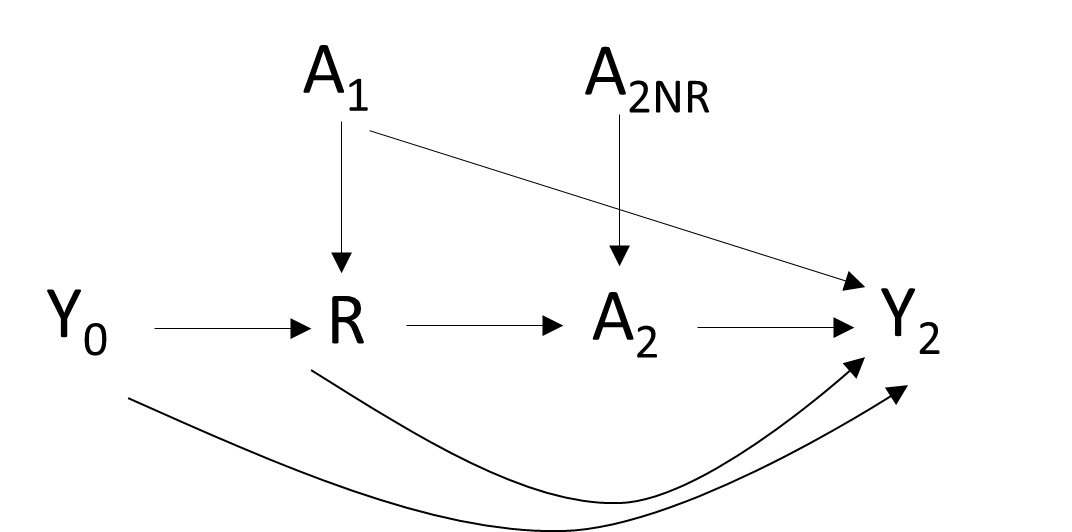
**Scenario and Notation for Simulation Study**

Suppose that the binary outcome at a given time point represents whether or not a desired event (e.g., negative drug test indicating abstinence for illicit drugs) is observed at that time point. Let be the outcome observed before initial treatment (i.e., pretest), observed at the beginning of stage 1. Suppose that the response status represents some healthy state or behavior, observed at the end of stage 1. Response status is likely to be positively correlated with , but they are not deterministically related (i.e., it is not always the case that for all participants). For example, response status could be abstinence during the last week of stage 1. Let the posttest be the outcome observed at the end of stage two of treatment (e.g., negative test result at end of study). For convenience, is said to be observed at time 0, at time 1, and at time 2.

Suppose that a dichotomous treatment , coded as or , is assigned after is observed but before is observed (i.e., immediately after time 0). Suppose that has a slight beneficial effect, so that the conditional expected value of tends to be higher for participants randomized to than for participants randomized to . Also, suppose that for nonresponders only, a dichotomous treatment , coded as or , is assigned. This is a restricted SMART, i.e., only nonresponders are rerandomized. Again, suppose the level of is beneficial, and thus associated with a slightly higher probability for relative to . The simulation model can be represented in a directed acyclic graph (DAG) by constructing a recoded variable

.

For responders, is 0 by convention, indicating that it cannot have an effect; for nonresponders, . With this notation, the graph can be written as follows:



To follow convention, , , and above are coded as versus , while and are coded as versus .

**Contrasts**

Suppose an investigator wishes to have adequate power for a significance test of the difference in expected values on between a pair of candidate adaptive interventions denoted , and . There are three, non-equivalent ways to operationalize this contrast. The contrast can be conditional on setting both and to the same value,

or it can be conditional on a particular value of only but average over ,

or it can be marginal over both and ,

The third contrast is probably of greatest practical interest because it would apply on average to the target population as a whole. Depending on the investigator, either the second or third contrast might be easiest to elicit. This paper has presented sample size and power formulas which use either the second or third contrast as the effect size. The first (fully conditional) contrast is the most complicated and least intuitive. Unfortunately, it is the unwanted first contrast which must necessarily be specified in order to do a simulation. That is because a simulation requires a fully specified data-generating model (DGM) which tells how all of the variables are related to each other.

