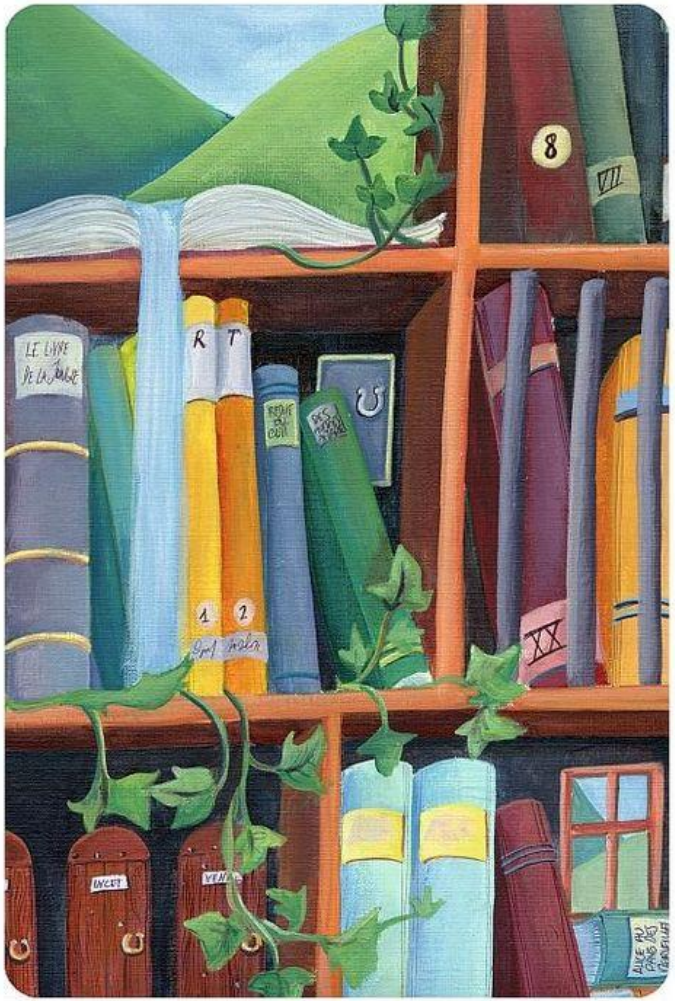
# Experiments in innovation support

Eszter Czibor
28 January 2022

# Agenda

- Intro to experiments in innovation support
- Pilots with Nesta Challenges
- INNOSUP trials
- Conclusions
- Resources
- Q&A



# Introduction

# Why support innovation?

**Innovation matters:**

- Driver of economic growth (Romer, 1990)
- Response to social & environmental challenges (Mazzucato, 2018)
- Distributional effects (Aghion et al. 2018)

**Rationale for policy intervention:**

- Market failures (Bloom et al, 2013, Williams, 2016)
- Inequality in benefits/costs and participation (Aghion et al. 2019, Cook, 2019)

# How to support innovation?

**Innovation policy levers:**

- Tax credits, research funding, R&D subsidies (Bloom et al. 2019)
- Support for innovative entrepreneurs + (local) innovation ecosystems (OECD, 2020)
- Intellectual property rights (Bloom et al. 2019)
- Education (Shambaugh et al, 2017)
- Immigration policies (Kerr, 2019)
- Antitrust / competition / trade policies (Federico et al, 2019)
- ...

→ *Open questions around how innovation works & what works to spur innovation*

# Why experiment?

***Experimental innovation policy*** (OECD, 2014):

- (Diagnostic) monitoring and evaluation, embedded at the design stage and throughout implementation
- Constant learning and adjustment

**Why *randomized* experiments?**

# Why experiment?

***Experimental innovation policy*** (OECD, 2014):

- (Diagnostic) monitoring and evaluation, embedded at the design stage and throughout implementation
- Constant learning and adjustment

**Randomized experiments**: replace *selection* into scheme with *random assignment*

- Constructing credible counterfactuals
  - Addressing selection bias
    - Who joins an accelerator
    - Who receives funding, ...
  - Additionality!
- Estimating/comparing returns on investment

