

## References

---

- Abadi, Martin, Andy Chu, Ian Goodfellow, H. Brendan Mironov, Ilya McMahan, Kunal Talwar, and Li Zhang. 2016. “Deep Learning with Differential Privacy.” *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*.
- Abrahamse, Peter. 2016. “How 8 Different A/B Testing Tools Affect Site Speed.” *CXL: All Things Data-Driven Marketing*. May 16. <https://conversionxl.com/blog/testing-tools-site-speed/>.
- ACM. 2018. *ACM Code of Ethics and Professional Conduct*. June 22. [www.acm.org/code-of-ethics](http://www.acm.org/code-of-ethics).
- Alvarez, Cindy. 2017. *Lean Customer Development: Building Products Your Customers Will Buy*. O'Reilly.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Apple, Inc. 2017. “Phased Release for Automatic Updates Now Available.” June 5. <https://developer.apple.com/app-store-connect/whats-new/?id=31070842>.
- Apple, Inc. 2018. “Use Low Power Mode to Save Battery Life on Your iPhone.” *Apple*. September 25. <https://support.apple.com/en-us/HT205234>.
- Athey, Susan, and Guido Imbens. 2016. “Recursive Partitioning for Heterogeneous Causal Effects.” *PNAS: Proceedings of the National Academy of Sciences*. 7353–7360. doi: <https://doi.org/10.1073/pnas.1510489113>.
- Azevedo, Eduardo M., Alex Deng, Jose Montiel Olea, Justin M. Rao, and E. Glen Weyl. 2019. “A/B Testing with Fat Tails.” February 26. Available at SSRN: <https://ssrn.com/abstract=3171224> or <http://dx.doi.org/10.2139/ssrn.3171224>.
- Backstrom, Lars, and Jon Kleinberg. 2011. “Network Bucket Testing.” *WWW '11 Proceedings of the 20th International Conference on World Wide Web*. Hyderabad, India: ACM. 615–624.
- Bailar, John C. 1983. “Introduction.” In *Clinical Trials: Issues and Approaches*, by Stuart Shapiro and Thomas Louis. Marcel Dekker.
- Bakshy, Eytan, Max Balandat, and Kostya Kashin. 2019. “Open-sourcing Ax and BoTorch: New AI tools for adaptive experimentation.” Facebook Artificial Intelligence. May 1. <https://ai.facebook.com/blog/open-sourcing-ax-and-botorch-new-ai-tools-for-adaptive-experimentation/>.

- Bakshy, Eytan, and Eitan Frachtenberg. 2015. "Design and Analysis of Benchmarking Experiments for Distributed Internet Services." *WWW '15: Proceedings of the 24th International Conference on World Wide Web*. Florence, Italy: ACM. 108–118. doi: <https://doi.org/10.1145/2736277.2741082>.
- Bakshy, Eytan, Dean Eckles, and Michael Bernstein. 2014. "Designing and Deploying Online Field Experiments." *International World Wide Web Conference (WWW 2014)*. <https://facebook.com/download/255785951270811/planout.pdf>.
- Barajas, Joel, Ram Akella, Marius Hotan, and Aaron Flores. 2016. "Experimental Designs and Estimation for Online Display Advertising Attribution in Marketplaces." *Marketing Science: the Marketing Journal of the Institute for Operations Research and the Management Sciences* 35: 465–483.
- Barrilleaux, Bonnie, and Dylan Wang. 2018. "Spreading the Love in the LinkedIn Feed with Creator-Side Optimization." *LinkedIn Engineering*. October 16. <https://engineering.linkedin.com/blog/2018/10/linkedin-feed-with-creator-side-optimization>.
- Basin, David, Soren Debois, and Thomas Hildebrandt. 2018. "On Purpose and by Necessity: Compliance under the GDPR." *Financial Cryptography and Data Security 2018*. IFCA. Preproceedings 21.
- Benbunan-Fich, Raquel. 2017. "The Ethics of Online Research with Unsuspecting Users: From A/B Testing to C/D Experimentation." *Research Ethics* 13 (3–4): 200–218. doi: <https://doi.org/10.1177/1747016116680664>.
- Benjamin, Daniel J., James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, et al. 2017. "Redefine Statistical Significance." *Nature Human Behaviour* 2 (1): 6–10. <https://www.nature.com/articles/s41562-017-0189-z>.
- Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman. 2011. *The Effect of Providing Peer Information on Retirement Savings Decisions*. NBER Working Paper Series, National Bureau of Economic Research. [www.nber.org/papers/w17345](http://www.nber.org/papers/w17345).
- Billingsly, Patrick. 1995. *Probability and Measure*. Wiley.
- Blake, Thomas, and Dominic Coey. 2014. "Why Marketplace Experimentation is Harder Than it Seems: The Role of Test-Control Interference." *EC '14 Proceedings of the Fifteenth ACM Conference on Economics and Computation*. Palo Alto, CA: ACM. 567–582.
- Blank, Steven Gary. 2005. *The Four Steps to the Epiphany: Successful Strategies for Products that Win*. Cafepress.com.
- Blocker, Craig, John Conway, Luc Demortier, Joel Heinrich, Tom Junk, Louis Lyons, and Giovanni Punzi. 2006. "Simple Facts about P-Values." *The Rockefeller University*. January 5. [http://physics.rockefeller.edu/luc/technical\\_reports/cdf8023\\_facts\\_about\\_p\\_values.pdf](http://physics.rockefeller.edu/luc/technical_reports/cdf8023_facts_about_p_values.pdf).
- Bodlewski, Mike. 2017. "When Slower UX is Better UX." *Web Designer Depot*. Sep 25. <https://www.webdesignerdepot.com/2017/09/when-slower-ux-is-better-ux/>.
- Bojinov, Iavor, and Neil Shephard. 2017. "Time Series Experiments and Causal Estimands: Exact Randomization Tests and Trading." *arXiv of Cornell University*. July 18. arXiv:1706.07840.
- Borden, Peter. 2014. "How Optimizely (Almost) Got Me Fired." *The SumAll Blog: Where E-commerce and Social Media Meet*. June 18. <https://blog.sumall.com/journal/optimizely-got-me-fired.html>.

- Bowman, Douglas. 2009. "Goodbye, Google." *stopdesign*. March 20. <https://stopdesign.com/archive/2009/03/20/goodbye-google.html>.
- Box, George E.P., J. Stuart Hunter, and William G. Hunter. 2005. *Statistics for Experimenters: Design, Innovation, and Discovery*. 2nd edition. John Wiley & Sons, Inc.
- Brooks Bell. 2015. "Click Summit 2015 Keynote Presentation." *Brooks Bell*. [www.brooksbell.com/wp-content/uploads/2015/05/BrooksBell\\_ClickSummit15\\_Keynote1.pdf](http://www.brooksbell.com/wp-content/uploads/2015/05/BrooksBell_ClickSummit15_Keynote1.pdf).
- Brown, Morton B. 1975. "A Method for Combining Non-Independent, One-Sided Tests of Significance." *Biometrics* 31 (4) 987–992. [www.jstor.org/stable/2529826](http://www.jstor.org/stable/2529826).
- Brutlag, Jake, Zoe Abrams, and Pat Meenan. 2011. "Above the Fold Time: Measuring Web Page Performance Visually." *Velocity: Web Performance and Operations Conference*.
- Buhrmester, Michael, Tracy Kwang, and Samuel Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?" *Perspectives on Psychological Science*, Feb 3.
- Campbell, Donald T. 1979. "Assessing the Impact of Planned Social Change." *Evaluation and Program Planning* 2: 67–90. [https://doi.org/10.1016/0149-7189\(79\)90048-X](https://doi.org/10.1016/0149-7189(79)90048-X).
- Campbell's law. 2018. *Wikipedia*. [https://en.wikipedia.org/wiki/Campbell%27s\\_law](https://en.wikipedia.org/wiki/Campbell%27s_law).
- Card, David, and Alan B Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84 (4): 772–793. <https://www.jstor.org/stable/2118030>.
- Casella, George, and Roger L. Berger. 2001. *Statistical Inference*. 2nd edition. Cengage Learning.
- CDC. 2015. *The Tuskegee Timeline*. December. <https://www.cdc.gov/tuskegee/timeline.htm>.
- Chamandy, Nicholas. 2016. "Experimentation in a Ridesharing Marketplace." *Lyft Engineering*. September 2. <https://eng.lyft.com/experimentation-in-a-risharing-marketplace-b39db027a66e>.
- Chan, David, Rong Ge, Ori Gershony, Tim Hesterberg, and Diane Lambert. 2010. "Evaluating Online Ad Campaigns in a Pipeline: Causal Models at Scale." *Proceedings of ACM SIGKDD*.
- Chapelle, Olivier, Thorsten Joachims, Filip Radlinski, and Yisong Yue. 2012. "Large-Scale Validation and Analysis of Interleaved Search Evaluation." *ACM Transactions on Information Systems*, February.
- Chaplin, Charlie. 1964. *My Autobiography*. Simon Schuster.
- Charles, Reichardt S., and Mark M. Melvin. 2004. "Quasi Experimentation." In *Handbook of Practical Program Evaluation*, by Joseph S. Wholey, Harry P. Hatry and Kathryn E. Newcomer. Jossey-Bass.
- Chatham, Bob, Bruce D. Temkin, and Michelle Amato. 2004. *A Primer on A/B Testing*. Forrester Research.
- Chen, Nanyu, Min Liu, and Ya Xu. 2019. "How A/B Tests Could Go Wrong: Automatic Diagnosis of Invalid Online Experiments." *WSDM '19 Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. Melbourne, VIC, Australia: ACM. 501–509. <https://dl.acm.org/citation.cfm?id=3291000>.

