Statistical Modeling and Causal Inference with R

Week 12: Field experiments

Manuel Bosancianu

Max Schaub

November 30, 2020

Hertie School of Governance

Today's focus

- ✓ Intro
- ✓ DIY field experiments
- ✓ Challenges in field experiments
- ✓ Innovations in RCT design and implementation

DIY field experiments

Perks of (DIY) field experiments

- ✓ Evidence of causal effects on real-world behaviors
- ✓ Advance theory, e.g. importance of social norms
- ✓ Inform social policy (cp. Baldassarri and Abascal (2017))
- ✓ Fun to plan and implement!

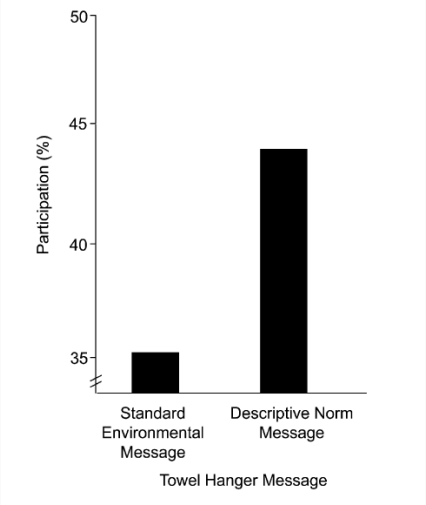
DIY field experiments

1. Nudging interventions and GOTV
2. Manipulation-of-context experiments
3. Lost letters
4. Correspondence tests/audit studies
5. Actor/conferderate-driven design
6. Small-incentive design
7. Lab-in-the-field experiments

1 Nudging interventions and GOTV

- ✓ Use small intervention ('nudge') to trigger behavioral change such as more environmentally friendly behavior (Nolan, Schultz, Cialdini, Goldstein, & Griskevicius, 2008), or higher turnout (get-out-the-vote or GOTV experiments) (Gerber, Green, & Larimer, 2008)
- ✓ Often uses i) social comparison, ii) 'opt in' instead of 'opt out', and/or iii) tests the effectiveness of different methods, e.g. in-person canvassing vs. phone calls (in GOTV experiments)
- ✓ Typically implemented within existing structures/procedures, e.g. existing business practices, planned election campaigns
- ✓ Cheap – if you can convince someone to randomize treatments

Example: Goldstein, Cialdini, and Griskevicius (2008)



2 Manipulation-of-context experiments

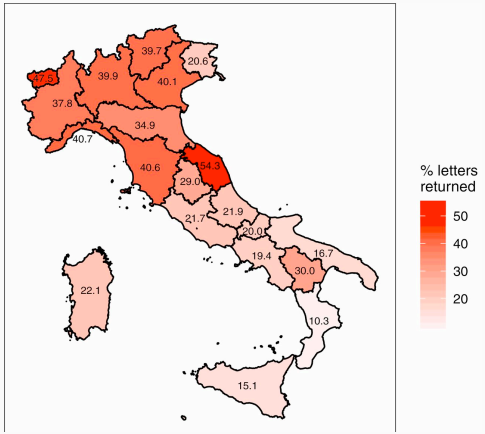
- ✓ Manipulate context to elicit changes in behavior
- ✓ Most famously used to test ‘broken windows’ theory, i.e. the idea that disorderly environments cause norm violations
- ✓ Fascinating, easy to implement – but limited applicability (?)

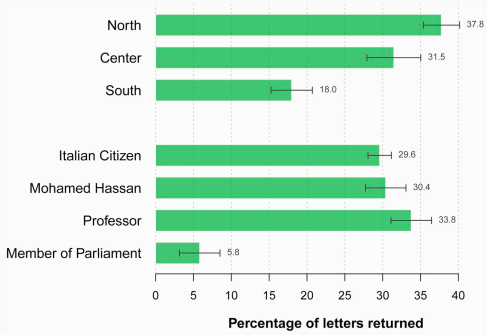
Example: Keizer, Lindeberg, and Steg (2008)

3 Lost letters

- ✓ Letters with an address and stamp are distributed in neighborhoods; outcome is return rate (Milgram, Mann, & Harter, 1965)
- ✓ Either used to test for the effect of context, e.g. ethnically diverse (Koopmans & Veit, 2014), high vs. low market integration (but: not randomized!), or to probe for discrimination e.g. against foreigners, politicians (easy to randomize)
- ✓ Fairly cheap to implement, needs careful planning (distance to post boxes, weather, etc.)

