Experimental Methods
University of Oxford DPIR Trinity 2022

Lecturer: Ray Duch `raymond.duch@nuffield.ox.ac.uk`

Tuesday, 14:00-16:00

Course Objectives

The course covers the design, implementation, and analytic tools necessary for conducting social science experiments and analysing experimental data. The course consists of two modules; Module 1 runs from week 1-4 and a Module 2 from week 5-8. Module 1 presents material that will be required for students who are being assessed for the course. Module 2 is entirely voluntary – this material will NOT be the basis for assessment. Lectures will be accompanied by lab sessions to verify recent experimental studies using R. There will be periodic guest speakers – they will give a seminar after the class on Tuesday from 16:00-17:00.

Participants will have the opportunity to present their own experimental research and receive feedback from an experienced team of instructors. Upon completion of the workshop participants should be able to (1) formulate research questions that can be addressed using experiments, (2) design and carry out experiments, and (3) analyse and interpret results from social sciences experiments.

The course is appropriate for participants from any discipline who expect to include experimental social research as part of their research agenda. It is also appropriate for participants who want to become informed consumers of experimental research scholarship.

Course Prerequisites

Participants should have a basic background in research design and statistics. For example, with respect to research design, they should understand basic concepts such as causal inference, exogeneity, control group, and confounding effects. With respect to basic statistics, they should understand the principals of ordinary least squares regression; how to calculate simple measures of association; and have some familiarity with a statistical software package. The hands-on experimental data analysis lab sessions will use R.

Course Materials and Logistics

Lectures and practice sessions take place in the Conference Room at Nuffield College. All materials will be available on the RayDuch Github. Core readings and the reading for verification exercises are mandatory. Participants are requested to have R up and running on their machines for the lab sessions, and participants should familiarize themselves with R prior to the first lab session if they have not already done so.

Participants will need to have installed:

- R and RStudio

Module 1			
Day	Time ($UTC+1$)	Place	Topic
26 April	14:00 - 16:00	CR Nuffield	Rubin Causal Model, ATE, alternative designs
3 May	14:00 - 16:00	CR Nuffield	Random Assignment, Covariates and Power
10 May	14:00 - 16:00	CR Nuffield	Compliance, Interference and Attrition
17 May	14:00 - 16:00	CR Nuffield	Heterogeneous Treatment Effects, Measurement, Mediation
Module 2			
24 May	14:00 - 16:00	CR Nuffield	Randomization Inference
31 May	14:00 - 16:00	CR Nuffield	Machine Learning: Generating and Analyzing Experimental Data
7 June	14:00 - 16:00	CR Nuffield	Discrete Choice Experiments
14 June	14:00 - 16:00	CR Nuffield	Designing Information Experiments

Week 1: Rubin Causal Model, ATE, Randomization Inference (Briefly), alternative designs

Core readings

- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc., New York, 2012.
  - Ch. 2
- Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principals for Social Research*. Cambridge University Press, 2007
  - Ch. 1-2

Additional readings

- Gary Charness, Uri Gneezy, and Michael A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1 – 8, 2012
- Charles Bellemare, Luc Bissonnette, and Sabine Kröger. Simulating power of economic experiments: the powerbbk package. *Journal of the Economic Science Association*, 2(2):157–168, 2016
- Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1):1–30, 2014
- Lauren E. Young. The psychology of state repression: Fear and dissent decisions in zimbabwe. *American Political Science Review*, 113(1):140–155, 2019
- Chapter 5 in Druckman and Green (2021) Advances in Experimental Political Science (CUP) - available via SOLO

Week 2: Random Assignment, Co-variates, and Power

Core readings

- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation.* W.W. Norton & Company, Inc., New York, 2012.
  - Ch. 3, 4

Additional readings

- Chapters 15, 17 in Druckman and Green (2021) Advances in Experimental Political Science (CUP) - available via SOLO

