Bayesian Analyses of Rubin Causal Model--Application to separating active treatment effects from placebo effects

Donald B. Rubin

Professor, Yau Mathematical Sciences Center, Tsinghua University

Murray Shusterman Senior Research Fellow, Fox School of Business

Can we disentangle causal effects of a treatment from "placebo effects" in placebo controlled randomized, double-blind, trails?

- "Placebo effects" = effects created when conscious humans think that they may be exposed to an active treatment (with probability π, as in 50/50 RCT), but are actually randomly exposed only to an inert placebo, i.e., a treatment condition with no possible real effect on the outcome Y.
- Important for understanding recommendations for medical practice.
  - —Careful formulation of the problem is needed
  - —Bayesian modeling ideas are needed
  - —Modern computing is needed to implement
  - —Presentation here is conceptual, limited technical details.

Prologue concerning personal influences on my approach to causal inference – idiosyncratic background but important for understanding source of ideas

- My introduction to causal inference as a kid:

--- Physics – John Wheeler 1961; Einstein, von Neumann

--- Psychology & consciousness – Julian Jaynes 1964 (JJ); Freud, Skinner

--- Experimental design – William Cochran 1968; Fisher, Neyman

Treatments, factors, levels of a factor, RCTs

- Clear Separation Between

--- Science = object of inference

--- What is done to learn about the science

  - Ideally: intervene to measure aspects at a point in time

- Same notation for and representation of science no matter how we try to learn about it or measure it – Science exists, we observe
- Missing data always complicate our inference from observations

  --- Cannot go back in time to observe the past science under a counterfactual intervention, but maybe can predict what past science would have been

  --- Natural to me because of previous exposure to Heisenberg uncertainty principle & observer effect ideas – Physics and quantum mechanics
- Start by defining estimands, not estimators – For causal inference, need treatments and outcomes
- Causal effects are comparisons of potential outcomes Y(1) versus Y(0), under different treatments, W=1 versus W=0, for each unit (person)
- With two factors, W and Z, [Y(0), Y(1)] replaced by [Y(W,Z): Y(0,0), Y(0,1), Y(1,0), Y(1,1)]. These define comparisons of scientific interest

Contributions of Rubin Causal model: RCM (1974, 1975, 1976, 1977, 1978)

- Causal estimands are comparisons of potential outcomes on a common set of experimental units (people), possibly conditional on covariates, age, sex,…
- Assignment mechanism, which creates missing and observed potential outcomes, is needed for causal inference:

  Probability for treatment indicator W given science; science = [X, Y(0), Y(1)]

  Pr(W|X,Y(0),Y(1)), general dependence on covariates X, *and* potential outcomes Y(0) and Y(1)

  Unconfounded = Pr(W|X), as in RCTs

  Ignorable = Pr(W|X, $\mathrm{Y_{obs}}$), as in sequential RCTs

  $\mathrm{Y_{obs}}$ = observed values of Y(0) and Y(1)

  —This is the basis for traditional Fisher-Neyman causal inference, although never formalized this way before Rubin (1974, 1975)
- Bayes = model the science (X,Y(0),Y(1)) in addition to the assignment mechanism
  —Artistic touch is needed here because all models are wrong: von Neumann; Box; Picasso
  —But models are needed to make progress in complicated problems

Historical Comments on these statistical insights originating with Neyman and Fisher, but with no formal Bayesian aspects from either statistician

- Potential outcomes: 20th Century insights, like those in quantum mechanics
  --- Causal estimands are defined in terms of measurable quantities, which are not simultaneously measurable, even theoretically (e.g., position, momentum; outcome Y(0) and outcome Y(1))
- After R. A. Fisher (1925, 1935) and Neyman (1923, 1935), RCTs quickly dominated agriculture and animal breeding – throughout Commonwealth and US
  --- More applied work (e.g., Kempthorne, Cochran & G Cox, Box, D Cox)
  --- Supporting mathematical work (e.g., ISI Mahalanobis, Bose, Nair, Rao, Roy)
- Subsequently RCTs entered western industry with physical objects for units
  --- Post WW II: (e.g., Deming Medal in Japan, 1951, for QC)
- But insights and applications limited to RCTs with non-conscious units (JJ)

Transportation of insights to double-blind, placebo-controlled RCTs with conscious units, humans, first used in medicine to estimate medical effects of treatments