- Chrystal, K. Alec, and Paul D. Mizen. 2001. *Goodhart's Law: Its Origins, Meaning and Implications for Monetary Policy*. Prepared for the Festschrift in honor of Charles Goodhart held on 15–16 November 2001 at the Bank of England. [http://cyberlibris.typepad.com/blog/files/Goodharts\\_Law.pdf](http://cyberlibris.typepad.com/blog/files/Goodharts_Law.pdf).
- Coey, Dominic, and Tom Cunningham. 2019. "Improving Treatment Effect Estimators Through Experiment Splitting." *WWW '19: The Web Conference*. San Francisco, CA, USA: ACM. 285–295. doi:<https://dl.acm.org/citation.cfm?doid=3308558.3313452>.
- Collis, David. 2016. "Lean Strategy." *Harvard Business Review* 62–68. <https://hbr.org/2016/03/lean-strategy>.
- Concato, John, Nirav Shah, and Ralph I Horwitz. 2000. "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs." *The New England Journal of Medicine* 342 (25): 1887–1892. doi:<https://www.nejm.org/doi/10.1056/NEJM200006223422507>.
- Cox, David Roxbee. 1958. *Planning of Experiments*. New York: John Wiley.
- Croll, Alistair, and Benjamin Yoskovitz. 2013. *Lean Analytics: Use Data to Build a Better Startup Faster*. O'Reilly Media.
- Crook, Thomas, Brian Frasca, Ron Kohavi, and Roger Longbotham. 2009. "Seven Pitfalls to Avoid when Running Controlled Experiments on the Web." *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1105–1114.
- Cross, Robert G., and Ashutosh Dixit. 2005. "Customer-centric Pricing: The Surprising Secret for Profitability." *Business Horizons*, 488.
- Deb, Anirban, Suman Bhattacharya, Jeremy Gu, Tianxia Zhuo, Eva Feng, and Mandie Liu. 2018. "Under the Hood of Uber's Experimentation Platform." *Uber Engineering*. August 28. <https://eng.uber.com/xp>.
- Deng, Alex. 2015. "Objective Bayesian Two Sample Hypothesis Testing for Online Controlled Experiments." Florence, IT: ACM. 923–928.
- Deng, Alex, and Victor Hu. 2015. "Diluted Treatment Effect Estimation for Trigger Analysis in Online Controlled Experiments." *WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. Shanghai, China: ACM. 349–358. doi:<https://doi.org/10.1145/2684822.2685307>.
- Deng, Alex, Jiannan Lu, and Shouyuan Chen. 2016. "Continuous Monitoring of A/B Tests without Pain: Optional Stopping in Bayesian Testing." *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. Montreal, QC, Canada: IEEE. doi:<https://doi.org/10.1109/DSAA.2016.33>.
- Deng, Alex, Ulf Knoblich, and Jiannan Lu. 2018. "Applying the Delta Method in Metric Analytics: A Practical Guide with Novel Ideas." *24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- Deng, Alex, Jiannan Lu, and Jonathan Litz. 2017. "Trustworthy Analysis of Online A/B Tests: Pitfalls, Challenges and Solutions." *WSDM: The Tenth International Conference on Web Search and Data Mining*. Cambridge, UK.
- Deng, Alex, Ya Xu, Ron Kohavi, and Toby Walker. 2013. "Improving the Sensitivity of Online Controlled Experiments by Utilizing Pre-Experiment Data." *WSDM 2013: Sixth ACM International Conference on Web Search and Data Mining*.

- Deng, Shaojie, Roger Longbotham, Toby Walker, and Ya Xu. 2011. "Choice of Randomization Unit in Online Controlled Experiments." *Joint Statistical Meetings Proceedings*. 4866–4877.
- Denrell, Jerker. 2005. "Selection Bias and the Perils of Benchmarking." (*Harvard Business Review*) 83 (4): 114–119.
- Dickhaus, Thorsten. 2014. *Simultaneous Statistical Inference: With Applications in the Life Sciences*. Springer. [https://www.springer.com/cda/content/document/cda\\_downloadaddocument/9783642451812-c2.pdf](https://www.springer.com/cda/content/document/cda_downloadaddocument/9783642451812-c2.pdf).
- Dickson, Paul. 1999. *The Official Rules and Explanations: The Original Guide to Surviving the Electronic Age With Wit, Wisdom, and Laughter*. Federal Street Pr.
- Djulgovic, Benjamin, and Iztok Hozo. 2002. "At What Degree of Belief in a Research Hypothesis Is a Trial in Humans Justified?" *Journal of Evaluation in Clinical Practice*, June 13.
- Dmitriev, Pavel, and Xian Wu. 2016. "Measuring Metrics." *CIKM: Conference on Information and Knowledge Management*. Indianapolis, In. <http://bit.ly/measuringMetrics>.
- Dmitriev, Pavel, Somit Gupta, Dong Woo Kim, and Garnet Vaz. 2017. "A Dirty Dozen: Twelve Common Metric Interpretation Pitfalls in Online Controlled Experiments." *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2017)*. Halifax, NS, Canada: ACM. 1427–1436. <http://doi.acm.org/10.1145/3097983.3098024>.
- Dmitriev, Pavel, Brian Frasca, Somit Gupta, Ron Kohavi, and Garnet Vaz. 2016. "Pitfalls of Long-Term Online Controlled Experiments." *2016 IEEE International Conference on Big Data (Big Data)*. Washington DC. 1367–1376. <http://bit.ly/expLongTerm>.
- Doerr, John. 2018. *Measure What Matters: How Google, Bono, and the Gates Foundation Rock the World with OKRs*. Portfolio.
- Doll, Richard. 1998. "Controlled Trials: the 1948 Watershed." *BMJ*. doi:<https://doi.org/10.1136/bmj.317.7167.1217>.
- Dutta, Kaushik, and Debra Vadermeer. 2018. "Caching to Reduce Mobile App Energy Consumption." *ACM Transactions on the Web (TWEB)*, February 12(1): Article No. 5.
- Dwork, Cynthia, and Aaron Roth. 2014. "The Algorithmic Foundations of Differential Privacy." *Foundations and Trends in Computer Science* 211–407.
- Eckles, Dean, Brian Karrer, and Johan Ugander. 2017. "Design and Analysis of Experiments in Networks: Reducing Bias from Interference." *Journal of Causal Inference* 5(1). [www.deaneckles.com/misc/Eckles\\_Karrer\\_Ugander\\_Reducing\\_Bias\\_from\\_Interference.pdf](http://www.deaneckles.com/misc/Eckles_Karrer_Ugander_Reducing_Bias_from_Interference.pdf).
- Edgington, Eugene S. 1972. "An Additive Method for Combining Probability Values from Independent Experiments." *The Journal of Psychology* 80 (2): 351–363.
- Edmonds, Andy, Ryan W. White, Dan Morris, and Steven M. Drucker. 2007. "Instrumenting the Dynamic Web." *Journal of Web Engineering*. (3): 244–260. [www.microsoft.com/en-us/research/wp-content/uploads/2016/02/edmondsjwe2007.pdf](http://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/edmondsjwe2007.pdf).
- Efron, Bradley, and Robert J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Chapman & Hall/CRC.