# Embedding experiments in innovation support schemes

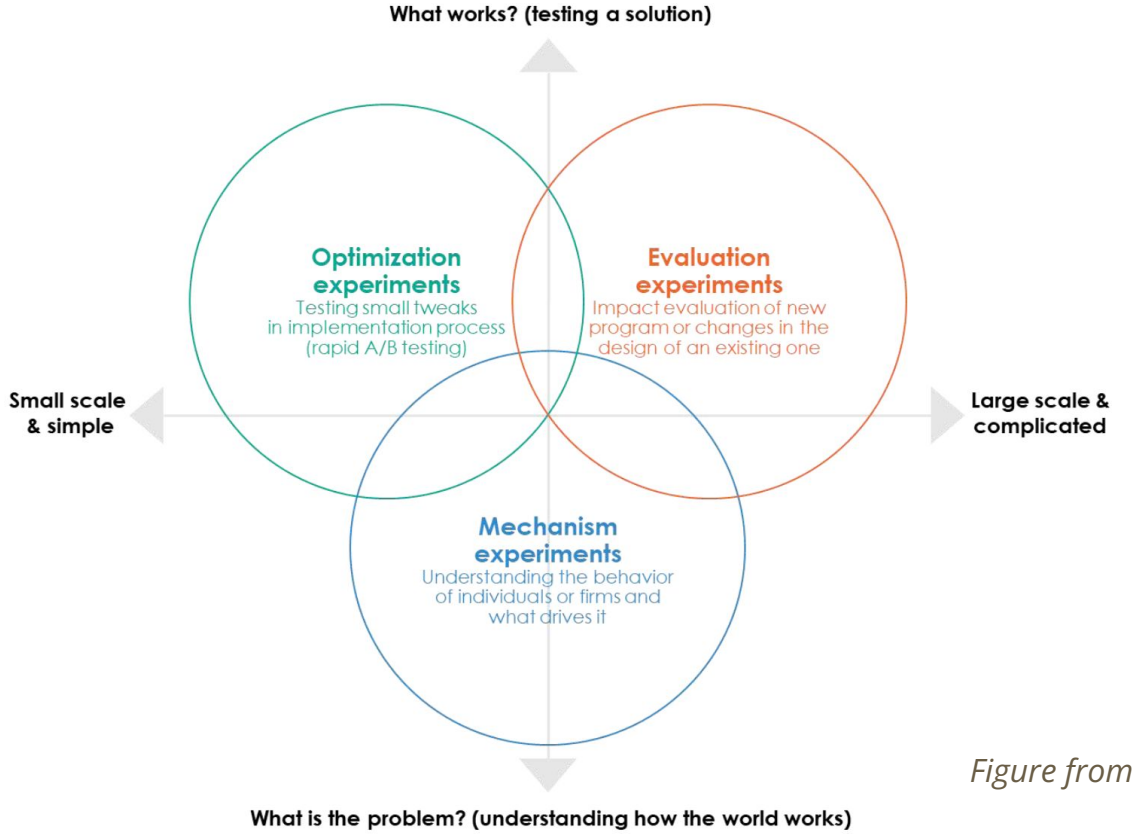



*Figure from Bravo-Biosca (2019)*

# Embedding experiments in innovation support schemes

**Mechanism:**

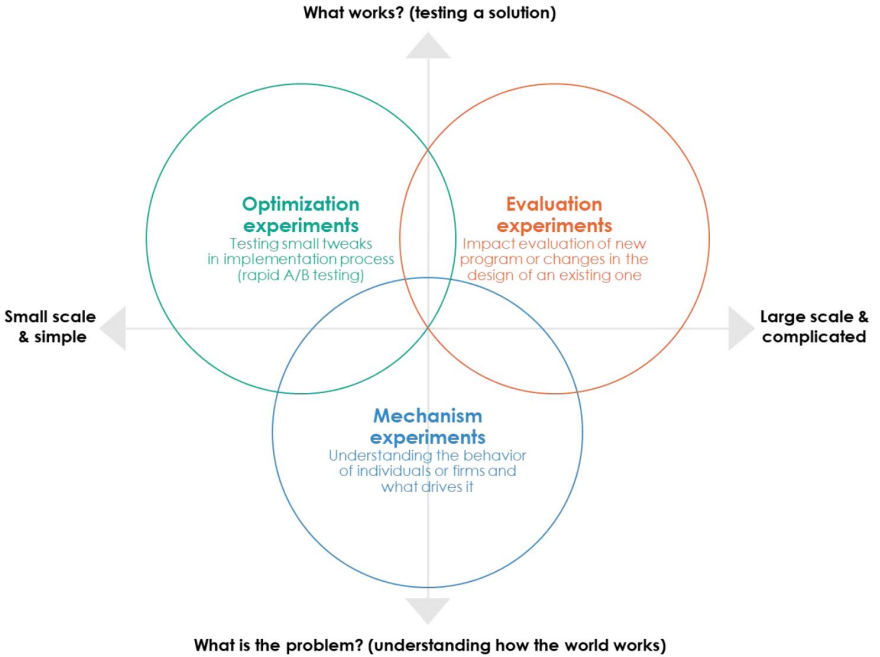
- Prize structure (Zivin & Lyons, 2021)
- Tournament size (Boudreau et al. 2016)
- Timing of disclosure (Boudreau et al. 2015)