- UK in 1946 MRC & Hill – strep
- Salk vaccine RCT in US – 1954
- US FDA and pharma – Paul Meier 1950s
- Overzealous adherence to using ITT in double-blind, placebo-controlled RCTs to estimate the *true effect* of *assignment* to drug rather than the medical effect of *assignment to and receip*t of drug – not medically wise?
- Incorrect allowance for consciousness of human beings

Defining “Active Treatment Effects” and “Placebo Effects” using a 2x2 Factorial Study

- W = blinding to treatment received, with levels: b=blinded, u=unblinded

  -- Example: Shadish et al. (JASA 2008)

- Z = treatment actually received, with levels: a=active, p=placebo
- More generally, levels of W are the (revealed to the subject) true probability π of being treated with active versus placebo

  - In simplest RCT, prob(active)=1/2, prob(placebo)=1/2

  - Actual effect of W=b versus W=u can depend on π

  - e.g., If π=.99, nearly certain you receive Z=active, so you react that way

- Y=outcome variable, e.g., blood pressure, test score
- Potential outcomes = Y(W=w, Z=z):
  - when blinded Y(b,a), Y(b,p)
  - when unblinded Y(u,a), Y(u,p)

For Drug Approval at FDA

- T(X) = “true” medical treatment effect relative to placebo as a function of patients’ covariates X, in blinded arm
  - —T(X)=[Y(b,a)-Y(b,p)|X]
  - —Condition on X to ensure that drug is safe (defined by non-Y outcomes, such as cardiac events) and effective (defined by Y) for subgroups defined by age, sex, mental status,...
  - —Information used for the label, warnings, to develop improved drugs
- T(X) ignores (or eliminates) “placebo effects”, effect of W=b versus W=u, because W fixed at blinded
  - —Placebo effect = P(X)=[Y(u,p)-Y(b,p)|X]
  - —Better called “blinding effect” than “placebo effect”?

To create more precise information on label or to inform better medical practice, we also need to estimate:

- M(X) = expected medical effect in practice, as a function of covariates X,
  - —M(X)=[Y(u,a)-Y(u,p)|X]
  - —Directly estimable in unblinded arm of 2x2 factorial RCT
  - —But this is not done in common practice because usually no 2x2 RCT
  - —Only results in blinded arm are needed for approval – Medically naive?

Improve medical practice

- Important implications for personalized medicine using covariates to improve patient outcomes in practice, which is unblinded
- Medical outcomes reflect placebo effects of treatment, because doctors prescribe and patients know the prescription – W=u
- All meds have side effects, so keep doses low
- FDA history of dose reductions in time, after approval
  - Initial approved dose and label based on double-blind, placebo-controlled RCTs, analyzed by Intention To Treat (ITT=as randomized)
  - ITT in blinded studies ignores non-compliance with assigned treatment and ignores results in unblinded comparison, usually not there

- Can M(X), and thus more precise medical recommendations for patient with X, be estimated from blinded arm of 2x2 factorial?
  - —Need medical/scientific assumptions to replace lack of data from unblinded arm
  - —Avoid collecting extra data from unblinded arm just to save $$$
- Key is to think about “placebo effect”, which connects scientific results in the unblinded arm and the blinded arm

P(X)=[Y(b,p)-Y(u,p)|X]=effect, when actually exposed to placebo, of your being told you might be exposed to active (with prob π) versus your knowing you are getting placebo

  - —P(X) directly estimable in placebo exposed arm of the 2x2 trial: compare Y among those blinded, and thus think that they are possibly receiving active, with those knowing they are getting placebo

General Modelling Strategy for Estimating M(X)=[Y(u,a)-Y(u,p)|X] from Studies with Only a Blinded Arm

- Explicate assumptions that can be used to “replace” unblinded data
- Key statistical tools: principal stratification (Frangakis and Rubin, 2002) and mixture modelling, specifically, approach of Jin and Rubin (2008) on non-compliance and dose-response
- Define continuous principal strata according to effect of blinding, often called the “placebo effect”, [Y(b,p)-Y(u,p)|X]
- Bayesian “Regression” models to predict the missing potential outcomes under assumptions about effect of blinding and smoothness of relationships, including placebo effect

Uncontroversial Assumptions for Estimation from Blinded Single-arm Studies

- Unconfounded treatment assignment mechanism – RCT
- SUTVA (Rubin 1980)
- Take baseline measurement of outcome (pre-randomization) and assume Y(u,p) = baseline;
- Effectively define Y to be change from baseline, assume Y(u,p)=0
  - —This assumption may be reasonable if short time span between baseline measurement and time of measurement of outcome in study
  - —Rationale: If you know you are getting nothing active, why should anything change?