- EGAP. 2018. "10 Things to Know About Heterogeneous Treatment Effects." *EGAP: Evidence in Government and Politics*. [egap.org/methods-guides/10-things-heterogeneous-treatment-effects](http://egap.org/methods-guides/10-things-heterogeneous-treatment-effects).
- Ehrenberg, A.S.C. 1975. "The Teaching of Statistics: Corrections and Comments." *Journal of the Royal Statistical Society. Series A* 138 (4): 543–545. <https://www.jstor.org/stable/2345216>.
- Eisenberg, Bryan 2005. "How to Improve A/B Testing." *ClickZ Network*. April 29. [www.clickz.com/clickz/column/1717234/how-improve-a-b-testing](http://www.clickz.com/clickz/column/1717234/how-improve-a-b-testing).
- Eisenberg, Bryan. 2004. *A/B Testing for the Mathematically Disinclined*. May 7. <http://www.clickz.com/showPage.html?page=3349901>.
- Eisenberg, Bryan, and John Quarto-vonTivadar. 2008. *Always Be Testing: The Complete Guide to Google Website Optimizer*. Sybex.
- eMarketer. 2016. "Microsoft Ad Revenues Continue to Rebound." April 20. <https://www.emarketer.com/Article/Microsoft-Ad-Revenues-Continue-Rebound/1013854>.
- European Commission. 2018. [https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules\\_en](https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en).
- European Commission. 2016. EU GDPR.ORG. <https://eugdpr.org/>.
- Fabijan, Aleksander, Pavel Dmitriev, Helena Holmstrom Olsson, and Jan Bosch. 2018. "Online Controlled Experimentation at Scale: An Empirical Survey on the Current State of A/B Testing." *Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. Prague, Czechia. doi:10.1109/SEAA.2018.00021.
- Fabijan, Aleksander, Pavel Dmitriev, Helena Holmstrom Olsson, and Jan Bosch. 2017. "The Evolution of Continuous Experimentation in Software Product Development: from Data to a Data-Driven Organization at Scale." *ICSE '17 Proceedings of the 39th International Conference on Software Engineering*. Buenos Aires, Argentina: IEEE Press. 770–780. doi:<https://doi.org/10.1109/ICSE.2017.76>.
- Fabijan, Aleksander, Jayant Gupchup, Somit Gupta, Jeff Omhover, Wen Qin, Lukas Vermeer, and Pavel Dmitriev. 2019. "Diagnosing Sample Ratio Mismatch in Online Controlled Experiments: A Taxonomy and Rules of Thumb for Practitioners." *KDD '19: The 25th SIGKDD International Conference on Knowledge Discovery and Data Mining*. Anchorage, Alaska, USA: ACM.
- Fabijan, Aleksander, Pavel Dmitriev, Colin McFarland, Lukas Vermeer, Helena Holmström Olsson, and Jan Bosch. 2018. "Experimentation Growth: Evolving Trustworthy A/B Testing Capabilities in Online Software Companies." *Journal of Software: Evolution and Process* 30 (12:e2113). doi:<https://doi.org/10.1002/smr.2113>.
- FAT/ML. 2019. *Fairness, Accountability, and Transparency in Machine Learning*. <http://www.fatml.org/>.
- Fisher, Ronald Aylmer. 1925. *Statistical Methods for Research Workers*. Oliver and Boyd. <http://psychclassics.yorku.ca/Fisher/Methods/>.
- Forte, Michael. 2019. "Misadventures in experiments for growth." The Unofficial Google Data Science Blog. April 16. [www.unofficialgoogledatascience.com/2019/04/misadventures-in-experiments-for-growth.html](http://www.unofficialgoogledatascience.com/2019/04/misadventures-in-experiments-for-growth.html).
- Freedman, Benjamin. 1987. "Equipose and the Ethics of Clinical Research." *The New England Journal of Medicine* 317 (3): 141–145. doi:<https://www.nejm.org/doi/full/10.1056/NEJM198707163170304>.

- Gelman, Andrew, and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* 9 (6): 641–651. doi:10.1177/1745691614551642.
- Gelman, Andrew, and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23 (2): 127–135. [www150.statcan.gc.ca/n1/en/pub/12-001-x/1997002/article/3616-eng.pdf](http://www150.statcan.gc.ca/n1/en/pub/12-001-x/1997002/article/3616-eng.pdf).
- Georgiev, Georgi Zdravkov. 2019. Statistical Methods in Online A/B Testing: Statistics for Data-Driven Business Decisions and Risk Management in e-Commerce. Independently published. [www.abtestingstats.com](http://www.abtestingstats.com)
- Georgiev, Georgi Zdravkov. 2018. "Analysis of 115 A/B Tests: Average Lift is 4%, Most Lack Statistical Power." *Analytics Toolkit*. June 26. <http://blog.analytics-toolkit.com/2018/analysis-of-115-a-b-tests-average-lift-statistical-power/>.
- Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton & Company. <https://www.amazon.com/Field-Experiments-Design-Analysis-Interpretation/dp/0393979954>.
- Goldratt, Eliyahu M. 1990. *The Haystack Syndrome*. North River Press.
- Goldstein, Noah J., Steve J. Martin, and Robert B. Cialdini. 2008. *Yes!: 50 Scientifically Proven Ways to Be Persuasive*. Free Press.
- Goodhart, Charles A. E. 1975. *Problems of Monetary Management: The UK Experience*. Vol. 1, in *Papers in Monetary Economics*, by Reserve Bank of Australia.
- Goodhart's law. 2018. *Wikipedia*. [https://en.wikipedia.org/wiki/Goodhart%27s\\_Law](https://en.wikipedia.org/wiki/Goodhart%27s_Law).
- Goodman, Steven. 2008. "A Dirty Dozen: Twelve P-Value Misconceptions." *Seminars in Hematology*. doi:<https://doi.org/10.1053/j.seminhematol.2008.04.003>.
- Google. 2019. *Processing Logs at Scale Using Cloud Dataflow*. March 19. <https://cloud.google.com/solutions/processing-logs-at-scale-using-dataflow>.
- Google. 2018. *Google Surveys*. <https://marketingplatform.google.com/about/surveys/>.
- Google. 2011. "Ads Quality Improvements Rolling Out Globally." *Google Inside AdWords*. October 3. <https://adwords.googleblog.com/2011/10/ads-quality-improvements-rolling-out.html>.
- Google Console. 2019. "Release App Updates with Staged Rollouts." *Google Console Help*. <https://support.google.com/googleplay/android-developer/answer/6346149?hl=en>.
- Google Developers. 2019. *Reduce Your App Size*. <https://developer.android.com/topic/performance/reduce-apk-size>.
- Google, Helping Advertisers Comply with the GDPR. 2019. *Google Ads Help*. <https://support.google.com/google-ads/answer/9028179?hl=en>.
- Google Website Optimizer. 2008. <http://services.google.com/fhts/siteoptimizer>.
- Gordon, Brett R., Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky. 2018. "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook (forthcoming at Marketing Science)." [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3033144](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3033144).
- Goward, Chris. 2015. "Delivering Profitable 'A-ha!' Moments Everyday." *Conversion Hotel*. Texel, The Netherlands. [www.slideshare.net/webanalisten/chris-goward-strategy-conversion-hotel-2015](http://www.slideshare.net/webanalisten/chris-goward-strategy-conversion-hotel-2015).
- Goward, Chris. 2012. *You Should Test That: Conversion Optimization for More Leads, Sales and Profit or The Art and Science of Optimized Marketing*. Sybex.



- Greenhalgh, Trisha. 2014. *How to Read a Paper: The Basics of Evidence-Based Medicine*. BMJ Books. <https://www.amazon.com/gp/product/B00IPG7GLC>.
- Greenhalgh, Trisha. 1997. "How to Read a Paper : Getting Your Bearings (deciding what the paper is about)." *BMJ* 315 (7102): 243–246. doi:10.1136/bmj.315.7102.243.
- Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. 2016. "Statistical Tests, P Values, Confidence Intervals, and Power: a Guide to Misinterpretations." *European Journal of Epidemiology* 31 (4): 337–350. <https://dx.doi.org/10.1007%2Fs10654-016-0149-3>.
- Grimes, Carrie, Diane Tang, and Daniel M. Russell. 2007. "Query Logs Alone are not Enough." *International Conference of the World Wide Web*, May.
- Grove, Andrew S. 1995. *High Output Management*. 2nd edition. Vintage.
- Groves, Robert M., Floyd J. Fowler Jr, Mick P. Couper, James M. Lepkowski, Singer Eleanor, and Roger Tourangeau. 2009. *Survey Methodology*, 2nd edition. Wiley.
- Gui, Han, Ya Xu, Anmol Bhasin, and Jiawei Han. 2015. "Network A/B Testing From Sampling to Estimation." *WWW '15 Proceedings of the 24th International Conference on World Wide Web*. Florence, IT: ACM. 399–409.
- Gupta, Somit, Lucy Ulanova, Sumit Bhardwaj, Pavel Dmitriev, Paul Raff, and Aleksander Fabijan. 2018. "The Anatomy of a Large-Scale Online Experimentation Platform." *IEEE International Conference on Software Architecture*.
- Gupta, Somit, Ronny Kohavi, Diane Tang, Ya Xu, and etal. 2019. "Top Challenges from the first Practical Online Controlled Experiments Summit." Edited by Xin Luna Dong, Ankur Teredesai and Reza Zafarani. *SIGKDD Explorations* (ACM) 21 (1). <https://bit.ly/OCESummit1>.
- Guyatt, Gordon H., David L. Sackett, John C. Sinclair, Robert Hayward, Deborah J. Cook, and Richard J. Cook. 1995. "Users' Guides to the Medical Literature: IX. A method for Grading Health Care Recommendations." *Journal of the American Medical Association (JAMA)* 274 (22): 1800–1804. doi:<https://doi.org/10.1001%2Fjama.1995.03530220066035>.
- Harden, K. Paige, Jane Mendle, Jennifer E. Hill, Eric Turkheimer, and Robert E. Emery. 2008. "Rethinking Timing of First Sex and Delinquency." *Journal of Youth and Adolescence* 37 (4): 373–385. doi:<https://doi.org/10.1007/s10964-007-9228-9>.
- Harford, Tim. 2014. *The Undercover Economist Strikes Back: How to Run – or Ruin – an Economy*. Riverhead Books.
- Hauser, John R., and Gerry Katz. 1998. "Metrics: You Are What You Measure!" *European Management Journal* 16 (5): 516–528. <http://www.mit.edu/~hauser/Papers/metrics%20you%20are%20what%20you%20measure.pdf>.
- Health and Human Services. 2018a. *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>.
- Health and Human Services. 2018b. *Health Information Privacy*. <https://www.hhs.gov/hipaa/index.html>.
- Health and Human Services. 2018c. *Summary of the HIPAA Privacy Rule*. <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>.
- Hedges, Larry, and Ingram Olkin. 2014. *Statistical Methods for Meta-Analysis*. Academic Press.