Week 3: Compliance, Interference, and Attrition

Core readings

- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation.* W.W. Norton & Company, Inc., New York, 2012.
  - Ch. 5, 6, 7

Additional readings

- Chapters 16 in Druckman and Green (2021) Advances in Experimental Political Science (CUP) - available via SOLO
- Kaitlin Anderson, Gema Zamarro, Jennifer Steele, and Trey Miller. Comparing performance of methods to deal with differential attrition in randomized experimental evaluations. *Evaluation Review*, 45(1-2):70–104, 2021
- Greg Ridgeway, Dan McCaffrey, Andrew Morral, Matthew Cefalu, Lane Burgette, Joseph Pane, and Beth Ann Griffin. Toolkit for weighting and analysis of nonequivalent groups: A guide to the twang package. R package, RAND, October 2021.

Week 4: Heterogenous Treatment Effects; Measurement; Mediation

Core readings

- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation.* W.W. Norton & Company, Inc., New York, 2012.
  - Ch. 9, 10

Additional readings

- Claudio Ferraz and Frederico Finan. Exposing corrupt politicians: The effects of brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2):703–45, 2008
- David E. Broockman, Joshua L. Kalla, and Jasjeet S. Sekhon. The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Analysis*, 25(4):435?464, 2017
- Scott Clifford, Geoffrey Sheagley, and Spencer Pisston. Increasing precision without altering treatment effects: Repeated measures designs in survey experiments. *American Political Science Review*, 115(3):1048–1065, 2021

- Alexander Coppock. Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods*, pages 1–16, 2018.

Week 5: Randomization Inference

Core readings

- Susan Athey and Guido Imbens. The econometrics of randomized experiments. *Handbook of Economic Field Experiments*, 1:73–140, 2017
- Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc., New York, 2012
  - Ch. 3
- Alberto Abadie and Matias D. Cattaneo. Econometric methods for program evaluation. *Annual Review of Economics*, 10(1):465–503, 2018.
  - Ch. 3-4

Week 6: Machine Learning; Generating and Analyzing Experimental Data

Core readings

- Raymond Duch, Denise Laroze, Thomas Robinson, and Pablo Beramendi. Multi-modes for detecting experimental measurement error. *Political Analysis*, page 1–21, 2020
- Molly Offer-Westort, Alexander Coppock, and Donald P. Green. Adaptive experimental design: Prospects and applications in political science. *American Journal of Political Science*, n/a(n/a), 2021
- Offer-Westort, Molly and Vitor Hadad and Susan Athey. Adaptive experimentation tutorial, 2021b. `https://mollyow.shinyapps.io/adaptive/`
- James Bisbee. Barp: Improving mister p using bayesian additive regression trees. *American Political Science Review*, page 1–6, 2020

Week 7: Discrete Choice Experiments

Core readings

- Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1):1–30, 2014
- Kirk Bansak, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. *Conjoint Survey Experiments*. Cambridge University Press, 2021
- Jens Hainmueller, Dominik Hangartner, and Teppei Yamamoto. Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, 112(8):2395–2400, 2015

- Thomas Robinson and Raymond Duch. How to detect heterogeneity in conjoint experiments, 2021. Working Paper

Week 8: Designing Information Experiments

Core readings

- Ingar Haaland, Christopher Roth, and Johannes Wohlfart. Designing information provision experiments, 2020. CESifo Working Paper No. 8406

- Sonja Vogt, Nadia Ahmed Mohmmed Zaid, Hilal El Fadil Ahmed, Ernst Fehr, and Charles Efferson. Changing cultural attitudes towards female genital cutting. *Nature*, 538(7626):506–509, 2016

- Alberto Cavallo, Guillermo Cruces, and Ricardo Perez-Truglia. Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, 9(3):1–35, July 2017

- Christian Fong and Justin Grimmer. Causal inference with latent treatments. *American Journal of Political Science*, n/a(n/a), 2022

- Roberto Cerina and Raymond Duch. Measuring public opinion via digital footprints. *International Journal of Forecasting*, 2020

