

# Understanding and replicating our published reports

**A guide for use with the Improving Access to  
Psychological Therapies dataset**

---

**Community and Mental Health Team  
July 2015**

**Version 1.4**

# Contents

---

<b>Version History</b>	<b>3</b>
<b>Purpose of this document</b>	<b>4</b>
How to use this document	4
<b>The IAPT dataset</b>	<b>5</b>
Submissions to the dataset	5
Structure of the dataset	6
<b>Published reports</b>	<b>8</b>
Monthly reports	8
Quarterly reports	9
Annual reports	9
<b>Data quality</b>	<b>10</b>
On-submission validations	10
Monthly data quality notices	12
Monthly data quality (VODIM) reports	12
Data quality reports	12
<b>Replicating our figures</b>	<b>14</b>
A note about KPIs	14
Using your post-deadline extracts	14
Recreating IC_ derivations	15
Recreating constructions	17
Calculating access figures	17
Calculating waiting times	18
Calculating recovery figures	19
<b>Common issues and misconceptions</b>	<b>22</b>

---

## Version History

Version	Date issued	Summary of changes
1.1	February 2015	Initial version
1.2	30/07/2015	Updated to reflect new developments in reporting
1.3	06/04/2016	Updated links to IAPT pages on NHS England website and HSCIC Contact Centre
1.4	23/08/2016	Rebranding to NHS Digital

## Purpose of this document

This document has been created in response to the interest of Providers, Commissioners and other data users in understanding the IAPT dataset and the reports that are published for it. We understand that it is important for service providers to understand how the data regarding the activity in their service is used and are keen to enable data users to investigate local data issues.

The Improving Access to Psychological Therapies dataset is a relational dataset and there are a number of complexities involved in its collection, processing and analysis which can create barriers to users being able to interpret the data. This document aims to cover these issues and the impact they have and provide guidance on how the information can be understood and replicated locally. We are always keen to receive feedback on the utility of the information we provide and would be happy to hear from you about any improvements or additional information you would like to have included in this document. Please send any queries or comments to the NHS Digital IAPT team at [enquiries@nhsdigital.nhs.uk](mailto:enquiries@nhsdigital.nhs.uk).

## How to use this document

This document is designed to explain how the IAPT dataset is collected, processed and analysed and should be a useful resource on its own. There are a number of documents available however, which can be used in conjunction with this guide, particularly when trying to replicate figures produced in published reports.

These documents are as follows:

- Dataset Specification – This document details each field of the IAPT dataset, including lists of valid codes to be used. There is also other technical documentation which data users may find useful. These documents can all be found [here](#).
- Technical Output Specification – This document gives more detailed breakdowns of the fields included in the IAPT dataset as well as containing information about the warnings, validations and diagnostics produced by the Bureau Service Portal upon submission of a file and also a full specification of the post-deadline extracts, including any derivations which are calculated by the IC. The document can be found [here](#).
- Constructions and derivations worksheets – These are found in the IAPT Metadata document held on the IAPT splash page [here](#) which details the methodology behind each measure.
- Methodological Change Papers – Whenever there is a significant change to the dataset or how measures are calculated, NHS Digital issues a Methodological Change Paper to explain the changes. It may be useful to examine recent papers, particularly when replicating or interpreting older figures (such as before the latest version of the dataset). Methodological change papers are stored on the NHS Digital website [here](#).

## The IAPT dataset

The IAPT dataset is a record-level dataset first mandated in April 2012, which is submitted to NHS Digital by IAPT providers on a monthly basis. The dataset contains information on each referral to IAPT services, the associated appointments and assessment scores recorded at those appointments and also information about the patient such as age, gender and ethnicity. The data is administrative data, meaning it is taken from the clinical systems used by IAPT providers to run their services. The creation of the dataset removed the requirement for services to calculate their data independently, as this would now be done centrally, thereby reducing burden to the NHS.

## Submissions to the dataset

Providers submit data to NHS Digital each month through a secure service known as the Bureau Service Portal.

### Submission window

There are two opportunities to submit each month's data, the Primary and the Refresh submissions. For each submission there is a period known as the submission window, during which the provider can make as many attempts at submission as required. Details about when the submission window is open for each month of data can be found [here](#).

### Last good file

Once the window closes the last good file is processed and added to the IAPT dataset. The last good file is the most recent submission made by the provider in the period which was not rejected during on-submission validation. Please bear in mind that the last good file does not take account of the number of validation warnings generated on submission, just that the file was not rejected. It is the provider's responsibility to ensure that the last good file is of the best quality possible and to address on-submission warnings where possible. There will be more information about the on-submission validation process later in this document.

