Update on Data Remediation Exercise for HSCIC Reporting of the IAPT dataset, April 2014

Background

IAPT services collect systematic data on the people who access the services, the treatments they receive, and their outcomes, using the IAPT minimum dataset. Starting in 2010, IAPT data was reported nationally to allow services to benchmark themselves against each other and to allow the government to monitor progress towards it's access and recovery targets.

Initially the national reports were based on data that was aggregated locally by the services themselves and submitted to the NHS Information Centre by commissioners. Data was submitted through the "omnibus" system and the NHS Information Centre (subsequently renamed the Health and Social Care Information Centre or HSCIC) collated and published the service/commissioner level key performance indicators (KPIs). Starting in 2012, it was agreed that national IAPT reports would move from being based on local aggregate data to a new system in which patient level data is uploaded to HSCIC on a monthly basis and the KPIs and other statistics are computed centrally. In order to maintain patient confidentiality, all patent records have to be pseudonymised at the HSCIC's Exeter Portal before flowing to the analytical team.

Central processing of patient level data has several potential advantages. In particular, there is consistency in the application of computations for KPIs and a much more nuanced set of reports covering access, process and outcome can be produced when the burden of calculation shifts from local services and commissioners to HSCIC. The Quarterly reports, the most recent of which was published in April 2014[1], nicely illustrate the more nuanced reporting that can be achieved. However, the requirements for data management are more challenging with the central processing method. In particular, all data is pseudonymised and the record of an individual's care over the course of treatment has to be built up by merging data that is uploaded to HSCIC in different months.

It was always envisioned that it would take a while to get the central processing to align with the locally generated aggregate data. For this reason, the switch over plans included a period in which the two types of processing system were run in parallel. Reports based on both types of processing were published by the Health and Social Care Information Centre (HSCIC) in 2012/2013. Differences between the two data sources were highlighted in the release of Q4 2012/13 IAPT KPI figures in August 2013[i]. The latter showed that the number of referrals to services were 15% less when calculated centrally, while the number of individuals who were recorded as having completed treatment or as having recovered were 34% and 39% less, respectively.

Since August 2014 HSCIC have been working closely with stakeholders (NHS England, IT system providers, IAPT services & IAPT commissioners) to understand these differences and to ensure that central reporting on IAPT services is as accurate as possible. This paper explains the key reasons for differences that have so far been identified (further details are given in the Appendix) and highlights areas of data quality where clinicians can influence the accuracy of central reporting about the services they deliver locally.

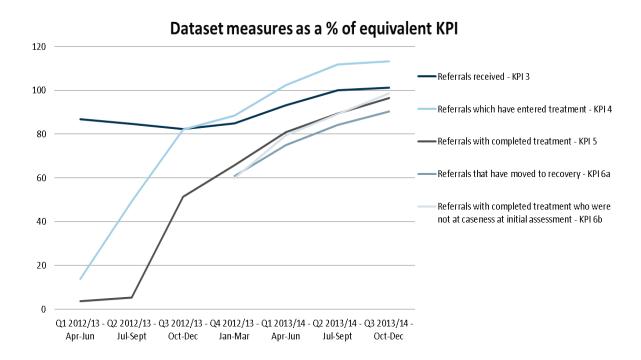[1] http://www.hscic.gov.uk/pubs/iapt1314q3exp

Key Reasons

1. Characteristics of a new collection of care pathways spanning multiple submissions

Most IAPT referrals last for several months. Only referrals that commenced in 2012-13, the first year for IAPT patient level data submissions to HSCIC, are available for analysis. Outcomes measurement requires information collected at both the start and the end of a referral – so referrals that ended in 2012/13 but started before the patient level dataset collection started could not be reported. As time has gone on, the percentage of closed cases that had a start date within the accumulating HSCIC database has gradually increased and is now approximately 98% for the latest quarterly statistics. This means that the total number of closed cases in the quarterly reports are now much closer to the estimates based on the local aggregate data published in August 2014. The figure below shows this point (see KPI5) and also illustrates the improvements that have been made in recording the numbers of people referred and the numbers who have entered treatment. (The Y axis is scaled as percentages of the local aggregate based figures reported in August 2014 and so do not necessarily reflect the current local aggregate data as some services are continuing to expand).

2. Differences between the ways in which recovery is measured locally and centrally

Although issue 1 above has had an impact on all measures (except KPI3) in the first year of reporting from the dataset, issues of analysis method and definition also caused differences between reporting of recovery locally and centrally. Discussions at HSCIC events highlighted the differences between how services have to be delivered locally, to meet users' needs and adapt to local circumstances. A survey of providers showed that there was variation between the constructions used to report recovery on different local systems and as compared with the construction used by HSCIC.