Example: Baldassarri (2020)





4 Correspondence tests/audit studies

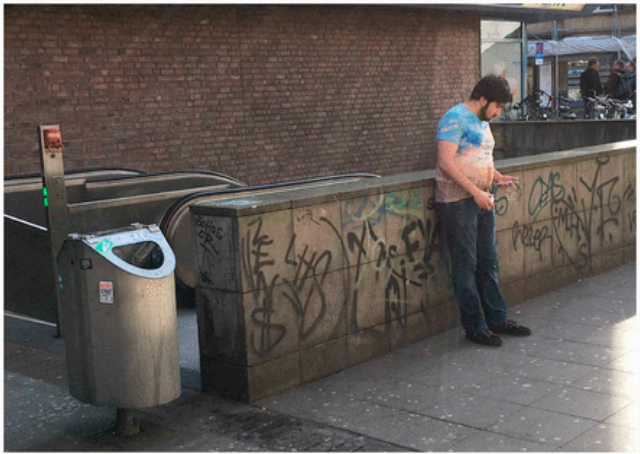
- ✓ Manipulate correspondence with potential employers (Bertrand & Mullainathan, 2004), landlords (Bartoš, Bauer, Chytilová, & Matějka, 2016), state institutions (Hemker & Rink, 2017)
- ✓ Classic method to test for discrimination in terms of ethnicity, sex, etc.
- ✓ Labor intense but fairly cheap, especially when using electronic correspondence

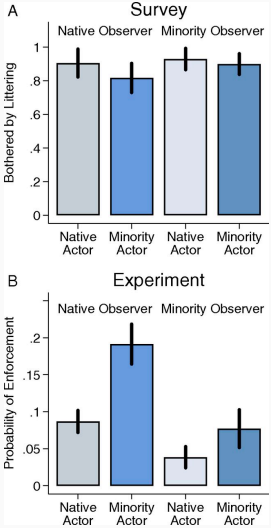
Example: Koopmans, Veit, and Yemane (2019)

5 Actor/confederate-driven design

- ✓ Have confederates (trained actors or layperson), record reaction of public
- ✓ Used to study social norms (Cohen & Nisbett, 1997; Winter & Zhang, 2018), anti-immigrant sentiments (Enos, 2014), etc.
- ✓ Effective method if done well, limited applications (?)

Example: Winter and Zhang (2018)

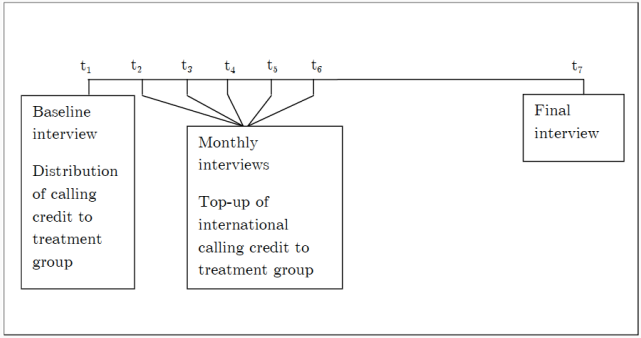




6 Small-incentive design

- ✓ Small monetary incentive to encourage change in behavior in large range of domains, from compliance with rules for picking up kids from daycare (Gneezy & Rustichini, 2000) to migration (Bryan, Chowdhury, & Mobarak, 2014)
- ✓ Incentive can we positive (reward) or negative (fine)
- ✓ Very flexible, but not so cheap in the case of rewards (small incentives can add up!)