### Primary and refresh submissions

As mentioned above, providers have two opportunities to submit data for a particular month. They should submit a Primary submission during the first period and can then submit a refresh submission if they wish. Once the refresh submission window is completed the data is processed and a final cut of the data is produced. The final data will always be the Refresh submission unless only a Primary submission is made in which case the Primary submission is used to create the final data. National activity reports are always calculated from final data.

Please note that if a Refresh file is submitted this will always overwrite the entirety of the Primary submission. This overwrite is at a file, not record, level. This means that if an individual referral is submitted in the Primary submission but was missed from the Refresh submission, it will not be included in final data and therefore in national figures.

## Structure of the dataset

The dataset itself is held as a relational dataset rather than a flat file. There are currently 7 tables in the data set, which can be joined together to give a full picture of activity in the IAPT programme. The tables are as follows:

**Referral** – This table gives details of each individual referral to IAPT (this is not the same as people as one person could have multiple referrals). It includes fields such as start and end dates, referral source, commissioner code, and information about whether the referral has been stepped across providers. This is also the central table that all other tables are linked to. To be included in the other tables a record must also have a corresponding entry in the Referral table.

**Person** – This table holds information about the service user behind each referral. This includes facts such as age, gender and ethnicity. The person table can be linked to the referral table using the IAPT\_RECORD\_NUMBER field. IAPT\_RECORD\_NUMBER is produced during processing and identifies every unique pairing of submission and person. Joining on this field allows the referral to be tied with the person record that was submitted alongside it for a given submission.

**Appointment** – The Appointment table holds details of every appointment in the IAPT service including facts such as whether it was attended, the type of appointment, therapy types provided and the assessment scores that were recorded at that appointment. The appointment table joins to the referral table using the IC\_PATHWAY\_ID field. The pathway ID uniquely identifies every combination of person and referral to allow the details of a referral to be tracked across submissions and its construction is covered in more detail later in this document. This ID is calculated by NHS Digital and is not available in the post-deadline extract available to providers and commissioners.

**Disability** – The Disability table joins to the Person table on IAPT\_RECORD\_NUMBER, and contains details of any disabilities linked to a person. Disability is a separate table rather than being part of the Person table as one person may have multiple disabilities. The Disability table is not always submitted by providers – it does not need to be submitted if there are no recorded disabilities in the submission.

**Waiting Time Pauses** – the Waiting Time Pauses table was newly added in version 1.5 of the dataset and contains details of patient instigated pauses. These could then be used in the calculation of waiting time measures to stop and start the waiting time clock. The intention behind this is to exclude patient initiated delays counting against providers in the calculation of waiting times. The Waiting Time Pauses table joins to the Referral table on IC\_PATHWAY\_ID.

**Assessment Questionnaire** – new in version 1.5, the Assessment Questionnaire table holds information on the patient's experience of their assessment process as reported by the patient themselves. This includes information such as when the questionnaire was completed. The Assessment Questionnaire table joins to the referral table on IC\_PATHWAY\_ID.

**Treatment Questionnaire** – The Treatment Questionnaire table, also new in version 1.5, holds information on patient experience questionnaires completed by the patient at the end of treatment. This table contains information on when the questionnaire was completed and how the patient rated the service provided. The Treatment Questionnaire table joins to the referral table on IAPT\_Record\_Number.

## **Constructing the time series**

IAPT data is submitted to the IAPT dataset as a snapshot of live systems each month. These individual snapshots then need to be pieced together in order to build a full picture of care over the periods. This is done through the development of an IAPT Person ID and IC\_Pathway\_ID which are covered in more detail later in this document. This feature of the dataset leads to two important consequences:

Firstly, as the dataset is made up of snapshots of data from live systems it is unlikely to match live systems exactly as these are dynamic and constantly updated. The data held in the dataset will only match live systems as they were when each extract was taken.

Secondly, the data quality of items used to join submissions across fields (such as those fields used to construct the Person ID or IC\_PATHWAY\_ID) must be high. Poor data quality in those fields can prevent linkage across periods leading to the inability for a pathway to be correctly traced. This will have an impact on a number of key measures including access and recovery. The issues around data quality, how it can be monitored and its impact are discussed fully later in this document.

Where the same person details and Service ID are submitted by more than one provider, all referrals and activity will be assigned the same Pathway ID.