This presents a small dilemma if we wish to use simulations to validate sample size or power formulas for a longitudinal SMART. It becomes necessary to “translate” from the fully conditional specification to at least a partly marginal specification. This is more difficult if a logistic link function is used in the DGM, because of noncollapsibility. Therefore, for the purposes of the DGM we will use a linear model structure to define the conditional means (i.e., conditional probabilities) for each binary variable. However, in the actual analysis of the simulated data, we fit both linear and logistic models and will compare the power of each.

**Data-Generating Model Parameters**

More specifically, suppose the probabilities follow a set of linear models as follows. First,

so about half the sample is abstinent at baseline. Second,

where and . Thus, the possible values of range from (for participants who were non-abstinent at baseline and received the low level of ) to (for participants who were abstinent at baseline and received the high level of ). Next,

where and . Thus, the possible values of range from (for someone who has not been abstinent at baseline, nor a responder, and who has received the weaker treatment at both stage) to (to someone who was abstinent at baseline, and received the high level of , and was a responder). These values seem implausibly extreme, but recall that there is no single assignment factor that gives a marginal mean of as low as 0.05 or as high as 0.85. Those extreme fitted values are only reasonable in the fully conditional model, conditioning on past values.

Under the assumption that these linear models hold, it becomes feasible to calculate the true value of a marginal contrast of interest between two regimens () and (), using only the coefficients of the fully conditional model. As described earlier, the fully conditional model (in which can depend on and ) must be specified in order to simulate data, but the marginal model (averaging over and ) is generally of much more interest in making policy decisions.

Let and . Then the marginal contrast is

The decomposition of the contrast into three terms can be intuitively explained by saying that there are three ways in which the adaptive intervention can have an effect: (1) indirectly by changing the proportion of responders, (2) directly by the effect of stage 1 treatment on expected regardless of response status, or (3) directly by the effect of the stage 2 treatment on expected for nonresponders. As an aside, this means that can be relevant even when comparing two regimens which prescribe the same . That is, even if , the third term above is still nonzero except in the special case . This is because, by increasing the proportion of responders, successful Stage One treatment can have the side effect of decreasing the proportion of individuals eligible to be affected by the Stage Two treatment option. In the special case where both and , then the contrast simplifies to the usual which would be expected in a factorial experiment. This complication is already a kind of noncollapsibility (i.e., does not cancel out even when it is not of interest), although it is still relatively tractable here because of the linear assumptions elsewhere in the model (i.e., at least cancels out by iterated expectation).

Suppose we are comparing two adaptive interventions, specifically to . We can use iterative expectations to calculate the true marginal probabilities of the regimens (which average over and ) from the fully conditional probabilities (which specify and ). First, the marginal response rates are calculated as

and

Next,

and

The probability difference is . Notice that this effect is not equal to , even though is the only randomized factor that differs between the arms. This is because the data-generating model was conditional on and , while the inference model of interest is not; it is not possible to simulate data directly from the marginal model.

Now that , , and have been obtained, power or sample size can be computed using the results in Kidwell et al (2019). That formula was based on a test of the log odds ratio, rather than directly on the probability difference, but the log odds ratio can be calculated from the two probabilities. Specifically

representing an odds ratio of about . The can be substituted into the Kidwell et al (2019) sample size formula for a desired power, say . Specifically,

> 2\*(((.8416+1.96)^2)/(1.0655^2)) \* ( (2-0.55)/(.615\*(1-.615)) + (2-.35)/(.355\*(1-.355)))

[1] 184.3168

where and are the marginal variances given treatment, is the significance level of the test, and and are normal quantiles.

These values also be substituted into the equivalent power formula for a given sample size, say, for ,

where is the cumulative normal distribution.

This shows that with the current approach, it is possible to simulate pretest-posttest binary SMART data and also use theoretical power formulas, so that the performance of the formulas can be checked empirically for this kind of data. The analysis of the simulated data need not use a linear link, because it was only required for a tractable data-generating function. It would be possible to compare the performance of the linear link, log-linear link, and logistic link, because although they are most suitable for use with different estimands (probability difference, log risk ratio, and log odds ratio), they test a logically equivalent null hypothesis.

**Discussion**

If a logistic model were being used, it would be necessary to use separate sets of parameters in the fully conditional model and in the marginal model, and to calibrate one set to the other via numerical approximation or trial and error. A log link function would allow a kind of collapsibility on the log-linear scale, but it would still be very complicated to translating between the fully conditional and marginal models for , because of the nonlinear role of .