The calculation of recovery using Anxiety Disorder Specific Measures (ADSMs) is one such example:.

ADSMs have been used to measure change and included as part of recovery calculations since 2012. The IAPT data Handbook includes guidance on how ADSM scores should be used in calculating recovery. However, it is now clear that this guidance has been interpreted in different ways. NHS England has therefore provided clarification which is now being implemented in national reports.

In particular, NHS England has confirmed that recovery calculations should be performed on all completed cases with paired scores. For cases with an ICD-10 anxiety disorder diagnosis the preferred pair of measures is the PHQ and the relevant ADSM. However, if the relevant ADSM is not available, paired scores based on the PHQ and GAD should be used. For cases without an ICD-10 anxiety disorder diagnosis, the PHQ and GAD should be used for the paired scores.

Prior to this clarification, HSCIC were operating an exclusion rule when considering cases that had an ICD-10 diagnosis of an anxiety disorder other than GAD. If the case had paired scores on the PHQ and the correct ADSM, recovery was assessed. If the case had paired scores on the PHQ and GAD but not on the relevant ADSM, recovery was not assessed. This issue was rectified from Quarter 2 of the National reports.

Another key difference has been in the definition of a treatment appointment. Due to widespread data quality issues regarding the appointment type, the national dataset reports have used the presence of a therapy type to identify a treatment appointment, rather than using the appropriate appointment type codes. Additionally, assessment appointments are currently identified as the first appointment with an outcome score recorded (e.g. PHQ-9, GAD-7 etc.)

In version 1.5 of the IAPT dataset the appointment type becomes mandatory and this is expected to relieve the data quality and completeness issues that affect this field. Therefore, with the implementation of version 1.5 the definition of treatment and assessment appointments will be determined by the appointment type. This will be implemented in reports from November 2014 and should make it easier to reliably identify individuals who have had a course of treatment (i.e two or more sessions with treatment as part of the appointment type code). It is important to note that the appointment type data item will become mandatory in version 1.5 of the IAPT data set, so if this field is not completed at EVERY appointment your submission will fail

3. The small but critical data quality issues that affect pathway analysis

Although providers submit patient identifiable information in their monthly IAPT dataset submission, including NHS number, DoB and postcode, the data received by HSCIC for analysis has been pseudonymised. The accuracy of this pseudonymisation, over the multiple submissions required to provide information about a full referral, will determine whether details of each individual pathway are available for analysis. Any gaps or inaccuracies in these key data items, and in the provision of a consistent unique Service Request Identifier for each referral, can break the pathway and make the referral ineligible for calculation of completing treatment and recovery numbers. In recent workshops, HSCIC has highlighted the vital importance of providing full information on a patients NHS number, DoB and postcode in time for that person's first data upload. The importance of keeping the same Service Request Identifier throughout an episode of treatment within a single provider (including step-up) was also emphasized. If the service Request ID changes within the provider then the HSCIC will not be able to join the referral across submissions, leading to a broken pathway. This should also be considered when changing IT systems, as if the service ID is not retained across systems in the provider then the pathway will be broken.

IAPT adopts a stepped care model of delivery as such it is acknowledged that a person may be stepped up or down and move between service providers dependent on the treatment being delivered. At present it is not possible to capture 'step' in your monthly data submission. However it will be possible to do so from July 2014 when version 1.5 of the IAPT data set is implemented. Until then, stepped care across providers in a single pathway cannot be tracked.

It is still important, however, that services endeavour to keep a consistent Service Request Identifier for a patient that is passing through a single course of stepped care treatment (i.e low intensity followed by high intensity or visa versa) within a single provider as failure to maintain the same Service Request Identifier across the two steps of a stepped care episode will tend to inflate access figures (because a single individual is being counted twice) and reduce recovery rates (because in general a person who recovers will only drop below the clinical/non-clinical threshold in one of the two steps of care)..

4. Timely recording of information about the service

The IAPT dataset and associated data flows have been designed to meet the needs of commissioners for timely, monthly information about the services delivered. Each submission is required to include the activity that occurred in that month and failure to provide information at the appropriate time means it will not be included in the dataset and therefore cannot be reported on. For example, the table below shows when activity occurring in May should be submitted. There are two submissions for each month, providing an opportunity to update or refresh data up to 7 weeks after the month end.