Additional readings

- Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016
- Justin Grimmer, Solomon Messing, and Sean J. Westwood. Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. *Political Analysis*, 25(4):413?434, 2017
- Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized programme evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013
-
- Donald P. Green and Holger L. Kern. Modeling heterogeneous treatment effects in survey experiments with bayesian additive regression trees. *Public Opinion Quarterly*, 76(3):491–511, 2012
- Raymond M. Duch, Denise Laroze, Thomas S. Robinson, and Pablo Beramendi. Multi-modes for detecting experimental measurement error. Nuffield College Centre for Experimental Social Sciences Working Paper Series, 2018.

Online

- Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, 20(3):351?368, 2012
- Daniel G. Goldstein, Siddharth Suri, R. Preston McAfee, Matthew Ekstrand-Abueg, and Fernando Diaz. The economic and cognitive costs of annoying display advertisements. *Journal of Marketing Research*, 51(6):742–752, 2014
- Ryan T. Moore and Sally A. Moore. Blocking for sequential political experiments. *Political Analysis*, 21(4):507?523, 2013
- Connor Huff and Dustin Tingley. "who are these people?" evaluating the demographic characteristics and political preferences of mturk survey respondents. *Research & Politics*, 2(3):2053168015604648, 2015
- Kevin J. Mullinix, Thomas J. Leeper, James N. Druckman, and Jeremy Freese. The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2):109–138, 2015
- Alexander Coppock and Donald P. Green. Assessing the correspondence between experimental results obtained in the lab and field: A review of recent social science research. *Political Science Research and Methods*, 3(1):113–131, 2015

Social Media

- Justin Grimmer, Solomon Messing, and Sean J. Westwood. How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation. *American Political Science Review*, 106(4):703?719, 2012

•

Field Experiments

- Esther Duflo, Rachel Glennerster, and Michael Kremer. Chapter 61 using randomization in development economics research: A toolkit. volume 4 of *Handbook of Development Economics*, pages 3895 – 3962. Elsevier, 2007
- Marianne Bertrand, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. Obtaining a driver's license in india: An experimental approach to studying corruption. *Quarterly Journal of Economics*, 122(4):1639–76, 2007
- Marianne Bertrand and Sendhil Mullainathan. Are emily and greg more employable than lakisha and jamal? *American Economics Review*, 94:991, 2004
- Karthik Muralidharan and Venkatesh Sundararaman. Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, 119(1):39–77, 2011
- R. Glennerster. Chapter 5 - the practicalities of running randomized evaluations: Partnerships, measurement, ethics, and transparency. In Abhijit Vinayak Banerjee and Esther Duflo, editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 175 – 243. North-Holland, 2017

Conjoint

- Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1):1–30, 2014

Sensitive Questions: Experiments

- Graeme Blair, Kosuke Imai, and Jason Lyall. Comparing and combining list and endorsement experiments: Evidence from afghanistan. *American Journal of Political Science*, 58(4):1043–1063, 2014

Adaptive Design

- Offer-Westort, Molly and Vitor Hadad and Susan Athey. Adaptive experimentation tutorial, 2021b. `https://mollyow.shinyapps.io/adaptive/`
- Molly Offer-Westort, Alexander Coppock, and Donald P. Green. Adaptive experimental design: Prospects and applications in political science. *American Journal of Political Science*, n/a(n/a), 2021
- Molly Offer-Westort, Leah R Rosenzweig, and Susan Athey. Optimal policies to battle the coronavirus "infodemic" among social media users in sub-saharan africa, Feb 2021

Interference Assumption Spillover Effects

- Stephanie Zonszein, Peter Aronow, and Cyrus Samii. interference, an r package for design-based estimation of spillover effects. `https://github.com/szonszein/interference`, 2019. Version 0.1.0

- Peter Aronow, Dean Eckles, Cyrus Samii, and Stephanie Zonszein. *Spillover Effects in Experimental Data*, page 526–543. Cambridge University Press, 2021
- Anna M. Wilke, Donald P. Green, and Jasper Cooper. A placebo design to detect spillovers from an education–entertainment experiment in uganda. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3):1075–1096, 2020

References

[1] Alberto Abadie and Matias D. Cattaneo. Econometric methods for program evaluation. *Annual Review of Economics*, 10(1):465–503, 2018.

