

The ethics of national routine outcomes monitoring policies: A case for taking action

Andy Fugard

Service users hope that psychological therapies will help them to feel better. Though challenging and controversial, there are signs that routine outcomes monitoring can be used to evaluate and improve therapy effectiveness. In this article, I briefly introduce three national policies in England which I think are ethically problematic and could be damaging to outcomes monitoring. I conclude with ideas for what might be done in response. The guiding principle is simple: outcomes monitoring should be focused on helping achieve its central aim, to learn about and improve care.

SERVICE USERS hope that psychological therapies will help them to feel better. Professionals and organisations such as the National Institute for Health and Care Excellence (NICE) have an ethical responsibility to evaluate therapies to ensure they are beneficial and not harmful. Randomised controlled trials are a crucial way to test the efficacy of interventions; however, they do not reveal how effective therapies are in routine practice. This has led to the development of practice-based evidence, typically using outcomes questionnaires asking clinicians and service users to rate symptoms and other information before, during and after therapy. Routine outcomes monitoring has been used by NHS services for many years; for example, through the Child Outcomes Research Consortium (Fleming, et al., 2014) and Improving Access to Psychological Therapies (IAPT; Clark, 2011; Wolpert et al., 2012).

Routine outcomes monitoring is challenging: a recent review enumerated 26 different sources of bias which could make a therapy appear in routine practice to be effective when it was not (Lilienfeld et al., 2014). It is also controversial. Mayo (2010), for example, argues that it uses 'oversimplified questions requiring tick-box answers', with the main aim of helping to implement policies for reducing human contact between therapists and service users and subsequently to reduce costs. One of the biggest impacts on clinical practice is that it is time consuming (Boswell et al., 2015). Some

have argued that this time is not well-spent since it creates a bureaucratic 'administratively created reality' (McLeod, 2001). And yet, there is emerging evidence that outcomes monitoring could be better than senseless bureaucracy.

Providing feedback to practitioners from questionnaires completed by service users seems to improve outcomes (Knaup et al., 2009), especially for people who are showing slower progress than expected compared to norms of change over time (Lambert & Shimokawa, 2011). Examining the content of questionnaires commonly used, it intuitively makes sense why feedback would help. For example, the GAD-7 (Spitzer et al., 2006), used across IAPT services, asks how often people have been bothered by 'worrying too much about different things' and 'feeling afraid as if something awful might happen'. A broad range of questionnaires are available for use, ideally chosen in partnership between therapist and service user depending on the goals of therapy. IAPT services also ask about the care received; for example, whether service users felt listened to by staff. Responses to these items can help clinicians understand to what extent services are helping people who engage in therapy.

It is early days for outcomes monitoring, and the generalisability of evidence to date concerning the benefits of feedback has been questioned (Davidson, Perry & Bell, 2014), but there are some promising signs that outcomes monitoring could improve the effectiveness of

therapy. Given the variety of ways outcomes evaluations are used, from branding interventions as 'evidence-based' to substantive uses of evidence to try to learn about and improve care, it is important that the ethical implications of outcomes policies are thoroughly debated. In this article, I wish to raise awareness of three national policies in England which I think are ethically problematic. I conclude with some ideas for what might be done in response.

Policy 1: Recommending an unreliable measure

In England, it is now recommended to use a Mental Health Clustering Tool (MHCT) to evaluate outcomes (Department of Health, 2013, p.12). This questionnaire, slightly extending the Health of the Nation Outcome Scales (Wing et al., 1998), is completed by clinicians and covers areas such as hallucinations, depression and relationship difficulties. The MHCT suffers from a very basic problem: the suggested scoring has poor internal consistency. Indeed, this problem was discovered by the people who proposed the approach (Speak & Hay, 2012, p.30): 'As a general guideline, alpha values of 0.70 or above are indicative of a reasonable level of consistency'. Their results are: 0.44, 0.58, 0.63, 0.57 – conspicuously smaller than 0.70. The authors also refer to previous studies explaining that this would always be the case, due to 'its original intended purpose of being a scale with independent items' (p.30). Closer inspection of the individual items reveals why the alphas are so low. For example, the lowest consistency (0.44) is for a two-item 'Severe Disturbance' scale which combines 'overactive aggressive behaviour' and 'hallucinations and delusions' – very different behaviours and experiences, and arguably unethical to put them together in one score. It is striking here that national recommendations appear to ignore the evidence. Once the governmental machinery starts rolling, there is no turning back.

