**Selection of Outcome measures**

**Consensus exercise to identify most important outcome domains**

***Background and issues***

The Delphi method has been widely used in healthcare research for establishing consensus from experts within particular topic areas to set research priorities, establish core professional competencies, develop criteria and measures for quality improvement, and validate the content of measurement tools.

***Method***

*Participants*

Delphi survey panel members were sought from an established ‘peer research group’ comprised of stakeholders with lived experience of offender services, together with professionals.

Participants were requested to score on a five-point Likert scale (1=*not important*, 2=*less important*, 3=*important*, 4=*very important*, and 5=e*ssential)* how important they thought each domain was.

Ratings from service users and professionals were calculated and compared to determine any differences in stakeholder perspectives, but consensus judgements were based on the responses of all participants.

***Results***

The questionnaire was circulated to 37 potential respondents, and 26 (70%) provided completed questionnaires. These comprised 16 professionals (academics, clinical academics and researchers) and 10 members of the peer research group.

The outcome domains of mental disorder presence and severity, self-harm or suicide acts, social inclusion, substance misuse, reoffending, and relationships were ranked as most important (Table 1). Satisfaction with services, economic stability and service usage were seen as least important areas for measurement.

**Table 1: Ranked order responses: What are the key outcome areas that should be measured among offenders with common mental health problems?**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | All respondents (n=26) | | | Professional respondents (n=16) | | | Stakeholder respondents (n=10) | | |
| Domain | Rank | Score:  Mean, Median | % Rating essential or  very important | Rank | Mean, median | % Rating essential or  very important | Rank | Mean, median | % Rating essential or  very important |
| Mental health symptoms (like depression, anxiety, PTSD) | 1 | 4.60, 5.0 | 92.3 | 1 | 4.67, 5.0 | 93.0 | =1 | 4.5, 5.0 | 90.0 |
| Self-harm or suicide acts | 2 | 4.23, 4.0 | 80.8 | 2 | 4.25, 4.0 | 81.3 | 3 | 4.2, 4.0 | 80.0 |
| Social inclusion (access to and use of resources, activities and services such as employment, adequate housing, health care, education and training) | 3 | 4.04, 4.0 | 76.9 | 3 | 4.13, 4.0 | 68.8 | =2 | 4.3, 4.0 | 90.0 |
| Substance misuse (harmful/ dependent use of alcohol or drugs) | 4 | 4.15, 4.0 | 76.9 | =5 | 3.94, 4.0 | 68.8 | =1 | 4.5, 5.0 | 90.0 |
| Reoffending (repeat offending behaviour) | =5 | 4.08, 4.0 | 73.1 | =5 | 3.94, 4.0 | 68.8 | =2 | 4.3, 4.5 | 90.0 |
| Relationships (with family, friends, others) | =5 | 4.08, 4.0 | 73.1 | =6 | 3.81, 4.0 | 62.5 | =1 | 4.5, 5.0 | 90.0 |
| Quality of life (well-being and life-satisfaction) | =6 | 4.04, 4.0 | 80.0 | =4 | 4.07,4.0 | 73.3 | =5 | 4.0, 4.0 | 90.0 |
| Social function (day-to-day activities that involve contact with people) | =6 | 4.04, 4.0 | 76.0 | =4 | 4.07, 4.0 | 73.3 | =5 | 4.0, 4.0 | 80.0 |
| Engagement with treatment | 7 | 3.92, 4.0 | 69.2 | =5 | 3.94, 4.0 | 68.8 | =6 | 3.9, 4.0 | 70.0 |
| Adherence or concordance with treatment (sticking to treatment plans/ or planned behaviour changes) | 8 | 3.88, 4.0 | 73.1 | =6 | 3.81, 4.0 | 68.8 | =5 | 4.0, 4.0 | 80.0 |
| Employment | 9 | 3.77, 4.0 | 57.7 | 7 | 3.44, 4.0 | 68.8 | =2 | 4.3, 4.0 | 90.0 |
| Stages/ readiness to change (preparedness for behaviour change) | 10 | 3.58, 4.0 | 53.8 | 10 | 3.25, 3.0 | 43.8 | 4 | 4.1, 4.0 | 70.0 |
| Physical health | 11 | 3.56, 4.0 | 52.0 | 8 | 3.33, 3.0 | 40.0 | =6 | 3.9, 4.0 | 70.0 |
| Service outcomes (how much someone uses national/ local services) | 12 | 3.42, 3.0 | 46.2 | 9 | 3.31, 3.0 | 37.5 | 8 | 3.6, 4.0 | 60.0 |
| Economic and financial stability (personal and societal costs and contributions) | 13 | 3.38, 3.0 | 42.3 | 11 | 3.19, 3.0 | 25.1 | 7 | 3.7, 4.0 | 70.0 |
| Satisfaction with services | 14 | 3.08, 3.0 | 26.9 | 12 | 2.88, 3.0 | 12.5 | 9 | 3.4, 3.5 | 50.0 |

**Field testing of candidate measures**

***Background***

This stage involved administering the candidate measures to men detained in prison who were near to release, to test for acceptability and assess the psychometric properties in the target population. We additionally sought to clarify certain statistical parameters as many of the measures had not been validated in a prisoner population