## Published reports

The data which is collected in the IAPT dataset is analysed centrally and from it a number of activity measures are calculated to give a picture of how the service is performing. Currently reports are available at National, Commissioner, Provider and Commissioner/Provider combination levels in the following reports and can be accessed at <http://www.digital.nhs.uk/iaptmonthly>.

## Monthly reports

Each month NHS Digital publishes data quality reports which give information on the validity of a number of fields submitted in the month. Full details of the data quality rules applied can be found [here](#). The reports give the number of records that fall under each of the following categories, known as VODIM:

**Valid** – The number of records where a value was provided that is one of a list of appropriate values. In most cases the IAPT data specification provides details of which codes are valid for a given field.

**Other** – The number of records for which a valid code is provided but where the code corresponds with “other” or similar. Full details of which codes are classed as “other” can be found in the data quality rules.

**Default** – The number of records where a default code such as 99 is used rather than a valid code being submitted. Again, further information on what is classed as a default code can be found in the data quality rules document.

**Invalid** – The number of records where a value which does not match the valid code list and is not classed as default or other is provided.

**Missing** – The number of records where a value would be expected but is not provided and is left null.

**Tip:** A common reason for a record to have an invalid code is the lack of or inclusion of a leading zero. Many codes defined in data dictionary require a leading zero and if these are missed then the record will be classed as invalid.

In addition other fields do not require leading zeroes and their inclusion will lead to the values being classed as invalid. It is important when inputting data to ensure that leading zeroes are included or excluded as appropriate.

Details of whether leading zeroes are required for a particular field can be found in the **IAPT data specification**.

In addition to the data quality reports, each month a csv file containing a wide range of activity measures is released at Clinical Commissioning Group (CCG), Provider and CCG-Provider combination level.

## Quarterly reports

### Future developments

From April 2015 the activity measures previously contained in quarterly reports began to be produced monthly instead and the quarterly reports as they stood were discontinued. The current monthly reports contain a quarterly supplement which provides a number of key measures such as access and recovery broken down by Age, Gender, Problem Descriptor, Disability and Ethnicity. Information on psychotropic medication status will also be provided. For more information on these changes please see the associated methodological change paper [here](#).

## Annual reports

Each year a detailed annual report is produced to give an overall picture of activity in the service over the preceding year. This tends to provide more in depth commentary regarding the data and allows more nuanced reporting to be carried out. For example, the 2013/14 annual report included analysis of the average number of sessions in a course of treatment, broken down by problem descriptor and therapy type provided. This report is accompanied by a series of Excel tables and a data quality report which summarises the key data quality issues seen throughout the year.

## Data quality

Data quality is a key element that needs to be understood when considering IAPT data and reports. Poor quality data can have a significant impact on the measures that are regularly reported on from the data set and can be a factor in disrupting the process of joining data across submissions. Providers and commissioners are given multiple opportunities to identify and address data quality issues and we also publish data quality reports so that data users can evaluate the data when interpreting its meaning.

### On-submission validations

The first opportunity providers have to assess their data quality is upon submission of their extract. Once the file has been submitted a number of warnings and errors will be generated.

If there is an error in a mandatory field (for example the appointment type is blank or invalid), or there is a problem with the structure of the dataset (for example if a mandatory table such as the referral table is missing) then the submission will be rejected with an error message explaining the reason for this. If the submission is rejected it **will not flow** to NHS Digital and there will be no data submitted for this organisation unless the issue that caused the rejection is remedied and the file is re-uploaded. A full list of validations that will cause the file to be rejected can be found in the Validations tab of the Technical Output Specification<sup>1</sup>

In addition to these file level rejections, a number of further validations are carried out on submissions that generate warnings if failed. Full details of the validations carried out at this point can be found in the Warnings tab of the Technical Output Specification<sup>1</sup> but they include checks such as the patient's postcode being invalid or referrals having an end code but no end date. These validation errors will not stop the data being processed and sent to NHS Digital but they may well have an impact on published figures if not rectified.

Although it is acknowledged that it may not be possible to address all the issues, or that in some cases the problems identified are expected or a known issue, it is still extremely important for providers to examine their warnings on submission and to attempt to minimise these issues. Although all of these validation warnings are important we have highlighted some of those which will have the largest impact on published figures below (More information on how these warnings are generated can be found in the Technical Output Specification.)