**Optimization:**

- Framing of messaging (Guzman et al. 2020)
- Establishing connections (Boudreau et al. 2017)
- Details of assessment (Boudreau et al. 2016)

**Evaluation:**

- Content of support (Tobro et al. 2019)



# RCTs can tackle various innovation policy priorities

- **Increase innovative activity and output**
  - E.g. Innovation vouchers impact evaluations (UK, NL)
- **Broaden participation in innovation**
  - More diverse pool (e.g. exposure to role models in STEM and entrepreneurship)
  - More equitable selection process (Tomkins et al. 2017)
- **Steer “quality” and direction of innovation**
  - Riskiness & novelty (Nane et al. 2021)
  - Who benefits / who is harmed? (Koning et al. 2021)

# Case study 1: Pilots with Nesta Challenges

# Context

## Context:

- Assessment of two *challenge prizes* run by Nesta Challenges in 2020 and 2021
  - Competitions offering a reward for the first/best solution to a (social) innovation problem

## Sample:

- 2020 – first judging round:
  - 60 proposals
  - 12 evaluators
  - 4 evaluators/proposal
- 2021 – first sift:
  - 148 proposals
  - 18 evaluators
  - 2 evaluators/proposal

# Research question & design

## Goal:

- Explore potential gender bias in funding decisions (Witteman et al, 2019)
- Test for **gender-based favoritism** in evaluation
  - Teaching evaluations (Boring, 2017, Mengel et al. 2019)
  - Hiring committees (Bagues & Esteve-Volart, 2010, Bagues et al, 2017)

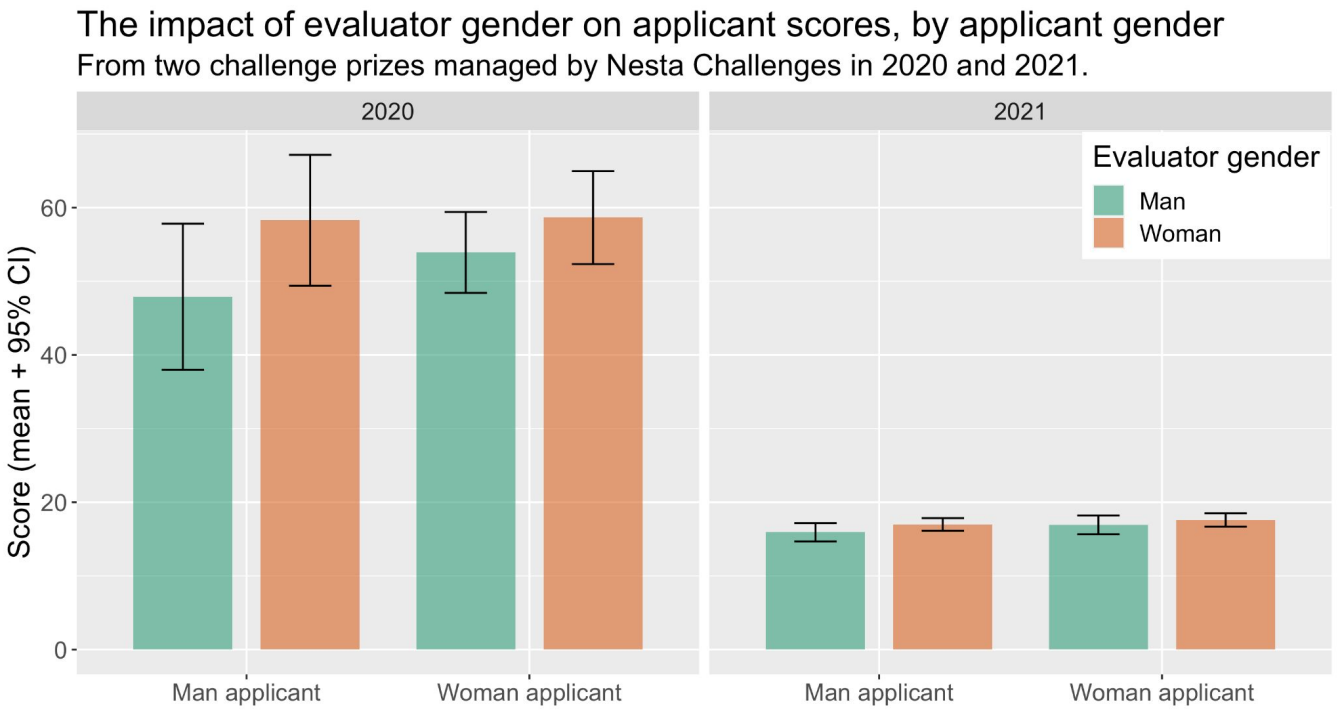
## Design:

