" is a realistic clinical alternative.

Example: Blood Pressure Study

- Week 1–4: On medication
- Week 5–8: Off medication

Note: *This is only a crossover if "no treatment i.e. off medication" is a legitimate clinical option. Otherwise, this is a self-controlled study.*

Patients as Their Own Control

Concept:

- Patients serve as their own control — measurements taken before, during, or after treatment are compared within the same subject.
- Common in early-phase or exploratory studies.
- Shares some similarities with crossover designs but introduces new challenges.

Examples of Self-Controlled Studies:

1. **Peck et al.** – 13-cis-retinoic acid for severe acne
   - Measured changes over time where spontaneous recovery was unlikely.
2. **Hypertension screening study**
   - Compared absenteeism before and after diagnosis.
3. **Prager et al.** – Nitroprusside withdrawal in heart failure
   - Self-controlled design ideal to test withdrawal effects.

Key Point: Many crossover and parallel studies include self-controlled elements, but new design issues arise.

Critical Issues in Self-Controlled Studies

Inherited Issues:

- All five critical issues from crossover designs apply fully.

New Issues Specific to Self-Controlled Studies:

1. Lack of symmetry
2. Nature of the problem studied
3. Study of patients with refractory disease
4. Absence of direct comparisons

Implication:

- These challenges can introduce bias, limit interpretability, and reduce external validity.

Lack of Symmetry and Nature of the Problem

1. **Lack of Symmetry:**
   - Observations during and after treatment differ in duration, intensity, or decision rules.
   - Leads to substantial bias (unequal opportunities to observe disease developments).
   - Well-designed crossover studies avoid this via treatment balancing.
2. **Nature of the Problem Studied:**
   - Often used early in clinical development of new treatments focusing on multiple lab measures rather than one or two measures.
   - Example: Dichloroacetate study — measured glucose, lactate, and biochemical indices before/after treatment.
   - Requires multivariate analysis due to multiple correlated outcomes.
   - Increases risk of multiple comparison issues.
   - Expert statistical help for multivariate method rarely used for all 28 studies.

Patients with Refractory Disease

3. Study of Patients with Refractory Disease:

- **Refractory Disease:** Condition that does not respond to standard therapies, responded inadequately or became resistant over time.
- Many self-controlled studies focused on patients with refractory disease (at least 7 of 28 studies).
- Often due to early-stage human trials and ethical considerations in patient enrollment.
    - Mithramycin for Paget's disease: patients responded poorly to prior treatments
    - 13-cis-retinoic acid for acne: minimal response to antibiotics, vitamin A, benzoyl peroxide, x-rays and other acne therapies.

Problems:

- Incomplete washout
- Regression toward the mean
- Nonresponders bias

Problems in the study of Patients with Refractory Disease

- **Incomplete washout (No evidence found in the 28 studies):**
  - Clinical urgency may prevent full recovery (due to intense, prolonged treatment) before next phase.
  - Not clear if any of the investigators looked for such evidence
- **Regression toward the mean:**
  - **Definition:**Tendency of an extreme value when it is remeasured to be closer to the mean, because the original value was likely to be have been unduly influence by random variation.
  - Patients often enrolled when their condition is unusually poor (standard treatment failing).
  - Subsequent improvement may occur naturally, not because of the new treatment.
  - **Addressing Regression to the Mean:**
    - Use first measurements after washout to establish eligibility
    - Use second measurements after "stabilizing" period for baseline
- **Nonresponders bias:** Patients may have atypical disease or poor general health.

Absence of Direct Comparisons

4. **Absence of Direct Comparisons:**
   - Self-controlled studies do not provide a direct comparison to standard therapies — results must be combined with other studies.
   - **Challenges in combining results:**
     - Differences in study design, patient demographics, and disease severity.
     - Use of subjective measures or inconsistent diagnostic criteria and outcome variables.
   - Adjustments for these differences are often uncertain or inadequate.

Authors' Conclusion:

> *"Despite the clear value of self-controlled studies in the initial investigation of new treatments, one must usually use a randomized controlled trial or another powerful design to determine with assurance whether the new treatment should be recommended for general use."*

Questions?