Often Reasonable Assumptions About Blinding

- One-sided effect of blinded versus unblinded when assigned placebo

  $[Y(b,p)|X] \geq [Y(u,p)|X]$

  —Reasoning: *Knowing* that you are assigned placebo (right side), should give smaller effect on Y than *knowing only that you might be assigned placebo* (left side)

- One-sided effect of blinded versus unblinded when assigned active

  $[Y(b,a)|X] \leq [Y(u,a)|X]$

  —Reasoning: *Knowing* you are assigned active (right side) should give larger effect on Y than *knowing only that you might be assigned active* (left side).

- These should hold for any fixed π, and should be monotonic in π:

  —Reasoning: effect of blinded versus unblinded grows as the probability of being exposed to placebo grows. George et al. (2017)

- Effect of blinding under placebo Y(b,p)-Y(u,p) should be considered a psychological characteristic of the unit, like compliance behavior (pill count) when blinded and assigned placebo – an important covariate
- Similar example: Efron-Feldman (1979), re-analyzed in Jin and Rubin (2008), but with two outcomes in one-factor blinded study; active versus placebo:
  -- Y_1 = Cholesterol Reduction from baseline, Y_1(a), Y_1(p)
  -- Y_2 = Compliance, proportion of assigned pills taken, Y_2(a), Y_2(p)