Policy 2: A path to payment by outcomes

Given evidence that collecting regular feedback might improve the quality of care people receive, it seems a good idea that the IAPT programme includes regular progress moni-

toring using questionnaires which are completed by service users. In contrast to the MHCT, the questionnaires used have adequate reliability. There is, however, another potential difficulty over and above that of the quality of questionnaires used, and that is how external influences such as Payment by Results (PbR) initiatives can change for the worse how data are gathered and used. And PbR initiatives are beginning to be used in practice. The main PbR system under development in mental health is based on predicted need (Jacobs, 2014; Yeomans, 2014). The basic idea is that service users are 'clustered' into one of twenty groups of equivalent cost. This, the story goes, should facilitate better planning and more transparent commissioning than is possible with 'block' contracts with providers. It also, however, facilitates marketisation and privatisation. IAPT desires something more. The IAPT PbR webpage (www.iapt.nhs.uk/pbr/payments-by-results) notes that: 'An outcome based payment and pricing system is being developed for IAPT services. This is unique as other systems of PbR are activity or needs based.' Initial pilot results were 'encouraging', says the web page, and another pilot is currently running.

The idea with this proposal is that the more improvement shown by service users, as partly determined by outcomes scores, the more money service providers would receive. This is a worry as linking measures to targets has a tendency to cause the measures to stop measuring what it is hoped that they measure. For instance, targets on ambulance response times have led to statistically unlikely peaks at exactly the target (Bevan & Hood, 2006), suggesting that times have been changed. The effect has been around for such a long time that it has a name, Goodhart's law, describing the effect that 'When a measure becomes a target, it ceases to be a good measure' (Strathern, 1997, p.308). This would seem to risk destroying the main purpose of outcomes monitoring (i.e. evaluating outcomes).

Faced with funding cuts, how many staff at already overstretched services will be forced to 'game' performance-based payment systems to ensure their service survives? It's not hard to do so; for example, people who drop out of ther-

apy tend to do so because they didn't think it was helping. It can be easy to justify not troubling people who leave therapy to complete questionnaires. People who stay in therapy and complete questionnaires tend to be the ones who show better outcomes; hence inflating the apparent effectiveness of a service (Clark, 2011, p.321). Another potential problem is so-called 'power-steering', whereby, given weekly progress monitoring, therapist and service user may work to decrease a symptom score rather than actual mental health difficulties (Evans, 2012). This might increase if clinicians are under pressure from management to obtain improved scores. It is difficult to see how any data gathered that has been subject to these difficulties could tell clinicians or service providers anything helpful about their services or the well-being of those who use them.

Policy 3: Use of public league tables

There is another problem with how measures are used by adult IAPT. The Health and Social Care Information Centre (HSCIC) has for the past few years presented recovery rates – 'key performance indicators' – for all services in England in the form of league tables. This has been done without any information on uncertainty, which is highly problematic. There can be vast variability in recovery rates entirely by chance, even if all services are of equivalent effectiveness (Fugard et al., 2014). Funnel plots, well-known from meta-analysis, are a better way to present the information and are already used in other areas of healthcare (Spiegelhalter et al., 2012). These take into consideration expected variation in outcomes and focus attention on services with outcomes scores reliably higher or lower than the national average. However, worst of all is HSCIC's presentation of named service data without any context or interpretation. We argue (Fugard et al., 2014) that data always needs to be interpreted in the local service context and suggest that clinicians, and ideally service users, are involved to make sense of differences from national norms.

What can be done?

I think that practice-based evidence using statistically reliable, high quality questionnaires could be helpful in clinical practice, if

thoughtfully used in collaboration with service users. A common analogy is that questionnaires are the thermometer or blood pressure measurements of mental health care: providing a general overview of health, and helpful when combined with other information jointly explored in a therapeutic relationship. You are not reduced to a blood pressure reading when you go to the GP and it is obvious that more interaction is needed to understand whether and why something is awry, and to formulate appropriate care. Using unreliable measures – especially when the developers know they are unreliable – is unethical for a national mental health programme. Linking questionnaire scores to payment raises even more complex ethical issues: poorer decisions about interventions could result because those decisions would be made on the basis of low quality, unreliable data. League tables have been known to be problematic for years. So what can be done? Below I offer some tentative ideas for action.

As a first step it seems important to raise awareness of the issues; for instance, at meetings in services. Clinicians and researchers are aware of the challenges of outcomes measurement and solutions. Senior managers and commissioners can be less well-informed. Issues for discussion include: the importance of considering the positive predictive value of questionnaires (derived from sensitivity, specificity and prevalence data) and using reliable change criteria; being aware of Goodhart's law; giving presentations on recent papers showing the use of practice-based evidence and current directions for research (e.g. Gyani et al., 2013; Lundh et al., 2013; Nilsen et al., 2015).

