Replicated N-of-1 RCTs for Rare Diseases

Patrick Onghena
Faculty of Psychology and Educational Sciences
KU Leuven, Belgium

Webinar 30 June 2023
DETERMINING OPTIMAL THERAPY — RANDOMIZED TRIALS IN INDIVIDUAL PATIENTS


Abstract Although the treatment of an individual patient in routine clinical practice has been likened to an experiment, the method is so susceptible to bias that we have come to demand multi-patient, double-blind, randomized controlled trials on matters of efficacy. Unfortunately, such trials have not or cannot be carried out for many clinical disorders; even when they have been executed their results may be difficult to extrapolate to individual patients.

To resolve this problem, we have begun to use double-blind randomized trials in which a single patient undergoes a series of pairs of treatments, consisting of one active and one placebo or alternative treatment per pair, with the order determined by random allocation. Appropriate treatment targets (signs, symptoms, or laboratory tests) are used as the measure of efficacy, and the trial is continued until efficacy is established or disproved. We describe such a trial, which resulted in a dramatically beneficial modification of treatment in a patient with partially reversible airflow limitation. We have established a clinical service that facilitates the widespread use of the method in our community. (N Engl J Med 1986; 314:889-92.)
SPECIAL ARTICLE

DETERMINING OPTIMAL THERAPY — RANDOMIZED TRIALS IN INDIVIDUAL PATIENTS


Abstract Although the treatment of an individual patient in routine clinical practice has been likened to an experiment, the method is so susceptible to bias that we have come to demand multi-patient, double-blind, randomized controlled trials on matters of efficacy. Unfortunately, such trials have not or cannot be carried out for many clinical disorders; even when they have been executed their results may be difficult to extrapolate to individual patients.

To resolve this problem, we have begun to use double-blind randomized trials in which a single patient undergoes a series of pairs of treatments, consisting of one active and one placebo or alternative treatment per pair, with the order determined by random allocation. Appropriate treatment targets (signs, symptoms, or laboratory tests) are used as the measure of efficacy, and the trial is continued until efficacy is established or disproved. We describe such a trial, which resulted in a dramatically beneficial modification of treatment in a patient with partially reversible airflow limitation. We have established a clinical service that facilitates the widespread use of the method in our community. (N Engl J Med 1986; 314:889-92.)
SPECIAL ARTICLE

DETERMINING OPTIMAL THERAPY — RANDOMIZED TRIALS IN INDIVIDUAL PATIENTS


Abstract Although the treatment of an individual patient in routine clinical practice has been likened to an experiment, the method is so susceptible to bias that we have come to demand multi-patient, double-blind, randomized controlled trials on matters of efficacy. Unfortunately, such trials have not or cannot be carried out for many clinical disorders; even when they have been executed their results may be difficult to extrapolate to individual patients.

To resolve this problem, we have begun to use double-blind randomized trials in which a single patient undergoes a series of pairs of treatments, consisting of one active and one placebo or alternative treatment per pair, with the order determined by random allocation. Appropriate treatment targets (signs, symptoms, or laboratory tests) are used as the measure of efficacy, and the trial is continued until efficacy is established or disproved. We describe such a trial, which resulted in a dramatically beneficial modification of treatment in a patient with partially reversible airflow limitation. We have established a clinical service that facilitates the widespread use of the method in our community. (N Engl J Med 1986; 314:889-92.)
SPECIAL ARTICLE

DETERMINING OPTIMAL THERAPY — RANDOMIZED TRIALS IN INDIVIDUAL PATIENTS


Abstract Although the treatment of an individual patient in routine clinical practice has been likened to an experiment, the method is so susceptible to bias that we have come to demand multi-patient, double-blind, randomized controlled trials on matters of efficacy. Unfortunately, such trials have not or cannot be carried out for many clinical disorders; even when they have been executed their results may be difficult to extrapolate to individual patients.

