Last update: 1/1/2025

1. Expert volunteers: Ronny Kohavi, Jakub Linowski, Lukas Vermeer, and the Center for Open Science. We provide our time, free of charge, to help design, execute, and analyze the experiment, in exchange for being able to share results that are approved by the companies online and in papers.
2. We are interested in companies that are willing to participate in A/B testing the patterns below. If you are interested, please contact one of us (e.g., Ronny Kohavi) over LinkedIn.
3. We have started an A/B Throwdown on 18 Dec 2024 and we are asking the community to make predictions about the first two patterns, which are now running live on two sites in Norway. If the treatment effect is statistically significant (p-value <0.05), then the person whose prediction is the closest will get to name a charity related to experimentation that will be awarded $1,000. See details and links to the prediction form in https://bit.ly/ABThrowdown .
    1/1/2025: Results from the first pattern are in: see https://bit.ly/ABThrowdown1ResultsPost
4. Benefits to companies
    a. Help in executing these A/B tests by the above named experts.
        Feedback and support from experts on design and analysis of experiments.
    b. Early access to results. Companies that share results will get access to the other results being analyzed. A paper is likely to take months to publish.
    c. Improved brand. The participating companies will be acknowledged while the experiments are happening and one person from each company will be a co-author on the paper we will draft and submit. All participating companies will be acknowledged.
5. Requirements from participating companies.
    a. The expert volunteers will each sign a mutual Non-Disclosure Agreement (NDA) that prohibits sharing any information.
    b. A separate specific disclosure document will be signed, where a "Draft Result" will be proposed, and the companies can approve it, or propose revisions, or terminate the agreement. The key paragraphs from the disclosure agreement are provided below.
6. We are focused on replicating six patterns below (last one added 12/24/2024)
    These are strong patterns that are relatively easy to introduce, so companies should be willing to try them and benefit from the analysis. Our goal is to understand the median relative treatment effects and their distribution, and perhaps conditions for when the effects are larger/smaller. The patterns are:
    a. Rounded/square corners – published result that some believe is highly exaggerated. See
        https://www.linkedin.com/posts/ronnyk_do-elements-with-rounded-shapes-enhance-click-through-activity-7183554027773734914--ugr

        Proposed MDE for power calculations on conversion metric: ~~2%.~~ Based on the results of the Coop experiment, we recommend an MDE lower than 0.56%, say 0.5%.
        This is a tough one to estimate, but several practitioners believe this is much closer to zero than to the 17%+ in the paper.

Relevant experiment by Evidoo:
https://www.linkedin.com/posts/jurjenjongejan_ecommerce-abtesting-activity-7275790931121901568-nXh7

b. Coupon code
Proposed MDE for power calculations on conversion metric: 2% for lowering prominence (70% of 2.8% below), 4% for removing field (70% of 7.8% would be 5.5%, but is high). The book https://experimentguide.com Chapter 2, which was based on a real experiment showed 2.8% and 7.8%, where the former was based on reducing prominence and the latter complete removal.
GoodUI has five tests, all small and borderline: 16K visits, p-value 0.07 negative 1.6% sales; 10K visits, p-value 0.29 positive 3.3% sales; 8K visits, p-value 0.01 positive 2.6% sales; 2K visits, p-value 0.61 positive 0.8% sales; 2K visits, p-value 0.21 positive 0.2.4% sales; unspecified VWO blog: positive 24% revenue.

https://www.evidoo.io/best-practices/136/ changed coupon to text (lower attraction) for mobile with the following statistics based on a total of 63K visitors:

  i. There were five A/B tests from electronics, B2C gaming
  ii. 40% win ratio, 40% loss ratio
  iii. Average treatment effect +4.0%

c. Page load time (performance)
Proposed MDE for power calculations: 1% for 250 msec slowdown (70% of 0.6%*250/100 msec) if done on all pages.
[Updated 3/2025] Lower MDE of 0.2% if done for a page (e.g., just search).

Detailed in Chapter 5 of https://experimentguide.com: 100 msec slowdown had 0.6% impact to revenue. Recording of talk:
https://vwo.com/events/convex-2019/sessions/test-user-experience-bing/

d. Sticky Call To Action

Proposed MDE: 2.5% (70% of the 3.6% claimed by Evidoo).
https://www.evidoo.io/best-practices/114/ . Evidoo claims:

  i. 14 A/B tests (from multiple domains, including Fashion, Furniture, Jewelry, Outdoor, Consumer Goods) with 57% win ratio, 7% loss (rest are flat).
  ii. Average treatment effect (purchase/transaction) is 3.6%, based on 3M users in total (over all experiments).

https://goodui.org/patterns/41/ shows 13 tests, but most underpowered.

https://goodui.org/patterns/41/tests/217/ had 80K visits with increase to engagement, tiny p-value below 0.001, but this was not conversions.

e. Authentic Product Photos
Proposed MDE: 4% (data is lacking here. 70% of the 5.6% below is 4%). Given the cost to implement something like this, you need a big boost to justify the effort.

