# Replication and Evidence Factors in Observational Studies

Paul Rosenbaum, Wharton, U of Pennsylvania

The talk follows [23,26] and [25, Chapter 20]. See [18] for replication, [11, 20, 21] for evidence factors.

**Causal effects are not identified without random assignment of treatments** An association between treatment received and outcome exhibited may reflect either an effect caused by the treatment or a bias in who is selected to receive treatment.

**Biases can replicate** Replication should study the same treatment, changing the likely biases, hoping the effect reappears.

**Can an observational study replicate itself?** If replication is not repetition, can one study replicate itself by analyzing the same data twice?

**One very simple example: treatment/control matched pairs, with varied doses of treatment among the pairs** A randomized experiment could randomly assign treatment/control within pairs and randomly assign doses among pairs, so the group of permutations that assigns treatments to individuals factors into the product of two subgroups.

**General structure** Individuals, $1, \ldots, n$, are assigned to $n$ treatment positions by a finite group $\mathfrak{G}$ of $n \times n$ permutation matrices $\mathbf{g}$, where $\mathfrak{G}$ has a subgroup $\mathfrak{H}$. In the example, $\mathfrak{G}$ keeps the pairs intact, changes treatment/control inside pairs and permutes doses among pairs. Subgroup $\mathfrak{H}$ permutes doses among pairs, leaving treatment/control unchanged. We may always find a set $\mathfrak{K} \subseteq \mathfrak{G}$ of distinct representatives of the cosets of $\{\mathfrak{H}\mathbf{g} : \mathbf{g} \in \mathfrak{G}\}$ in $\mathfrak{G}$, so that each $\mathbf{g} \in \mathfrak{G}$ has a *unique* representation as $\mathbf{g} = \mathbf{hk}$ with $\mathbf{h} \in \mathfrak{H}$ and $\mathbf{k} \in \mathfrak{K}$.

**Probability of various treatment assignments** Write $|\mathcal{S}|$ for the number of elements in a finite set $\mathcal{S}$. A probability distribution on $\mathbf{k} \in \mathfrak{K}$ is a vector $\mathbf{p} = \left(p_{\mathbf{k}_1}, \ldots, p_{\mathbf{k}_{|\mathfrak{K}|}}\right)$, with $p_{\mathbf{k}_\ell} \geq 0$, $1 = \sum_{\ell=1}^{|\mathfrak{K}|} p_{\mathbf{k}_\ell}$. A distribution on $\mathbf{h} \in \mathfrak{h}$ is a vector $\mathbf{p}' = \left(p_{\mathbf{h}_1}, \ldots, p_{\mathbf{h}_{|\mathfrak{H}|}}\right)$. A distribution on the treatment assignments $\mathbf{g} \in \mathfrak{G}$ is a vector $\mathbf{p}'' = \left(p_{\mathbf{g}_1}, \ldots, p_{\mathbf{g}_{|\mathfrak{G}|}}\right)$. Randomized assignment uses $\mathbf{p}'' = \left(|\mathfrak{G}|^{-1}, \ldots, |\mathfrak{G}|^{-1}\right)$.

**The null hypothesis of no treatment effect, $H_0$** The hypothesis $H_0$ of no effect asserts that changing your treatment does not change your outcome. Under $H_0$, write a test statistic as a function of the treatment assignment, $t(\mathbf{g})$, as outcomes are merely permuted. In a randomized experiment with realized assignment $\mathbf{G} \in \mathfrak{G}$, the $P$-value using $t(\mathbf{g})$ is the random variable $\left|\{\mathbf{g} \in \mathfrak{G} : t(\mathbf{g}) \geq t(\mathbf{G})\}\right| / |\mathfrak{G}|$. A level $\alpha$ test of $H_0$ may be inverted to obtain a $1-\alpha$ confidence set, an equivalence test, a 3-sided test.

**Sensitivity analysis for a test of $H_0$** If we do not randomly assign treatments, then we do not know the distribution of treatment assignments. Consider a set $\mathcal{P}_\Gamma$ whose elements are distributions $\mathbf{p}$ on $\mathfrak{K}$. Can we reject $H_0$ for all $\mathbf{p} \in \mathcal{P}_\Gamma$? For each $\mathbf{p} \in \mathcal{P}_\Gamma$, we may compute a $P$-value as the probability that $t(\mathbf{k})$ exceeds its realized value $t(\mathbf{K})$, and we reject at level $\alpha$ if the largest of these, say $\overline{P}_\Gamma$, is $\overline{P}_\Gamma \leq \alpha$; then, if $\mathbf{p} \in \mathcal{P}_\Gamma$, the probability that we falsely reject $H_0$ is at most $\alpha$. A sensitivity analysis does this for a nested sequence of $\mathcal{P}_\Gamma$'s, with $\mathcal{P}_\Gamma \subseteq \mathcal{P}_{\Gamma^*}$ if $\Gamma \leq \Gamma^*$, reporting the results to be sensitive to the smallest $\Gamma$ that leads to acceptance of $H_0$.