To resolve this problem, we have begun to use double-blind randomized trials in which a single patient undergoes a series of pairs of treatments, consisting of one active and one placebo or alternative treatment per pair, with the order determined by random allocation. Appropriate treatment targets (signs, symptoms, or laboratory tests) are used as the measure of efficacy, and the trial is continued until efficacy is established or disproved. We describe such a trial, which resulted in a dramatically beneficial modification of treatment in a patient with partially reversible airflow limitation. We have established a clinical service that facilitates the widespread use of the method in our community. (N Engl J Med 1986; 314:889-92.)
SPECIAL ARTICLE

DETERMINING OPTIMAL THERAPY — RANDOMIZED TRIALS IN INDIVIDUAL PATIENTS


Abstract Although the treatment of an individual patient in routine clinical practice has been likened to an experiment, the method is so susceptible to bias that we have come to demand multi-patient, double-blind, randomized controlled trials on matters of efficacy. Unfortunately, such trials have not or cannot be carried out for many clinical disorders; even when they have been executed their results may be difficult to extrapolate to individual patients.

To resolve this problem, we have begun to use double-blind randomized trials in which a single patient undergoes a series of pairs of treatments, consisting of one active and one placebo or alternative treatment per pair, with the order determined by random allocation. Appropriate treatment targets (signs, symptoms, or laboratory tests) are used as the measure of efficacy, and the trial is continued until efficacy is established or disproved. We describe such a trial, which resulted in a dramatically beneficial modification of treatment in a patient with partially reversible airflow limitation. We have established a clinical service that facilitates the widespread use of the method in our community. (N Engl J Med 1986; 314:889-92.)
1. The Guyatt et al. (1986) example

2. What is an N-of-1 trial? *Definition*

3. Importance for health and life sciences: In general and for rare diseases in particular

4. Validity and methodological quality of N-of-1 RCTs

5. Data analysis in N-of-1 RCTs
1. The Guyatt et al. (1986) example

- Patient: 65-year, male
- Diagnose: Uncontrolled asthma
- Current treatment:
  - Albuterol (2 puffs 4 times a day)
  - Theophylline (300 mg by mouth 3 times a day)
  - Ipratropium bromide (2 puffs 4 times a day)
  - Prednisone (25 mg alternating with 10 mg daily)
1. The Guyatt et al. (1986) example

- Patient: 65-year, male
- Diagnose: Uncontrolled asthma
- Current treatment:
  - Albuterol (2 puffs 4 times a day)
  - Theophylline (300 mg by mouth 3 times a day)
  - Ipratropium bromide (2 puffs 4 times a day)
  - Prednisone (25 mg alternating with 10 mg daily)
1. The Guyatt et al. (1986) example

- Double-blind N-of-1 RCT of theophylline versus placebo
- Randomized Block Design: pairs of both treatment in random order
- Treatment periods of 10 days
- At the end of each 10-day period: 7-point scale
  7 = optimal function, 1 = severe symptoms
  - Shortness of breath on (1) bending, (2) hurrying and (3) climbing stairs
  - Perceived need for inhaler during the day
  - Extent to which breathlessness disturbed his sleep
<table>
<thead>
<tr>
<th></th>
<th>score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortness of breath</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Need for inhaler</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

*The patient rated his symptoms on a 7-point scale in which 7 represented optimal function and 1 represented severe symptoms.*
Table 1. An N of 1 Randomized Controlled Trial of Theophylline.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Pair 1</th>
<th></th>
<th></th>
<th>Pair 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Period 1</td>
<td>Period 2</td>
<td></td>
<td>Period 1</td>
<td>Period 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Drug)</td>
<td>(Placebo)</td>
<td></td>
<td>(Drug)</td>
<td>(Placebo)</td>
<td></td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>3</td>
<td>6</td>
<td></td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7</td>
<td></td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Need for inhaler</td>
<td>3</td>
<td>5.5</td>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>5</td>
<td>5.5</td>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

score*

*The patient rated his symptoms on a 7-point scale in which 7 represented optimal function and 1 represented severe symptoms.
1. The Guyatt et al. (1986) example

