

THE CAUSAL IMPACT OF
ANTI-DEPRESSANTS AT THE
INDIVIDUAL LEVEL

NOAH DIXON

9-4-24

Question: Is it possible to deduce causality, at the individual level, when administering antidepressants for depression?

Motivation: At the individual level, mental health is diagnosed by weekly tracking an individuals responses to the HAM-D (Hamilton Depression Rating Scale).¹ At the population level, tests are validated using an RCT, which is subsequently estimated using an MMRM model. However, the HAM-D measure (or other alternative measures) is currently utilized such that *improvements in baseline scores are treated as causal*. Clearly, at the individual level, this test does not account for interdependencies across time, individual-level shocks, any form of biometric tracking, most forms of measurement error, and only a small subset of adverse effects. Further, the consensus in the literature is that antidepressants are only beneficial, in the long run, for individuals with severe depression (Fournier et al. 2010; Kirsch et al. 2008), with the effect sizes being statistically small (Cipriani et al. 2018) or negligible for those with small-to-mild depression. Around 11-13% of adults report using antidepressants, but only 2-3% of individuals report having severe depression (Brody & Gu 2020; Villarroel & Terlizzi 2020). Further, Wong, Motulsky, and Eugale (2016) report that over 50% of individuals may be using antidepressants for something other than depression. *And all of this assumes that*

¹Although recently, new tests – like the MADRS (Montgomery-Asberg Depression Rating Scale), QIDS-C (Clinician-rated Quick Inventory of Depressive Symptomatology), and PHQ-9 (Patient Health Questionnaire) – have gained traction, each successive test placing more emphasis on the individual.

an individual's utility is inextricably linked to their score on the test. There are many reasons to believe that the individual-level data neglected under an MMRM model is actually an integral piece of the puzzle, and thus there is a necessity, in the literature and in the market, to introduce rigorous statistical methodology to deduce the causal impact of antidepressants on a *specific* individual.

Ideal Experiment (periodic randomization is possible): Run a double-blind N-of-1 trial (an experiment on a single individual, in which neither the individual receiving treatment nor the prescribing therapist know the level of medication being received, at any point in time) with varying levels of medication (low, standard, high) on a week-to-week basis. For individuals who stop taking medication, information should still be tracked, to properly evaluate the washout phase. Analysis, in this instance, does not require individual-level covariates (but does require an abundance of pre-treatment values), although certain covariates should be tracked *for the purpose of dealing with missing data*. The outcomes tracked are tri-daily (morning, afternoon, evening) “mood” scores and weekly PHQ-9 scores, in conjunction with richer measures (sleep, step count, mood diary, adverse events, and biomarkers). The exact blend of covariates requires meticulous thought.

Identification: Hinges on classic causal inference principles – consistency, sequential ignorability, positivity (Hernn & Robins 2020) – adapted to N-of-1 settings. Design features like washout periods (or proper lagged modeling),

randomization, and blinding (or double-blinding, in this instance) are essential for internal validity in single-patient trials (Lillie et al. 2011; Duan et al. 2013; Hawksworth et al. 2024). For time-series treatment modeling, Bayesian Structural Time-Series (BSTS) methods impose stationarity and pre-intervention stability/post-intervention stability – assumptions that must be carefully justified (Brodersen et al. 2015). Under these combined conditions, the Personal Average Treatment Effect (PATE) may be identified. Identification under these assumptions requires a rigorous proof, as it is not exactly a standard model setup.

Estimation: Should be conducted using a manipulation of the Bayesian Structural Time Series (BSTS) model, as shown in Brodersen (2015). Rather than evaluating each estimated measure, in isolation, a conglomerate measure of utility should be estimated, based on an individuals ranking of each respective metric. For instance,

$$U_i = W_A \cdot \text{Sleep} + W_B \cdot \text{PHQ-9} + W_C \cdot \text{Fulfillment},$$

where W_A represents arbitrary weights such that the $\sum_{j=1}^J W_J = 1$ and j represents the number of categories. In practice, the individuals themselves determine the weights *as well as the categories*. However, it may be that an individual’s utility functions changes, as they take the drug.

Potential Issues: The biggest potential issue, of course, is that periodic randomization (altering, in no particular order, from low dosage to medium

dosage to high dosage) is an ethical issue that may not pass IRB design. However, even if periodic randomization fails, one-shot randomization or an observational setting (with saturated data) may still allow identification – although there needs to be careful consideration in constructing the individual-level synthetic control. Another issue is that the process may put “undue burden” on the participant itself, which could cause the MAR assumption to be violated.

Economic Importance: The goal of causal inference is evaluate the average effect on an arbitrary member of the population. However, one question is whether this metric always sufficient, in application. Sometimes, the goal may not be generalizability but individualizability. Different between these goals is critical, in many settings.

Data: Not sure how to facilitate something like this.

Sources: In Zotero.