The loudest political voices at the moment come from the extremes: those who wish to determine funding using outcomes indicators and those who reject outcomes monitoring entirely. A reflective evidence-informed alternative voice is needed. Possible pathways to ensure this voice is heard include:

1. Writing to your MP (e.g. via www.writetothem.com). This, though relatively easy, seems to have a low probability of changing anything. However, it can't hurt to increase

awareness of the problems and low probability is better than zero. Coordinated action, as supported by organisations like 38 Degrees (or perhaps the BPS?), increases the likelihood of success.

2. Keeping an eye out for, making others aware of, and responding to consultation documents. This has been successful in the past; for instance, encouraging the introduction of clinically useful outcomes measures in children and young people's IAPT.
3. Joining government advisory groups. NHS England and Monitor advertise positions on their websites and through e-mail – contact them to be added to the relevant lists. These are unpaid, time consuming, and frustrating; however, again, they provide an opportunity to ensure issues are at least discussed – a helpful first step.
4. Writing briefing documents which can be used to defend against unethical decisions. This has been successful in the past for helping a service prevent the introduction of Commissioning for Quality and Innovation (CQUIN) targets on outcome scores.
5. Coordinated protests; for instance, only using questionnaires which are useful for evaluation and clinical decisions, refusing to use measures with poor psychometric properties and explaining why. (I am unaware of any attempts to do this.)

These are only some ideas to get started or to react against – I am sure others have additional ideas. Whatever your view on the utility of outcome measurement, I hope I have made a case for political engagement. Service users are most affected by policy, often bearing the burden of questionnaire completion and benefitting or suffering from how analyses are used. We all – clinicians,

researchers, and others involved in outcomes evaluations – must take responsibility for ensuring policies are ethical, so that outcomes monitoring can be focussed on helping achieve its central aim: to learn about and improve care.

Author

Dr Andy Fugard, Educational Psychology Group, Research Department of Clinical, Educational and Health Psychology, University College London; a.fugard@ucl.ac.uk; Twitter: @inductivestep

Acknowledgments

Thanks very much to Martha Pollard, Justine McMahon and Stephen Weatherhead for helpful comments on earlier drafts of this article.

Declaration of interest

I am a social scientist and lecturer in research methods at University College London. My training is in computer science and psychology and my current research areas include practice-based evidence. I worked for Child Outcomes Research Consortium, helped develop the outcomes framework for CYP IAPT, and was involved in analyses for a CAMHS PbR project. I attempt to influence national policy concerning practice-based evidence as a member of the Monitor and NHS England 'Quality and Cost Benchmarking Group'. I have volunteered for a non-directional confidential telephone listening service and been a 'befriender' at a psychiatric in-patient unit. More personally, my father experienced depression for much of his life. I also have experience of mental health difficulties and engaged in brief therapy, which I found helpful. I once refused as a client to complete outcomes measures as I felt the results would be used for marketing purposes. This article represents my own inevitably biased views.

References

- Bevan, G. & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, 84, 517–538.
- Boswell, J.F., Kraus, D.R., Miller, S.D. & Lambert, M.J. (2015). Implementing routine outcome monitor-

ing in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, 25, 6–19. DOI: 10.1080/10503307.2013.817696