- Hemkens, Lars, Despina Contopoulos-Ioannidis, and John Ioannidis. 2016. "Routinely Collected Data and Comparative Effectiveness Evidence: Promises and Limitations." *CMAJ*, May 17.
- HIPAA Journal. 2018. *What is Considered Protected Health Information Under HIPAA*. April 2. <https://www.hipaajournal.com/what-is-considered-protected-health-information-under-hipaa/>.
- Hochberg, Yosef, and Yoav Benjamini. 1995. "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing Series B." *Journal of the Royal Statistical Society* 57 (1): 289–300.
- Hodge, Victoria, and Jim Austin. 2004. "A Survey of Outlier Detection Methodologies." *Journal of Artificial Intelligence Review*. 85–126.
- Hohnhold, Henning, Deirdre O'Brien, and Diane Tang. 2015. "Focus on the Long-Term: It's better for Users and Business." *Proceedings 21st Conference on Knowledge Discovery and Data Mining (KDD 2015)*. Sydney, Australia: ACM. <http://dl.acm.org/citation.cfm?doid=2783258.2788583>.
- Holson, Laura M. 2009. "Putting a Bolder Face on Google." *NY Times*. February 28. <https://www.nytimes.com/2009/03/01/business/01marissa.html>.
- Holtz, David Michael. 2018. "Limiting Bias from Test-Control Interference In Online Marketplace Experiments." *DSpace@MIT*. <http://hdl.handle.net/1721.1/117999>.
- Hoover, Kevin D. 2008. "Phillips Curve." In R. David Henderson, *Concise Encyclopedia of Economics*. <http://www.econlib.org/library/Enc/PhillipsCurve.html>.
- Huang, Jason, David Reiley, and Nikolai M. Raibov. 2018. "David Reiley, Jr." *Measuring Consumer Sensitivity to Audio Advertising: A Field Experiment on Pandora Internet Radio*. April 21. <http://davidreiley.com/papers/PandoraListenerDemandCurve.pdf>.
- Huang, Jeff, Ryen W. White, and Susan Dumais. 2012. "No Clicks, No Problem: Using Cursor Movements to Understand and Improve Search." *Proceedings of SIGCHI*.
- Huang, Yanping, Jane You, Iris Wang, Feng Cao, and Ian Gao. 2015. *Data Science Interviews Exposed*. CreateSpace.
- Hubbard, Douglas W. 2014. *How to Measure Anything: Finding the Value of Intangibles in Business*. 3rd edition. Wiley.
- Huffman, Scott. 2008. *Search Evaluation at Google*. September 15. <https://googleblog.blogspot.com/2008/09/search-evaluation-at-google.html>.
- Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
- Ioannidis, John P. 2005. "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research." *(The Journal of the American Medical Association)* 294 (2).
- Jackson, Simon. 2018. "How Booking.com increases the power of online experiments with CUPED." *Booking.ai*. January 22. <https://booking.ai/how-booking-com-increases-the-power-of-online-experiments-with-cuped-995d186fff1d>.
- Joachims, Thorsten, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. "Accurately Interpreting Clickthrough Data as Implicit Feedback." *SIGIR*, August.
- Johari, Ramesh, Leonid Pekelis, Pete Koomen, and David Walsh. 2017. "Peeking at A/B Tests." *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Halifax, NS, Canada: ACM. 1517–1525. doi:<https://doi.org/10.1145/3097983.3097992>.

- Kaplan, Robert S., and David P. Norton. 1996. *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business School Press.
- Katzir, Liran, Edo Liberty, and Oren Somekh. 2012. "Framework and Algorithms for Network Bucket Testing." *Proceedings of the 21st International Conference on World Wide Web* 1029–1036.
- Kaushik, Avinash. 2006. "Experimentation and Testing: A Primer." *Occam's Razor*. May 22. [www.kaushik.net/avinash/2006/05/experimentation-and-testing-a-primer.html](http://www.kaushik.net/avinash/2006/05/experimentation-and-testing-a-primer.html).
- Keppel, Geoffrey, William H. Saufley, and Howard Tokunaga. 1992. *Introduction to Design and Analysis*. 2nd edition. W.H. Freeman and Company.
- Kesar, Alhan. 2018. *11 Ways to Stop FOOC'ing up your A/B tests*. August 9. [www.widerfunnel.com/stop-fooc-ab-tests/](http://www.widerfunnel.com/stop-fooc-ab-tests/).
- King, Gary, and Richard Nielsen. 2018. *Why Propensity Scores Should Not Be Used for Matching*. Working paper. <https://gking.harvard.edu/publications/why-propensity-scores-should-not-be-used-formatching>.
- King, Rochelle, Elizabeth F. Churchill, and Caitlin Tan. 2017. *Designing with Data: Improving the User Experience with A/B Testing*. O'Reilly Media.
- Kingston, Robert. 2015. *Does Optimizely Slow Down a Site's Performance*. January 18. <https://www.quora.com/Does-Optimizely-slow-down-a-sites-performance/answer/Robert-Kingston>.
- Knapp, Michael S., Juli A. Swinnerton, Michael A. Copland, and Jack Monpas-Huber. 2006. *Data-Informed Leadership in Education*. Center for the Study of Teaching and Policy, University of Washington, Seattle, WA: Wallace Foundation. <https://www.wallacefoundation.org/knowledge-center/Documents/1-Data-Informed-Leadership.pdf>.
- Kohavi, Ron. 2019. "HiPPO FAQ." *ExP Experimentation Platform*. <http://bitly.com/HIPPOExplained>.
- Kohavi, Ron. 2016. "Pitfalls in Online Controlled Experiments." *CODE '16: Conference on Digital Experimentation*. MIT. <https://bit.ly/Code2016Kohavi>.
- Kohavi, Ron. 2014. "Customer Review of A/B Testing: The Most Powerful Way to Turn Clicks Into Customers." *Amazon.com*. May 27. [www.amazon.com/gp/customer-reviews/R44BH2HO30T18](http://www.amazon.com/gp/customer-reviews/R44BH2HO30T18).
- Kohavi, Ron. 2010. "Online Controlled Experiments: Listening to the Customers, not to the HiPPO." *Keynote at EC10: the 11th ACM Conference on Electronic Commerce*. [www.exp-platform.com/Documents/2010-06%20EC10.pptx](http://www.exp-platform.com/Documents/2010-06%20EC10.pptx).
- Kohavi, Ron. 2003. *Real-world Insights from Mining Retail E-Commerce Data*. Stanford, CA, May 22. <http://ai.stanford.edu/~ronnyk/realInsights.ppt>.
- Kohavi, Ron, and Roger Longbotham. 2017. "Online Controlled Experiments and A/B Tests." In *Encyclopedia of Machine Learning and Data Mining*, by Claude Sammut and Geoffrey I Webb. Springer. [www.springer.com/us/book/9781489976857](http://www.springer.com/us/book/9781489976857).
- Kohavi, Ron, and Roger Longbotham. 2010. "Unexpected Results in Online Controlled Experiments." *SIGKDD Explorations*, December. <http://bit.ly/expUnexpected>.
- Kohavi, Ron and Parekh, Rajesh. 2003. "Ten Supplementary Analyses to Improve E-commerce Web Sites." *WebKDD*. <http://ai.stanford.edu/~ronnyk/supplementaryAnalyses.pdf>.