**Method** Test $H_0$ twice in two separate sensitivity analyses, assuming only what each sensitivity analysis separately assumes. Assume $\mathbf{p} \in \mathcal{P}_\Gamma$, and test $H_0$ using a statistic $t(\mathbf{g}) = t(\mathbf{hk})$ that is invariant with respect to $\mathfrak{H}$ in the sense that $t(\mathbf{hk}) = t(\mathbf{k})$ for all $\mathbf{h}$, $\mathbf{k}$, obtaining the maximum $P$-value, $\overline{P}_\Gamma$. Test $H_0$ again using the conditional distributions $\mathbf{p}' \in \mathcal{P}'_{\Gamma'}$ of $t'(\mathbf{g}) = t'(\mathbf{hk})$ of $\mathbf{H}$ given $\mathbf{K} = \mathbf{k}$, obtaining the maximum $\overline{P}'_{\Gamma'}$ of the conditional $P$-values. Pointedly, *do not* assume $\mathbf{H}$ and $\mathbf{K}$ are independent.

**Proposition**: If $H_0$ is true, if $\Pr(\mathbf{K} = \mathbf{k})$ is one of the distributions $\mathbf{p} \in \mathcal{P}_\Gamma$, if $\Pr(\mathbf{H} = \mathbf{h} \,|\, \mathbf{K} = \mathbf{k})$ is one of the distributions $\mathbf{p}' \in \mathcal{P}'_{\Gamma'}$, then $\left(\overline{P}_\Gamma, \overline{P}'_{\Gamma'}\right)$ is stochastically larger than the uniform distribution on the unit square, so that $\Pr\left(\overline{P}_\Gamma \leq \alpha, \overline{P}'_{\Gamma'} \leq \alpha'\right) \leq \alpha\alpha'$ for all $0 \leq \alpha \leq 1$, $0 \leq \alpha' \leq 1$.

**At the keyboard**

```
library(DOS2)
data("periodontal")
attach(periodontal)
y<-pcteither[z==1]-pcteither[z==0]
x<-cigsperday[z==1]
senWilcox(y,gamma=2.75)
crosscutplot(x,y,ct=.2)
crosscut(x,y,ct=.2,gamma=1.6)
sensitivitymv::truncatedP(c(0.04651554,0.04433723))
```

# Bibliography


[1] Bailar, JC, Gornik, HL. Cancer undefeated. NEJM 1997;336:1569-1574. Quote: p1571.

[2] Card D, Krueger AB. Minimum wages and employment. Am Econ Rev 1994;84:772-93.

[3] Cochran, WG. Planning of observational studies of human populations (with Discussion). JRSS-A 1965;128:234-266. Quote: pp 252-3.

[4] Crama Y, Spieksma FC. Approximation algorithms for three-dimensional assignment problems with triangle inequalities. Euro J Oper Res 1992;60(3):273-9.

[5] Dwass, M. Some k-sample rank-order tests. In: I. Olkin, ed., *Contributions to Probability and Statistics in Honor of Hotelling*, Stanford, 1960.

[6] Fisher RA. Design of Experiments. 1935.

[7] Haack, S. Evidence and Inquiry. Blackwell 1995.

[8] Hubbard, RL. et al. Overview of 1-year follow-up... Psych Addict Behav 1997;11;261-78.

[9] Karmakar B, French B, Small DS. Integrating the evidence from evidence factors in observational studies. Biometrika. 2019;106:353-367.

[10] Karmakar, B, Small, DS, Rosenbaum, PR. Using evidence factors to clarify exposure biomarkers. Am J Epidemiol 2020;189:243-249.

[11] Karmakar B, Small DS, Rosenbaum PR. Reinforced designs: Multiple instruments plus control groups as evidence factors... JASA, 2020.

[12] Karmakar B, Small DS, Rosenbaum PR. Using approximation algorithms to build evidence factors and related designs for observational studies. J Comp Graph Statist 2019;28;698-709.

[13] Kopjar N, Garaj-Vrhovac V. Application of the alkaline comet assay in human biomonitoring for genotoxicity. Mutagenesis 2001;16:71-8.

[14] Manski, CF et al. Assessment of Two Cost-Effectiveness Studies on Cocaine Control Policy. Washington, DC: NAS Press, 1999.

[15] Marden JI. Use of nested orthogonal contrasts in analyzing rank data. JASA 1992;87:307-18.

[16] Olmstead PS, Tukey JW. A corner test for association. Ann Math Statist 1947;18:495-513.

[17] Polya, G. Heuristic reasoning and the theory of probability. Am Math Monthly, 1941;48:450-465.

[18] Rosenbaum PR. Replicating effects and biases. Am Statist 2001;55:223-227.

[19] Rosenbaum PR, Silber JH. Amplification of sensitivity analysis in observational studies. JASA 2009;104:1398-1405. `amplify` in `sensitivitymv`

[20] Rosenbaum PR. Evidence factors in observational studies. Biometrika. 2010;97:333-345.

[21] Rosenbaum PR. Some approximate evidence factors in observational studies. JASA 2011;106:285-295.

[22] Rosenbaum PR. The cross-cut statistic and its sensitivity to bias in observational studies with ordered doses of treatment. Biometrics 2016;72:175-183.

[23] Rosenbaum PR. The general structure of evidence factors in observational studies. Statist Sci 2017;32:514-530.

[24] Rosenbaum PR. Sensitivity analysis for stratified comparisons .... Ann Appl Statist 2018;12:2312-2334.

[25] Rosenbaum PR. Design of Observational Studies. Springer, 2020, $2^{nd}$ ed. §20. `R` Package `DOS2`

[26] Rosenbaum PR. Replication and Evidence Factors in Observational Studies. Chapman and Hall/CRC, in prepartion. `R` Package `evident`

[27] Tomar SL, Asma S. Smoking-attributable periodontitis in the US. J Periodont 2000;71:743-751.

[28] Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS. Truncated product method of combining P-values. Genet Epidemiol 2002;22:170-185.