Recommendations following the response to national consultation on
the standardisation of outcomes measurement for adult psychological therapy
services across Scotland: a discussion and consultation paper.

1.0 Introduction

Work to implement the new psychological therapies HEAT target recognises that clinical outcomes and access to treatment are both important to service users. Given the potential interaction between these two parameters, a paper was prepared in order to inform discussion and facilitate consultation with services to address the following questions:

1. Should we seek to establish one or more standardised outcome measures that are used across adult psychological therapy services in Scotland?
2. What additional support might Boards require to support them in routinely measuring and monitoring outcomes?
3. What, if anything, should Boards be asked to report nationally with regards to the use of outcome measures locally.

The consultation paper was circulated to all NHS Boards through a variety of routes. This report summarises the outcomes of that consultation process, making recommendations on next steps.

Twenty four responses were received and Appendix A of this document collates the responses under each question for information. The recommendations have been put together in discussion with a number of clinicians who sit on the Psychological Therapies HEAT Implementation Group. (Moira Connolly (Principal Medical Officer for Mental Health, Scottish Government); Mike Henderson (Consultant Clinical Psychologist); Kevin Powers (Consultant Clinical Psychologist) and Michael Smith (Consultant Psychiatrist)).

2.0 Recommendations

2.1 Standardising outcome measures across adult psychological therapy services in Scotland.

There was a general consensus that embedding the use of clinical outcome data in the delivery of adult psychological therapy services is desirable, albeit the challenges of operationalising this are significant.

The majority opinion from this consultation was that we should standardise to CORE1 as the main measure across adult psychological therapy services, with other measures used in conjunction with it as clinically appropriate. The benefits for standardising include:

---

1 The CORE System (Clinical Outcomes for Routine Evaluation) has been designed in the UK for use in psychotherapy, counselling and other psychological therapies to measure outcome and provide for service audit, evaluation and performance management. The CORE Outcome Measure (CORE-OM) is a 34-item questionnaire designed to measure a pan-theoretical ‘core’ of clients’ global distress, including subjective well-being, commonly experienced problems or symptoms, and life/social functioning. In addition, items on risk to self and others are included to assist risk assessment in NHS and other sectors.
• Allows clinicians and managers to benchmark outcomes across similar services. This then enables significant variation (both positive and negative) to be identified and explored. This will support reflective practice and the replication of best practice across Scotland.
• Makes it easier for patients and staff who move between systems if services are using common outcome measures.
• Makes it easier and more efficient to provide relevant and high quality training and support on the use of clinical outcomes measures nationally.
• Will provide efficiencies around IT system developments to enable electronic capture and reporting of outcomes.

However, within the responses there were some key issues raised about the appropriateness of this measure for all client groups. Therefore we are recommending:

• Adult psychological therapy services should standardise to using CORE as the main measure. In learning disabilities and forensic services, the evidence base in support of CORE is less well established. Services in these specialist areas may consider using other measures where these show superiority over CORE. Any service may complement CORE with other outcome measures applicable to local needs and context. There are currently no CORE tools aimed at individuals with dementia.

• Further work is needed to establish whether it is possible and appropriate to standardise measures for adult psychological therapy services working in the fields of dementia, learning disabilities and forensic services.

• Further, even across generic adult psychological therapy services, there will be individuals for whom CORE is not appropriate. It is recommended that a similar approach is taken to the ICP standards: CORE is the recommended outcome measure across Scotland for adult psychological therapies but deviation from this norm is acceptable at an individual patient level where there is sound clinical justification.

• No recommendation is being made, at this stage, as to which of the CORE tools to use. **However rating at every client contact is recommended** as a range of studies show this to have beneficial impact on data validity (eg Clark et al, 2009) and post treatment outcomes (eg Lambert 2010 and Duncan 2010). Lambert’s work also suggests that session by session monitoring by therapists can reduce the failure to attend (“DNA”) rate.

One model services may wish to consider is the method recommended by CORE which involves using CORE OM at the start and end of treatment and CORE10 (10 item questionnaire) or CORE5 (5 item questionnaire) at each of the first 4-6 sessions and monthly thereafter. Other approaches in use within Scotland include using CORE OM at start, middle and end of treatment and the CGI at every appointment.

• We are **not** recommending that adult psychological therapy services only use CORE - it will often be appropriate to supplement CORE with relevant diagnostic specific clinical outcome measures and services should be encouraged to do this as clinically appropriate.