Warning Reference Number	Warning text	Why is it important
W01	Patient had more than one record in the person table (duplicate NHS Number)	NHS Number is used to create a unique person ID for an individual, which is then used to trace a person across multiple submissions as part of the pathway ID. This issue will cause problems in the tracing of the person through submissions and disrupt the pathway which will have implications for all activity measures for that individual.
W05/W06	Patient had an invalid postcode/ Patient had a null postcode	When reporting by CCG, if there is no valid Commissioner code or GP code provided the postcode of residence is used to derive the

<sup>1</sup> [http://www.digital.nhs.uk/media/13514/IAPT-v15-Technical-Output-Specification/xls/IAPT\\_Data\\_Set\\_v1.5\\_Technical\\_Output\\_Specification.xls](http://www.digital.nhs.uk/media/13514/IAPT-v15-Technical-Output-Specification/xls/IAPT_Data_Set_v1.5_Technical_Output_Specification.xls)

		CCG. If this is not available then the referral cannot be attributed to a CCG.
W07/W08	Patient with invalid GMPC/Patient with null GMPC	When reporting by CCG, if no valid organisation code of commissioner is provided then CCG is derived from the GP. If this is not valid then the postcode of the patient is used to derive CCG instead.
W09/W10	Referral with invalid organisation code of commissioner as defined by ODS/ Referral with null organisation code of commissioner as defined by ODS	When reporting by CCG the organisation code of commissioner field is used as the primary source for CCG. If this is not present then CCG must be derived from the GP code or the postcode of residence instead, which may have lower accuracy. In addition CCGs can only receive data in their extracts which has their CCG code as the organisation code of commissioner. If a provider has entered a null or invalid code of commissioner, the records will not be available to the CCG in the Commissioner extract, and the activity may not be recorded for the CCG in published reports.
W11	Referral with no diagnosis	The problem descriptor is used to determine the correct anxiety measure to be used in measuring recovery. In addition, reports also include information on activity such as access and recovery broken down by problem descriptor.
W13	Referral with an end code but no end date	If no end date is provided the referral will not be assessed as having ended. This means that the referral will not be counted towards the number finishing a course of treatment, and will not be assessed for recovery. A common reason for this error is data being submitted in the wrong period. The end date must occur in the period being submitted for it to flow. This means that the end date must be recorded on the system in a timely manner, so that it can be processed in the correct submission.
W20	Attended appointment without a valid PHQ-9 score	The PHQ-9 should be recorded at every appointment. If a referral does not have at least two PHQ-9 scores recorded then it does not have paired scores and cannot be assessed for recovery.
W21	Attended appointment without a valid GAD7 score	The GAD-7 should be recorded at each appointment. Although recovery may be calculated using an anxiety measure, if the anxiety measure cannot be used for any reason GAD7 is the default measure used in recovery calculations. If a referral does not have at least two GAD7 scores recorded, and there is no appropriate anxiety measure which can be used then recovery cannot be calculated for the referral.

Although these fields are key in the calculation of activity such as access and recovery, all the warnings generated on submission should be examined by the provider and rectified where possible as others may also contribute to published reports either now or in the future. For example, reporting by therapy type requires valid therapy types to be submitted for each treatment appointment (Warnings W17, W39, W40, W41 and W42) and any future reporting on stepped care will require fields regarding this to be well completed (Warnings W29, W12 and W33).

## Monthly data quality notices

Once the data has been submitted and the window has been closed the data is processed by NHS Digital. At this point it is not possible for providers to alter a Refresh submission although they may still address any issues with a Primary submission in the Refresh period for that month's data.

Once the data has been processed, NHS Digital issues providers with detailed data quality notices. These allow providers to have early sight of the quality of their refresh data and also provide detailed information of any issues with the primary submission which can be addressed during the refresh period.

The list of validations covered by the data quality notices can be found in the data quality reports that accompany our quarterly releases.

These notices also contain information for providers on how the measures can impact published figures and provide an opportunity for them to give feedback on the reasons behind the issues. This feedback can then be incorporated in the quarterly data quality notices if required or can be used to help refine the issues included in the notices themselves.

## Monthly data quality (VODIM) reports

Each month NHS Digital also produces a monthly Data Quality report which is published on the website as described in the Monthly reports section earlier in this document. This report gives providers the facility to look into the validity of key fields in their data submission but it is also publicly available and so can be used by other data users to help assess the validity of published data. This is the first time in the process that data quality information for a given month is made public, all other feedback is available only to the provider who submitted the data and so these reports are an important resource for those wanting to understand the impact data quality may have on the data for a given provider or on national figures.