- Kohavi, Ron, and Stefan Thomke. 2017. "The Surprising Power of Online Experiments." *Harvard Business Review* (September–October): 74–92. <http://exp-platform.com/hbr-the-surprising-power-of-online-experiments/>.
- Kohavi, Ron, Thomas Crook, and Roger Longbotham. 2009. "Online Experimentation at Microsoft." *Third Workshop on Data Mining Case Studies and Practice Prize*. <http://bit.ly/expMicrosoft>.
- Kohavi, Ron, Roger Longbotham, and Toby Walker. 2010. "Online Experiments: Practical Lessons." *IEEE Computer*, September: 82–85. <http://bit.ly/expPracticalLessons>.
- Kohavi, Ron, Diane Tang, and Ya Xu. 2019. "History of Controlled Experiments." *Practical Guide to Trustworthy Online Controlled Experiments*. <https://bit.ly/experimentGuideHistory>.
- Kohavi, Ron, Alex Deng, Roger Longbotham, and Ya Xu. 2014. "Seven Rules of Thumb for Web Site." *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. <http://bit.ly/expRulesOfThumb>.
- Kohavi, Ron, Roger Longbotham, Dan Sommerfield, and Randal M. Henne. 2009. "Controlled Experiments on the Web: Survey and Practical Guide." *Data Mining and Knowledge Discovery* 18: 140–181. <http://bit.ly/expSurvey>.
- Kohavi, Ron, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, and Ya Xu. 2012. "Trustworthy Online Controlled Experiments: Five Puzzling Outcomes Explained." *Proceedings of the 18th Conference on Knowledge Discovery and Data Mining*. <http://bit.ly/expPuzzling>.
- Kohavi, Ron, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. 2013. "Online Controlled Experiments at Large Scale." *KDD 2013: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Kohavi, Ron, David Messner, Seth Eliot, Juan Lavista Ferres, Randy Henne, Vignesh Kannappan, and Justin Wang. 2010. "Tracking Users' Clicks and Submits: Trade-offs between User Experience and Data Loss." *Experimentation Platform*. September 28. [www.exp-platform.com/Documents/TrackingUserClicksSubmits.pdf](http://www.exp-platform.com/Documents/TrackingUserClicksSubmits.pdf).
- Kramer, Adam, Jamie Guillory, and Jeffrey Hancock. 2014. "Experimental evidence of massive-scale emotional contagion through social networks." *PNAS*, June 17.
- Kuhn, Thomas. 1996. *The Structure of Scientific Revolutions*. 3rd edition. University of Chicago Press.
- Laja, Peep. 2019. "How to Avoid a Website Redesign FAIL." *CXL*. March 8. <https://conversionxl.com/show/avoid-redesign-fail/>.
- Lax, Jeffrey R., and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53 (1): 107–121. [www.columbia.edu/~jhp2121/publications/HowShouldWeEstimateOpinion.pdf](http://www.columbia.edu/~jhp2121/publications/HowShouldWeEstimateOpinion.pdf).
- Lee, Jess. 2013. *Fake Door*. April 10. [www.jessyoko.com/blog/2013/04/10/fake-doors/](http://www.jessyoko.com/blog/2013/04/10/fake-doors/).
- Lee, Minyong R, and Milan Shen. 2018. "Winner's Curse: Bias Estimation for Total Effects of Features in Online Controlled Experiments." *KDD 2018: The 24th ACM Conference on Knowledge Discovery and Data Mining*. London: ACM.
- Lehmann, Erich, L., and Joseph P. Romano. 2005. *Testing Statistical Hypothesis*. Springer.

- Levy, Steven. 2014. "Why The New Obamacare Website is Going to Work This Time." [www.wired.com/2014/06/healthcare-gov-revamp/](http://www.wired.com/2014/06/healthcare-gov-revamp/).
- Lewis, Randall A, Justin M Rao, and David Reiley. 2011. "Here, There, and Everywhere: Correlated Online Behaviors Can Lead to Overestimates of the Effects of Advertising." Proceedings of the 20th ACM International World Wide Web Conference (WWW). 157–166. <https://ssrn.com/abstract=2080235>.
- Li, Lihong, Wei Chu, John Langford, and Robert E. Schapire. 2010. "A Contextual-Bandit Approach to Personalized News Article Recommendation." *WWW 2010: Proceedings of the 19th International Conference on World Wide Web*. Raleigh, North Carolina. <https://arxiv.org/pdf/1003.0146.pdf>.
- Linden, Greg. 2006. *Early Amazon: Shopping Cart Recommendations*. April 25. <http://glinde.blogspot.com/2006/04/early-amazon-shopping-cart.html>.
- Linden, Greg. 2006. "Make Data Useful." December. <http://sites.google.com/site/glinde/Home/StanfordDataMining.2006-11-28.ppt>.
- Linden, Greg. 2006. "Marissa Mayer at Web 2.0 ." *Geeking with Greg* . November 9. <http://glinde.blogspot.com/2006/11/marissa-mayer-at-web-20.html>.
- Linowski, Jakub. 2018a. *Good UI: Learn from What We Try and Test*. <https://goodui.org/>.
- Linowski, Jakub. 2018b. *No Coupon*. <https://goodui.org/patterns/1/>.
- Liu, Min, Xiaohui Sun, Maneesh Varshney, and Ya Xu. 2018. "Large-Scale Online Experimentation with Quantile Metrics." *Joint Statistical Meeting, Statistical Consulting Section*. Alexandria, VA: American Statistical Association. 2849–2860.
- Loukides, Michael, Hilary Mason, and D.J. Patil. 2018. *Ethics and Data Science*. O'Reilly Media.
- Lu, Luo, and Chuang Liu. 2014. "Separation Strategies for Three Pitfalls in A/B Testing." *KDD User Engagement Optimization Workshop*. New York. [www.ueo-workshop.com/wp-content/uploads/2014/04/Separation-strategies-for-three-pitfalls-in-AB-testing\\_withacknowledgments.pdf](http://www.ueo-workshop.com/wp-content/uploads/2014/04/Separation-strategies-for-three-pitfalls-in-AB-testing_withacknowledgments.pdf).
- Lucas critique. 2018. *Wikipedia*. [https://en.wikipedia.org/wiki/Lucas\\_critique](https://en.wikipedia.org/wiki/Lucas_critique).
- Lucas, Robert E. 1976. *Econometric Policy Evaluation: A Critique*. Vol. 1. In *The Phillips Curve and Labor Markets*, by K. Brunner and A. Meltzer, 19–46. Carnegie-Rochester Conference on Public Policy.
- Malinas, Gary, and John Bigelow. 2004. "Simpson's Paradox." *Stanford Encyclopedia of Philosophy*. February 2. <http://plato.stanford.edu/entries/paradox-simpson/>.
- Manzi, Jim. 2012. *Uncontrolled: The Surprising Payoff of Trial-and-Error for Business, Politics, and Society*. Basic Books.
- Marks, Harry M. 1997. *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990*. Cambridge University Press.
- Marsden, Peter V., and James D. Wright. 2010. *Handbook of Survey Research*, 2nd Edition. Emerald Publishing Group Limited.
- Marsh, Catherine, and Jane Elliott. 2009. *Exploring Data: An Introduction to Data Analysis for Social Scientists*. 2nd edition. Polity.
- Martin, Robert C. 2008. *Clean Code: A Handbook of Agile Software Craftsmanship*. Prentice Hall.
- Mason, Robert L., Richard F. Gunst, and James L. Hess. 1989. *Statistical Design and Analysis of Experiments With Applications to Engineering and Science*. John Wiley & Sons.