2.2 National Reporting Recommendations
• It is recommended that initially NHS Boards are asked to report what % of services delivering psychological therapies are routinely using clinical outcome measures at the
individual patient level. This will be included as part of the quarterly risk assessment returns for the delivery of the psychological therapies HEAT target (on the basis that if services are not routinely collecting and using clinical outcome data there is a risk that improved access could be delivered by actions which impact negatively on clinical outcomes).

- In future years, this criteria will be added to so that NHS Boards are asked to indicate what % of services are routinely using clinical outcomes data within the clinical supervision process and what % of services are aggregating and monitoring clinical outcomes data at a service level.

- Once the use of clinical outcome data is embedded at service level, it is recommended that the Mental Health Benchmarking Implementation Group progresses work to include aggregated data on clinical outcomes as part of the overall balanced scorecard returns.
2.3 Additional support to help embed the routine use of clinical outcome data

The responses indicated some issues that would benefit from further consideration nationally. The following table summarises these and the action that has now been put in place to address them.

<table>
<thead>
<tr>
<th>Issue</th>
<th>Action to be taken nationally</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The issue of access to IT systems that support clinicians to efficiently use and analyse data was raised.</strong></td>
<td>Work is progressing nationally to scope the current capacity of NHS Board’s IT systems to support routine analysis and reporting of clinical outcome data at both a patient and service level.</td>
</tr>
<tr>
<td>This was also a consistent theme raised by clinicians attending the Mental Health Improvement Game (MHIG) which has been run with over 400 staff across Scotland. The MHIG promoted the routine use of clinical outcome measures and the day included the opportunity for individuals to discuss the barriers to this progressing.</td>
<td>Each NHS Board will be contacted and asked to self assess their current clinical outcomes reporting capability against a list of useful analysis.</td>
</tr>
<tr>
<td>The issue is about both the existence of an IT system that supports efficient use and analysis of data and access to the IT system.</td>
<td>Where this functionality does not exist, this work will also assess the potential costs of developing it either within existing systems or through the purchase of a system such as CORE-NET. This work will include looking at the availability of IT to enable data input and use.</td>
</tr>
<tr>
<td><strong>There was some confusion over the financial costs attached to using CORE.</strong></td>
<td>This issue will be taken on nationally as part of the work identified to scope the costs of putting in place IT systems which support clinicians to use outcome data as part of their everyday practice.</td>
</tr>
<tr>
<td>The CORE forms and manuals may be copied and used by an organisation or individual without payment of any copyright fee on the condition that: the forms are not altered or changed in any way; due acknowledgement of copyright is reproduced; and reproduction is not for commercial gain.</td>
<td>Work will also be undertaken nationally to outline the case for change – ie the potential quality and efficiency gains attached to routine use of clinical outcome measures. This case for change will be a vital part of progressing work which requires financial investment.</td>
</tr>
<tr>
<td>NHS Boards can put the scores into their electronic systems without breaching copyright as long as they don’t reproduce the CORE text in any way. Entering the text of the CORE forms into a computer, would be considered a breach of copyright unless a licence agreement is in place. Applications can be made to the CORE System Trust for a licence which would normally be granted on payment of an annual fee.</td>
<td></td>
</tr>
<tr>
<td>In addition to the issues raised within the consultation, NES are also looking at any training issues around the routine use of clinical outcome measures including how to effectively integrate clinical outcome data into clinical supervision processes.</td>
<td>NES will ensure that training in the use of clinical outcome data to inform psychological therapies supervision is built into the training of psychological therapies supervisors.</td>
</tr>
</tbody>
</table>
3.0 Summary and actions required

3.1 By NHS Boards
In summary, it is recommended that all adult psychological therapy services across NHSScotland standardise to use CORE as the main measure; with other measures used in conjunction with it as clinically appropriate.

In response to this recommendation NHS Boards will need to:
- Assess which clinical outcome measures are currently being used by adult psychological therapy services.
- Where no measure is being used, plans should be put in place to introduce clinical outcome measurement (please note that this is already a requirement of ICP Standard 21) and for general adult psychological therapy services this should include CORE.

As part of the Psychological Therapies HEAT Risk Assessment, NHS Boards are required to report on what % of staff delivering psychological therapies are routinely using clinical outcomes data. At present Boards are simply asked to report whether 0-25%, 25-50% or over 50%

- Therefore NHS Boards will need to ensure systems are in place locally for identifying which staff are delivering psychological therapies and what percentage of these are routinely using clinical outcomes data.
- NHS Boards will also need to start work to enable recording and reporting of clinical outcomes scores.