Example: Batista and Narciso (2013)



7 Lab-in-the-field experiments

- ✓ Use of behavioral games (dictator game, trust game, PGG) typically employed in econ labs in field settings
- ✓ Used to measure cooperative (Baldassarri & Grossman, 2011), spiteful (Prediger, Vollan, & Herrmann, 2014), intergroup (Abascal, 2015), and other types of behavior in abstract (and putatively generalizable) way
- ✓ Blend of controlled lab environment with field setting (i.e. 'bringing the lab to the field')
- ✓ Somewhat specialized type of research; needs careful preparation and buy in; protocols for standard games widely available; but not for free (behavioral games usually entail incentives).

Example: Habyarimana et al. (2007)

Challenges to RCT-based inference

RCT-based inference: vulnerabilities

Evidence of impressive creativity and flexibility in applying field experiments to substantive questions.

Some hesitation about the potential of turning into a glorified form of policy analysis (Bates, 2006).

Concerns about inference (Humphreys & Scacco, 2020):

- ✓ external validity: Metaketa initiative;
- ✓ external validity: “patient-preference”/selective trials;
- ✓ aggregation challenge: structural models

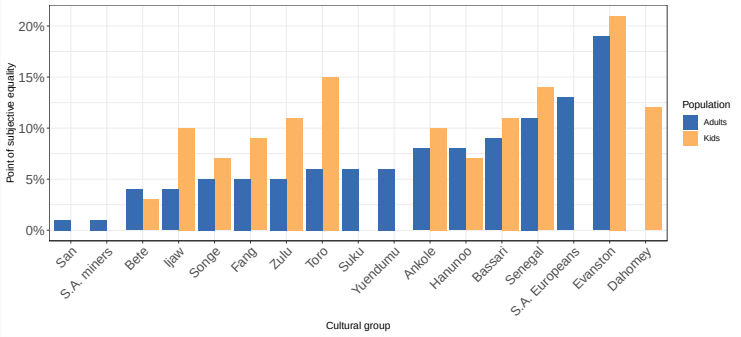
Müller-Lyer illusion

Which one is longer?

We've known about this illusion since 1889. However, since 1966 we also know that not all cultures experience this in the same way.

Cross-cultural variance

Segall, Campbell, and Herskovits (1966) find differences between cultures in how different people perceive these lines to be.



Adapted from McCauley and Henrich (2006)

External validity concerns

Validity (Campbell & Stanley, 1966):

- ✓ internal: ability of design to capture actual magnitude of relationship in real life
- ✓ external: extent to which we can generalize measured effect to other populations, contexts, and ways of measuring phenomenon

Many choices in RCTs: treatment arms and delivery, study population, ways of measuring outcomes (attitudes, behavior, task-based).

Professional incentives for innovation in study design (“plant the flag”), rather than replication.

Metaketa initiative

Challenges	Features
Potential for confounding in observational research	RCTs
Limited external validity of RCTs	Multiple studies across sites
Heterogeneous, scattered findings	Meta-analysis for findings
Diversity of interventions	“Common arm” intervention
Noncomparable measurement	Harmonized measurement of inputs, outcomes, and controls
Researcher incentives for innovation	“Alternative arm” intervention
Private data	Open data and replication code
Errors in data or code	Third-party data analysis
Data fishing	Pre-analysis plans
Publication bias	Publication of all registered analyses

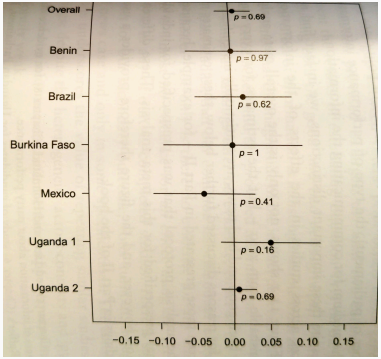
Adapted from Dunning, Grossman, et al. (2019)