Clark, D.M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *Interna-*

- tional Review of Psychiatry*, 23, 318–227. DOI: 10.3109/09540261.2011.606803
- Davidson, K., Perry, A. & Bell, L. (2014). Would continuous feedback of patient's clinical outcomes to practitioners improve NHS psychological therapy services? Critical analysis and assessment of quality of existing studies. *Psychology and Psychotherapy* [online]. DOI: 10.1111/papt.12032
- Department of Health (2013). *Mental Health Payment by Results Guidance for 2013–14*. London: Author. Retrieved from www.gov.uk/government/uploads/system/uploads/attachment_data/file/232162/Mental_Health_PbR_Guidance_for_2013-14.pdf
- Evans, C. (2012). Cautionary notes on power steering for psychotherapy. *Canadian Psychology*, 53(2), 131–139. DOI: 10.1037/a0027951
- Fleming, I., Jones, M., Bradley, J. & Wolpert, M. (2014). Learning from a Learning Collaboration: The CORC Approach to Combining Research, Evaluation and Practice in Child Mental Health. *Administration and Policy in Mental Health*. DOI: 10.1007/s10488-014-0592-y
- Fugard, A.J.B., Stapley, E., Ford, T., Law, D., Wolpert, M. & York, A. (2014). Analysing and reporting UK CAMHS outcomes: An application of funnel plots. *Child and Adolescent Mental Health*. DOI: 10.1111/camh.12086
- Gyani, A., Shafran, R., Layard, R., & Clark, D.M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy*, 51, 597–606. DOI: 10.1016/j.brat.2013.06.004
- Jacobs, R. (2014). Payment by results for mental health services: Economic considerations of case-mix funding. *Advances in Psychiatric Treatment*, 20, 155–164. DOI: 10.1192/apt.bp.113.011312
- Knap, C., Koesters, M., Schoefer, D., Becker, T. & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: Meta-analysis. *British Journal of Psychiatry*, 195, 15–22. DOI: 10.1192/bjp.bp.108.053967
- Lambert, M.J. & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, 48, 72–79. DOI :10.1037/a0022238
- Lilienfeld, S.O., Ritschel, L.A., Lynn, S.J., Cautin, R.L., & Latzman, R.D. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science*, 9, 355–387. DOI: 10.1177/1745691614535216
- Lundh, A., Forsman, M., Serlachius, E., Lichtenstein, P. & Landén, M. (2013). Outcomes of child psychiatric treatment. *Acta Psychiatrica Scandinavica*, 128, 34–44. DOI: 10.1111/acps.12043
- Mayo, R. (2010). Dark times: Auditing the psychotherapy profession. In L. King & C. Moutsou (Eds.) *Rethinking Audit Cultures: A critical look at evidence-based practice in psychotherapy and beyond* (pp.61–73). Ross-on-Wye: PCCS Books.
- McLeod, J. (2001). An administratively created reality: Some problems with the use of self-report questionnaire measures of adjustment in counselling/psychotherapy outcome research. *Counselling and Psychotherapy Research*, 1, 215–226. DOI: 10.1080/14733140112331385100
- Nilsen, T.S., Handegård, B-H., Eisemann, M. & Kvernmo, S. (2015). Evaluating change in symptomatic and functional level of children and youth with emotional disorders: A naturalistic observation study. *European Child & Adolescent Psychiatry*. DOI: 10.1007/s00787-014-0671-3
- Speak, B. & Hay, P. (2012). *The Utility of the Mental Health Clustering Tool (MHCT) as a Generic Outcome Measure for the Payment by Results Quality and Outcomes Framework in Working Age and Older Adult Mental Health Services*. Wakefield: Care Pathways and Packages Project. Retrieved from [www.cppconsortium.nhs.uk/admin/files/The utility of the MHCT as a generic outcome measure for the PbR quality and outcomes framework for working age and older adult mental health V2.pdf](http://www.cppconsortium.nhs.uk/admin/files/The%20utility%20of%20the%20MHCT%20as%20a%20generic%20outcome%20measure%20for%20the%20PbR%20quality%20and%20outcomes%20framework%20for%20working%20age%20and%20older%20adult%20mental%20health%20V2.pdf)
- Spiegelhalter, D., Sherlaw-Johnson, C., Bardsley, M., Blunt, I., Wood, C., & Grigg, O. (2012). Statistical methods for healthcare regulation: Rating, screening and surveillance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175, 1–47. DOI: 10.1111/j.1467-985X.2011.01010.x
- Spitzer, R.L., Kroenke, K., Williams, J.B.W. & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166, 1092–1097. DOI: 10.1001/archinte.166.10.1092
- Strathern, M. (1997). 'Improving ratings': Audit in the British university system. *European Review*, 5, 305–321. DOI: 10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4
- Wing, J.K., Beevor, A.S., Curtis, R.H., Park, S.B., Hadden, S. & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS): Research and development. *British Journal of Psychiatry*, 172, 11–18. DOI: 10.1192/bjp.172.1.11
- Wolpert, M., Fugard, A.J.B., Deighton, J. & Görzig, A. (2012). Routine outcomes monitoring as part of children and young people's Improving Access to Psychological Therapies (CYP IAPT) – improving care or unhelpful burden? *Child and Adolescent Mental Health*, 17, 129–130. DOI: 10.1111/j.1475-3588.2012.00676.x
- Yeomans, D. (2014). Clustering in mental health payment by results: A critical summary for the clinician. *Advances in Psychiatric Treatment*, 20, 227–234. DOI: 10.1192/apt.bp.113.011320