Work is currently in place to summarise the reports that need to be available to clinicians and managers to enable clinical outcome data to be embedded as part of individual patient reviews, clinical supervision processes and service level monitoring.

- NHS Boards will be contacted and asked to self assess their current capability against this ideal set of reporting and, where the current systems do not have the functionally, the potential costs of developing it. This will enable further work to be undertaken nationally on the options, costs and benefits of moving to a situation where all the relevant reports are routinely available to clinicians and managers.

3.2 Summary and actions required nationally
Throughout this consultation there has been a general consensus that embedding the use of clinical outcome data in the delivery of adult psychological therapy services is desirable, albeit the challenges of operationalising this are significant.

It is recognised that delivering the routine measurement of clinical outcomes in practice is challenging and that it will not be delivered overnight. However, the routine use of a professionally validated tool to measure outcomes has been a standard for mental health services since 2007 (ICP Standard 21) and, by putting an increased emphasis on this nationally, our aim is to support those tasked with delivering this. In addition the following action will be taken nationally:

- NES will ensure that training in the use of clinical outcome data to inform psychological therapies supervision is built into the training of psychological therapies supervisors.
• Work will progress nationally to scope the current capacity of NHS Board’s IT systems to support routine analysis and reporting of clinical outcome data at both a patient and service level. This will include scoping the costs of putting in place IT systems which support clinicians to use outcome data as part of their everyday practice.

• Work will also be undertaken nationally to outline the case for change – ie the potential quality and efficiency gains attached to routine use of clinical outcome measures. This case for change will be a vital part of progressing work which requires financial investment.

• Work will progress nationally to conduct a similar exercise to this one focused on Child and Adolescent Mental Health services. Further work will progress to assess the viability of identifying one or more standardised clinical outcome measure for psychological therapists working with: individuals with a forensic background, individuals with a learning disability and individuals with dementia.

Reshaping Care and Mental Health Division
Scottish Government
Sept 2011
APPENDIX A – Collation of Responses to Consultation

1. Do you agree with the key criteria identified against which any measure assessing outcomes in PT should be assessed? If not, what would you change?

<table>
<thead>
<tr>
<th></th>
<th>23 but a number of responses made suggestions for additional criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
</tr>
</tbody>
</table>

Additional Criteria Suggested
1. Provision of useful and meaningful data, system and funding in place for data analysis, management and dissemination.
2. Arguably there is a case for recording a small amount of qualitative data (e.g. an "Any other comments about how you feel" box) for possible thematic analysis, but I appreciate this is substantially more time-consuming both to record and analyse.
3. but in LD a carer/ other involved professional perspective may be relevant
4. The list of criteria represents an appropriate range of outcomes to consider but omits a carer perspective where this is relevant.
5. however since most people who attend adult psychology will usually have some symptoms of anxiety & depression I think it's important that these are also measured.
6. Key criteria all relevant, any measures needs reliability & validity
7. Essentially, yes. I would suggest that “can track change over time” should be amended to “sensitive to change”, as the two are not necessarily the same. For example, someone might go from a score of 100 (bad) to 0 (good), showing the tool can indeed track change, but it may be completely unable to detect intermediate states that are of clinical relevance.

Further, I’m not totally convinced that the ICP standards should necessarily be the benchmark; although in most cases the needs of psychotherapy outcome measures and the ICP standards are likely to be similar. Once particular tools become embedded in everyday use, they don’t change easily. However, ICP standards will probably change/ evolve over time whereas the requirement for good clinical outcomes measures will remain the same regardless of what the standards are.

The only other criterion I would add would be that we should have a good knowledge of the psychometrics of the scale (such as sensitivity to change) as well as having good normative data so that we know how our patients’ scores compare to other similar patients.
2. Do you agree with the measures chosen to consider in more detail against the key criteria?

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>18</td>
</tr>
<tr>
<td>No</td>
<td>5</td>
</tr>
</tbody>
</table>

**Yes – additional comments**

1 - The list of measures represents an appropriate range of outcomes to consider but need to include some which can be used with less cognitively able groups and rated by carers. Also, given the numbers of people with severe and enduring conditions, the items within HONOS may also be relevant for significant numbers of people referred for psychological interventions with long term conditions.

2 - Looks fairly comprehensive, and can’t see any stand-out omissions

**No – additional comments**

3 - Wondering why SCL90 was excluded - cheap, easily administered, enormous credibility,

4 - No, I think the Depression, Anxiety and Stress Scale (DASS) should be considered

5 - Needs to consider inclusion of outcomes which can be used by those with cognitive limitation and by carers.