Metaketa I: information and accountability

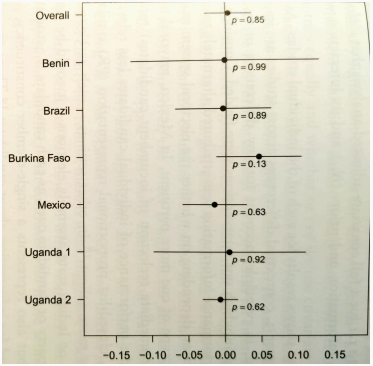
Site	Common arm	Alternative arm
Benin	Legislative performance	Civic lesson about legislative performance
Brazil	Accounting irregularities	Municipal education outcomes
Burkina Faso	Quality of municipal services	Invitation to municipal gov't meetings
India	Criminal background of candidates (info. from enumerators)	Criminal background (info. from local brokers)
Mexico	Unauthorized / misallocated spending	Misallocated spending (loudspeakers, or no benchmark)
Uganda 1	Voter–candidate policy alignment (campaign videos)	Public provision of information
Uganda 2	Budget irregularities (vias SMS)	Quality of service provision

Adapted from Dunning, Grossman, et al. (2019)

Results: information on turnout



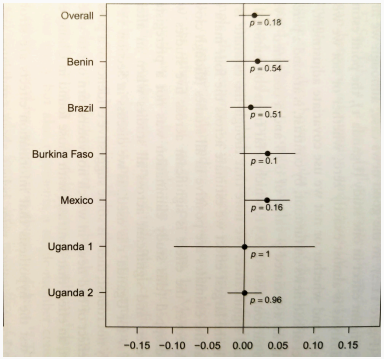
(a) Good news



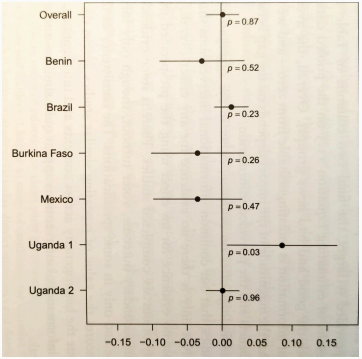
(b) Bad news

Taken from Dunning, Bicalho, et al. (2019)

Results: information on vote choice



(a) Good news



(b) Bad news

Taken from Dunning, Bicalho, et al. (2019)

More Metaketa rounds

Currently-ongoing projects on:

- ✓ taxation (drivers of formalization, esp. information)
- ✓ natural resource governance (effects of strengthening community monitoring capacity)
- ✓ community policing (effectiveness in weak-state contexts)
- ✓ women involvement in consultative processes (mobilization factors)

These, and similar other projects, take some steps toward greater generalizability of findings from RCTs.

Self-selection everywhere

In an RCT researchers have high degree of control over treatment assignment.

A great benefit, but represents a vulnerability as well: treatment is assigned to people who in real life wouldn't take it.

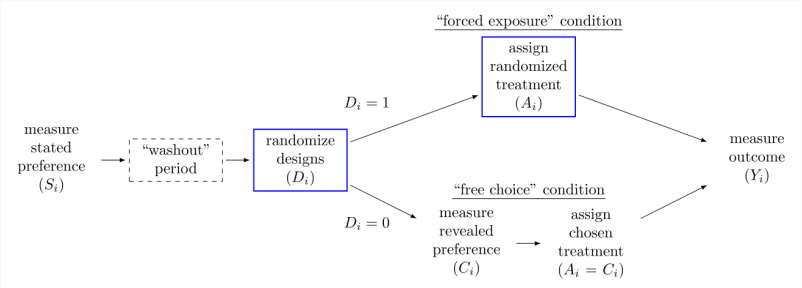
Examples: media exposure effects on candidates for President, effects on tolerance from contact with out-group individuals.

This isn't how people consume media or interact with others in real life—self-selection.

Patient preference trials

The ATE estimated from a RCT might only approximate poorly the effect in the population.

