# cNORM Weighting Manuscript: Editorial Discussion

Before I start editing, I want to make sure I understand the method of the study, and also address nomenclature for the various aspects of the method.

# Descriptive Labels

One thing that would make the manuscript more readable is to replace the current *non-descriptive labeling* of various aspects of the study.

* Stratification variables: variable names are 1,2,3, and variable levels are 1,2,3
* Sampling conditions are labeled 1-6
* The 27 strata have 1,1,1 (etc.) subscripts
* In the post-hoc analysis, the ANOVAs are referred to as the first and second ANOVAs

The use of non-descriptive labels (integers) for all conditions and levels can make it hard to follow the narrative as you describe this complex methodology.

Here are some ideas for descriptive labels:

## Stratification Variables:

* Line 698 refers to the SVs in this study being modeled after *education*, *ethnicity*, *region* – what if we used these as labels?
  + Levels could then be:
    - *no high school diploma, high school grad, college degree* 
      * To address a international readers more general, we would like to use the SV “education” with the levels “low”, “medium” and “high”
    - *White, Black, Hispanic*
      * *Proposal* 1: SV “socio-economic status” with levels “low”, “medium” and “high”
      * *Prooposal 2: SV “ethnicity” with levels “native”, “mixed” and “non-native”*
    - *South, North-East, West*
      * *Yes*
    - (or something similar)

## Sampling Conditions:

In accordance with the adjustments of the SV labels and levels, we suggest to use the following labels for the different sampling conditions:

1. Unbiased
2. High education moderately underrepresented
3. High education largely underrepresented
4. Low and high education underrepresented
5. Bias joint probabilities
6. Clustered sampling

As you already mentioned, we think, this change will make the different types of biases more clearly for the readers.

# Clarification of Methods:

## Reference Population

The simulated reference population of 24 million persons is first described in detail starting on line 342. Then starting on line 360 you refer to the creation of five additional population samples “in the same way”.

* Does this mean that the five “population samples” (description starts on line 366) are also 24 million persons, with 4 million per age cohort?
* If this is the case, then I suggest we refer to all six “population samples” as “six simulated populations”,one unbiased simulated population and five additional biased simulated populations. This maintains the conventional distinction between a “population” (the entire collection of people) from which we draw “samples” (which may be more or less representative of the “population”).
  + We would suggest to use the term “six simulated populations”: one unbiased simulated population and five additional biased populations
* If what I've said so far is correct, then the clearest description of the data hierarchy is:
  + Six simulated populations (n = 24 million???).
  + 100 normative samples (n = 600) are drawn from each simulated population.

## Stratification Variables

**Table 1**

*Marginal proportions of the levels of the three stratification variables in the reference population*

|  |  |  |  |
| --- | --- | --- | --- |
| Stratification variable | Level 1 | Level 2 | Level 3 |
| Education | Low (40 %) | Medium (20 %) | High (40 %) |
| Ethnicity | Native (30 %) | Mixed (40 %) | Non-native (30 %) |
| Region | South (60 %) | East (20 % | North-west (20 %) |

The current table represents the marginals of the single SVs for the unbiased simulated population, and, therefore, the single proportions sum up to 100% for every line. To make this point more clearly, we change the table labeling from “proportions” to “marginals”.

Table 1 anchors much of the method description, but it is hard to decipher.