## Data quality reports

The monthly reports described above give detailed information on activity over the course of the month at Provider, Commissioner and Provider/Commissioner combination levels and are also accompanied by a Data Quality Report. This document covers all the data quality issues that may affect the data in these reports. These issues range from common problems such as end dates of referrals being missed, issues specific to the period such as providers who missed a submission in that period and elements in the processing of data which may cause a difference in the time series or affect the data included, such as methodological changes in the calculation of measures.

This document covers 9 European Statistical Systems quality dimensions and principles<sup>2</sup>, in order to comply with the UK Statistics Authority Code of Practice for Official Statistics<sup>3</sup>, which are:

---

<sup>2</sup> The original quality dimensions are: relevance, accuracy and reliability, timeliness and punctuality, accessibility and clarity, and coherence and comparability; these are set out in Eurostat Statistical Law. However more recent quality guidance from Eurostat includes some additional quality principles on: output quality trade-offs, user needs and perceptions, performance cost and respondent burden, and confidentiality, transparency and security.

- Relevance
- Accuracy and reliability
- Timeliness and Punctuality
- Accessibility and clarity
- Coherence and Comparability
- Trade-offs between output quality components
- Assessment of User Needs and Perceptions
- Performance, Cost and respondent burden
- Confidentiality, Transparency and Security

---

<sup>3</sup> UKSA Code of Practice for Statistics: <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>

## Replicating our figures

One of the most common enquiries we receive is from providers and commissioners looking to understand our published figures and to reconcile them with local reports. The aim of the next section is to provide some guidance for those data users, including information on how to go about using post-deadline extracts to match our figures and a walk-through guide to replicating two of the most commonly asked about measures, namely access and recovery. The common issues and misconceptions section may also be useful as this identifies some of the key pitfalls users experience when using and replicating our data.

### A note about KPIs

It is important to discuss the difference between Key Performance Indicators and central reports. The KPI reports ended at the end of 2012/13 and the central reports are now the authoritative source of information. Although many of our published measures are based on the old KPI figures, a number of the constructions and methodologies have been updated over time and so central figures may not exactly replicate the old figures. For example KPI4 which regarded the number of referrals entering treatment counted the patient as having entered treatment if a second appointment had been received within 28 days of the first. The definition of entering treatment has since changed and is represented in central reports as being the number of referrals with a first treatment appointment in the period, regardless of when a second treatment was received. This has been confirmed in recent NHS England waiting times guidance<sup>4</sup>.

There are a number of reasons for the change to central reporting including a decrease in burden, but one issue that has become clear is that KPIs were not always calculated in the same way between providers as guidance was applied in differing ways. The central reports apply the same methodology across all providers and commissioners and the methodology used has been assured by the IAPT Central team at NHS England to ensure that the logic applied is correct.

Since the KPI reports were discontinued many services continued to extract KPI reports from their systems for local monitoring. It is important for services to note however that while these reports may be useful for local monitoring they are no longer relevant with regards to central processing. Particularly services should be aware that the reports produced by NHS Digital will be used to assess performance against the mandate commitments laid out in Everyone Counts: planning for patients 2013/14<sup>5</sup>.

### Using your post-deadline extracts

In order to successfully recreate NHS Digital reports, providers and commissioners must use their post-deadline extracts, which are available for download from the Bureau Service Portal once the window has closed. If users are attempting to recreate data from live clinical systems they may not get the same results, as these systems are constantly updated. In addition live systems may also hold data which has not been processed by the portal, such as appointments or end dates that were not submitted in the appropriate period and therefore will not be included in the data which is passed to NHS Digital.

---

<sup>4</sup> with <https://www.england.nhs.uk/mentalhealth/adults/iapt/>

<sup>5</sup> <http://www.england.nhs.uk/everyonecounts/>

Each extract contains the data for that month. In order to replicate figures it is important to join extracts together to create a time series. This will also require data users to join referrals over periods and replicate the IC\_Pathway\_ID, in order to be able to then replicate the figures in our reports. Details of how to calculate IC\_Pathway\_ID and many other useful derivations are contained in our IC\_Derivations tab at the end of each excel spreadsheet for our reports.

The post-deadline extracts do contain some derivations currently however we would not recommend using the pathway derivations (data items such as first and last scores) produced by Exeter and included in provider and commissioner extracts. Whilst these can be used for approximate calculations, in most cases they cannot be used to replicate NHS Digital analysis accurately as they are very sensitive to data quality issues. For example in version 1.0 key issues were leading zeroes missing from APPOINTMENT\_PURPOSE and missing or invalid ORG\_CODE\_OF\_COMMISSIONER entries. We think these affect submissions to varying degrees.