- Random assignment of proposals to evaluators (subject to constraints) → Within-proposal random variation in "match" b/w applicant & evaluator gender

## Hypothesis:

- Proposals submitted by women receive higher *relative* scores (as compared to men) when evaluated by a woman (rather than by a man)

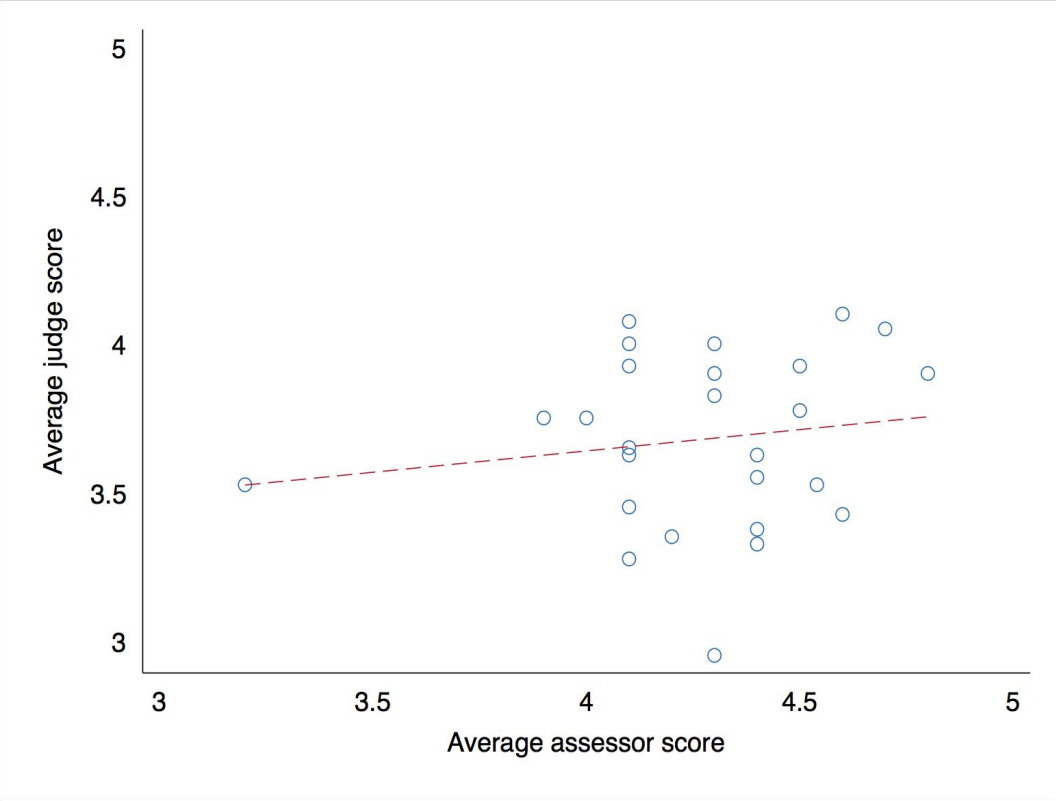
# Result I: No evidence of favoritism by gender