* Each SV has three levels
  + What does “level” mean in this context? I'm not clear as to whether it means something like a demographic category (e.g., black vs. white ethnicity) or what the manuscript refers to as “performance” (1 = above, 2 = average, 3 = below)?
    - With the term “levels”, something like demographic or socio-economic category is addressed. This point should be more clear now, since we now use concrete SVs like “eduction” instead of SV1, SV2 and SV3.
* The next paragraph refers to “3 x 3 x 3 = 27 strata”, but this description is at odds with the structure of three SVs, each with three levels, which would lead to 3 x 3 = 9 strata or joint categories.
  + Since we use three SVs, we have 3x3x3=27 strata
* Part of the reason why Table 1 is confusing is that it combines information on both *structure* (in the table margins) and *distribution* (in the cells).
* The 3 x 3 x 3 structure can be shown more clearly in a nested table that shows only classification structure:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | low education 40% | medium education  20% | high education  40% |
| ethnicity: native  30 % | region: south 60% | 40 % x 30% x 60 % | 2,1,1 | 3,1,1 |
| region: east 20% | 1,1,2 | 2,1,2 | 3,1,2 |
| region: north-west 20% | 1,1,3 | 2,1,3 | 3,1,3 |
| ethnicity: mixed  40% | region: south 60% | 1,2,1 | 2,2,1 | 3,2,1 |
| region: east 20% | 1,2,2 | 2,2,2 | 3,2,2 |
| region: north-west 20% | 1,2,3 | 2,2,3 | 3,2,3 |
| ethnicity:  Non-native  30% | region: south 60% | 1,3,1 | 2,3,1 | 3,3,1 |
| region: east 20% | 1,3,2 | 2,3,2 | 3,3,2 |
| region: north-west 20% | 1,3,3 | 2,3,3 | 3,3,3 |

* Here, the cell values (1,1,1 . . . etc.) represent the 27 possible combinations of the levels of the three SVs. Of course “collegeGrad-black-west” would be a lot more meaningful and vivid than “3,2,3” (for example).
  + That’s a quite good point. We would like to add a table like this for the unbiased simulated population to make clear that the biased populations and, therefore, the corresponding norming samples, differ from the unbiased population in terms of the joint proprotions.
* Additional implicit three-category classifications related to the SVs:
  + Effect of SV on the latent cognitive ability (*large*, *medium*, *small*). As I understand it, this correlation between SV and ability differs between SVs, but is constant within the levels of each SV?
    - Misconception: for each single level of the SVs, there cannot be a correlation between SV and cognitive ability, because the level of the SV is constant.
  + *Above* vs. *average* vs. *below* performance. This presumably refers to the latent cognitive ability, in the form of the group mean of the raw test score from the simulated test?
    - Yes.
  + Distribution of levels of SVs within the reference population (*40-20-40* vs. *30-40-30* vs. *60-20-20*, as shown in Table 1).
* With all of this information flying around, it's not clear to me what values belong in the cells of my 27-cell nested table. To me, this is the most important issue to clarify in the methods section of this paper.
* Another way to configure the 27-cell table is to relate it back to the descriptions of the sampling methods.
  + The primary thing that's being manipulated to differentiate the six sampling conditions is the distribution of the levels of SV1 (large effect on cognitive ability).
    - In sampling conditions 5 and 6, the probabilities of the other SVs were manipulated, too.
  + Thus, if you applied my 27-cell table to each of the six sampling conditions, what numbers should we put in the 27 cells to most clearly show the difference between the sampling conditions?
    - Since the SV were independent from each other, the numbers in the cells simply correspond to the product of the respective marginal probabilities (e.g. 40 % x 30 % x 20 % = 2.4 %)

### Interaction between weighting and person location: First and Second ANOVAs

Here's what I can figure out (but need confirmation) about this aspect of the design.

ANOVA configuration:

* First: 11 (location) X 2 (norming method: WCN biased vs. SCN biased)
* Second: 11 (location) x 2 (WCN biased vs. SCN unbiased (benchmark))

ANOVA outcome:

* 1. First: Compare WCN to SCN across the range of person abilities: which method yielded norm scores that were closer to the benchmark (unbiased population sample) 🡪 Which method yielded lower norming errors?Second: How did the accuracy of WCN (with respect to the benchmark) change across the range of person abilities? 🡪 How close to the benchmark is WCN across the range of person abilities? Can it compensate for all the bias in the norm sample or not?

Do I have all this correct?