***Method***

Participants were 61 incarcerated men recruited from two prisons in England. Participants were between 4 and 12 weeks from release from prison and were being released to the geographical area of the study. All participants screened in for having, or likely to have on release, a common mental health problem (see Lennox et al, 2018 for details).1

All participants completed a demographic measure together with various selected outcome measures. For the candidate measures, participants were subdivided into three groups of 20 to complete the measures on either mental health, social inclusion, or substance misuse. All participants completed the PHQ-9 and GAD-7 as these were used as screening tools for determining the presence of common mental health problems. All measures were suitable for self-completion; however, to avoid literacy problems, a researcher read the questionnaire items to each participant.

In relation to acceptability, participants rated candidate measures on a 6-point Likert scale to assess views on length (scored ‘*not long*’ to ‘*very long*’), difficulty of understanding (scored ‘*not difficult*’ to ‘*very difficult*’), and perceived relevance (scored ‘*not relevant*’ to ‘*very relevant*’).Mean scores on rating of length, difficulty, and perceived relevance were calculated.

***Results***

Tables 2-4 detail the results in relation to the different measures.

**Table 2: Measure characteristics, mental health and illness features**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measure (score range) | n  (% complete) | n (%) reaching standard cut-point [≥x] | Mean (SD) | Minimum Clinically Important Difference (MCID) | N (%) scoring ≤ MCID | Acceptability | | | Readability | |
| Length  (1: *not long* - 6: *very long*) | Difficulty  (1: *not difficult* – 6: *very difficult*) | Relevance (1: *not relevant* – 6: *very relevant*) | Flesch Reading Ease | Flesch-Kincaid Grade Level |
| PHQ-9  (0-27) | 61 (100%) | 32 (53%)  [≥10] | 10.5 (6.4) | 5 | 16 (26%) | 1.3 | 1.7 | 4.3 | 68.3 | 6.5 |
| GAD-7  (0-21) | 61 (100%) | 28 (46%)  [≥10] | 9.2 (6.2) | 4 | 17 (28%) | 1.3 | 1.9 | 4.2 | 65.2 | 6.0 |
| GHQ-12  (0-36) | 20 (100%) | 8 (40%)  [>15] | 15.5 (6.8) | 3 | 0 (0%) | 1.7 | 2.5 | 4.4 | 68.7 | 6.0 |
| CORE-OM  (0-40) | 20 (100%) | 14 (70%)  [≥10] | 13.1 (5.6) | 5 | 2 (10%) | 2.0 | 2.5 | 4.7 | 77.1 | 4.6 |

**Table 3: Measure characteristics, social inclusion**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Measure (score range) | n  (% completing) | Mean (SD) | Acceptability | | | Readability | |
| Length  (1: *not long* - 6: *very long*) | Difficulty  (1: *not difficult* – 6: *very difficult*) | Relevance (1: *not relevant* – 6: *very relevant*) | Flesch Reading Ease | Flesch-Kincaid Grade Level |
| APQ-6 (n/a) | 21 (100%) | (n/a) | 1.3 | 1.4 | 4.7 | 76.3 | 5.4 |
| CIM (10-50) | 21 (100%) | 41 (6.4) | 1.3 | 4.4 | 4.6 | 74.1 | 5.8 |
| SIQ (20-80) | 21 (100%) | 54.6 (15.4) | 1.3 | 1.6 | 4.5 | 71.9 | 6.0 |
| SCOPE (n/a) | 21 (100%) | n/a | 2.1 | 2.1 | 4.7 | 53.1 | 9.8 |

**Table 4: Measure characteristics, drug and alcohol misuse**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measure (score range) | n  (% completing) | n (%) reaching standard cut-point [≥x] | Mean (SD) | Minimum Clinically Important Difference (MCID) | N (%) scoring ≥ MCID | Acceptability | | | Readability | |
| Length  (1: *not long* - 6: *very long*) | Difficulty  (1: *not difficult* – 6: *very difficult*) | Relevance (1: *not relevant* – 6: *very relevant*) | Flesch Reading Ease | Flesch-Kincaid Grade Level |
| TOP  (n/a) | 20 (100%) | n/a | n/a |  |  | 1.5 | 1.8 | 5.1 | 68.8 | 8.0 |
| AUDIT  (0-40) | 20 (100%) | 14 (70%)  [≥8] | 17.1 (14.6) |  |  | 1.2 | 1.3 | 3.8 | 68.0 | 8.1 |
| LDQ  (0-30) | 20 (100%) | 16 (80%)  [≥10] | 19.9 (9.6) |  |  | 1.4 | 1.5 | 5.0 | 74.3 | 8.5 |

**References**

1. Lennox, C., Kirkpatrick, T., Taylor, R.S., Todd, R., Greenwood, C., Haddad,et al. Pilot randomised controlled trial of the ENGAGER collaborative care intervention for prisoners with common mental health problems, near to and after release. Pilot and Feasibility Studies. 2018;7(4):15. doi: 10.1186/s40814-017-0163-6