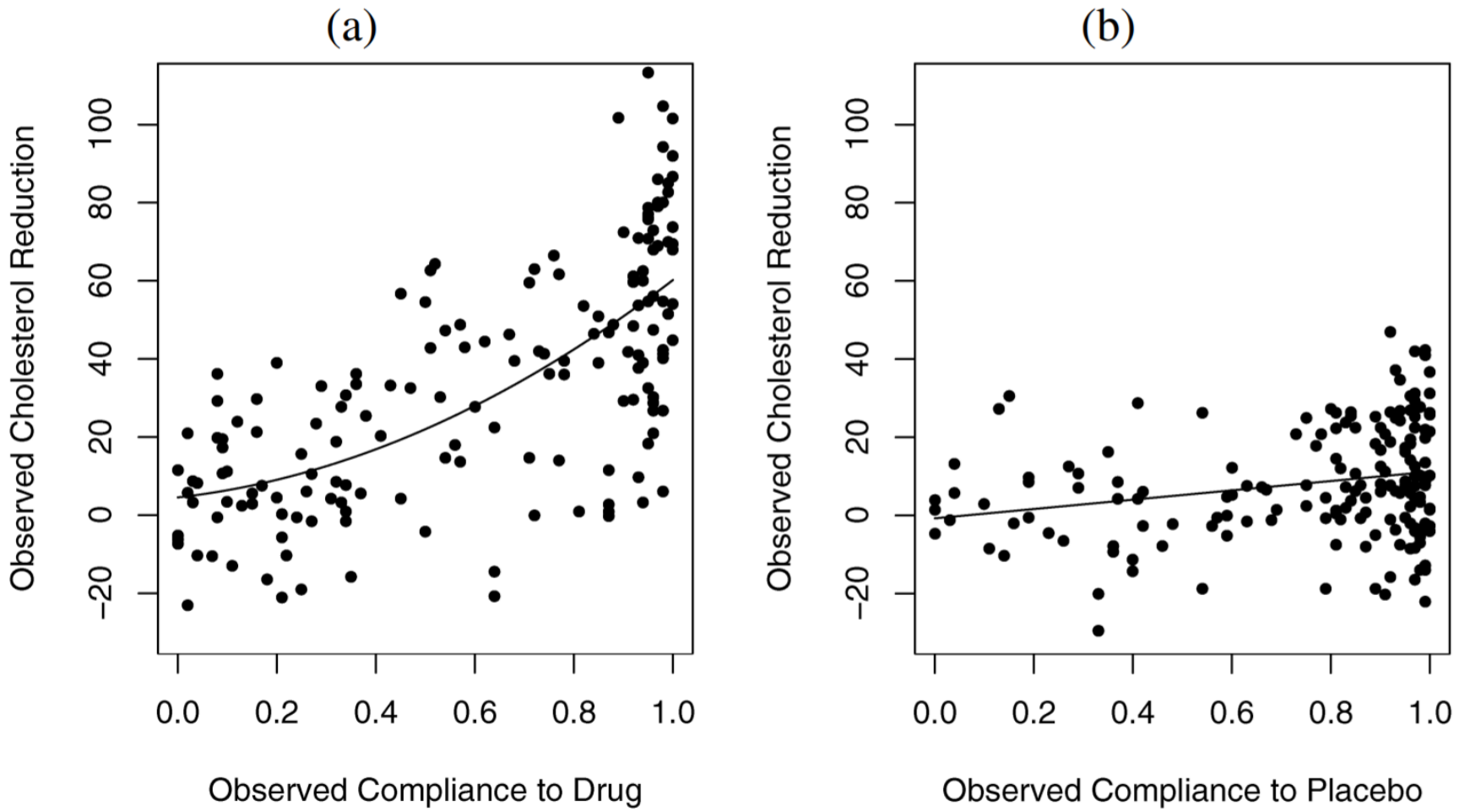


Figure 1. From Efron-Feldman: Y_1 versus Y_2
Dose-response. (a) active treatment; (b) placebo group.

Imperfect Blind Revealed

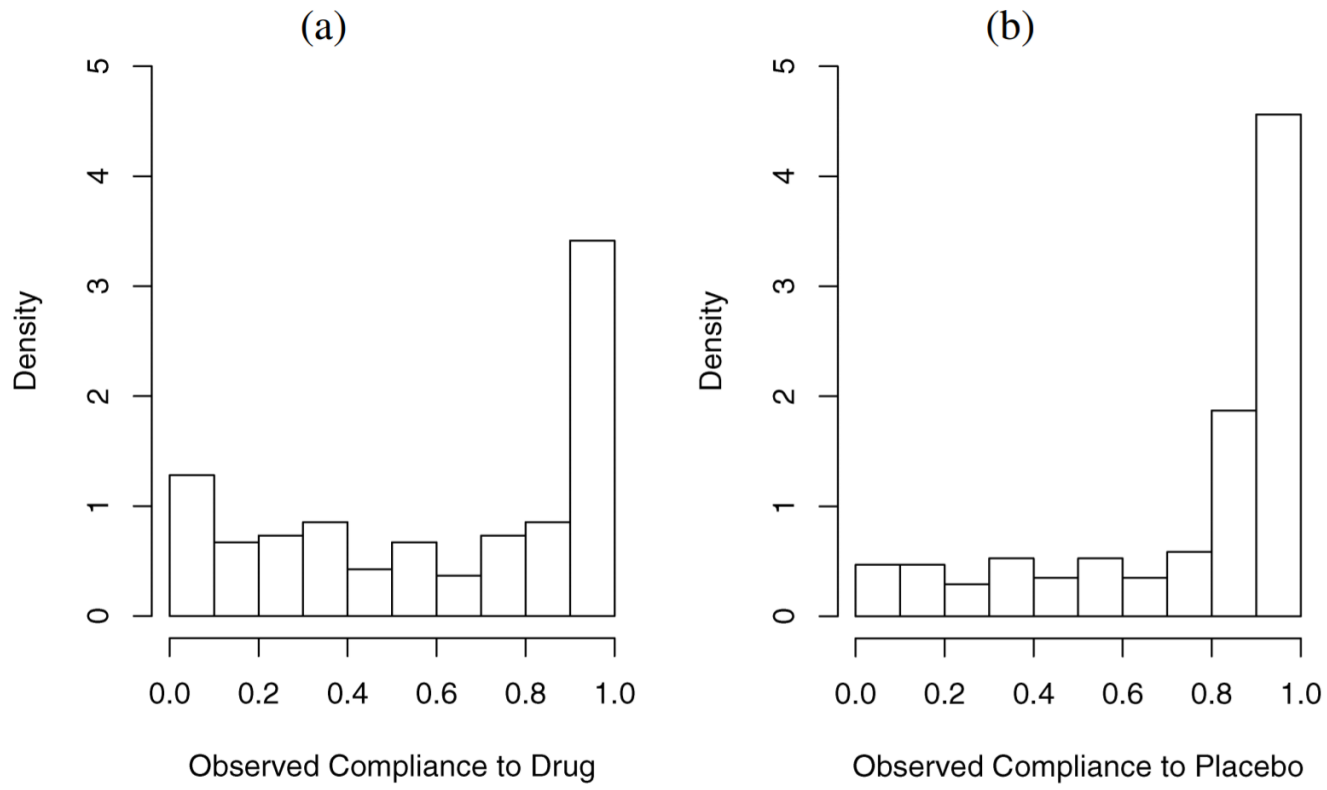


Figure 2. Histograms of observed drug compliance and observed placebo compliance in blinded study. (a) treatment group; (b) placebo group.

