My personal difficulties

When I see notations involving potential outcomes I feel uncomfortable; it never clear to me if these are fixed unknown constants or random variables.

The analogy with missing data problems is only apparent; or at least this is a very special kind of missing data; so, I do not see where it helps to see the problem this way.

Assignment mechanism

A notation like $pr(Z \mid Y, X)$ is very confusing to me: I would be tempted to read this as the probability of a given treatment conditionally on having a given response which depends on the treatment.

But even if I read Y as meaning the set of potential outcomes, I still do not understand: these potential outcomes are fixed ? But the treatment is fixed too, so what are we talking about ? This provides me evidence that only people with superior mental power can master these concepts.

Randomization

I have a similar mental difficulty in understanding what is meant by saying that the treatment is independent from the outcome.

Latent class models

I believe that latent class models may provide an alternative for handling causal inference within a framework which is perhaps more familiar to most statisticians.

Suppose that the response Y is a random variable which depends on treatment, Z, covariates, X and on a latent variable, U. What is being said probably is that we want to assign the treatment in such a way that Z is independent of U conditionally on X.

Likelihood inference

Personally, I feel philosophically uneasy with an approach where the probability distribution of the response is simply (if I understand correctly) induced by sampling.

Is this not a metaphysical attitude, like pretending that we are God ?

I feel more at home thinking that the individual heterogeneity (which can be modelled by latent variables), covariates and treatment can only change the probability of responding within a set of different possible ways.

Strongly ignorable assignments

I assume that the puzzling (to me) statement $Pr[Z_i = 1 \mid Y_i(0), Y_i(1), X = x] = e(x)$ is another way of saying that, given covariates, the treatments assignment is independent of the latent U which may affect the response Y.

The important ingredients within a probabilistic modelling approach are the parameters which describe the conditional distribution of the response. For instance, if Y was binary, the effect of Z, X and U on Y may or not be additive.

If the model is identifiable, the overall averaged effect of Z may also be assessed, but this is only a side product of a probabilistic description of the system.

Modelling non compliance

A latent class model for this context might be as follows.

There are two types of subjects indexed by a binary latent U. The probability of compliance depends on U which also affects the probability of response.

A desirable model Let for the moment ignore identifiability constraints; so let

Probabilities	parameters
$P/U = 1)$	λ
$P(D = 1 \mid Z = 1, U = 0)$	ψ
$P(D = 1 \mid Z = 1, U = 1)$	ϕ
$P/Y = 1 \mid D = d, U = u)$	$\frac{\exp(\alpha+\beta u+\delta d)}{1+\exp(\alpha+\beta u+\delta d)}$

The compliers-never takers model

This model seems to be a special case of the above if we make the following very simplified assumptions:

- U is independent of Z, that is the treatment is assigned at random;
- $\psi = 0$, that is people with $U = 0$ never take the treatment, even if assigned;
- $\phi = 1$, that is people with $U = 1$ always take the treatment if assigned;

Then, simple method of moment estimation gives:
$\hat{\lambda}$ = $P(U = 1)$ = $\frac{9675}{12094}$ = 0.80
$P(Y = 1 \mid D = 0, U = 0)$ may be estimated by $\frac{34}{2419}$ = 0.0141
$P(Y = 1 \mid D = 1, U = 1)$ may be estimated by $\frac{12}{9663}$ = 0.0012
finally, an estimate of $P(Y = 1 \mid D = 0, U = 1)$ may be obtained by noting that $\frac{74}{11514}$ = 0.0064 estimates the mixture
$P(Y = 1 \mid D = 0, U = 0)(1 - \lambda) + P(Y = 1 \mid D = 0, U = 1)\lambda$