- Double-blind N-of-1 RCT of ipratropium versus placebo
- Randomized Block Design: pairs of both treatment in random order
- Treatment periods of 10 days
- At the end of each 10-day period: 7-point scale
  - 7 = optimal function, 1 = severe symptoms
    - Shortness of breath on (1) bending, (2) hurrying and (3) climbing stairs
    - Perceived need for inhaler during the day
    - Extent to which breathlessness disturbed his sleep
- 3 repeated ratings during each period
<table>
<thead>
<tr>
<th></th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortness of breath</td>
<td>6 6 6  6 6 6  6 6 6  5 5 5</td>
</tr>
<tr>
<td></td>
<td>6 6 6  6 6 6  6 5 5  5 5 5</td>
</tr>
<tr>
<td></td>
<td>4 5 4  5 5 5  5 4 5  4 5 5</td>
</tr>
<tr>
<td>Need for inhaler</td>
<td>4 4 4  5 5 5  5 5 5  4 4 4</td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>6 5 5  5 7 7  7 7 7  4 4 4</td>
</tr>
</tbody>
</table>
Table 2. An N of 1 Randomized Controlled Trial of Ipratropium.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Pair 1</th>
<th></th>
<th>Pair 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Period 1 (Placebo)</td>
<td>Period 2 (Drug)</td>
<td>Period 1 (Drug)</td>
<td>Period 2 (Placebo)</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Need for inhaler</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>
2. What is an N-of-1 trial? *Definition*

N-of-1 trial = a prospective, multiple crossover trial in a single patient
N-of-1 RCT = N-of-1 trial + randomization of the treatment sequence
Replicated N-of-1 RCTs = N-of-1 RCTs + replication across patients

≠ Case studies, case series, case reports, observational time series studies
= A specific design in a broader family of single-case experimental designs
(Tate et al., 2016)
Figure 1. Simplified design of probiotics in fibromyalgia N-of-1 trial with timeline. A—denotes active supplementation; B—denotes placebo supplementation.

(Bradbury et al., 2020)
3. Importance for health and life sciences: In general and for rare diseases in particular

- RCTs can answer the clinical research question: What works?
- Large-scale group-comparison RCTs: What works on average?
  ≠ What works in general?
  ≠ What works for the majority of patients?
- N-of-1 RCTs can answer the clinical research question:

  What works for this particular patient?
N-of-1 trials for personalized treatment

THE CASE OF MUSCLE CHANNELOPATHIES

BAS C. STUNNENBERG

Building an evidence-base for the treatment of rare diseases

Rare diseases constitute a heterogeneous group of over 6,000 disorders with a prevalence of <1 per 2,000 per disease. In Europe, 30 million patients (6 to 8% of the population) are affected with one of these rare diseases. Since neurological symptoms are present in about 75% of rare diseases, neurologists are familiar with the difficulties in determining the optimal therapy in patients with a rare disease while facing the paradox of evidence-based medicine (EBM) (see textbox 1).

International regulatory authorities such as the Food and Drug Administration (FDA) and European Medical Agency (EMA) accept that it is unreasonable to demand the standard level of evidence (level 1) of multiple Randomized Controlled Trials (RCTs) in building an evidence-base for treatment of rare diseases.[6] The ability to conduct RCTs in rare diseases is hampered by low numbers of patients and large clinical heterogeneity. However, relying simply on case reports or case series incurs a considerable risk of selection and ascertainment bias. Currently, it is unclear which concessions can be accepted towards the level 1 evidence needed for registration and coverage decisions in case of rare diseases.[2]

Textbox 1: The paradox of evidence-based medicine (EBM)

