**eAppendix for “Constructed measures and causal inference: towards a new model of measurement for psychosocial constructs”**

*Example of Causal Inference Under Multiple Versions of Treatment: BMI and Mortality*

One paradigmatic example to which multiple version of treatment (MVT) theory is applicable concerns attempts to assess the effect of being overweight or obese, as assessed by body mass index (BMI), on for example 10-year mortality risk.1,2 There is no unique hypothetical intervention on BMI. One might accomplish a change in BMI from 30 to 25 by exercise, or dietary changes, or surgery, or through other means. Each of these might have very different effects on mortality risk. It then becomes difficult to speak unambiguously of *the* effect of BMI on mortality. The theory of causal inference under multiple versions of treatment was intended to provide a more precise interpretation of quantitative causal effect estimates in settings like this.1,2

Consider the setting in which A=1 corresponds to BMI=30 and A=0 corresponds to BMI=25 and Y is mortality after 10 years. In the case of BMI, a particular “version of treatment” k might correspond to a particular set of life-style choices from study entry onwards, which would in turn lead to a particular level of BMI.

To illustrate and make more concrete this interpretation, consider data analyses presented by Bhaskaran et al.3 In their paper, they used UK primary care data on 3.6 million persons from the Clinical Practice Research Datalink (CPRD), linked to national mortality registration data from January 1998 through March 2016. For never-smokers aged 16 and above, they used proportional hazards models for mortality, modeling BMI with splines and additionally controlling for baseline age, sex, alcohol use, diabetes, socioeconomic status, and calendar period. Their estimated hazard ratio comparing BMI=30 to BMI=25 was HR = 1.21 (95% CI: 1.20, 1.22). As noted above, there is no well-defined intervention on BMI. The underlying “versions of treatment” K for each person may correspond here to a complex trajectory of lifestyle choices that brings an individual to a particular BMI upon study entry. Full information on such lifestyle trajectories is of course unavailable in the data and it is effectively impossible to even fully enumerate all the possible trajectories. However, by the MVT result in the text, if we were willing to assume that the covariates in the model controlled for confounding of the effect of version of treatment on the outcome, then we could still potentially interpret the HR=1.21 using the MVT theory. The estimate HR of 1.21 would then correspond to the hazard ratio we would obtain in a hypothetical randomized trial which, within strata of covariates C, the participants in one arm are further each randomly assigned a lifestyle trajectory “version of treatment” K from the actual distribution of K in the subpopulation with BMI=30 and C=c, and in the other arm, persons are instead further randomly assigned a lifestyle trajectory “version of treatment” K from the actual distribution of K in the subpopulation with BMI=25 and C=c.

*Multiple Versions of Treatment Theory and Nested Randomization*

In many analyses of BMI, it is categories of BMI (e.g. 20-25 versus 25-30) that are compared, rather than actual values of BMI (e.g. BMI=30 versus BMI=25). The MVT theory can also be applied to categories of the exposure variable using a nested randomization. For example, if A=1 corresponded to BMI in the range of 25-30 and A=0 corresponded to BMI in the range of 20-25, then the estimate under the MVT theory could be interpreted as follows. Assuming the covariates controlled for confounding, the estimate would correspond to the hazard ratio we would obtain in a hypothetical randomized trial which, within strata of covariates C, the participants in one arm were first each randomly assigned to a BMI value between 25 and 30 according to the distribution of BMI between 25 and 30 from the actual subpopulation with BMI in the range 25-30 and covariates C=c and then further randomly assigned a lifestyle trajectory “version of treatment” K from the actual distribution of K in the subpopulation with C=c and the level of BMI to which they had been assigned; and in the other arm, persons would instead be randomly assigned to a BMI value between 20 and 25 according to the actual distribution of BMI between 20 and 25 from the subpopulation with BMI in the range 20-25 and covariates C=c and then further randomly assigned a lifestyle trajectory “version of treatment” K from the actual distribution of K in the subpopulation with C=c and the level of BMI to which they had been assigned. The MVT interpretation would still potentially be applicable but becomes yet more complex.

Likewise, in the social integration example given in the paper, the MVT interpretation would require nested randomization wherein, within strata of covariates C=c, for a given social integration quantile, the participants would be randomized to a value of the components of social integration (religious service attendance, community group participation, number of close friends, and marital status) according to the actual distribution of the components in that quantile among those with C=c, but then would be further randomized to a set of life choices and relationship trajectories K (including e.g. quality of relationship, type of community, style of religious service etc., represented by multi-dimensional η in Figures 5 and 7) drawn from the actual distribution of these “versions of treatment” K among the subpopulation with actual levels of religious service attendance, community group participation, number of close friends, and marital status that they were hypothetically randomized to and with covariates C=c.