- Proposals *submitted* by women receive higher scores on avg
- Proposals *evaluated* by women receive higher scores on avg
- **No gender interaction effect** in linear regression with proposal & evaluator fixed effects
- No favoritism for own gender ≠ no bias! (Card et al. 2019)

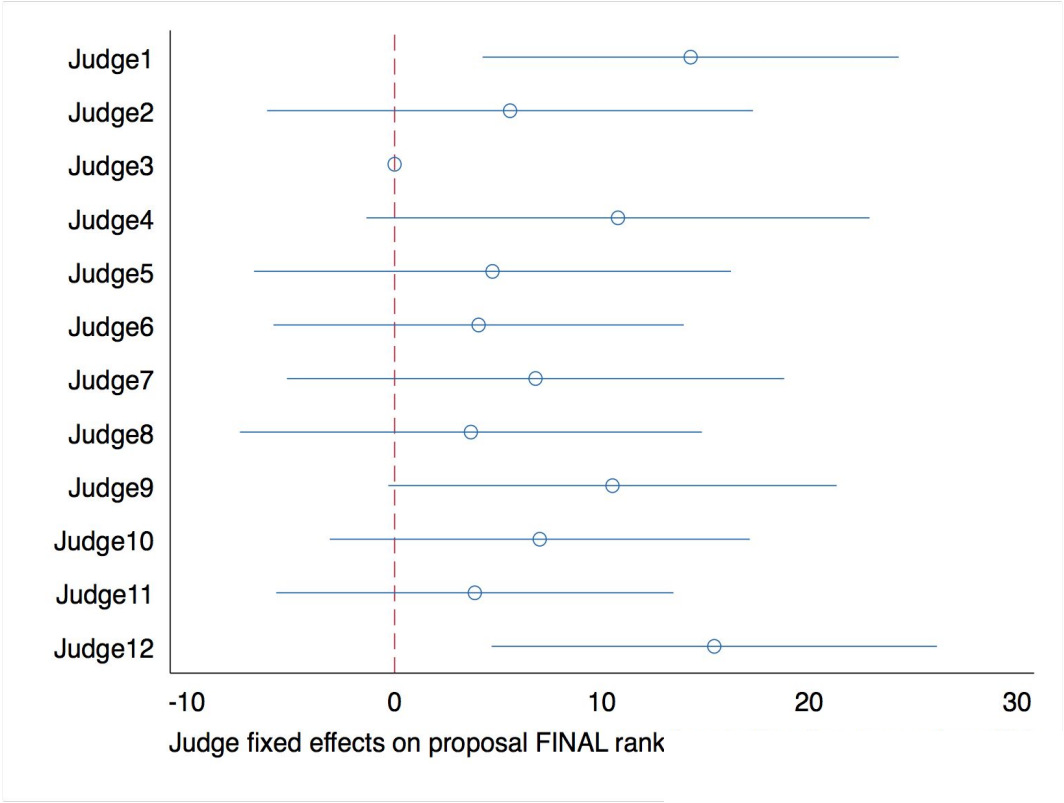
Note: In 2020, evaluators used a 0 - 100 scale, whereas in 2021 they used a 0 - 25 scale.

# Result II: Noise in scoring




- **First-sift assessor scores weak predictors of judge scores**
  - **Judges may not even see proposals that they otherwise might like**
- Evaluators differ in their leniency and the dispersion of their scores
  - Sizable within-proposal score variation

# Result II: Noise in scoring




- First-sift assessor scores weak predictors of judge scores
  - Judges may not even see proposals that they otherwise might like
- **Evaluators differ in their leniency and the dispersion of their scores**
  - **Sizable within-proposal score variation**

# Insights for the funder

- Test for favoritism now standard part of assessment process
  - No pushback from PMs / evaluators
  - Need to improve demographic data collection from applicants and evaluators
    - Other characteristics
    - Going beyond lead applicant
- Even when bias is not an issue – noise is!
  - Random matching + normalization can help
  - **The case for (partial) randomization**

# Insights for researchers

- Randomization process not straightforward
  - Generating random pairs until constraints satisfied takes forever with large samples...
    - Better way of randomizing?
- Analysis:
  - SE calculations need to account for assignment mechanism and dependence of observations
    - Randomization inference-based SE?
- Power calculations
  - Sample size justification based on studying the entire (small) population (Lakens, 2021)
  - Estimates from pilot can inform design of future studies
- Using evaluator leniency as IV might work to estimate impact of funding
  - Requires random assignment of proposals to evaluators!