The paradox of EBM, first described by Guyatt et al. in 1996 in the New England Journal of Medicine, describes how physicians struggle to determine the optimal therapy for an individual patient in a scientific fashion while dealing with the evidence gap between clinical care and science. Physicians cannot trust their own uncontrolled therapeutic trials, but neither can they often look to large-scale randomized trials for definitive treatment recommendation. Guyatt and colleagues explain that, on the other hand, the uncontrolled therapeutic trials in clinical practice (where a treatment is provided for a certain time while the effect is estimated based on the subjective recollection of a response by the patient, sometimes supported by changes in physical examination or ancillary tests) are attributable to all kinds of sources of bias, such as disease fluctuations and the placebo (or nocebo) effect. But, on the other hand, the RCT, that deals with these important sources of bias by introduction of randomization and a placebo, only provides evidence on the effectiveness of a drug on a population level, i.e. an estimate of treatment effectiveness for the fictive 'average' patient. For numerous reasons, extrapolation of RCT trial results to inform treatment decisions in an individual patient in clinical practice can be inappropriate (e.g., often the patient does not match the trials' inclusion criteria because of complex comorbidity or co-medication). Finally, negative result from an RCT do not discard the possibility of some patients actually benefiting from the treatment (or vice versa).
Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders

The Power of 1

Annelieke R. Müller, MSc, Marion M.M.G. Brands, MD, PhD, Peter M. van de Ven, PhD, Kit C.B. Roes, PhD, Martina C. Cornel, MD, PhD, Clara D.M. van Karnebeek, MD, PhD, Frits A. Wijburg, MD, PhD, Joost G. Daams, MA, Erik Boot, MD, PhD, and Agnies M. van Eeghen, MD, PhD

Correspondence
Dr. van Eeghen
a.m.vaneeghen@amsterdamumc.nl

Neurology® 2021;96:529-540. doi:10.1212/WNL.0000000000011597
4. Validity and methodological quality of N-of-1 RCTs
Level 1: Systematic review of randomized trials or n-of-1 trials

Level 2: Randomized trial or observational study with dramatic effect

Level 3: Nonrandomized controlled cohort/follow-up study

Level 4: Case-series, case-control studies, or historically controlled studies

Level 5: Mechanism-based reasoning
ICN MISSION STATEMENT AND OBJECTIVES

Vision: A world where personalised clinical studies (for both single and groups of individuals) are an integral part of clinical practice and health research.

Mission: To promote, support and advance the use of personalised clinical studies, and to share relevant knowledge, experience, expertise, resources, and data through our global network.
Symposium: Small is beautiful {once more}

Small is beautiful
{once more}

The third international N=1 Symposium
April 2023

Within-person research has exponentially increased in recent years. Clinicians and researchers alike are benefiting from the flexibility and potential of research methods focusing on the individual. The result of such staggering interest has culminated in the current symposium, which strives to bring together experts and novices in this growing field of research. We aim to continue expanding awareness, knowledge, and expertise of this increasingly prominent research methodology.
CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement

Sunita Vohra a,*, Larissa Shamseer b, Margaret Sampson c, Cecilia Bukutu d, Christopher H. Schmid e, Robyn Tate f, Jane Nikles g, Deborah R. Zucker h, Richard Kravitz i, Gordon Guyatt j, Douglas G. Altman k, David Moher b, the CENT Group

SPIRIT extension and elaboration for n-of-1 trials: SPENT 2019 checklist

Antony J Porcino,1 Larissa Shamseer,2 An-Wen Chan,3,4 Richard L Kravitz,5 Aaron Orkin,6,7 Salima Punja,8 Philippe Ravaud,9,10,11 Christopher H Schmid,12 Sunita Vohra,8,13

on behalf of the SPENT group

For numbered affiliations see end of the article.
Correspondence to: SVohra
svohra@ualberta.ca
(ORCID 0000-0002-6210-7933)
Additional material is published online only. To view please visit the journal online.
Cite this as: BMJ 2020;368:m122
http://dx.doi.org/10.1136/bmj.m122
Accepted: 11 December 2019
The reporting quality of N-of-1 trials and protocols still needs improvement

Zhipeng Wei¹,²,³,⁴  |  Xiajing Chu¹,²,³,⁴  |  Jiani Han¹,²,³,⁴  |  Na Zhang¹,²,³,⁴  
Yanfei Li¹,²,³,⁴  |  Chaoqun Yang¹,²,³,⁴  |  Qi Wang⁵,⁶,⁷  |  Jiang Li¹,⁴,⁸  
Ahmed Atef Belal⁵,⁶,⁷  |  Peijing Yan⁹  |  Xiuxia Li¹,²,³,⁴  |  Kehu Yang¹,²,³,⁴
5. Data analysis in N-of-1 RCTs