6 - I think the HADS should be included. I agree that the CORE should be included but depts need support in purchasing the license to facilitate scoring the measure as it is pretty time consuming to score which can put people off using it.

7 - I think that it is important to have a good understanding of the intended purpose of each scale, and how the scale performs under certain conditions. For example, the HADS was primarily developed as a screening tool for anxiety and depression (Zigmond and Snaith, 1983), although it is frequently used to measure severity and progress and in such situations it may perform satisfactorily (Bjelland et al., 2002).

Further, despite the wish for a ‘one size fits all’ solution, different scales perform differently depending on the use of the scale. For example, Ian Reid has presented data on caseness depending on which depression scale is used. Similar findings have been reported elsewhere (Kendrick et al., 2009). This is an essential issue to address since the choice of scale may impact not only on the apparent prevalence of a disorder such as depression, but also the organisational and national responses.

Otherwise, the scales selected (EQ-5D, GAF, WASAS, CGI, CORE, and HoNOS) are commonly-used scales which are not specific to a particular disorder. However, how the scales behave (and their usefulness/ appropriateness) does seem to be dependent on the underlying disorders being rated and some further awareness of this seems appropriate; particularly since the utility of the HoNOS in psychotherapy situations is questionable (Audin et al., 2001). There are also issues about its sensitivity to change (Trauer et al., 1999). (classed as a no)
3. Do you agree with the analysis of measures against key criteria? If not, please specify what you would change.

<table>
<thead>
<tr>
<th>Yes</th>
<th>21 – with 3 of those responses qualified, please see additional comments for detail.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>2</td>
</tr>
</tbody>
</table>

**Yes additional comments**

1. Broadly yes, although I would argue that CORE-10 and CORE-OM/34 are very different measures both from a client and clinician point of view, with different strengths/weaknesses. I do not think it is entirely correct to say that CORE-OM/34 can be completed in less than three minutes.

2. On the whole yes. My understanding though is that CORE only requires a licence when being used electronically. However if it becomes standardized, the use of the software and electronic forms would be the most beneficial and efficient way of using the CORE system.

3. More or less. Whether a scale “tracks change over time” is not a dichotomous measure and this needs to be addressed. As discussed above, the sensitivity to change is a crucial issue and different scales perform differently depending on the population being used and the severity of illness of that population. For example, there is evidence that the EQ-5D has ceiling effects (Brazier et al., 2004) and this may affect its appropriateness for a particular population. The responsiveness of the EQ-5D in depression has also been questioned (Günther et al., 2008). This doesn’t make it a bad tool but using a tool without understanding its limitations and addressing these as part of a comprehensive outcome assessment is problematic.

The WASAS has sufficient evidence of sensitivity to change (Mundt et al., 2002) and this box should be ticked.

The CGI can be completed by patients and often provides a useful way of triangulating clinician-rated outcomes (Fieve et al., 1986). It’s called the Patients’ Global Impression of Change (PGIC) scale and uses identical scoring. It’s frequently used in research and we have used it alongside clinician ratings and found strong correlations: r=0.925 for patient-ratings and r=0.804 for next-of-kin rated CGI-I scores (Christmas, 2006).

I’m not sure about the validity of the “aids clinical decision-making” criterion since it would be strange if ratings of quality of life and functioning (e.g. GAF, WASAS) had no bearing on how we make clinical decisions. I would therefore challenge if this criterion is clear in terms of what is being measured. The document seems to suggest that only the CORE is able to affect clinical decision-making and this is clearly wrong. I suspect this is because the CORE has cut-offs but these depend on the context (Connell et al., 2007) and it would be disappointing if rating scale scores alone are used to determine suitability for services or the clinical response/ treatment decisions.

I can’t understand how the CORE-OM is able to tick a box for the S33 depression ICP standard (“A validated measure of depression is used at initial assessment and repeated at regular intervals to monitor progress and outcome”) since CORE is not a depression scale, despite the fact that it asks about symptoms of low mood. It doesn’t even have a depression domain. I suspect this reflects misunderstandings between a depression rating scale and a scale that asks about depressive symptoms.