*Limitations of the Multiple Versions of Treatment Causal Interpretation*

As discussed in the text, the MVT interpretation is subject to a number of limitations and challenges. First, when the set of versions of treatment, K, is unknown, this limits the precise understanding of the interpretation. One does not know all the various complex trajectories that may give rise a particular value of the measure A, nor, of course, does one then know the distribution of these various trajectories. Second, and relatedly, with the set of underlying versions unknown, it would then effectively be impossible to implement the hypothetical randomized trials embedded within the interpretation. The usefulness of the interpretation is thus, in this regard, somewhat limited. However, as discussed further below, the causal MVT interpretation may nevertheless provide clues as to where to best intervene. Third, the interpretation will vary depending on what is included in the measured covariates C. This is because, once C is fixed, this may limit the range of potential “versions of treatment” that are possible. For example, suppose in the context of BMI, exercise was included in the measured covariates C that were adjusted for in the analysis. In this case, because the hypothetical randomized trial in the interpretation is stratified by the measured covariates C, the distribution of the lifestyle trajectory “version of treatment” variable K will no longer vary by exercise across the BMI arms (or by whatever aspect of it is captured by the measured covariate) because the hypothetical randomized trial is effectively stratified by exercise. Fourth, with the versions of treatment unknown, it becomes difficult to substantively assess the unconfoundedness assumption and thus to know whether the proposed interpretation is reasonable.

Although the MVT interpretation has limitations, it may be the best we can do with respect to a formal potential-outcomes based interpretation of the quantitative effect estimate of a composite exposure.4,5 Even reflective and formative models that assume an underlying univariate latent variable face the challenge of what one means by intervening upon the latent variable, and different interventions on it may again result in different causal effects on the outcome. Moreover, these reflective and formative models typically also additionally impose what is the often unrealistic assumption of an underlying *univariate* latent variable, which the MVT approach circumvents. Another alternative is of course abandoning a causal interpretation entirely, but what that often results in, in practice, is that the researcher/author/analyst/reader implicitly imposes a vague ill-defined causal interpretation. The MVT interpretation at least makes clear – and forces the interpreter to think about – the caveats.

While the limitations of MVT interpretation are important and need to be taken seriously, attempting to interpret associations causally, even if imprecisely, can still sometimes be valuable with regard to gaining insight into potential interventions. Analyses that suggest causal effects of phenomena related to our psychosocial constructs may help identify potential intervention targets. Attempts at estimating causal effects of phenomena related to our constructs of interest (e.g. using MVT theory) can be useful in informing on what to try to intervene first, i.e. concerning what we ought to begin with as potential intervention targets, whilst keeping us mindful (through the very limitations and assumptions of the MVT interpretation) that there are indeed likely multiple complex trajectories involved, and that the effects of our actual interventions may thus not correspond to what we had estimated. Nevertheless, when attempting to develop interventions to improve population health and well-being, one must begin somewhere. Attempting, as best as we can, to estimate causal effects of the phenomena of interest may provide important clues as to where to begin. For example, analyses that suggest that measures of purpose in life have considerably stronger associations with all-cause mortality than do measures of affective happiness (Trudel-Fitzgerald et al.6) might suggest that we ought to focus on the former as a potential intervention target. Of course, decisions concerning intervention development are shaped not only by the potential causal effects of the intervention target, but also by the plausibility that that target can be changed, and by the costs and ethical considerations of trying to change it. But analyses examining putative causal effects, as best as possible, that correspond to phenomena related to our constructs of interest may be useful information that can help inform such decision-making.

*Analysis and Interpretation of Associations with Time-Varying Indicators or Latents*

As noted in the text, in the context of time-varying indicators or latents, a possibly attractive analysis alternative is using the indicators at time t as the exposure, while simultaneously controlling for past values of the indicators (along with confounders C) at time t-1.7,8 At least in Figure 6 this has the advantage of simplifying confounding control to avoid the potential complications arising from time-dependent confounding. In Figure 6, each of the indicators at time t could be considered one at a time as the exposure, while simultaneously controlling, when data allow, for past values of the entire set of indicators . The associations between the time t indicators are then more easily interpreted as corresponding to the effects of present exposure on the outcomes. Moreover, control for prior exposure levels can help rule out reverse causation by prior outcome7,8 and potentially helps rule out residual unmeasured confounding, since an unmeasured confounder would have to substantially affect current exposure through pathways independent of prior exposure to generate considerable bias. These remarks pertain directly to Figure 6 with causally efficacious indicators, or approximately in Figure 7, to the extent that the indictors capture what is causally relevant in the underlying reality with respect to the outcome Y. Unfortunately, relatively large sample sizes will be needed to deal with the potential collinearity arising between current and prior exposure levels, especially if the individual indicators themselves are being used. In cases of tens of thousands of participants, as in the Nurses’ Health Study described in the text, this may be possible. The analyses of Li et al.9 and VanderWeele et al.10 discussed in the text in fact did control for past exposure levels. In principle, in Figure 6, it would be possible to regress the outcome Y on the entire set of indicators at time t, , while also controlling for the indicators at time t-1, , but this would require even larger sample sizes to deal with collinearity. Another alternative, when sample sizes are more limited, would be to control for a summary measure of the indicators, e.g. their mean, at time t-1, . This will only partially control for confounding by the past indicators, but, depending on how strongly correlated the indicators are, it may suffice to remove most of the confounding. Which of these strategies is to be preferred will depend on a combination of the sample size available, the correlation among the indicators, and the extent to which there are differential effects of the indicators on one another and on the outcome of interest.

