

Supplementary Online Content

**Experience of Inpatient Mental Health Care Assessed With
Service User–Developed and Conventional Patient-Reported
Outcome Measures**

Ioannis Bakolis, Ph.D., Prashant Gupta, M.D., Til Wykes, Ph.D.

Table S1: Glossary of Measurement Terms.

| Term | Definition |
|--|---|
| Ability to detect change | Evidence that an instrument can identify differences in scores over time in individuals or groups who have changed with respect to the measurement concept. Effect sizes and 95% confidence intervals can be used to assess this property. |
| Construct validity | Evidence that relationships among items, domains, and concepts conform to a priori hypotheses concerning logical relationships that should exist with other measures or characteristics of patients and patient groups. |
| Convergent validity | An aspect of construct validity. Evidence that relationships between results gathered using the instrument and results gathered using other measures are consistent with pre-existing hypotheses concerning those relationships. Correlations can be used to assess this property. |
| Exploratory Factor Analysis | A statistical method used to uncover the underlying structure of a relatively large set of measured variables. |
| Confirmatory Factor Analysis | A statistical method that is used to test how well the measured variables represent the number of constructs. |
| Instrument | A means to capture data (i.e., a questionnaire) plus all the information and documentation that supports its use. Generally, that includes clearly defined methods and instructions for administration or responding, a standard format for data collection, and well-documented methods for scoring, analysis, and interpretation of results in the target patient population. |
| Internal consistency | An aspect of reliability. A measurement of the extent to which items comprising a scale measure of the same concept. This can be assessed using Cronbach's alpha coefficient (α) or KR-20. |
| Item | An individual question, statement, or task (and its standardized response options) that is evaluated by the patient to address a particular concept. |
| Measurement properties | All of the attributes relevant to the application of an instrument including construct validity, reliability, and ability to detect change. These attributes are specific to the measurement application and cannot be assumed to be relevant to all measurement situations, purposes, populations, or settings in which the instrument is used. |
| Reliability | The ability of an instrument to yield consistent, reproducible estimates of the construct. |
| Scale | The system of numbers or verbal anchors by which a value or score is derived for an item. Examples include visual analogue scales, Likert scales, and numeric rating scales. |
| Score | A number derived from a person's response to items in a questionnaire. A score is computed based on a prespecified, validated scoring algorithm and is subsequently used in statistical analyses of clinical trial results. Scores can be computed for individual items, domains, or concepts, or as a summary of items, domains, or concepts. |
| Stepped wedge randomised controlled trial | A randomised controlled trial in which clusters (e.g. group of patients or hospitals) receive the intervention at different time points, the order in which they receive it is randomised, and data are collected from clusters over time. |

Note: Source for measurement terms: (Food and Drug Administration, 2009)

TABLE S2 Intercorrelations of total scores of VOICE and SSS-Res and sub-scales of VOICE-S, VOICE-T, SSS-Res-E, SSS-Res-C

| Scale | VOICE | SSS-Res | VOICE-T | VOICE-I | SSS-RES-E | SSS-RES-C |
|--|-------|---------|---------|---------|-----------|-----------|
| At Baseline (n=670) | | | | | | |
| VOICE | 1 | | | | | |
| SSS-Res | 0.85 | 1 | | | | |
| VOICE-T | 0.88 | 0.80 | 1 | | | |
| VOICE-I | 0.88 | 0.72 | 0.06 | 1 | | |
| SSS-Res-E | 0.79 | 0.90 | 0.78 | 0.62 | 1 | |
| SSS-Res-C | 0.74 | 0.89 | 0.65 | 0.67 | 0.06 | 1 |
| At the end of intervtnion (n=438) | | | | | | |
| VOICE | 1 | | | | | |
| SSS-Res | 0.84 | 1 | | | | |
| VOICE-T | 0.91 | 0.78 | 1 | | | |
| VOICE-I | 0.89 | 0.73 | 0.06 | 1 | | |
| SSS-Res-E | 0.81 | 0.93 | 0.77 | 0.73 | 1 | |
| SSS-Res-C | 0.76 | 0.93 | 0.70 | 0.69 | 0.07 | 1 |

All $p < 0.001$ for all correlations. VOICE = total score of VOICE questionnaire; SSS-Res = total score of SSS-Res Questionnaire; VOICE-T: trust subscale of the VOICE scale; VOICE-I: involvement subscale of the SSS-Res scale; SSS-Res-E: Environment subscale of the SSS-Res subscale; SSS-Res-C: Care subscale of the SSS-Res scale.

Panel S1. The setting

- Borough 1 serves an inner-city population that has a high deprivation index. Five 18-bedded wards participated in this study, three for men and two for women.
- Borough 2 serves a more suburban affluent area. Three wards participated in our study, one for men and two for women. Two wards had 22 beds and one, a women's ward, had eight beds and did not admit patients under any legal sanction.
- Borough 3 has a high deprivation score and four 18-bedded wards provide acute inpatient care. There were two single gender and two mixed gender wards (a triage ward and an early intervention unit).
- Borough 4 was mixed socioeconomically with two 18-bedded mixed gender wards serving an area with a high deprivation score.
- Borough 5 serves a suburban and affluent area and had two mixed gender, 18-bedded wards.