Further, the CGI has been demonstrated to have adequate sensitivity to change in depressed populations and compares favourably with the HAM-D and MADRS (Fieve et al., 1986, Khan et al., 2002). The suggestion that the CGI is not appropriate for measuring depression is inconsistent with the literature. It depends what you want to measure. Further, it is worth noting that the CGI
has been modified for use in Bipolar Disorder (Spearing et al., 1997), Schizophrenia (Haro et al., 2003), and Borderline Personality Disorder (Perez et al., 2007). This is where the ICP standards are not necessarily the gold standard – having to use a “professionally rated assessment tool” (Standard 24b) should not be prioritised if a patient-rated tool is perfectly adequate and non-inferior. The lack of clarity about what is being mandated by the ICP standards is indicated by the suggestion to use the MADRS or YMRS which are not the most meaningful indicators of outcome in bipolar disorder when most patients will spend little time in a (hypo)manic state and cognition and social functioning are often neglected in the management of Bipolar Disorder.

4 - Just one thing: are you sure that the CORE system requires purchase of a license? My understanding was that you need to buy a license if you want to use CORE-NET but the measures, themselves, were free to use

5 – Agree. In addition CORE offers risk of self harm to be highlighted by patients, even in brief version.

No additional comments

6 - No. SCL90 as above is preferable. CORE has not to my knowledge appeared in any outcome study that was not supported by the originators unlike SCL90 which is everywhere!

7 - No. I would include an anxiety and depression measure.
4. Do you agree that Scotland should adopt CORE as the standard outcome measure for psychological therapies across adult services, with additional diagnostic specific measures being used as appropriate?

<table>
<thead>
<tr>
<th>Yes</th>
<th>20 – with 3 responses qualifying that not suitable for all client populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>4</td>
</tr>
</tbody>
</table>

Yes
1 - Yes for those populations for whom it is appropriate. The needs of other populations such as people with learning disabilities, dementia or limited literacy are not met by the standard CORE-OM. Further work is needed to address the requirements of these populations.
2 - Yes for majority of population to whom it applies, but not for those who cannot complete it, or for whom goals of intervention are broader than emotional change, and for whom carer rating may be significant in determining outcomes.
3 - Yes with additional specific measures being used as appropriate which might include adapted measures for LD.
4 - Yes, though -- as the paper makes clear -- there is an important question about which version to use.
5 - Yes - but some colleagues have suggested that HONOS be also used as clinician/observer-rating in addition to patient self-rating based on CORE; but that is a minority view, majority view is CORE is enough.
6 - Agree -would recommend use of CORE-Net because this is a web based resource and would alliviate many of the fixed asset issues in remote and rural settings. Also important that services are encouraged to develop the use of specific additional outcome measures (eg HONOS, AVON Mental Health Measure) that may be more appropriate for those with "severe mental illness".
7 - Yes - and in the first instance the diagnostic specific measures should be optional. It will be hard enough getting the CORE used routinely, without adding to the task. Once clinicians and services are accustomed to gathering the information, and can appreciate the benefits for their clinical practice and service management, we can move on to something more ambitious.
8 - Yes. I think it represents a 'best fit' standardised tool. I think that practitioners should be encouraged to use it alongside profession specific outcome measures (where and when appropriate) to demonstrate the scope and finer detail of their treatment i.e. CORE doesn't in itself demonstrate all of the outcomes of Music Therapy. However when used across the whole of PT in Scotland , it offers an opportunity to draw together lots of strands of work to create a bigger picture.
9 - Yes particularly like that this is service user completed. Clinicians have opportunity to use own outcome measures in addition.
10 - Yes, we already use it in our dept.
11 - Yes, although there is a financial implication which Boards may be unable to commit to due to the current economic climate
12 - Definitely we need an evaluation of therapies tool, and in clinical supervision. Also as audit for whole service & in future development.

13 - Yes, very much. It is used widely across practice in the UK, and, both from research and in my own clinical experience, is an excellent tool which provides a valuable overall indicator of a client's levels of distress, while also acting as a useful tool for stimulating clinical dialogue: including the use of risk items. It is also part of a family of systems, which has the advantage that
there are related tools that easily fit into it, though I think a clear decision should be made about whether or not the CORE-10 or CORE-34 is used (in my experience, the outcomes are not entirely comparable). Also, I think it is imperative that the standard is that clinicians use it EVERY SESSION, so that there is always last observation data that can be carried forward. If it only intended for use at beginning and end of intervention, the missing data will be such that the meaning of the nation-wide data will be pretty much unintelligible. Although clinicians may balk at this, most - for instance, within IAPT - get used to it, and clients are generally very positive about it.

14 - yes. Within CMHT I feel using the full form pre, (mid, if long term therapy) and post would be appropriate.