In summary, to obtain the cleanest possible interpretation of causal effect estimates, it may be desirable to regress outcome Y on the individual indicators one-by-one, along with past values of these indicators, and potentially confounding covariates. Associations with summary measures might still be interpreted as under the MVT theory but, as seen in the social integration example, this has the potential to obscure the more subtle relationships concerning each indicator.

*Discussion of the Independence Assumption for Multiple Versions of Treatment*

The independence assumption in the text and in the derivation given in the paper’s print Appendix was that Y is independent of A conditional on (K,C). This was a weakening of the assumption employed in Proposition 8 of VanderWeele and Hernán1 on the original multiple versions of treatment theory which was that the mapping from K to A was a deterministic many-to-one map. The assumption is weaker insofar as if the mapping from K to A was a deterministic many-to-one map then it immediately follows that Y is independent of A conditional on (K,C) since, conditional on K, A is a constant. However, the weakening of this assumption to Y being independent of A conditional on (K,C) also gives rise to ways that this assumption can be violated that would be precluded if the mapping from K to A were a deterministic many-to-one map. For example, if in Figure 2(b), with K=η, there were causal effects of one or more of the indictors ( on Y, then the assumption would be violated. This might arise, for example, if the actual measurement of a particular indicator (or e.g. a physician’s comment upon it to a patient) could potentially lead to change in behavior that affected the outcome itself. Likewise, the assumption would be violated if there were a common cause U that might affect one or more of the indicators ( and also the outcome Y. Such unmeasured common causes might include, for example, intelligence, education, or ethnicity. These possibilities, however, do not ultimately threaten the applicability of the MVT interpretation insofar as K can itself be defined to include one or more of the indicators ( or the common causes U of the indicators ( and the outcome. The underlying constituents of reality η that are relevant for the outcome can all be defined as being part of the multivariate latent variable K in the MVT theory. By defining K in this manner, one effectively guarantees the validity of a causal diagram in which there no causal effects from the indicators ( on the diagram to the outcome Y, since, if these effects were there, they would already be captured by the arrow corresponding to the causal effect from K=η to Y. The mediated or indirect effect of K=η on Y through ( is effectively 0 since these effects are captured by the direct effect of K=η on Y. This is so by the very definition of K=η. Note that in the application of the MVT theory to the models in Figure 4 which had causal effects of the indicators themselves on the outcome Y, the variable K was taken to be the entire set of indicators ( in the proposed interpretation in the text.

*References for eAppendix*

1. VanderWeele TJ, Hernán MA. Causal inference under multiple versions of treatment. Journal of Causal Inference, 2013;1:1-20.

2. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. Epidemiology, 2011;22:368-377.

3. Bhaskaran K, dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3• 6 million adults in the UK. The Lancet Diabetes & Endocrinology. 2018 Dec 1;6(12):944-53.

4. VanderWeele TJ. On well-defined hypothetical interventions in the potential outcomes framework. Epidemiology (Cambridge, Mass.). 2018 Jul;29(4):e24.

5. VanderWeele TJ. On causes, causal inference, and potential outcomes. International Journal of Epidemiology, 2016;45:1809-1816.

6. Trudel-Fitzgerald C, Millstein R, von Hippel C, Howe R, Tomasso LP, Wagner G, VanderWeele TJ (2019). Psychological well-being as part of the public health debate? Insight into dimensions, interventions, and policy. BMC Public Health, 15;9(12):e033697.

7. VanderWeele TJ, Jackson JW, Li S. Causal inference and longitudinal data: a case study of religion and mental health. Social Psychiatry and Psychiatric Epidemiology, 2016;51:1457-1466.

8. VanderWeele TJ, Mathur MB, Chen Y. Outcome-wide longitudinal designs for causal inference: a new template for empirical studies. Statistical Science, in press. Available Early Access: https://www.e-publications.org/ims/submission/STS/user/submissionFile/38456?confirm=800c6919

9. Li S, Stampfer MJ, Williams DR, VanderWeele TJ. Association of religious service attendance with mortality among women. JAMA internal medicine. 2016;176(6):777-85.

10. VanderWeele TJ, Li S, Tsai AC, Kawachi I. Association between religious service attendance and lower suicide rates among US women. Jama Psychiatry. 2016;73(8):845-51.