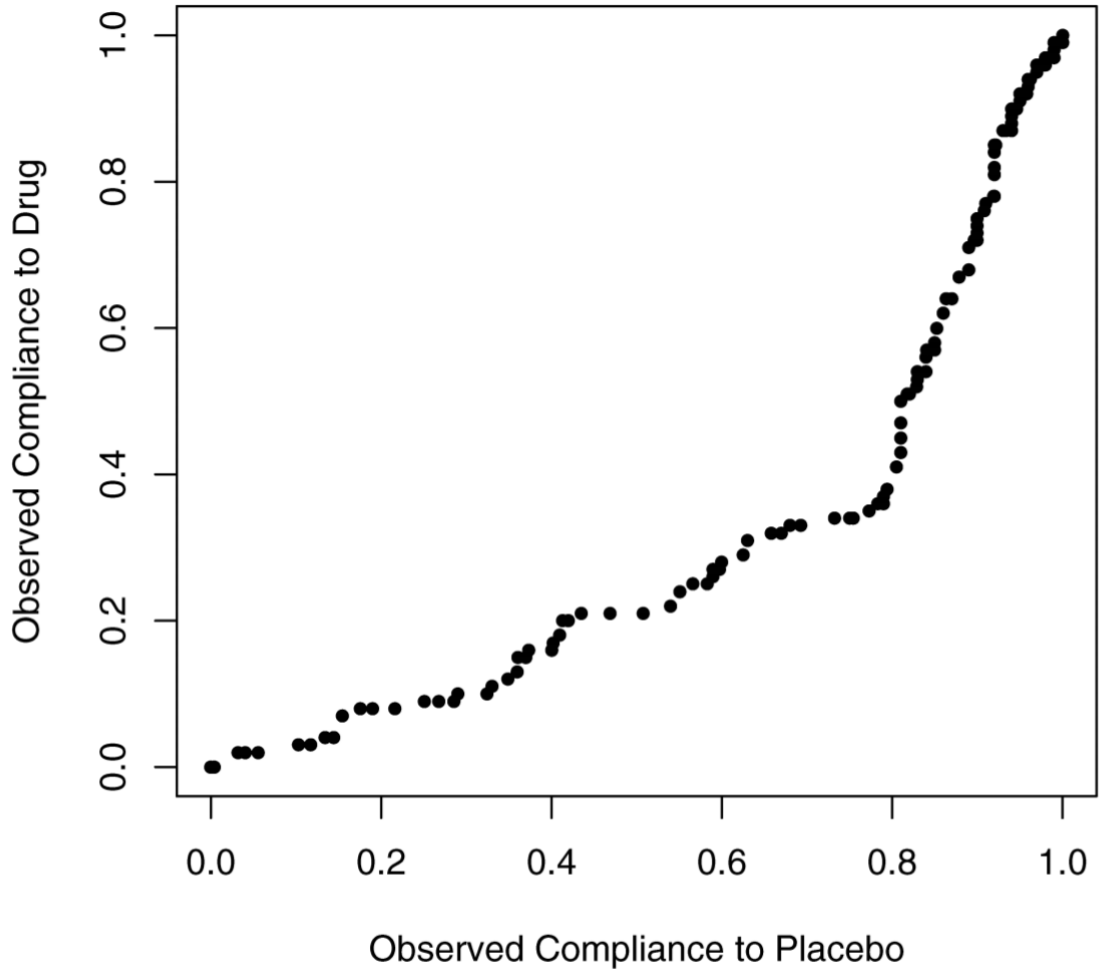


Figure 3. Q-Q plot of observed drug compliance and observed placebo compliance
– imperfect blind obvious because not 45 degree line as in randomized trial with perfect blind

Model-based Bayesian analyses of Efron-Feldman data in Jin+Rubin

- JR used dose-response parametric model to multiply – impute placebo compliance, which is observed for units in placebo group but missing for those in treatment group – for details of model, see JASA article
- Medical conclusions
  - – Good placebo compliers, those who take high doses of placebo pills (when blinded to placebo versus active), benefit little from taking active treatment rather than placebo
  - – Bad placebo compliers, those who take low doses of placebo pills (when blinded to placebo versus active), benefit much from taking active treatment rather than placebo
  - – Does this make sense? THINK!