No – additional comments

13 - No. CORE publications relate to primary care counselling services and then they are not good results. They were developed on psychotherapy clientele of 10 years ago (I participated in this and our client group is very different now)

14 - There is no doubt that the CORE is a popular scale, particularly in psychological therapy services. However, not all psychological therapy is delivered by psychologists and there are a number of potential issues with the CORE:
1. It was developed for psychological therapy services, not mental health services in general.
2. The CORE has really only been demonstrated to be a useful tool in psychological therapy services. Less is known about its performance in chronically depressed, psychotic, or personality-disordered populations. Of course, most such patients are not necessarily treated in psychological therapy services and there needs to be a robust approach to ensuring that meaningful data can be obtained from different patients within different parts of the service.
3. Barkham et al (1998) didn’t seem to intend the measure to be the only outcome measure to be used by services. The popularity of the scale and external drivers increases the risk that it might be used exclusively and this isn’t appropriate.

An illustration of the problems being treated within services using the CORE comes from a recent publication by Barkham et al (2011). Ten percent of the sample was deemed to be ‘subclinical’ and the reasons for treatment were:

- Anxiety 68.1%
- Depression 63.4%
- Interpersonal/relationship 44.7%
- Self-esteem 41.1%
- Bereavement/loss 28.1%
- Trauma/abuse 18.2%
- Work/academic 17.5%
- Living/welfare 13.9%

I would like to suggest that outcome measures for psychotherapy need to be applicable to all patients being seen in MH services, and this problem list suggests that two potentially different populations exist. We can’t have different approaches to outcome assessment for patients that might move relatively freely between different clinical services, depending on need.

I think one of the problems arises from the statement: “CORE and HONOS are probably therefore the only two measures suitable for the monitoring of PT outcomes in Scotland.” It suggests homogeneity among those having psychotherapy that is probably not there, and it is not possible to automatically conclude that the same outcome-measurement approach for psychotherapy is generalisable to all disorders and all interventions. This is problematic since it would be unusual for most patients to be receiving only one intervention (psychotherapy). The more appropriate aim would arguably be how to measure outcomes in particular disorders rather than how do we
measure outcomes from a particular intervention, which itself might consist of different modalities. Unless psychological therapy simply means CBT these days?
The weaknesses of the CORE are discussed above, particularly the uncertain utility in measuring outcomes in patients who are not necessarily the ‘bread and butter’ of psychologists. There is minimal information about how it performs in patients typically seen in CMHTs. It may perform well, but I would argue that we should attempt to find this out before adopting it. An additional issue relates to how complete a record might be obtained from the CORE since completion rates (post therapy) appear to be less than 30% in a service allied to the developers of the tool (Barkham et al., 2011).

Similarly, the HoNOS has its problems; some of which are discussed above. It’s worth noting that the developers of the HoNOS stated that: “Sequential ratings are not a good method for assessing outcome” and that “there are serious problems in using HoNOS as a routine measure of clinical status in busy psychiatric services” (Bebbington et al., 1999). Another study concluded “Certain items, notably 11 and 12, were unreliable. The absence of evidence of sensitivity to change may be due to the short re-rating interval, little real change in the clients, or the characteristics of the scale itself.” (Trauer et al., 1999) Inter-rater reliability, along with test-retest reliability may be poor (Trauer et al., 1999).

The justification for stating that the HoNOS is suitable for psychotherapy outcomes is unclear and I’m not convinced that the case has been made.

15 - There is concern about the implication that one measure will be able to capture the differences between the different services delivering psychological therapies. Different services at different levels of service delivery have different aims- is there a single scale which captures all these differences? . Outcomes in the patient group seen in psychotherapy departments are typically multi-dimensional and we are not convinced that one scale is enough to measure outcomes from all patients in Scotland. It has been suggested that Core does this but it was developed at a time when psychotherapy departments were seeing different populations than they now see.

Since the development of primary care mental health services the less complex patients are seen in primary care and the more complex, co morbid patients are seen in psychotherapy services and secondary care psychology. The psychotherapy departments who have been using core for some time for long term treatments are not happy with it and feel it fails to capture both something of the level of disturbance and something of intrapsychic change in these more complex patients. Some departments have stopped using Core for this reason and a large number of other departments are actively looking at alternatives as they are unhappy with Core capturing these aspects of these very complex individuals.