Data item	Date of activity	Which submission	Primary Submission Deadline	Provisional monthly data publication	Refresh submission deadline	Final monthly date publication
End of IAPT Referral	08/05/2014	May 2014	27/06/2014	29/07/2014	25/07/2014	22/08/2014
Appointment	29/05/2014	May 2014	27/06/2014	29/07/2014	25/07/2014	22/08/2014

Although local practice in respect of recording patient activity may vary across and within different services, timely recording and submission of the data, in line with the submission requirements[ii], is essential to support accurate central reporting.

This is particularly important for recording the closing of cases

In order for a case to be classified as having completed treatment or to be assessed for recovery or reliable improvement it is necessary to identify that the case has been closed. Local IT systems can manage this identification quite easily as they use a live data management system that can be continuously updated with information about a particular case and the retention of detailed personal information means that staff can confidently ascribe the right information about case closure to the right person.

Identifying closure is more tricky at the central processing level as the database is not live and instead is updated *en masse* every month. It is important that successive updates don't "double count" an event (such as case closure), so the closure of a given case will only be credited to the database in one update period. To ensure that this happens, the Exeter portal will strip out of a case record any case closure dates that appear in a current month's data upload but have a date that comes within a previous (or future) month's upload period. When this happens, HSCIC will not be able to identity that

the case has closed, as it will not be included in the dataset, and will therefore not be assess it for recovery or reliable improvement. **To avoid this problem, services need to make sure that they close a case on their IT systems as soon as they decide to discharge the case**. The discharge date may correspond to the last treatment session if discharge was planned but with other cases it will be sometime after the last treatment session.

It is important that therapists are aware of this issue and ensure administrative processes are in place to ensure case closures are not missed.

Although in the first year of the dataset key access and outcome measures were reported once a quarter, some of these measures are now being reported monthly.

5. Summary of Progress and Actions for Services

As Figure 1 shows, considerable progress has been made in reconciling IAPT indices calculated on local systems with those calculated from central processing of the pseudonymised patient level data that HSCIC receives from the Exeter Portal. The Quarterly reports published by HSCIC also nicely illustrate the more nuanced, and clinically informative data that can be made available to services through central processing. In particular, pre-post treatment outcome data completeness rates for people who have finished a course of treatment are now reported. The average rates (for example 96.8% in the Quarterly data published in January 2014) are almost certainly world leading and reflect the immense amount of effort that IAPT services put into data collection. Detailed breakdowns of numbers referred and numbers completing treatment by ethnicity, gender, age, and provisional diagnosis are also provided. These will help services as they move forward in their attempts to provide equitable access and treatment for all.

Despite the progress shown in Figure 1, HSCIC are aware that some services are still reporting that there are significant discrepancies between IAPT indices based on local versus central processing. HSCIC is continuing to work with stakeholders to tie down the reasons for these discrepancies and to educate services about the more strict requirements for timely and complete data entry that underpin central processing. HSCIC is keen to hear from services that are aware of substantial continuing discrepancies.

In the meantime, some clear messages have emerged for services. These are:

1) Close cases in a timely fashion and ensure that you enter a case closure date on your IT system in the initial upload (or refresh) period that includes that date. Entering a date that does not fall into the period that is being submitted will result in the date being excluded from the dataset and the case being ignored for calculating completing treatment and recovery.

2) Ensure that when a case is entered on your IT system you include complete information in the following fields:

- NHS number
- Date of birth
- Postcode
- Commissioner Code
- GP Practice Code

This will ensure that the patient can be identified across submissions, an essential part of rebuilding the pathway in the central dataset.

3) Ensure that a consistent unique Service Request Identifier is used for an individual when he/she moves from one step to another within a single episode of stepped care treatment within a single provider to ensure that the pathway can be joined.

Appendix

HSCIC routinely monitors and publishes information about the quality of the data it receives from health care providers, but over the last 6 months HSCIC has undertaken additional activities to build confidence in the IAPT dataset.

Data remediation activities undertaken Oct 2013- Jan 2014

- Review testing undertaken on the Open Exeter portal for IAPT submissions
- Data quality workshops in London and Leeds attended by 110 individuals from 73 provider organisations[iii]
- A survey of provider organisations about their local reporting arrangements
- An expert user groups of experienced data analysts from providers/ commissioner/ system suppliers
- Support to individual providers to investigate accuracy of referral level analysis
- Discussions with systems suppliers and policy leads

Draft April 2014. Jointly produced by HSCIC & NHSEngland

---

[i] See: http://www.hscic.gov.uk/pubs[add reference]
[ii] See data specification and guidance here: http://www.hscic.gov.uk/iapt