Evidoo has https://www.evidoo.io/best-practices/300 for contextual imagery, and claims

    i.   2 A/B tests from Fashion & shoes
    ii.   100% win ratio, but only 233K users for two experiments on mobile
    iii.   Average treatment effect for mobile transactions was 13%

Evidoo has https://www.evidoo.io/best-practices/231 for AI models and claims

1. One A/B test from fashion & shoes
2. 100% win ratio (for one) with only 233K users
3. Average treatment effect for mobile transactions was 5.6%

https://goodui.org/patterns/30/ shows three tests, but the largest has just 12K visits and a p-value of 0.09

f. Open link in new tab
See
https://www.linkedin.com/posts/ronnyk_experimentguide-abtest-openinnewtab-activity-7272140955506339841-qGsb for a summary of results

    i.   MSN UK home page: 8.9% increase to clicks/user to users triggered by clicking on hotmail
    ii.   MSN US home page: replication run, similar to above
    iii.   MSN Search: +5% increase to clicks/user
    iv.   Airbnb search: 3-4% increase to bookings
    v.   Sunweb: +7.6%
    vi.   GoodUI/job site: +6.4% to successful applications
    vii.   GoodUI/VivaReal: +2.2% to form submissions

Proposed MDE: 2%
This is a trivial change and one of the most successful patterns we have ever seen.
It is a trivial change: one only needs to add: target="_blank" as explained in
https://www.w3schools.com/tags/att_a_target.asp .

7. FAQ (from meetings)
a. Does a participating company need to run all patterns, or can they choose?
They can choose. Given the overhead of joining the project, we would love to see companies running a few patterns, not just a single one.
b. What is the timeline?
We recognize that planning/designing/running/analyzing/certifying takes time.
We don't have a pre-set timeline, but we assume this will take several months.

This is our V1, and when we have sufficient replications for an interesting result to share, we might drop some planned runs, and perhaps do round 2.

c. Does disclosing sample size not reveal confidential information not otherwise available?
Because we are not specifying the experiment runtime (e.g., the experiment could have run for one week or eight weeks), the sample size provides very limited information.
In addition, we do not specify that the experiment was run on 100%, or was triggered to a smaller sample. For example, an A/B pattern could be tested as an A/B/C experiment, with A/B our goal pattern, but C is a related pattern that a company wants to test, so we're effectively running on 66% of traffic. All these things make estimates of traffic from these results imprecise.

d. Are there examples where companies published experiment results?
Microsoft (where Ronny Kohavi worked) published multiple experiments:

- https://bit.ly/HBR_AB
- https://bit.ly/expRulesOfThumb

Booking (where Lukas worked) published multiple experiments:

- https://www.ueo-workshop.com/wp-content/uploads/2014/04/noulasKDD2014.pdf
- https://dl.acm.org/doi/abs/10.1145/2766462.2776777
- https://dl.acm.org/doi/10.1145/3292500.3330744

GoodUI.org (headed by Jakub) has multiple experiments

- https://goodui.org

Vivli has 56 companies that pay to share their information about all clinical trials. See https://bit.ly/vivliRCTSharing

An earlier version of this doc is available here.

Appendix: Disclosure Agreement

Here are the two key paragraphs of the disclosure agreement.

WHEREAS, Kohavi and his colleagues, including Lukas Vermeer and Jakub Linowski, have each entered into a confidentiality, non-disclosure, or similar agreement with Company (the "**NDAs**") in connection with the Project and the development, implementation, evaluation, and other consulting related to certain A/B tests to be conducted by the Company (the "**Company A/B Tests**"); and

WHEREAS, Kohavi and Company desire to publish certain summary statistics, insights, and aggregated data generated or learned from the Company A/B Tests (the "**Published Results**"), subject to the terms and conditions of this Agreement.

NOW, THEREFORE, in consideration of the mutual promises and covenants herein contained, the Parties agree as follows:

1.   Publication Review. From time-to-time during the Term, Kohavi will provide to Company a proposed draft of the Published Results that Kohavi proposes be made publicly available ("**Draft Results**"). Within fourteen (14) days following its receipt of the Draft Results, Company will, acting in good faith, either: (i) approve publication of the Draft Results or (ii) propose revisions or other comments to the Draft Results. If Company approves the Draft Results, Kohavi and his designees may publish and otherwise publicly disclose the Draft Results and such Draft Results will be considered Published Results hereunder and available to the public for purposes of, and not otherwise subject to, the NDAs. If Company proposes revisions or comments to the Draft Results, Kohavi may revise the Draft Results and resubmit them for Company's review in accordance with the process set forth in this Section 1.

2.   Scope of Published Results. The Parties currently anticipate that the Published Results would include the following information, except to the extent such information is competitively sensitive or otherwise subject to a *bona fide* confidentiality concern:

   a. Company's identity as a participant in the Project;

   b. Company's identity as the source of the applicable Company A/B Test;

   c. Description of the applicable Company A/B Test with a level of detail sufficient to enable replication from such description by an outside observer (e.g., screenshots of control and treatments);

   d. For each relevant metric tested by the applicable Company A/B Test, the relative treatment effect (e.g., lift), confidence interval, p-value, sample size, and the statistical power with such metric's associated relative minimum detectable effect (MDE);

   e. Summary statistics related to trust in the applicable Company A/B Test, such as sample-ratio and caveats related to trust (e.g., logging issues); and

   f.   Subject to the Company's reasonable approval, such other notable or differentiated segments and the impact of such segments on relevant metrics tested by the applicable Company A/B Test.