# Harmonia Activity Responding to GBV Training Pre/Post Analysis Documentation

Cory Spencer | [cns.spencer@gmail.com](mailto:cns.spencer@gmail.com)

## Data cleaning

#### Data entry

We utilized RedCap to manage and enter our data. Cory and Xylia created the data entry form on RedCap (Project name: **Responding to GBV Health Provider Pre/Post Questionnaires**). Surveys were completed by participants on paper forms and data was entered into RedCap by HAI/HAMNASA staff. Data entry was checked for accuracy through a 10% random sample of questionnaires. Laurensius Amir ([Laurensius.amir@hamnasa.org](mailto:Laurensius.amir@hamnasa.org)) is the best point of contact for any questions related to data collection and entry. One of the redcap entries is a “key” corresponding to the correct answers for each question (participant\_id == “KEY”).

#### Known issues and how they were addressed

* RedCap records include records which were conducted for the training of trainers (standardization training conducted in Dili). These were excluded for the purposes of the pre/post training analysis.
* Redcap records include duplicate records resulting from erroneous duplicated import (same participant information imported twice to RedCap but assigned different record IDs). These were excluded by looking for data records with the exact same values across all fields (besides the unique record ID).
* Participants entered their unique identifiers themselves, which resulted in inconsistencies between pre and post data entry forms and necessitated matching records by hand in some cases by comparing identifiers, demographic information, site of training, and date of training. Any IDs which did not have clear matches were dropped from the analytical dataset.
* Demographic information was sometimes entered inconsistently from pre to post for the same participant. If age, sex, position, position years, or patient volume was not entered consistently from pre- to post-training, a participant was flagged for having a mismatch. To be used in the analytical dataset, participants could not have more than 2 demographic info mismatches from pre- to post-, due to concerns about data quality. For participants with two or fewer mismatches, the inconsistent field (i.e., age or sex) was replaced with an NA value.

#### Scripts used for cleaning data

##### 00\_Prep\_data.R

* **Summary**: This script reads in raw data downloaded from the RedCap project, standardizes data format, creates participant matches where needed, and re-codes variables according to the desired scoring framework.
* The scores used in analysis and in the script are outlined in the attached excel document (“**harmonia\_project\_derived\_variables\_data\_dictionary.xlsx**”)
* The output (‘hp\_pre\_post\_clean.csv’) is a long-formatted dataset with two rows for each participant included in the analysis (one for the pre-time point and one for the post-time point, indicated by the time\_point variable)

##### 00b\_composite\_indicator\_prep.R

* Summary: This script reads in the cleaned data saved in the previous script and creates composite variables by topic (e.g., attitudes, knowledge, etc.) before calculating indicators of performance by participant:

|  |  |
| --- | --- |
| Indicator name in dataset | Description |
| Pre | Raw variable score for pre-training module |
| Post | Raw variable score for post-training module |
| Change | Raw change in points from pre to post-training (post score – pre score) |
| Pre\_pct | Proportional score for variable on pre-training survey (i.e., points achieved / points possible) |
| Post\_pct | Proportional score for variable on post-training survey (i.e., points achieved / points possible) |
| Change\_pct | Absolute difference in proportional scores from pre- to post-training (post\_pct – pre\_pct) |
| Improve | Binary indicator of whether participant improved (scored more points) after training |

* There are two outputs from this script:
  + **Prop\_data.csv**: wide-formatted dataset where each row represents a unique participant/domain combination, with each of the above scores (pre, post, change, pre\_pct, post\_pct, etc.)
  + **Prop\_data\_long.csv**: long-formatted dataset where each row represents a unique participant/domain/time\_point combination, with “variable” representing the time point (pre or post) and “value” representing the score for the given domain/participant

## Data Analysis

For the pre-/post-intensive training, we chose to analyze individually matched training data.

### Internal consistency

For the initial report, we analyzed all domains included in the evaluation questionnaire. However, the WHO evaluation tool and Toronto empathy questionnaire (TEQ) have not been validated in Timor-Leste. Kayli Wild had a secondary aim of validating the TEQ in our project’s context; however, we observed very low internal consistency among the empathy scale (measured via alpha’s Cronbach) when trying to do so. We also saw the same for several other sub-scales used in the WHO evaluation questionnaire, which raises concerns about how participants understood and/or filled out the questionnaire. For this reason, we chose only to analyze domains for which the alpha’s Cronbach was 0.6 or higher for our publication. This will likely be a limitation in your analyses as well (since they rely upon these data) and is something to check in the follow-up data.

### Pre- / post-training improvements

For the initial evaluation report, I utilized parametric tests (the paired t-test) since our sample size was greater than 50 participants. For publication, I slightly changed this approach and applied the Wilcoxon Signed Rank (non-parametric) test. This did not change the overall patterns and significance of our results, but it makes our findings/reporting more directly comparable with [Arora et al](https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-021-12042-7) and [Jayatilleke et al](https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-1674-9). From my understanding, checking for normality is usually only recommended when a sample size is <50; which is the assumption I had relied upon when using the paired t-test. However, since the other authors found evidence of non-normal distribution in their variables, I checked and found this within our data as well. Ultimately, I think both tests are justifiable, but the non-parametric approach is more conservative and makes for easy comparison with other published findings.

### Factors associated with improvement

To investigate factors potentially associated with improvement, I used a mixed effects, multivariate logistic regression models with gender, age, position type, municipality, and month of training as explanatory variables. The dependent variable was operationalized as a binary indicator of score improvement pre- to post-training ([post-training score – pre-training score] > 0). Each model also included a random intercept for each unique participant. I ran separate models for each construct, so that we could assess the effects of each covariate of interest independently by construct. To account for ceiling effects, observations from participants with pre-training, domain-specific proportional scores of greater than 90% were excluded from datasets used in the models.

Here is the official write-up from our paper appendix:

“The probability of improvement in score, was estimated:

*Where:*

* is a binary indicator of participant sex;
* is a binary indicator of participant age group (dichotomized to <35 and 35+);
* is a dummy variable indicating the specific position type, to which the observation belongs;
* is a binary indicator of municipality;
* is a dummy variable indicating the month of training, to which the observation belongs;
* is a binary indicator of observation sub-domain; and
* is a participant identifier random intercept.

Age categories and positions were condensed into broader groups due to small numbers of participants in terminal age categories and positions other than doctor, midwife or nurse (main text, table 2). Month of training was considered as a covariate to account for potential increases in facilitator confidence and competency in delivering trainings as they gained more practice over time. It was not possible to consider years of clinical practice and average weekly patient volume as covariates due to a high degree of missingness in these variables. For attitudes and perceived preparedness constructs (in which multiple domains were assessed), we included an additional covariate to indicate the sub-domain to which the observation belonged.”

### Scripts

1. 01\_cronbach's alpha.R
2. 02\_wilcoxon.R
3. 03\_regression.R