[2] Kaitlin Anderson, Gema Zamarro, Jennifer Steele, and Trey Miller. Comparing performance of methods to deal with differential attrition in randomized experimental evaluations. *Evaluation Review*, 45(1-2):70–104, 2021.

[3] Peter Aronow, Dean Eckles, Cyrus Samii, and Stephanie Zonszein. *Spillover Effects in Experimental Data*, page 526–543. Cambridge University Press, 2021.

[4] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

[5] Susan Athey and Guido Imbens. The econometrics of randomized experiments. *Handbook of Economic Field Experiments*, 1:73–140, 2017.

[6] Kirk Bansak, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. *Conjoint Survey Experiments*. Cambridge University Press, 2021.

[7] Charles Bellemare, Luc Bissonnette, and Sabine Kröger. Simulating power of economic experiments: the powerbbk package. *Journal of the Economic Science Association*, 2(2):157–168, 2016.

[8] Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, 20(3):351?368, 2012.

[9] Marianne Bertrand, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. Obtaining a driver's license in india: An experimental approach to studying corruption. *Quarterly Journal of Economics*, 122(4):1639–76, 2007.

[10] Marianne Bertrand and Sendhil Mullainathan. Are emily and greg more employable than lakisha and jamal? *American Economics Review*, 94:991, 2004.

[11] James Bisbee. Barp: Improving mister p using bayesian additive regression trees. *American Political Science Review*, page 1–6, 2020.

[12] Graeme Blair, Kosuke Imai, and Jason Lyall. Comparing and combining list and endorsement experiments: Evidence from afghanistan. *American Journal of Political Science*, 58(4):1043–1063, 2014.

[13] David E. Broockman, Joshua L. Kalla, and Jasjeet S. Sekhon. The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Analysis*, 25(4):435?464, 2017.

[14] Alberto Cavallo, Guillermo Cruces, and Ricardo Perez-Truglia. Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, 9(3):1–35, July 2017.

[15] Roberto Cerina and Raymond Duch. Measuring public opinion via digital footprints. *International Journal of Forecasting*, 2020.

[16] Gary Charness, Uri Gneezy, and Michael A. Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1 – 8, 2012.

[17] Scott Clifford, Geoffrey Sheagley, and Spencer Pisston. Increasing precision without altering treatment effects: Repeated measures designs in survey experiments. *American Political Science Review*, 115(3):1048–1065, 2021.

[18] Alexander Coppock. Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods*, pages 1–16, 2018.

[19] Alexander Coppock and Donald P. Green. Assessing the correspondence between experimental results obtained in the lab and field: A review of recent social science research. *Political Science Research and Methods*, 3(1):113–131, 2015.

[20] Raymond Duch, Denise Laroze, Thomas Robinson, and Pablo Beramendi. Multimodes for detecting experimental measurement error. *Political Analysis*, page 1–21, 2020.

[21] Raymond M. Duch, Denise Laroze, Thomas S. Robinson, and Pablo Beramendi. Multi-modes for detecting experimental measurement error. Nuffield College Centre for Experimental Social Sciences Working Paper Series, 2018.

[22] Esther Duflo, Rachel Glennerster, and Michael Kremer. Chapter 61 using randomization in development economics research: A toolkit. volume 4 of *Handbook of Development Economics*, pages 3895 – 3962. Elsevier, 2007.

[23] Claudio Ferraz and Frederico Finan. Exposing corrupt politicians: The effects of brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2):703–45, 2008.

[24] Christian Fong and Justin Grimmer. Causal inference with latent treatments. *American Journal of Political Science*, n/a(n/a), 2022.