Looking at the research published by the Core group it is clear that the vast majority of the work comes from primary care services and a tiny amount from secondary or tertiary care services (Stiles et al Psychological medicine 2006). The most recent paper from the CORE group (Barkham et al 2011 psychology and psychotherapy: Theory research and practise), shows a pitiful return rate after treatment has been completed- and this in services which are fully signed up to the CORE model. Only 9,761 of 33,587 patients were completers. This means that we only have reasonable outcome data in a service allied to the developers of the CORE in 29.1% of patients. Ten percent of patients were from a ‘subclinical’ sample, so what this tells us about the performance of the CORE in mainstream adult mental health services and secondary and tertiary psychology and psychotherapy services is unclear.

We feel that we need to have greater confidence in the measures to use them as single outcome measures across the board. We do not feel the evidence is there at a sufficiently strong level to adopt this as a measure for these services.
There is also a concern about the resources needed to input CORE data - the admin staff in the departments which use it find that they are duplicating a lot of the work which also needs to be inputted into patient information systems.

Some colleagues have raised the question about the CGI and have suggested that there is good evidence for this as well as it being focused on the individual and hence more closely targeting change. Being focused on the individual addresses the one size fits all conundrum. It is also extensively used in mental health services. The Patients Global Impression of Change scale has been suggested by a number of clinicians. Using the CGI would also allow comparison between patients receiving pharmacological treatments as it is well validated in this sample. Colleagues have asked about how the data will be used and how meaningful will comparisons be given the wide variety in services and the different resources in different areas? Should we be trying to identify the questions we are trying to answer and then find the most appropriate scales to address the questions rather than assuming that one scale will give us all the information we are looking for? Are we interested for instance in quality of life scales as well as symptom scales? It has been suggested that we pilot a number of measures over 3-6 months, which could probably get sufficient information on outcomes on e.g. CORE, CGI, a disorder-specific scale, EQ-5D (along with diagnostic data and a baseline CGI-S score) to tell us something useful about whether we should be adopting it. Colleagues have also raised the question about the impact of the scales chosen on the work then delivered in a department. Will services become tailored to scale outcome not patient need?
5. What clinical outcomes data do you think should be reported nationally?

Everyone who responded thought that some national outcome reporting should take place. In terms of the specific nature of this

<table>
<thead>
<tr>
<th>Respondents</th>
<th>10</th>
<th>Respondents thought we should be reporting whether outcome monitoring is routinely in place for all services which delivery psychological therapies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - National reporting should state whether outcomes monitoring by psychological therapy services is in place in health boards.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 - Outcome data consistent with ICPs requirements.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 - Reporting whether outcome monitoring is routinely in place for all services which deliver psychological therapies.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 - Information that monitoring is taking place.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 - Reporting whether outcome monitoring is routinely in place for all services which deliver psychological therapies.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 - I am not sure. There are pros and cons for detailed reporting, the cons mainly being the potential for misinterpretation. On balance, I think reporting whether outcome monitoring is routinely in place for services is the best option.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 - Whether outcome monitoring is in place.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 - Reporting of whether routine outcome data is collected in services.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 - Reporting whether outcome monitoring is in place with examples of how results are analysed and utilized.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10- Perhaps what % services use the outcome tool, % completion.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 4 | Initially, Reporting whether outcome monitoring is routinely in place for all services which deliver psychological therapies, preferably moving to more detailed reporting in time. |
|-----------------------------------------------|
| 1 - Initially perhaps that outcomes are being monitored routinely rather than specific data. Later on - may be specific projects that require more detailed data |
| 2 - Whether outcome monitoring is in place. Once we have 90% compliance across Scotland, move to requesting a report on the % completion rates the service is achieving. Once this is in place then actual scores can be requested and national benchmarking of services can begin. If we try to gather it all, straight away, we'll get into a muddle, engender huge resistance and end up abandoning the whole project. |
| 3 - Preference for more full reporting. If you have the data why not report on it. More likely to be used. We can work towards this. |
| 4 - I feel the minimum level required should be reporting whether outcome monitoring is routinely in place for all services which deliver psychological therapies. I feel this will be crucial in ensuring that using outcome measures are actually implemented cores at assessment and discharge, together. Detailed reporting would be preferable including which disciplines are using the outcome measures if the information is collated it can be possibly used at later dates for other purposes. |
More detailed reporting, such as that required in IAPT, which includes actual scores at assessment and discharge, together with % completion rates – though not everyone agreed this should be using the CORE measure