- Another Example: Emotional Brain (EB)

  -- EB is focused on developing female Viagra, huge potential market

- Y = Increase from baseline, per week of self-reported Sexually Satisfying Events (SSEs)
- Huge placebo effects anticipated because Y is self-reported!

  -- Was event satisfying?

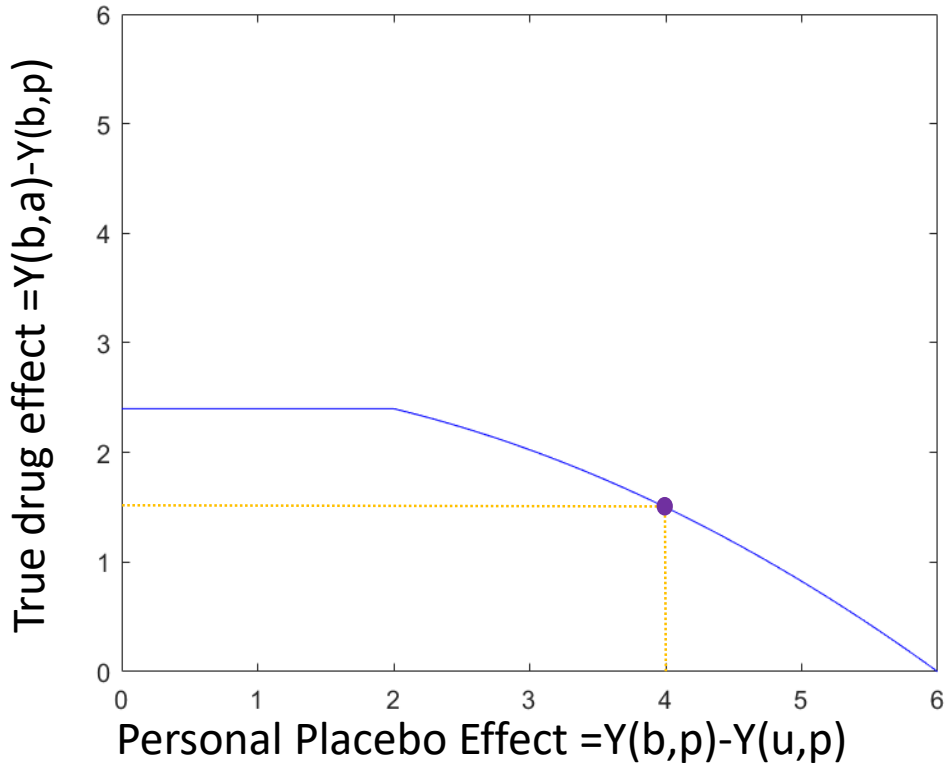
  -- Conscious units!

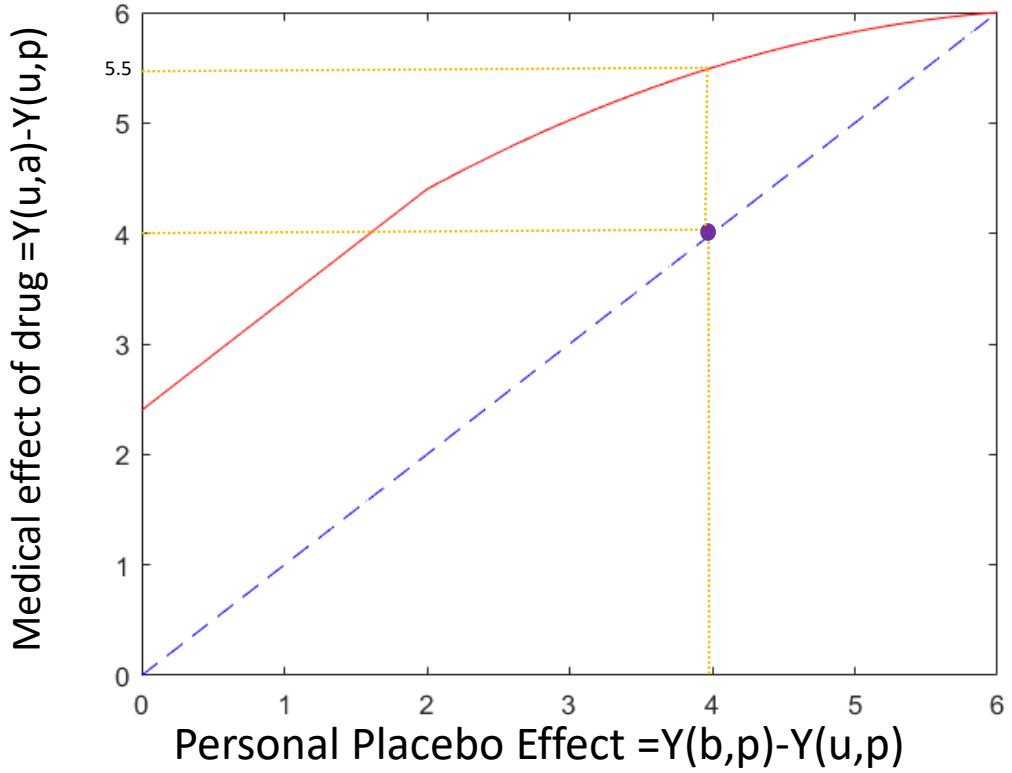
  -- I’m not going to argue this with you!