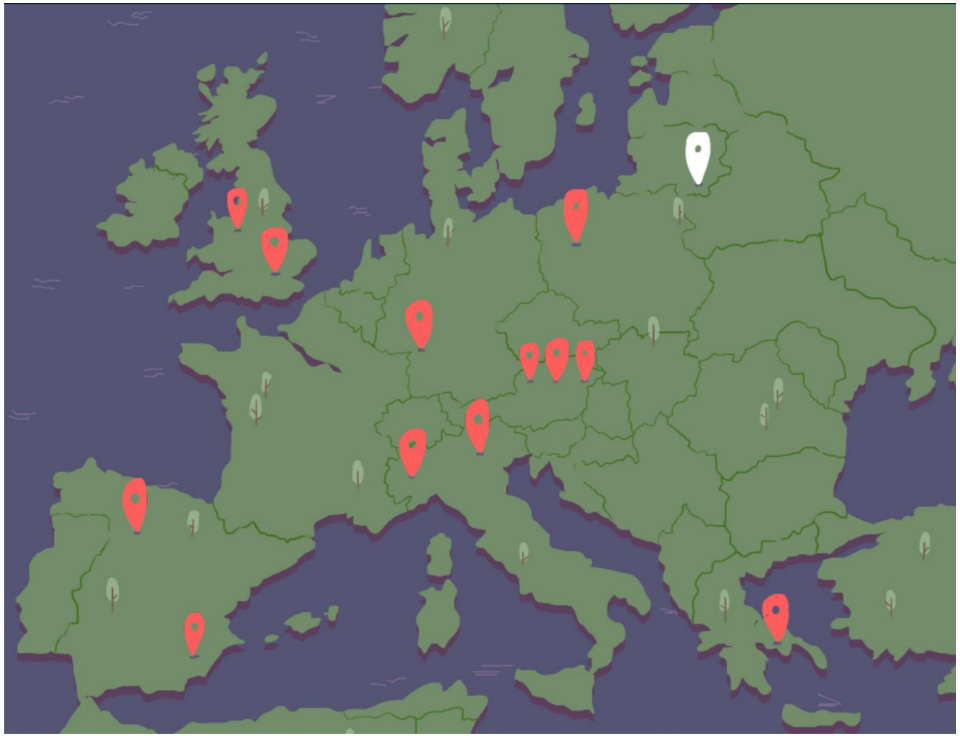
# Case study 2: INNOSUP-06-2018 trials

# INNOSUP-06-2018: EU Horizon 2020 program

**Aim:** encourage innovation agencies across Europe to experiment in their policy schemes supporting SME innovation.

**13 pilots and trials funded**:

- 27 national and regional agencies participating
- Encouraging co-creation, user-centered design, digital transformation, social innovation collaborations, age-inclusive leadership; etc.
- Budgets €60k - €700k



# Addressing common challenges

## What does the control group receive?

- Nothing
- Business as usual
- “Placebo” treatment
- Pared-down version of the treatment
- Same treatment, later (wait list CG or phase-in design)

→ *Note: choice affects the research question!*

## Measuring the outcomes of interest:

- Incentives for survey completion:
  - Monetary
  - Personalised feedback
- Moving beyond surveys:
  - Text analysis of social media activity
  - Revealed preference: participating in related activity

# What worked well

- Innovative schemes
- Building evaluation capacity
  - Better outcome data collection
  - Upfront planning
  - Emphasis on theory of change
  - Experimental mindset
- Peer learning
  - Timely access to best practices and insights from peer organizations
- Agile response to COVID pandemic
  - Flexibility from Commission re: timelines
  - Move to online support → unexpected benefits

# What worked well – and what didn't

- Innovative schemes
- Building evaluation capacity
  - Better outcome data collection
  - Upfront planning
  - Emphasis on theory of change
  - Experimental mindset
- Peer learning
  - Timely access to best practices and insights from peer organizations
- Agile response to COVID pandemic
  - Flexibility from Commission re: timelines
  - Move to online support → unexpected benefits

