* + I want to keep things as simple as possible because we have no ground truth (two or three different simulation methods, but none of them are verified, and they take different inputs and give different outputs; and a formula with an unrealistic assumption, therefore also untrusted)
  + But I still need a way to check the robustness of the formula, in cases where Y0, R, and Y1 are all correlated (which the formula officially can’t handle, although in reality it will always be the case)
  + I’m also worried that for at least some datasets, the longitudinal analysis is less powerful than the end-of-study analysis due to collinearity and noncollapsibility; but I don’t know if this is freakishly rare or extremely common, because I am too confused about the conflicting simulation methods
  + You can only simulate from the fully conditional model, but the estimands are for the marginal model.
    - If the data-generating model is linear (even for binary outcomes) this is very easy because the mean structure is mostly collapsible (slight complications due to issues with responders vs. nonresponders). The resulting formula gives insight.
    - If the data-generating model is log-linear, then it’s possible to get an explicit formula to convert conditional to marginal probabilities, but the formula is difficult and complicated and gives no insight.
    - If the data-generating model is logistic, no workable formula is possible, and you can only find the true value by running a preliminary simulation with a giant sample size.
    - I would prefer to use a linear link in the data-generating model because it’s the only one where I can be reasonably confident of whether I got the right answer or not. But it is the least realistic.
    - In the analysis of the data we can try all three link functions – I don’t think it would change power very much.