- Use model-based assumptions, as in Jin+Rubin
- Quadratic dose-response for treatment effect as function of placebo compliance in Jin+Rubin
- Quadratic dose-response in EB as function of placebo response

Illustrating idea with no covariates and additive true effect of drug and placebo (blinding) effect

Personal Placebo Effect	0	1	2	3	4	5	6
True Drug Effect of dose=1mg	2.4	2.4	2.4	2.04	1.5	0.85	0
Medical Drug Effect in reality	2.4	3.4	4.4	5.04	5.5	5.85	6



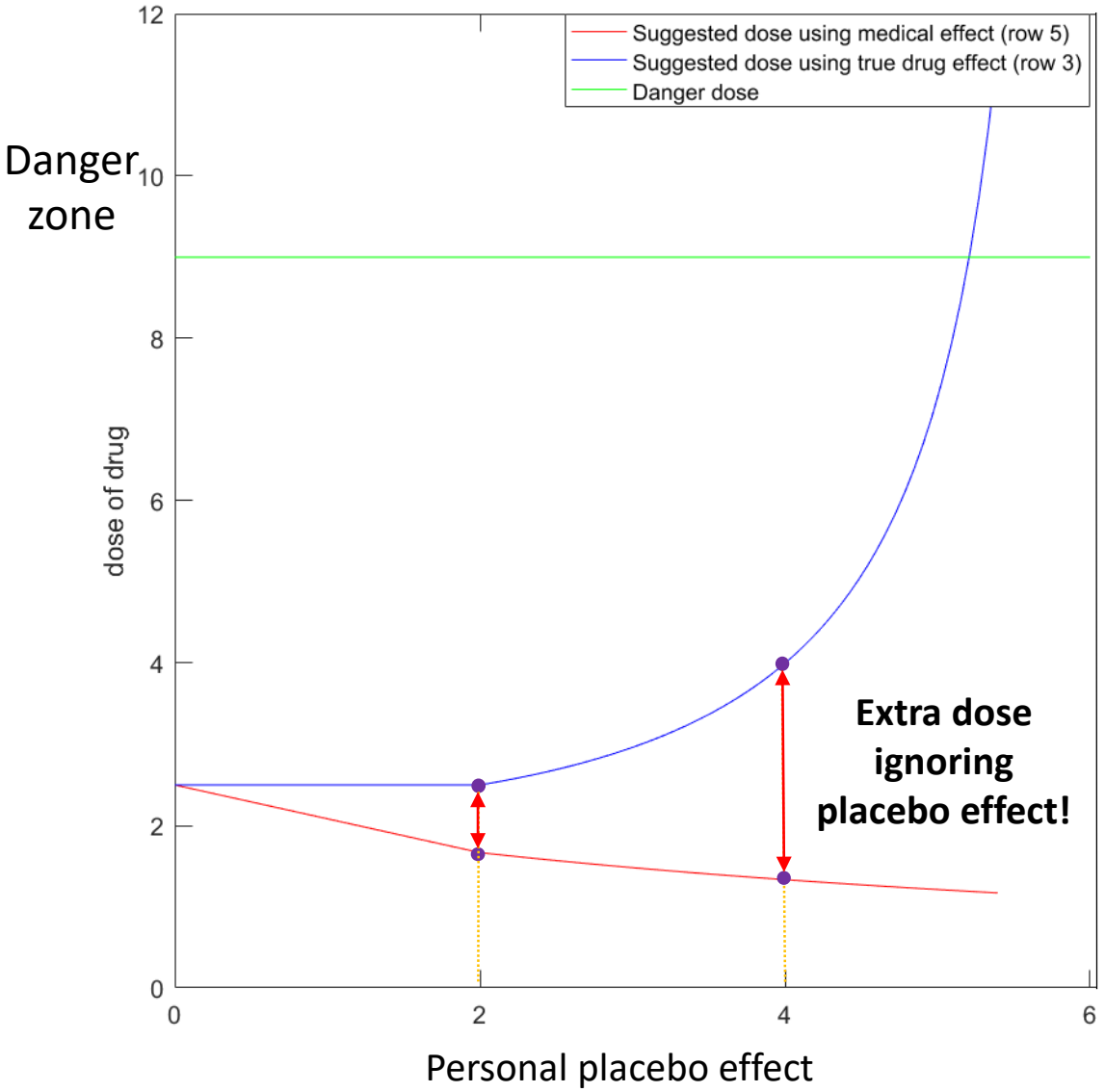


Exploratory Results from EB

- Averaged over all patients (ITT), ignoring their placebo effects, there appears to be little effect of active treatment over placebo
- However,
  - Patients with placebo effect larger than ≈ 1.7 SSE, are estimated to have essentially zero active treatment effects
  - Higher level placebo responders benefit minimally from active treatment
  - Yet patients who do NOT respond to placebo have larger increases in SSE due to taking active
- Important for drug development and for accurate prescribing in practice, especially if placebo response can be well predicted from X or baseline covariates, such as genetic markers

Exploratory Results from EB from an experiment where p=0 mg and a=1 mg and the objective is to achieve increase in SSE equal to 6 (Sunday is a holiday!)

1	Placebo Response	0	1	2	3	4	5	6
2	Assume True Drug Effect of dose=1mg for person with stated placebo response	2.4	2.4	2.4	2.04	1.5	0.85	0
3	Suggested Dose Using True Drug Effect, where effect is linear in dose (6 ÷ row 2)	2.5	2.5	2.5	2.94	4	7.06	Infinity (the drug doesn't work)
4	Extra SSE, above placebo response, needed to achieve 6 SSE (6 minus row 1)	6	5	4	3	2	1	0
5	Suggested Dose Using Medical Effect, where effect is linear in dose (row 4 ÷ row 2)	2.5	2.08	1.67	1.47	1.46	1.18	0 (no active drug needed)



Assume that the objective is to achieve final effect, after taking drug, of an increase in SSE= 6, and the drug effect has linear increase with the dose of drug. Then we can consider two doses: one allowing placebo effect and a second ignoring placebo effect.