- Pilots needed to ensure demand and consistency in delivery
  - The catch-22 of novelty
- Recruitment, retention and survey response challenges
- No flexibility around requirement to randomize – even when realized sample size way too small
- Collaboration with experimental researchers:
  - When it happened, it was super valuable
  - But it didn't happen often enough

# Conclusions

# When (not) to run innovation support RCTs

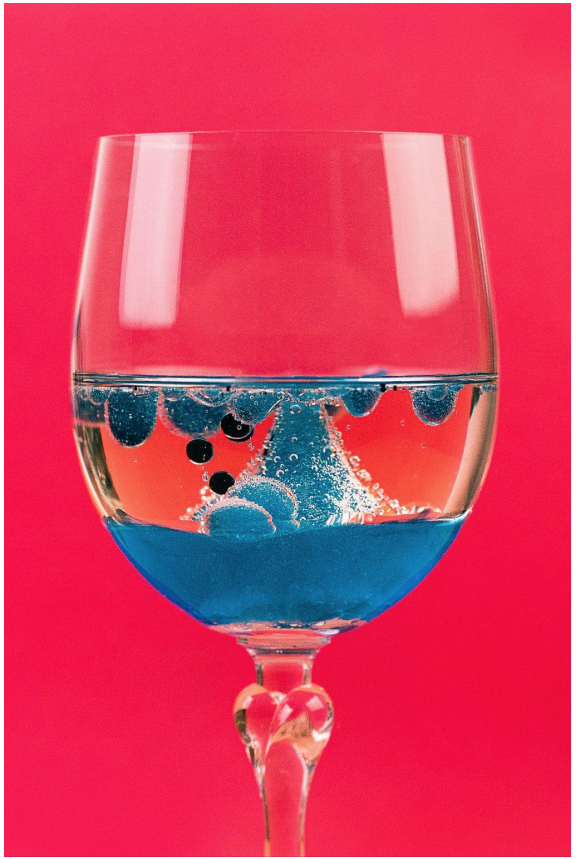
+ Narrow, well-defined questions
  + Testing theory-backed hypotheses
  + Comparing two clear alternatives
  + Generating data to shift our priors
+ Clear, easily measurable outcomes
+ Optimizing delivery
+ When schemes are over-subscribed and merit/need hard to judge
+ Before scaling up an (expensive) individual-based support programs

- Evaluating ecosystem-wide transformation or change in legislation
- Main outcome of interest is very skewed
- Never-before tested and implemented policy schemes (pilot first!)
- Impossible/unethical to ration access
- Impact eval for inexpensive program with convincing non-RCT evidence
- When predictability and stability of support landscape very important

# Other approaches

Alternatives & complements to RCTs:

- Shadow experiments
- Difference-in-differences
- Regression discontinuity design
- Qualitative research
- Etc.



# Resources

Review papers/reports:

- OECD (2014): Making innovation policy work – Learning from experimentation
- Karim R. Lakhani & Kevin J. Boudreau (2016): Innovation Experiments: Researching Technical Advance, Knowledge Production, and the Design of Supporting Institutions
- Albert Bravo-Biosca (2019): Experimental innovation policy
- Sandra Bendiscioli et al. (2021): The experimental research funder's handbook

Trial database:

- IGL's innovation and entrepreneurship trial database

Newsletters:

- Matt Clancy's What's New Under the Sun (weekly)
- Innovation Growth Lab newsletter (monthly)
- Experimental notes (quarterly)