1 - CORE
2 - Would favour - "More detailed reporting, such as that required in IAPT, which includes actual scores at assessment and discharge, together with % completion rates". Rational: this will enable benchmarking of services across Scotland and contribute to service improvement.
3 - It would be useful to eventually have databases with CORE results for specific diagnoses and therapy undertaken.
4 - I think in order to demonstrate this 'bigger picture' it would be ideal to report in detail i.e. scores and completion rates.
5 - The HADS or another anxiety and depression measure and the CORE.
6 - Detailed reporting at national level. With actual scores at assessment and discharge with % completion rate
7 - How well services are doing what outcomes are compared to waiting times & numbers of staff.
8 - I would like to see a case being made for the universal reporting of some kind of utility measure, which would potentially enable the calculation of important metrics such as cost-per-QALY. The two most commonly-used scales are probably the EQ-5D and the SF-6D (derived from the SF-36). However, the SF-36 is a licensed scale and the cost may be prohibitive for most services.

Until we are able to collect such data, we cannot answer such questions as:
1. Do our interventions provide value-for-money?
2. In which groups of patients do we see the most QALYs gained?
3. Should we prioritise those interventions which demonstrate the best cost-utility scores and deprioritise those interventions which are unable to show meaningful gains?
4. If we struggle to see useful gains from particular treatments, is it because: a) we are using the wrong measures of outcome; b) we are using the wrong treatment; c) we are poor at patient selection?

I would suggest that rather than trying to come up with an imperfect solution to a poorly-defined question, we should attempt to define the question more clearly and determine the answers empirically. The decisions that are made now will probably go unchallenged for many years and we need to know that the tools endorsed are up to the job. Over 3-6 months, we could probably get sufficient information on outcomes for a range of populations on the CORE, CGI, a disorder-specific scale, EQ-5D (along with diagnostic data and a baseline CGI-S score) which would tell us something useful about whether we should be adopting it. This would be a better use of NHS time and money.

9- I definitely think there should be nation-wide reporting of outcome scores at assessment and last session (whether planned or LOCF [last observation carried forward]) as per IAPT. Critically, though, I also think it is ESSENTIAL that services report the percentage of clients/patients who were seen for two or more sessions from whom the assessment/endpoint scores are derived. Without knowing that, as above, the pre- to post- scores are relatively meaningless: We've found, as have Clark and colleagues, that lower completion rates are associated with artificially inflated outcome scores, because of missing data from clients who drop out.
6. Any other comments

1 - This is a welcome proposal but the financial impact requires consideration at a national level

2 - General Comments:
The consultation says it refers to adults and only excludes CAMHS. Those working with adults with Learning Disabilities represented on these groups have made the following points:

- Whilst outcome measurement is welcomed, and the national CORE system is impressive in those situations and for those client groups where it will be applicable, appropriate measures which can be used by adults who are unable to self report using CORE need to be adopted also. This may include people with learning disabilities or perhaps people with dementia.
- CORE-LD is being evaluated nationally, led by Art Therapy and CBT Nurse Specialists initially, and can be utilized when psychotherapy is being delivered. This will take some time. The Glasgow Anxiety and Depression Scales for people with a learning disability are also commonly used as appropriate.
- However much of what is delivered psychologically in learning disability is delivered via carers so a carer rated outcome measure is also essential at times.
- Behavioural interventions (which will be part of the next version of the Matrix) also require objective and individualised measures of behavioural change.
- HONOS-LD is used by some multi-disciplinary services as a measure of a range of health and quality of life issues.
- Thus a small range of measures may be adopted depending on the mode and medium for intervention.

This makes reporting more complex. Perhaps Boards could be asked to simply report how they are measuring outcome and how they are utilising this information to evaluate performance for their services.

3 - There are problems with NHS Highlands IT system so there will need to be discussions re compatability. We have tried to use the paper CORE for years and clinicians cant be bothered. A PC version is preferable but IT can be a big issue.

4 - Despite the desire for some kind of unified and ubiquitous outcome measure, I would have to agree with Dickerson (1997) who wrote: “It is unlikely that any one instrument can meet the many demands for outcomes data across various settings. The potential utility of any instrument needs to be evaluated in terms of the specific questions that are being asked within a health system and in terms of how the outcomes information will be used.”

I am not convinced that key questions about what we are wanting to know have been addressed, and this means we are trying to select a tool without fully understanding what we want to ask of it. The choice of tool will ultimately be determined by the questions we want to ask. For example, the CORE cannot tell us if our interventions are improving quality of life or the cost-utility of our interventions whilst the EQ-5D could. Conversely, the EQ-5D can’t tell us if someone symptoms have improved as a result of an intervention. We need to work out what it is we really want to know.
References