Possible Medical Conclusion

- Isn’t it wiser to conduct 2x2 factorial and avoid the reliance on assumptions and the cost of statistical consultants, who can be expensive?
- Also, avoid future legal action and associated costs of damages claims.

Roots of Principal Stratification

- History in economics, medicine, evaluation literature
- Tindbergen (1930): “Determination and Interpretation of Supply Curves: an Example”
- Haavelmo (1944): “The Probability Approach in Econometrics”
- Sommer and Zeger (1991): “On Estimating Efficacy in Clinical Trials”
- Angrist, Imbens, and Rubin (1991): “Identification of Causal Effects Using Instrumental Variables”
- Bloom (1984): “Accounting for No-Shows in Experimental Evaluation Designs” – smoking cessation world
- Frumento, Mealli, Pacini, and Rubin (2012): “Evaluating the Effect of Training on Wages in the Presence of Non compliance, Nonemployment, and Missing Outcome Data”
  - Reveals meaningful principal strata
  - Total N>10^4

References

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American statistical Association, 91(434):445-455.
- Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. Biometrics, 58(1):21-29.
- George, B. J., Li, P., Lieberman, H. R., Pavela, G., Brown, A. W., Fontaine, K. R., Jeansonne, M. M., Dutton, G. R., Idigo, A. J., Parman, M. A., et al. (2017). Randomization to randomization probability: Estimating treatment effects under actual conditions of use.
- Jin, H. and Rubin, D. B. (2008). Principal stratification for causal inference with extended partial compliance. Journal of the American statistical Association, 103(481):101-111.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology, 66(5):688.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. The Annals of statistics, pages 34-58.
- Shadish et al. (2008). JASA. “The Design and Analysis of Gold Standard Randomized Experiments.” Comment on “Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random to Nonrandom Assignment”