- McChesney, Chris, Sean Covey, and Jim Huling. 2012. *The 4 Disciplines of Execution: Achieving Your Wildly Important Goals*. Free Press.
- McClure, Dave. 2007. *Startup Metrics for Pirates: AARRR!!!*. August 8. [www.slide  
share.net/dmc500hats/startup-metrics-for-pirates-long-version](http://www.slideshare.net/dmc500hats/startup-metrics-for-pirates-long-version).
- McClure, Dave. 2007. *Startup Metrics for Pirates: AARRR!!!*. August 8. [www.slide  
share.net/dmc500hats/startup-metrics-for-pirates-long-version](http://www.slide<br/>share.net/dmc500hats/startup-metrics-for-pirates-long-version).
- McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* (142): 698–714.
- McCullagh, Declan. 2006. *AOL's Disturbing Glimpse into Users' Lives*. August 9. [www.cnet.com/news/aols-disturbing-glimpse-into-users-lives/](http://www.cnet.com/news/aols-disturbing-glimpse-into-users-lives/).
- McFarland, Colin. 2012. *Experiment!: Website Conversion Rate Optimization with A/B and Multivariate Testing*. New Riders.
- McGue, Matt. 2014. *Introduction to Human Behavioral Genetics, Unit 2: Twins: A Natural Experiment*. Coursera. [https://www.coursera.org/learn/behavioralge  
netics/lecture/u8Zgt/2a-twins-a-natural-experiment](https://www.coursera.org/learn/behavioralge<br/>netics/lecture/u8Zgt/2a-twins-a-natural-experiment).
- McKinley, Dan. 2013. *Testing to Cull the Living Flower*. January. [http://mcfunley.com/  
testing-to-cull-the-living-flower](http://mcfunley.com/<br/>testing-to-cull-the-living-flower).
- McKinley, Dan. 2012. *Design for Continuous Experimentation: Talk and Slides*. December 22. <http://mcfunley.com/design-for-continuous-experimentation>.
- Mechanical Turk. 2019. *Amazon Mechanical Turk*. <http://www.mturk.com>.
- Meenan, Patrick. 2012. "Speed Index." *WebPagetest*. April. [https://sites.google.com/a/  
webpagetest.org/docs/using-webpagetest/metrics/speed-index](https://sites.google.com/a/<br/>webpagetest.org/docs/using-webpagetest/metrics/speed-index).
- Meenan, Patrick, Chao (Ray) Feng, and Mike Petrovich. 2013. "Going Beyond Onload – How Fast Does It Feel?" *Velocity: Web Performance and Operations* conference, October 14–16. [http://velocityconf.com/velocityny2013/public/sched  
ule/detail/31344](http://velocityconf.com/velocityny2013/public/sched<br/>ule/detail/31344).
- Meyer, Michelle N. 2018. "Ethical Considerations When Companies Study – and Fail to Study – Their Customers." In *The Cambridge Handbook of Consumer Privacy*, by Evan Selinger, Jules Polonetsky and Omer Tene. Cambridge University Press.
- Meyer, Michelle N. 2015. "Two Cheers for Corporate Experimentation: The A/B Illusion and the Virtues of Data-Driven Innovation." *13 Colo. Tech. L.J.* 273. <https://ssrn.com/abstract=2605132>.
- Meyer, Michelle N. 2012. *Regulating the Production of Knowledge: Research Risk–Benefit Analysis and the Heterogeneity Problem*. 65 *Administrative Law Review* 237; Harvard Public Law Working Paper. doi:[http://dx.doi.org/10.2139/  
ssrn.2138624](http://dx.doi.org/10.2139/<br/>ssrn.2138624).
- Meyer, Michelle N., Patrick R. Heck, Geoffrey S. Holtzman, Stephen M. Anderson, William Cai, Duncan J. Watts, and Christopher F. Chabris. 2019. "Objecting to Experiments that Compare Two Unobjectionable Policies or Treatments." *PNAS: Proceedings of the National Academy of Sciences* (National Academy of Sciences). doi:<https://doi.org/10.1073/pnas.1820701116>.
- Milgram, Stanley. 2009. *Obedience to Authority: An Experimental View*. Harper Perennial Modern Thought.
- Mitchell, Carl, Jonathan Litz, Garnet Vaz, and Andy Drake. 2018. "Metrics Health Detection and AA Simulator." *Microsoft ExP (internal)*. August 13. <https://aka.ms/exp/wiki/AASimulator>.

- Moran, Mike. 2008. *Multivariate Testing in Action: Quicken Loan's Regis Hadiaris on multivariate testing*. December. [www.biznology.com/2008/12/multivariate\\_testing\\_in\\_action/](http://www.biznology.com/2008/12/multivariate_testing_in_action/).
- Moran, Mike. 2007. *Do It Wrong Quickly: How the Web Changes the Old Marketing Rules*. IBM Press.
- Mosavat, Fareed. 2019. *Twitter*. Jan 29. <https://twitter.com/far33d/status/1090400421842018304>.
- Mosteller, Frederick, John P. Gilbert, and Bucknam McPeck. 1983. "Controversies in Design and Analysis of Clinical Trials." In *Clinical Trials*, by Stanley H. Shapiro and Thomas A. Louis. New York, NY: Marcel Dekker, Inc.
- MR Web. 2014. "Obituary: Audience Measurement Veteran Tony Twyman." *Daily Research News Online*. November 12. [www.mrweb.com/drno/news/20011.htm](http://www.mrweb.com/drno/news/20011.htm).
- Mudholkar, Govind S., and E. Olusegun George. 1979. "The Logit Method for Combining Probabilities." Edited by J. Rustagi. Symposium on Optimizing Methods in Statistics." Academic Press. 345–366. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a049993.pdf>.
- Mueller, Hendrik, and Aaron Sedley. 2014. "HaTS: Large-Scale In-Product Measurement of User Attitudes & Experiences with Happiness Tracking Surveys." *OZCHI*, December.
- Neumann, Chris. 2017. *Does Optimizely Slow Down a Site's Performance?* October 18. <https://www.quora.com/Does-Optimizely-slow-down-a-sites-performance>.
- Newcomer, Kathryn E., Harry P. Hatry, and Joseph S. Wholey. 2015. *Handbook of Practical Program Evaluation (Essential Tests for Nonprofit and Publish Leadership and Management)*. Wiley.
- Neyman, J. 1923. "On the Application of Probability Theory of Agricultural Experiments." *Statistical Science* 465–472.
- NSF. 2018. *Frequently Asked Questions and Vignettes: Interpreting the Common Rule for the Protection of Human Subjects for Behavioral and Social Science Research*. [www.nsf.gov/bfa/dias/policy/hsfaqs.jsp](http://www.nsf.gov/bfa/dias/policy/hsfaqs.jsp).
- Office for Human Research Protections. 1991. *Federal Policy for the Protection of Human Subjects ('Common Rule')*. [www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html](http://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html).
- Optimizely. 2018. "A/A Testing." *Optimizely*. [www.optimizely.com/optimization-glossary/aa-testing/](http://www.optimizely.com/optimization-glossary/aa-testing/).
- Optimizely. 2018. "Implement the One-Line Snippet for Optimizely X." *Optimizely*. February 28. [https://help.optimizely.com/Set\\_Up\\_Optimizely/Implement\\_the\\_one-line\\_snippet\\_for\\_Optimizely\\_X](https://help.optimizely.com/Set_Up_Optimizely/Implement_the_one-line_snippet_for_Optimizely_X).
- Optimizely. 2018. *Optimizely Maturity Model*. [www.optimizely.com/maturity-model/](http://www.optimizely.com/maturity-model/).
- Orlin, Ben. 2016. *Why Not to Trust Statistics*. July 13. <https://mathwithbaddrawings.com/2016/07/13/why-not-to-trust-statistics/>.
- Owen, Art, and Hal Varian. 2018. *Optimizing the Tie-Breaker Regression Discontinuity Design*. August. <http://statweb.stanford.edu/~owen/reports/tiebreaker.pdf>.
- Owen, Art, and Hal Varian. 2009. *Oxford Centre for Evidence-based Medicine – Levels of Evidence*. March. [www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/](http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/).



- Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 375–385.
- Parmenter, David. 2015. *Key Performance Indicators: Developing, Implementing, and Using Winning KPIs*. 3rd edition. John Wiley & Sons, Inc.
- Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*. 2nd edition. Cambridge University Press.
- Pekelis, Leonid. 2015. "Statistics for the Internet Age: The Story behind Optimizely's New Stats Engine." *Optimizely*. January 20. <https://blog.optimizely.com/2015/01/20/statistics-for-the-internet-age-the-story-behind-optimizelys-new-stats-engine/>.
- Pekelis, Leonid, David Walsh, and Ramesh Johari. 2015. "The New Stats Engine." *Optimizely*. [www.optimizely.com/resources/stats-engine-whitepaper/](http://www.optimizely.com/resources/stats-engine-whitepaper/).
- Pekelis, Leonid, David Walsh, and Ramesh Johari. 2005. *Web Site Measurement Hacks*. O'Reilly Media.
- Peterson, Eric T. 2005. *Web Site Measurement Hacks*. O'Reilly Media.
- Peterson, Eric T. 2004. *Web Analytics Demystified: A Marketer's Guide to Understanding How Your Web Site Affects Your Business*. Celilo Group Media and CafePress.
- Pfeffer, Jeffrey, and Robert I Sutton. 1999. *The Knowing-Doing Gap: How Smart Companies Turn Knowledge into Action*. Harvard Business Review Press.
- Phillips, A. W. 1958. "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957." *Economica, New Series* 25 (100): 283–299. [www.jstor.org/stable/2550759](http://www.jstor.org/stable/2550759).
- Porter, Michael E. 1998. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press.
- Porter, Michael E. 1996. "What is Strategy." *Harvard Business Review* 61–78.
- Quarto-vonTivadar, John. 2006. "AB Testing: Too Little, Too Soon." *Future Now*. [www.futurenowinc.com/abtesting.pdf](http://www.futurenowinc.com/abtesting.pdf).
- Radlinski, Filip, and Nick Craswell. 2013. "Optimized Interleaving For Online Retrieval Evaluation." *International Conference on Web Search and Data Mining*. Rome, IT: ASM. 245–254.
- Rae, Barclay. 2014. "Watermelon SLAs – Making Sense of Green and Red Alerts." *Computer Weekly*. September. <https://www.computerweekly.com/opinion/Watermelon-SLAs-making-sense-of-green-and-red-alerts>.
- RAND. 1955. *A Million Random Digits with 100,000 Normal Deviates*. Glencoe, Ill: Free Press. [www.rand.org/pubs/monograph\\_reports/MR1418.html](http://www.rand.org/pubs/monograph_reports/MR1418.html).
- Rawat, Girish. 2018. "Why Most Redesigns fail." *freeCodeCamp*. December 4. <https://medium.freecodecamp.org/why-most-redesigns-fail-6ecaaf1b584e>.
- Razali, Nornadiah Mohd, and Yap Bee Wah. 2011. "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests." *Journal of Statistical Modeling and Analytics*, January 1: 21–33.
- Reinhardt, Peter. 2016. *Effect of Mobile App Size on Downloads*. October 5. <https://segment.com/blog/mobile-app-size-effect-on-downloads/>.
- Resnick, David. 2015. *What is Ethics in Research & Why is it Important?* December 1. [www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm](http://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm).
- Ries, Eric. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown Business.