Solution: incorporate stated and revealed preferences over treatment arms in design (Knox, Yamamoto, Baum, & Berinsky, 2019).



Patient preference trials

Design allows estimating multiple quantities:

- ✓ ATE: the standard treatment effect
- ✓ ACTE: average choice-specific treatment effect—effect of treatment t for those who would opt for something else in real life

Can provide information on what effects we might notice if we scale up an intervention.

Selective trials

Design related to that of *selective trials* (Chassang, Padró I Miquel, & Snowberg, 2012).

Hard to say why a treatment's measured effect is, for example, low:

- ✓ its true effect really is low
- ✓ people *believe* it's low $\Rightarrow$ don't use it

We cannot observe *effort* with standard designs.

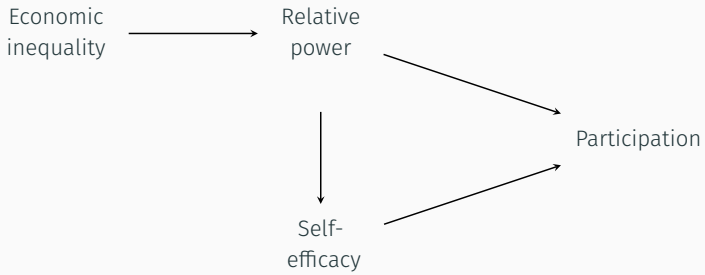
Selective trials

In these designs, participants can self-select into their preferred treatment arm at a cost (with probability dependent on willingness to pay).

Provides information on treatment effects at different levels of belief in the treatment's benefits.

From RCT estimates to macro dynamics

Relative power theory (updated)



RCT result: accurate information about inequality in district ⇒ lowered belief that voting can change anything.

What can we say about the macro-level phenomenon?

Alternative mediating pathways

If information is sole channel, we could get a sense of dynamics at macro level.

If alternative pathways operate, e.g. parties changing mobilization strategies, we run into major difficulties (Humphreys & Scacco, 2020).

General equilibrium effects come into play: if everyone's crop increases, prices might *decrease*.

Structural models

Moving from “effects of causes” to “causes of effects” (Holland, 1986).

$$\tau_i = Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \tag{1}$$

Trying to better grasp the network of causal dynamics that generate Y, and how causal pathways in the network combine to produce effects in reality.

Valuable initial work by Pearl and Bareinboim (2014).

It was a pleasure having you in the course!

References i

Abascal, M. (2015, August). Us and Them Black-White Relations in the Wake of Hispanic Population Growth. *American Sociological Review*, *80*(4), 789–813.

Baldassarri, D. (2020, February). Market integration accounts for local variation in generalized altruism in a nationwide lost-letter experiment. *Proceedings of the National Academy of Sciences*, *117*(6), 2858–2863.

Baldassarri, D., & Abascal, M. (2017). Field Experiments Across the Social Sciences. *Annual Review of Sociology*, *43*(1), 41–73.

Baldassarri, D., & Grossman, G. (2011, July). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences*, *108*(27), 11023–11027.

Bartoš, V., Bauer, M., Chytilová, J., & Matějka, F. (2016, June). Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition. *American Economic Review*, *106*(6), 1437–1475.

References ii

Bates, R. H. (2006). Banerjee's approach might teach us more about impact but at the expense of larger matters. *Boston Review of Books*, *31*(4). Retrieved 2020-11-24, from `https://bostonreview.net/bates-expense-larger-matters`

Batista, C., & Narciso, G. (2013). Migrant remittances and information flows: Evidence from a field experiment. *The World Bank Economic Review*.

Bertrand, M., & Mullainathan, S. (2004, September). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, *94*(4), 991–1013.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, *82*(5), 1671–1748.

References iii

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handbook of Research on Teaching* (pp. 171–246). Chicago, IL: Rand McNally.

Chassang, S., Padró I Miquel, G., & Snowberg, E. (2012). Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments. *American Economic Review*, *102*(4), 1279–1309.