- Graphical Data Analysis
- Descriptive Statistics
- Inferential Statistics
Visual inspection alone: only for clinical use

“The t-test is routinely used for N-of-1 RCTs, and is universally included in statistical packages.” (p. 1289)
N-of-1 Randomized Trials

Reza D. Mirza, Sunita Vohra, Richard Kravitz, and Gordon H. Guyatt

© Springer Nature Switzerland AG 2022
S. Piantadosi, C. L. Meinert (eds.), Principles and Practice of Clinical Trials,
https://doi.org/10.1007/978-3-319-52636-2_97

Fig. 1 N-of-1 RCT results: Mean daily Likert score

Fig. 2 N-of-1 RCT mean period score
Fig. 3  N-of-1 RCT treatment and placebo difference scores

Fig. 4  N-of-1 RCT t-test results

N of 1 RCT - Ms. A.D.

- **Targets** (data--symptom means)
  - Active:
    - Pair 1: 5.00
    - Pair 2: 5.095
    - Pair 3: 4.62
    - Pair 4: 4.38
  - Placebo:
    - Pair 1: 3.56
    - Pair 2: 1.98
    - Pair 3: 2.83
    - Pair 4: 2.83
  - Diff.:
    - Pair 1: 1.44
    - Pair 2: 3.18
    - Pair 3: 1.79
    - Pair 4: 1.55

- **Analysis** (2 tailed paired t-test)

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>D</th>
<th>t</th>
<th>P</th>
<th>C.I. (90%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.99</td>
<td>4.94</td>
<td>0.016</td>
<td>(1.041, 2.937)</td>
</tr>
</tbody>
</table>
Problems with the routine use of parametric $t$-tests in the analysis of N-of-1 RCT data

1. The pairs are not independent
2. The distributional assumptions of the test are implausible
3. The variability within a period is ignored
4. Missing data are ignored
5. Optional stopping requires additional Type I error rate control
Inferential data-analysis (Onghena et al., 2018, 2020)

- What is the statistical inference about?
  - Population = one particular patient
  - Sample = the repeated measures
  - Causal inference = demonstrations of a cause-and-effect relation for that specific patient

- Which statistical model?
  - Segmented linear and nonlinear regression models
  - Interrupted time series models – Borckard’s Simulation Modeling Analysis
  - Multilevel models – Meta-analysis

- Which inferential procedure / logic?
  - Ordinary least squares and maximum likelihood criteria
  - Design-based – Randomization-based inference
  - Bayesian inference
Inferential data-analysis (Onghena et al., 2018, 2020)

• What is the statistical inference about?
  o Population = one particular patient
  o Sample = the repeated measures
  o Causal inference = demonstrations of a cause-and-effect relation for that specific patient

• Which statistical model?
  o Segmented linear and nonlinear regression models
  o Interrupted time series models – Borckard’s Simulation Modeling Analysis
  o Multilevel models – Meta-analysis

• Which inferential procedure / logic?
  o Ordinary least squares and maximum likelihood criteria
  o Design-based – Randomization-based inference
  o Bayesian inference
### Targets
(data--symptom means)

<table>
<thead>
<tr>
<th></th>
<th>Pair 1</th>
<th>Pair 2</th>
<th>Pair 3</th>
<th>Pair 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>5.00</td>
<td>5.095</td>
<td>4.62</td>
<td>4.38</td>
</tr>
<tr>
<td>Placebo</td>
<td>3.56</td>
<td>1.98</td>
<td>2.83</td>
<td>2.83</td>
</tr>
<tr>
<td>Diff.</td>
<td>1.44</td>
<td>3.18</td>
<td>1.79</td>
<td>1.55</td>
</tr>
</tbody>
</table>