- Rodden, Kerry, Hilary Hutchinson, and Xin Fu. 2010. "Measuring the User Experience on a Large Scale: User-Centered Metrics for Web Applications." *Proceedings of CHI*, April. <https://ai.google/research/pubs/pub36299>
- Romano, Joseph, Azeem M. Shaikh, and Michael Wolf. 2016. "Multiple Testing." In *The New Palgrave Dictionary of Economics*. Palgram Macmillan.
- Rosenbaum, Paul R, and Donald B Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55. doi:<http://dx.doi.org/10.1093/biomet/70.1.41>.
- Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. 2004. *Evaluation: A Systematic Approach*. 7th edition. Sage Publications, Inc.
- Roy, Ranjit K. 2001. *Design of Experiments using the Taguchi Approach : 16 Steps to Product and Process Improvement*. John Wiley & Sons, Inc.
- Rubin, Donald B. 1990. "Formal Mode of Statistical Inference for Causal Effects." *Journal of Statistical Planning and Inference* 25, (3) 279–292.
- Rubin, Donald 1974. "Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.
- Rubin, Kenneth S. 2012. *Essential Scrum: A Practical Guide to the Most Popular Agile Process*. Addison-Wesley Professional.
- Russell, Daniel M., and Carrie Grimes. 2007. "Assigned Tasks Are Not the Same as Self-Chosen Web Searches." HICSS'07: 40th Annual Hawaii International Conference on System Sciences, January. <https://doi.org/10.1109/HICSS.2007.91>.
- Saint-Jacques, Guillaume B., Sinan Aral, Edoardo Airoldi, Erik Brynjolfsson, and Ya Xu. 2018. "The Strength of Weak Ties: Causal Evidence using People-You-May-Know Randomizations." 141–152.
- Saint-Jacques, Guillaume, Maneesh Simpson, Jeremy Varshney, and Ya Xu. 2018. "Using Ego-Clusters to Measure Network Effects at LinkedIn." *Workshop on Information Systems and Economics*. San Francisco, CA.
- Samarati, Pierangela, and Latanya Sweeney. 1998. "Protecting Privacy When Disclosing Information: k-anonymity and its Enforcement through Generalization and Suppression." *Proceedings of the IEEE Symposium on Research in Security and Privacy*.
- Schrage, Michael. 2014. *The Innovator's Hypothesis: How Cheap Experiments Are Worth More than Good Ideas*. MIT Press.
- Schrijvers, Ard. 2017. "Mobile Website Too Slow? Your Personalization Tools May Be to Blame." *Bloomreach*. February 2. [www.bloomreach.com/en/blog/2017/01/server-side-personalization-for-fast-mobile-pagespeed.html](http://www.bloomreach.com/en/blog/2017/01/server-side-personalization-for-fast-mobile-pagespeed.html).
- Schurman, Eric, and Jake Brutlag. 2009. "Performance Related Changes and their User Impact." Velocity 09: Velocity Web Performance and Operations Conference. [www.youtube.com/watch?v=bQSE51-gr2s](http://www.youtube.com/watch?v=bQSE51-gr2s) and [www.slideshare.net/dyninc/the-user-and-business-impact-of-server-delays-additional-bytes-and-http-chunking-in-web-search-presentation](http://www.slideshare.net/dyninc/the-user-and-business-impact-of-server-delays-additional-bytes-and-http-chunking-in-web-search-presentation).
- Scott, Steven L. 2010. "A modern Bayesian look at the multi-armed bandit." *Applied Stochastic Models in Business and Industry* 26 (6): 639–658. doi:<https://doi.org/10.1002/asmb.874>.
- Segall, Ken. 2012. *Insanely Simple: The Obsession That Drives Apple's Success*. Portfolio Hardcover.

- Senn, Stephen. 2012. "Seven myths of randomisation in clinical trials." *Statistics in Medicine*. doi:10.1002/sim.5713.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. 2nd edition. Cengage Learning.
- Simpson, Edward H. 1951. "The Interpretation of Interaction in Contingency Tables." *Journal of the Royal Statistical Society, Ser. B*, 238–241.
- Sinofsky, Steven, and Marco Iansiti. 2009. *One Strategy: Organization, Planning, and Decision Making*. Wiley.
- Siroker, Dan, and Pete Koomen. 2013. *A/B Testing: The Most Powerful Way to Turn Clicks Into Customers*. Wiley.
- Soriano, Jacopo. 2017. "Percent Change Estimation in Large Scale Online Experiments." *arXiv.org*. November 3. <https://arxiv.org/pdf/1711.00562.pdf>.
- Souders, Steve. 2013. "Moving Beyond window.onload()." *High Performance Web Sites Blog*. May 13. [www.stevesouders.com/blog/2013/05/13/moving-beyond-window-onload/](http://www.stevesouders.com/blog/2013/05/13/moving-beyond-window-onload/).
- Souders, Steve. 2009. *Even Faster Web Sites: Performance Best Practices for Web Developers*. O'Reilly Media.
- Souders, Steve. 2007. *High Performance Web Sites: Essential Knowledge for Front-End Engineers*. O'Reilly Media.
- Spitzer, Dean R. 2007. *Transforming Performance Measurement: Rethinking the Way We Measure and Drive Organizational Success*. AMACOM.
- Stephens-Davidowitz, Seth, Hal Varian, and Michael D. Smith. 2017. "Super Returns to Super Bowl Ads?" *Quantitative Marketing and Economics*, March 1: 1–28.
- Sterne, Jim. 2002. *Web Metrics: Proven Methods for Measuring Web Site Success*. John Wiley & Sons, Inc.
- Strathern, Marilyn. 1997. "'Improving ratings': Audit in the British University System." *European Review* 5 (3): 305–321. doi:10.1002/(SICI)1234-981X (199707)5:33.0.CO;2-4.
- Student. 1908. "The Probable Error of a Mean." *Biometrika* 6 (1): 1–25. <https://www.jstor.org/stable/2331554>.
- Sullivan, Nicole. 2008. "Design Fast Websites." *Slideshare*. October 14. [www.slideshare.net/stubbornella/designing-fast-websites-presentation](http://www.slideshare.net/stubbornella/designing-fast-websites-presentation).
- Tang, Diane, Ashish Agarwal, Deirdre O'Brien, and Mike Meyer. 2010. "Overlapping Experiment Infrastructure: More, Better, Faster Experimentation." *Proceedings 16th Conference on Knowledge Discovery and Data Mining*.
- The Guardian. 2014. *OKCupid: We Experiment on Users. Everyone does*. July 29. [www.theguardian.com/technology/2014/jul/29/okcupid-experiment-human-beings-dating](http://www.theguardian.com/technology/2014/jul/29/okcupid-experiment-human-beings-dating).
- The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. *The Belmont Report*. April 18. [www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html](http://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html).
- Thistlewaite, Donald L., and Donald T. Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment." *Journal of Educational Psychology* 51 (6): 309–317. doi:<https://doi.org/10.1037%2Fh0044319>.
- Thomke, Stefan H. 2003. "Experimentation Matters: Unlocking the Potential of New Technologies for Innovation."