Cohen, D., & Nisbett, R. E. (1997, November). Field Experiments Examining the Culture of Honor: The Role of Institutions in Perpetuating Norms about Violence. *Personality and Social Psychology Bulletin*, *23*(11), 1188–1199.

Dunning, T., Bicalho, C., Chowdhury, A., Grossman, G., Humphreys, M., Hyde, S. D., … Nellis, G. (2019). Meta-Analysis. In T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, & G. Nellis (Eds.), *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I* (pp. 315–374). Cambridge: Cambridge University Press.

References iv

Dunning, T., Grossman, G., Humphreys, M., Hyde, S. D., & McIntosh, C. (2019). The Metaketa Initiative. In T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, & G. Nellis (Eds.), *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I* (pp. 16–49). Cambridge: Cambridge University Press.

Enos, R. D. (2014, March). Causal effect of intergroup contact on exclusionary attitudes. *Proceedings of the National Academy of Sciences*, *111*(10), 3699–3704.

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008, February). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *American Political Science Review*, *102*(1), 33–48.

Gneezy, U., & Rustichini, A. (2000). A Fine Is a Price. *Journal of Legal Studies*, *29*, 1.

References v

Goldstein, N. J., Cialdini, R. B., Griskevicius, V., & served as associate editor for this article., J. D. s. a. e. a. M. F. L. (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research*, *35*(3), 472–482.

Habyarimana, J., Humphreys, M., Posner, D. N., & Weinstein, J. M. (2007). Why Does Ethnic Diversity Undermine Public Goods Provision? *American Political Science Review*, *101*(04), 709–725.

Hemker, J., & Rink, A. (2017, October). Multiple Dimensions of Bureaucratic Discrimination: Evidence from German Welfare Offices. *American Journal of Political Science*, *61*(4), 786–803.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

References vi

Humphreys, M., & Scacco, A. (2020). The aggregation challenge. *World Development*, *127*, 104806. doi: 10.1016/j.worlddev.2019.104806

Keizer, K., Lindenberg, S., & Steg, L. (2008, December). The Spreading of Disorder. *Science*, *322*(5908), 1681–1685.

Knox, D., Yamamoto, T., Baum, M. A., & Berinsky, A. J. (2019). Design, Identification, and Sensitivity Analysis for Patient Preference Trials. *Journal of the American Statistical Association*, *114*(528), 1532–1546.

Koopmans, R., & Veit, S. (2014). Cooperation in Ethnically Diverse Neighborhoods: A Lost-Letter Experiment. *Political Psychology*, *35*(3), 379–400.

Koopmans, R., Veit, S., & Yemane, R. (2019, December). Taste or statistics? A correspondence study of ethnic, racial and religious labour market discrimination in Germany. *Ethnic and Racial Studies*, *42*(16), 233–252.

References vii

McCauley, R. N., & Henrich, J. (2006). Susceptibility to the Müller–Lyer Illusion, Theory-Neutral Observation, and the Diachronic Penetrability of the Visual Input System. *Philosophical Psychology*, *19*(1), 79–101.

Milgram, S., Mann, L., & Harter, S. (1965). The lost-letter technique: A tool of social research. *Public Opinion Quarterly*, *29*(3), 437.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008, July). Normative Social Influence is Underdetected. *Personality and Social Psychology Bulletin*, *34*(7), 913–923.

Pearl, J., & Bareinboim, E. (2014). External Validity: From Do-Calculus to Transportability Across Populations. *Statistical Science*, *29*(4), 579–595.

Prediger, S., Vollan, B., & Herrmann, B. (2014, November). Resource scarcity and antisocial behavior. *Journal of Public Economics*, *119*, 1–9.

References viii

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The Influence of Culture on Visual Perception*. Indianapolis, IN: Bobbs-Merrill Company.

Winter, F., & Zhang, N. (2018, March). Social norm enforcement in ethnically diverse communities. *Proceedings of the National Academy of Sciences*, *115*(11), 2722–2727.