# Your feedback is much appreciated

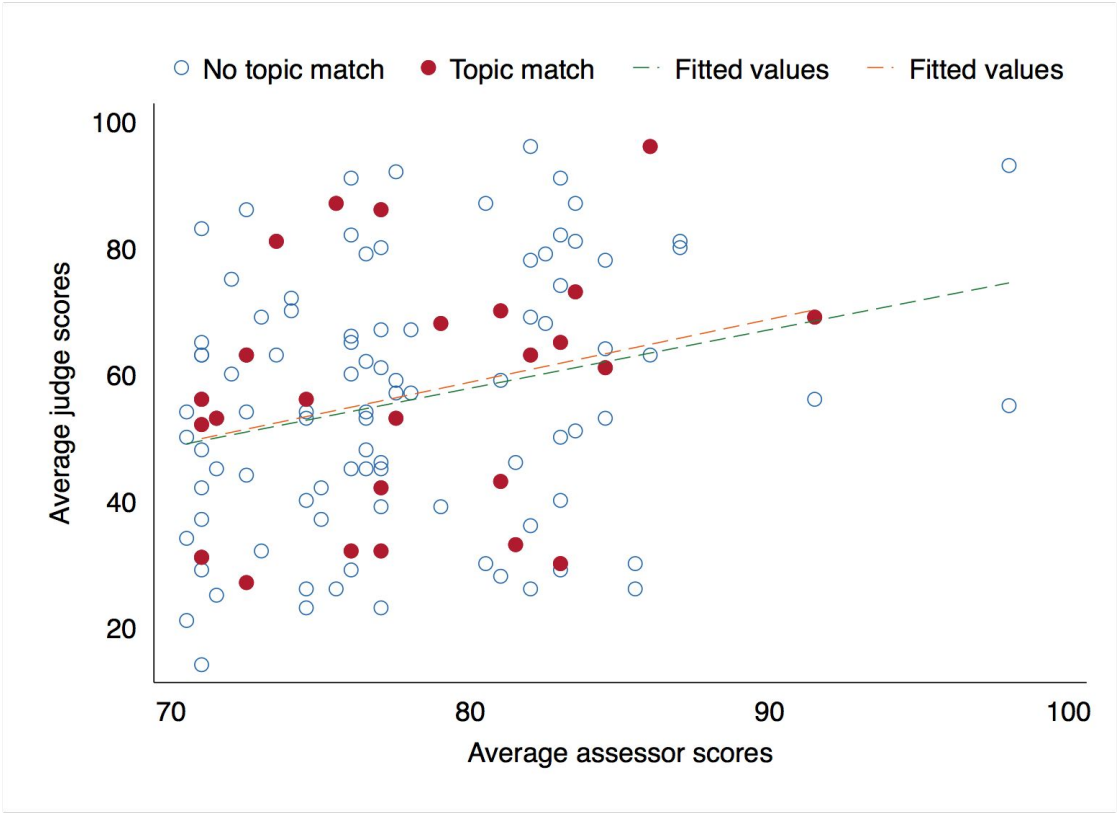
Share with me now –
and/or get in touch!


 czibore@gmail.com

 @ECzibor

 Eszter Czibor

# Topic match and scores



# Challenges: Measurement

- **Patents**: “Patents are a very imperfect measure of innovation; there is heterogeneity across countries, firms, and industries in the propensity to patent.” (Branstetter et al. 2019) + citations takes a long time to materialize (“the maximum probability of a citation occurs 10 to 12 years after the initial R&D investment”)
- **Business outcomes,** incl. R&D expenditure, employment, turnover, propensity to export
  - R&D spending measure tricky b/c of “relabeling” (Hall & Van Reenen (2000), Chen et al. (2019))
  - Is exit always bad? Benchmarking induced exit of low performers by resolving uncertainty over ability (Hou & Png, 2021)
  - Accessing VC funding: subject to huge biases
- **Self-reported measures**:
  - hard to collect meaningful survey data from a large sample
  - Relationship between intermediate and final outcomes not as clear and established as in other fields (e.g. education)
- **Direction of innovation**? Text analysis (can feel arbitrary)
- **Demographic characteristics of innovators**
  - Often unavailable or not detailed enough: US and UK: Race, ethnicity, and gender are not recorded in patent data, classification based on names possible but imperfect (Cook et al, 2021, Nathan (2015))
- “**Riskiness**” (potentially groundbreaking, but high chance of failure) or **novelty**:
  - “Most researchers who study risk depend on partial measures that look at the degree to which research results deviate from past results and/or look at the building blocks upon which the research is based”

# Limitations of innovation policy RCTs

- Can't randomize institutions, culture & legal environment → fact of life :)
- Selection is a crucial determinant of innovative outcomes → randomize after initial selection
- Changing a single dimension ("all else equal") unrealistic → when complementarities matter, design more complex treatment
- SUTVA rarely holds: spillovers, GE effects → choose design that allows to measure them
- Sample size troubles → collect data in multiple rounds / countries (and/or intensive treatment w/ large expected effect size + precise measurement)
- Cost ("100 NIH grants → $50mn") → program cost ≠ cost of experimentation!
- Outcomes:
  - Time horizon → intermediate outcomes (relationship needs to be verified!) from ToC
  - Measurement → methodological innovations needed