- Tiffany, Kaitlyn. 2017. "This Instagram Story Ad with a Fake Hair in It is Sort of Disturbing." *The Verge*. December 11. [www.theverge.com/tldr/2017/12/11/16763664/sneaker-ad-instagram-stories-swipe-up-trick](http://www.theverge.com/tldr/2017/12/11/16763664/sneaker-ad-instagram-stories-swipe-up-trick).
- Tolomei, Sam. 2017. *Shrinking APKs, growing installs*. November 20. <https://medium.com/googleplaydev/shrinking-apks-growing-installs-5d3fcb23ce2>.
- Tutterow, Craig, and Guillaume Saint-Jacques. 2019. *Estimating Network Effects Using Naturally Occurring Peer Notification Queue Counterfactuals*. February 19. <https://arxiv.org/abs/1902.07133>.
- Tyler, Mary E., and Jerri Ledford. 2006. *Google Analytics*. Wiley Publishing, Inc.
- Tyurin, I.S. 2009. "On the Accuracy of the Gaussian Approximation." *Doklady Mathematics* 429 (3): 312–316.
- Ugander, Johan, Brian Karrer, Lars Backstrom, and Jon Kleinberg. 2013. "Graph Cluster Randomization: Network Exposure to Multiple Universes." *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 329–337.
- van Belle, Gerald. 2008. *Statistical Rules of Thumb*. 2nd edition. Wiley-Interscience.
- Vann, Michael G. 2003. "Of Rats, Rice, and Race: The Great Hanoi Rat Massacre, an Episode in French Colonial History." *French Colonial History* 4: 191–203. <https://muse.jhu.edu/article/42110>.
- Varian, Hal. 2016. "Causal inference in economics and marketing." *Proceedings of the National Academy of Sciences of the United States of America* 7310–7315.
- Varian, Hal R. 2007. "Kaizen, That Continuous Improvement Strategy, Finds Its Ideal Environment." *The New York Times*. February 8. [www.nytimes.com/2007/02/08/business/08scene.html](http://www.nytimes.com/2007/02/08/business/08scene.html).
- Vaver, Jon, and Jim Koehler. 2012. *Periodic Measurement of Advertising Effectiveness Using Multiple-Test Period Geo Experiments*. Google Inc.
- Vaver, Jon, and Jim Koehler. 2011. *Measuring Ad Effectiveness Using Geo Experiments*. Google, Inc.
- Vickers, Andrew J. 2009. *What Is a p-value Anyway? 34 Stories to Help You Actually Understand Statistics*. Pearson. [www.amazon.com/p-value-Stories-Actually-Understand-Statistics/dp/0321629302](http://www.amazon.com/p-value-Stories-Actually-Understand-Statistics/dp/0321629302).
- Vigen, Tyler. 2018. *Spurious Correlations*. <http://tylervigen.com/spurious-correlations>.
- Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 13 (523): 1228–1242. doi:<https://doi.org/10.1080/01621459.2017.1319839>.
- Wagner, Jeremy. 2019. "Why Performance Matters." *Web Fundamentals*. May. [https://developers.google.com/web/fundamentals/performance/why-performance-matters/#performance\\_is\\_about\\_improving\\_conversions](https://developers.google.com/web/fundamentals/performance/why-performance-matters/#performance_is_about_improving_conversions).
- Wasserman, Larry. 2004. *All of Statistics: A Concise Course in Statistical Inference*. Springer.
- Weiss, Carol H. 1997. *Evaluation: Methods for Studying Programs and Policies*. 2nd edition. Prentice Hall.
- Wider Funnel. 2018. "The State of Experimentation Maturity 2018." *Wider Funnel*. [www.widerfunnel.com/wp-content/uploads/2018/04/State-of-Experimentation-2018-Original-Research-Report.pdf](http://www.widerfunnel.com/wp-content/uploads/2018/04/State-of-Experimentation-2018-Original-Research-Report.pdf).

- Wikipedia contributors, Above the Fold. 2014. *Wikipedia, The Free Encyclopedia*. Jan. [http://en.wikipedia.org/wiki/Above\\_the\\_fold](http://en.wikipedia.org/wiki/Above_the_fold).
- Wikipedia contributors, Cobra Effect. 2019. *Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/wiki/Cobra\\_effect](https://en.wikipedia.org/wiki/Cobra_effect).
- Wikipedia contributors, Data Dredging. 2019. *Data dredging*. [https://en.wikipedia.org/wiki/Data\\_dredging](https://en.wikipedia.org/wiki/Data_dredging).
- Wikipedia contributors, Eastern Air Lines Flight 401. 2019. *Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/wiki/Eastern\\_Air\\_Lines\\_Flight\\_401](https://en.wikipedia.org/wiki/Eastern_Air_Lines_Flight_401).
- Wikipedia contributors, List of .NET libraries and frameworks. 2019. [https://en.wikipedia.org/wiki/List\\_of\\_.NET\\_libraries\\_and\\_frameworks#Logging\\_Frameworks](https://en.wikipedia.org/wiki/List_of_.NET_libraries_and_frameworks#Logging_Frameworks).
- Wikipedia contributors, Logging as a Service. 2019. *Logging as a Service*. [https://en.wikipedia.org/wiki/Logging\\_as\\_a\\_service](https://en.wikipedia.org/wiki/Logging_as_a_service).
- Wikipedia contributors, Multiple Comparisons Problem. 2019. *Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/wiki/Multiple\\_comparisons\\_problem](https://en.wikipedia.org/wiki/Multiple_comparisons_problem).
- Wikipedia contributors, Perverse Incentive. 2019. [https://en.wikipedia.org/wiki/Perverse\\_incentive](https://en.wikipedia.org/wiki/Perverse_incentive).
- Wikipedia contributors, Privacy by Design. 2019. *Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/wiki/Privacy\\_by\\_design](https://en.wikipedia.org/wiki/Privacy_by_design).
- Wikipedia contributors, Semmelweis Reflex. 2019. *Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/wiki/Semmelweis\\_reflex](https://en.wikipedia.org/wiki/Semmelweis_reflex).
- Wikipedia contributors, Simpson's Paradox. 2019. *Wikipedia, The Free Encyclopedia*. Accessed February 28, 2008. [http://en.wikipedia.org/wiki/Simpson%27s\\_paradox](http://en.wikipedia.org/wiki/Simpson%27s_paradox).
- Wolf, Talia. 2018. "Why Most Redesigns Fail (and How to Make Sure Yours Doesn't)." *GetUplift*. <https://getuplift.co/why-most-redesigns-fail>.
- Xia, Tong, Sumit Bhardwaj, Pavel Dmitriev, and Aleksander Fabijan. 2019. "Safe Velocity: A Practical Guide to Software Deployment at Scale using Controlled Rollout." *ICSE: 41st ACM/IEEE International Conference on Software Engineering*. Montreal, Canada. [www.researchgate.net/publication/333614382\\_Safe\\_Velocity\\_A\\_Practical\\_Guide\\_to\\_Software\\_Deployment\\_at\\_Scale\\_using\\_Controlled\\_Rollout](http://www.researchgate.net/publication/333614382_Safe_Velocity_A_Practical_Guide_to_Software_Deployment_at_Scale_using_Controlled_Rollout).
- Xie, Huizhi, and Juliette Auriisset. 2016. "Improving the Sensitivity of Online Controlled Experiments: Case Studies at Netflix." *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM. 645–654. <http://doi.acm.org/10.1145/2939672.2939733>.
- Xu, Ya, and Nanyu Chen. 2016. "Evaluating Mobile Apps with A/B and Quasi A/B Tests." *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA: ACM. 313–322. <http://doi.acm.org/10.1145/2939672.2939703>.
- Xu, Ya, Weitao Duan, and Shaochen Huang. 2018. "SQR: Balancing Speed, Quality and Risk in Online Experiments." *24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. London: Association for Computing Machinery. 895–904.
- Xu, Ya, Nanyu Chen, Adrian Fernandez, Omar Sinno, and Anmol Bhasin. 2015. "From Infrastructure to Culture: A/B Testing Challenges in Large Scale Social Networks." *KDD '15: Proceedings of the 21th ACM SIGKDD International*

- Conference on Knowledge Discovery and Data Mining*. Sydney, NSW, Australia: ACM. 2227–2236. <http://doi.acm.org/10.1145/2783258.2788602>.
- Yoon, Sangho. 2018. *Designing A/B Tests in a Collaboration Network*. [www.unofficialgoogledatascience.com/2018/01/designing-ab-tests-in-collaboration.html](http://www.unofficialgoogledatascience.com/2018/01/designing-ab-tests-in-collaboration.html).
- Young, S. Stanley, and Allan Karr. 2011. “Deming, data and observational studies: A process out of control and needing fixing.” *Significance* 8 (3).
- Zhang, Fan, Joshy Joseph, and Alexander James, Zhuang, Peng Rickabaugh. 2018. Client-Side Activity Monitoring. *US Patent US 10,165,071 B2*. December 25.
- Zhao, Zhenyu, Miao Chen, Don Matheson, and Maria Stone. 2016. “Online Experimentation Diagnosis and Troubleshooting Beyond AA Validation.” *DSAA 2016: IEEE International Conference on Data Science and Advanced Analytics*. IEEE. 498–507. doi:<https://ieeexplore.ieee.org/document/7796936>.