

Evaluating Mental Health Crisis Training Outcomes Among Police, Firefighters, and Public Health Staff

Statistical Analysis Plan

Acronym	TAIWAN-CIT
Principal investigator	Hua Ho, MD Department of Emergency Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan Email: femg97657@femh.org.tw
Study coordinator / contact	Chiao-Yin Cheng, PhD, EMT-P Department of Emergency Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan Email: femhr9969@femh.org.tw
Institution	Far Eastern Memorial Hospital No. 21, Section 2, Nanya South Road, Banqiao District, New Taipei City 220, Taiwan
Ethics approval	Research Ethics Review Committee, Far Eastern Memorial Hospital; approval number 114067-E; approved 15 May 2025
Approved research period	15 May 2025 to 30 April 2027
Document version	Version 1.0, 6 July 2026
Basis of this English version	Prepared from the IRB-approved protocol Version 2 dated 5 May 2025, questionnaire information page Version 2 dated 5 May 2025, e- questionnaire Version 1 dated 4 April 2025, and the study manuscript statistical analysis section.

1. Purpose

This Statistical Analysis Plan describes the planned and manuscript-aligned analyses for the single-group pre-post evaluation of a 4-hour mental health crisis and behavioral emergency response training programme in New Taipei City, Taiwan.

2. Analysis population

- Primary analysis population: all respondents who completed the pre-training questionnaire and are included in the imputed paired dataset.
- Observed matched analysis population: respondents whose pre- and post-training questionnaires can be linked using available non-identifying variables.
- Post-training satisfaction population: respondents who completed the post-training satisfaction items.

Because the questionnaire is anonymous and direct personal identifiers are not collected, true one-to-one linkage may be incomplete. Matching and imputation procedures are therefore prespecified to preserve transparency.

3. Variables and scoring

Item	Description
Self-efficacy	Items 6-13; 5-point confidence scale from 1 to 5. The scale score is the mean of component items. Higher scores indicate greater self-efficacy.
Cognition toward behavioral emergencies	Items 15-20; 5-point agreement scale from 1 to 5. Item 20 is reverse-coded because endorsement of routine force use reflects less favorable cognition. The scale score is the mean of component items.
Behavioral emergency stigma	Items 21-26; dichotomous Yes/No items. Items reflecting recovery or interaction are reverse-coded so that higher values indicate greater stigma. The scale score ranges from 0 to 1.
Substance-use stigma	Items 27-32; dichotomous Yes/No items using the same coding approach. The scale score ranges from 0 to 1, with higher values indicating greater stigma.
Satisfaction	Five post-training Likert-type items assessing work helpfulness, instructor/course content, practical exercise value, course flow, and overall satisfaction.

4. Matching and imputation

Observed pre- and post-training responses will be matched deterministically using available non-identifying variables, including training date, sex, occupation category, birth year, professional experience category, prior mental health training status, and within-stratum timestamp order, where available.

For pre-training respondents without an observed matched post-training questionnaire, post-training outcome values will be estimated using subgroup mean-change imputation. Observed mean changes among matched participants within available subgroups will be applied to the corresponding baseline scores of unmatched participants. If a subgroup is sparse, broader occupation- and sex-based strata may be used. The imputed paired dataset will be clearly distinguished from the observed matched dataset.

Because single imputation can underestimate uncertainty, imputed analyses will be interpreted as sensitivity-oriented estimates rather than purely observed paired evidence.

5. Descriptive analysis

- Continuous variables will be summarised using mean and standard deviation. If distributions are markedly skewed, median and interquartile range may also be reported.
- Categorical variables will be summarised using counts and percentages.
- Participant flow will report total attendees, exclusions before the pre-training questionnaire, pre-training respondents, post-training respondents, observed matched pairs, and imputed post-training outcomes.

6. Reliability analysis

Internal consistency will be assessed using Cronbach's alpha for multi-item Likert scales. For dichotomous stigma scales, the same alpha coefficient will be reported as equivalent to KR-20 under binary scoring. Because alpha is influenced by scale length, dimensionality, and item heterogeneity, short dichotomous stigma scales will be interpreted descriptively rather than judged against a strict acceptability threshold.

7. Primary pre-post analysis

For each outcome scale, post-training minus pre-training change will be calculated. The primary analysis will use paired t tests in the imputed paired sample. Mean differences will be reported with 95% confidence intervals and two-sided p values. Effect sizes for paired data will be expressed as Cohen's dz.

8. Sensitivity analyses

- Wilcoxon signed-rank tests will be used as non-parametric sensitivity analyses for pre-post changes.
- Where relevant, observed matched-pair analyses will be reported or compared with the imputed paired analyses to examine consistency of findings.
- Interpretation will explicitly consider the limitations of anonymous matching and single imputation.

9. Satisfaction analysis

Satisfaction items will be summarised as mean and standard deviation and as the proportion reporting satisfaction or strong satisfaction. Associations between satisfaction variables and changes in self-efficacy or cognition will be evaluated using Pearson correlation coefficients. If distributions are not appropriate for Pearson correlation, Spearman rank correlation may be used as a sensitivity analysis.

10. Subgroup or exploratory analyses

Exploratory analyses may compare baseline scores and pre-post changes by occupation group, sex, years of service, and prior mental health training. These analyses will be interpreted as exploratory and hypothesis-generating.

11. Statistical significance and software

A two-sided p value less than 0.05 will be considered statistically significant. Analyses may be conducted using standard statistical software such as R, SPSS, Stata, SAS, or Python. The software and version used will be reported in the final manuscript where required.

12. Deviations from this plan

Any material deviation from this SAP will be documented and justified in the final study report or manuscript.