B1 Active 5
B1 Placebo 3.56
B2 Placebo 1.98
B2 Active 5.095
B3 Active 4.62
B3 Placebo 2.83
B4 Active 4.38
B4 Placebo 2.83

\[ t = 5.10 \]
<table>
<thead>
<tr>
<th>Group</th>
<th>Condition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>Active</td>
<td>5</td>
</tr>
<tr>
<td>B1</td>
<td>Placebo</td>
<td>3.56</td>
</tr>
<tr>
<td>B2</td>
<td>Placebo</td>
<td>1.98</td>
</tr>
<tr>
<td>B2</td>
<td>Active</td>
<td>5.095</td>
</tr>
<tr>
<td>B3</td>
<td>Active</td>
<td>4.62</td>
</tr>
<tr>
<td>B3</td>
<td>Placebo</td>
<td>2.83</td>
</tr>
<tr>
<td>B4</td>
<td>Active</td>
<td>4.38</td>
</tr>
<tr>
<td>B4</td>
<td>Placebo</td>
<td>2.83</td>
</tr>
</tbody>
</table>

**t_{\text{OBS}} = 5.10**

<table>
<thead>
<tr>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>t8</th>
<th>t9</th>
<th>t10</th>
<th>t11</th>
<th>t12</th>
<th>t13</th>
<th>t14</th>
<th>t15</th>
</tr>
</thead>
</table>
\[ P = \frac{\text{Number of test statistic values that are equal to, or more extreme, than the observed value}}{\text{Total number of test statistic values}} \]

\[ P = \frac{2}{16} \]

\[ P = 0.125 \]
> oneway_test(V3 ~ V2 | V1 , alternative='two.sided', distribution='exact',
+   data=Dataset)

**Exact Two-Sample Fisher-Pitman Permutation Test**

data:  V3 by V2 (Active, Placebo)
stratified by V1
Z = 1.8936, p-value = 0.125
alternative hypothesis: true mu is not equal to 0
Example: $p_1 = .30, p_2 = .20 \rightarrow S = .50$

$P(S \leq .50)$?

Under $H_0$: Uniform distribution

$P(S \leq .50) = (.50)^2/2 = .125$

$P(S \leq S_{obs}) = \frac{(S_{obs})^2}{2}$

(as long as the observed sum is not larger than 1)
Example: \( p_1 = .55, p_2 = .95 \rightarrow S = 1.50 \)

\[ P(S \leq 1.50) \]

Under \( H_0 \): Uniform distribution

\[ P(S \leq 1.50) = \frac{(1.50)^2}{2} - (2) \frac{(0.50)^2}{2} \]

\[ = 0.875 \]

\[ P(S \leq S_{obs}) = \frac{(S_{obs})^2}{2} - (S_{obs} - 1)^2 \]
\[ P(S \leq S_{obs}) = \sum_{k=0}^{\tilde{S}} (-1)^k \binom{n}{k} \frac{(S_{obs} - k)^n}{n!} , \]

with \( n \) = the number of \( P \) – values to be combined, and \( k \) = a counter up to the largest integer smaller than the observed sum \( \tilde{S} = \max (k < S_{obs}) \).

\[ .30, .30, .20, .20 \to .0417 \]

(Edgington, 1972; Onghena & Edgington, 2005)
Select the design type
- ABAB Phase Design

X-axis label
- Measurement Times

Y-axis label
- Scores

A1 phase label
- A1

B1 phase label
- B1

A2 phase label
- A2

B2 phase label
- B2

Y-axis minimum

Y-axis maximum

Plot
1. The randomization tests do not assume independent data
2. The randomization tests are distribution-free
3. Variability within a period may be included by using other designs
4. Missing data are taken into account (even MNAR)
5. Optional stopping *not yet included*
1. Replicated N-of-1 RCTs have a long history, but only recently have been gaining popularity in the health sciences.

2. Replicated N-of-1 RCTs are appealing for research on rare diseases because of their feasibility and because of their validity to test treatment effects at the individual level.

3. Routine statistical analysis of N-of-1 RCT data needs to be improved.

4. We need more user-friendly statistical tools and an effort in statistics education to move beyond the parametric t-test.
Thank you!

@PatrickOnghena@mastodon.nl

@PatrickOnghena

patrick.onghena@kuleuven.be
References and further reading