[25] Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, Inc., New York, 2012.

[26] R. Glennerster. Chapter 5 - the practicalities of running randomized evaluations: Partnerships, measurement, ethics, and transparency. In Abhijit Vinayak Banerjee and Esther Duflo, editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 175 – 243. North-Holland, 2017.

[27] Daniel G. Goldstein, Siddharth Suri, R. Preston McAfee, Matthew Ekstrand-Abueg, and Fernando Diaz. The economic and cognitive costs of annoying display advertisements. *Journal of Marketing Research*, 51(6):742–752, 2014.

[28] Donald P. Green and Holger L. Kern. Modeling heterogeneous treatment effects in survey experiments with bayesian additive regression trees. *Public Opinion Quarterly*, 76(3):491–511, 2012.

[29] Justin Grimmer, Solomon Messing, and Sean J. Westwood. How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation. *American Political Science Review*, 106(4):703?719, 2012.

[30] Justin Grimmer, Solomon Messing, and Sean J. Westwood. Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. *Political Analysis*, 25(4):413?434, 2017.

[31] Ingar Haaland, Christopher Roth, and Johannes Wohlfart. Designing information provision experiments, 2020. CESifo Working Paper No. 8406.

[32] Jens Hainmueller, Dominik Hangartner, and Teppei Yamamoto. Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, 112(8):2395–2400, 2015.

[33] Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1):1–30, 2014.

[34] Connor Huff and Dustin Tingley. "who are these people?" evaluating the demographic characteristics and political preferences of mturk survey respondents. *Research & Politics*, 2(3):2053168015604648, 2015.

[35] Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized programme evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.

[36] Ryan T. Moore and Sally A. Moore. Blocking for sequential political experiments. *Political Analysis*, 21(4):507?523, 2013.

[37] Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principals for Social Research*. Cambridge University Press, 2007.

[38] Kevin J. Mullinix, Thomas J. Leeper, James N. Druckman, and Jeremy Freese. The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2):109–138, 2015.

[39] Karthik Muralidharan and Venkatesh Sundararaman. Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, 119(1):39–77, 2011.

[40] Molly Offer-Westort, Alexander Coppock, and Donald P. Green. Adaptive experimental design: Prospects and applications in political science. *American Journal of Political Science*, n/a(n/a), 2021.

[41] Molly Offer-Westort, Leah R Rosenzweig, and Susan Athey. Optimal policies to battle the coronavirus "infodemic" among social media users in sub-saharan africa, Feb 2021.

[42] Offer-Westort, Molly and Vitor Hadad and Susan Athey. Adaptive experimentation tutorial, 2021b. `https://mollyow.shinyapps.io/adaptive/`.

[43] Greg Ridgeway, Dan McCaffrey, Andrew Morral, Matthew Cefalu, Lane Burgette, Joseph Pane, and Beth Ann Griffin. Toolkit for weighting and analysis of nonequivalent groups: A guide to the twang package. R package, RAND, October 2021.

[44] Thomas Robinson and Raymond Duch. How to detect heterogeneity in conjoint experiments, 2021. Working Paper.

[45] Sonja Vogt, Nadia Ahmed Mohmmed Zaid, Hilal El Fadil Ahmed, Ernst Fehr, and Charles Efferson. Changing cultural attitudes towards female genital cutting. *Nature*, 538(7626):506–509, 2016.

[46] Anna M. Wilke, Donald P. Green, and Jasper Cooper. A placebo design to detect spillovers from an education–entertainment experiment in uganda. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3):1075–1096, 2020.

[47] Lauren E. Young. The psychology of state repression: Fear and dissent decisions in zimbabwe. *American Political Science Review*, 113(1):140–155, 2019.

[48] Stephanie Zonszein, Peter Aronow, and Cyrus Samii. interference, an r package for design-based estimation of spillover effects. `https://github.com/szonszein/interference`, 2019. Version 0.1.0.