To help providers follow our analysis methodology, we have produced a report called PAVE (Provider Analysis Validation Extract). This is delivered to providers monthly after submission of the final (Refresh) data for a period. The report is available via Data Depot only for provider contacts who have a user logon to the OpenExeter Bureau Service Portal.

PAVE is a referral-level report and shows a 1 or a 0 in a series of columns which signify whether a specific referral has been counted in a specific category. The totals of the main columns (after small number suppression rules have been taken into account) will match the figures in the Monthly Data File csv. This allows a provider to identify where a specific referral is counted locally, but not nationally or vice-versa and we can then work with providers to show the reason for this.

Providers will need the portal post-deadline extract to link the PAVE report back to local data.

## Recreating IC\_ derivations

In order to help those using their post-deadline extracts to replicate our figures, and to provide other data users with in-depth descriptions of how our figures are calculated we provide logical constructions and derivations for each of our measures in the 'IAPT Metadata document' held on our website at [www.digital.nhs.uk/iaptmonthly](http://www.digital.nhs.uk/iaptmonthly). The constructions list the fields and tables involved in each measure, give details of what data is included or excluded and provide information on the specific conditions applied to create each measure.

These constructions rely on a number of fields in the dataset and also a number of derivations. Any field referred to in these constructions which are prefixed with IC are derived fields which need to be calculated before they can be used in creating the measure. They are not included in the post-deadline extract. Details of how these derivations are calculated is also included in the excel spreadsheets on the IAPT 1.5 Derivations tab. There are a number of derivations which are useful such as first and last GAD7 and PHQ9 scores and the date of first therapeutic session. We would encourage all data users to look at these constructions and to attempt to replicate them but we are also happy to offer further

**Note:** Remember that if you have data from both a primary and refresh submission, the refresh submission is always used rather than the primary submission. If only a primary submission is available then this will be used instead

guidance on their interpretation if unclear. We have also chosen to cover some key derivations required for analysis of data below to provide extra guidance on how these can be calculated.

## Key Derivations

### - IC\_PATHWAY\_ID

This Derivation is one of the most fundamental used in the dataset as it allows a referral to be traced across submissions. All activity measures in our quarterly and monthly reports are counts of referrals which involves counting unique IC\_Pathway\_IDs. The derivation for IC\_Pathway\_ID is:

Incremental value assigned to new SERVICEID and IAPT\_PERSON\_ID combinations. If such a combination has been received in a previous reporting period the record is assigned the same A.IC\_PATHWAY\_ID value.

This means that when a unique combination of SERVICEID and person first occurs it is given a unique reference number known as its PathwayID. When that unique combination occurs again in a subsequent submission it is assigned the same PathwayID, allowing the records in different periods to be connected.

### - IC\_USE\_QTR\_REFERRAL\_FLAG

This flag allows the identification of which referral record should be used to produce quarterly figures. Within a quarter the details for a referral may be submitted up to three times (once per month) and if the details change between these periods (for example if data was missing when the referral was first opened but was then added at a later date) this could affect counts of activity. The purpose of this flag is to identify which details are the most recent and will prioritise these over any past information. For example in the case of a referral that opened in April but which did not have a problem descriptor at the point the referral was received, if a problem descriptor was then added following an assessment in May this flag would ensure that the most recent record was used and so any activity would be attributed to the correct problem descriptor in reporting. The derivation for IC\_USE\_QTR\_REFERRAL\_FLAG is:

IC\_USE\_QTR\_REFERRAL\_FLAG = 'Y' where RANKING = 1 (highest ranked referral instance based on the following subset).

Partition referrals by QUARTER\_ID and A.IC\_PATHWAY\_ID and order by MONTH\_ID descending.

The top ranked referral is assigned a flag.

This means that the referral record for the most recent month within that quarter is flagged. Note that the data is partitioned by QUARTER\_ID. This means that the record flagged is the most recent for the months in that quarter. For example, if further information was received in July this would not change how data for Quarter 1 was reported instead the updated information would be incorporated into Quarter 2 data.

## IAPT Person ID

Please note that IAPT\_PERSON\_ID is a field that is not provided in post-deadline extracts nor is it listed as an IC derivation. This is because the generation of the Person ID is a process of anonymisation that occurs on submission. The IAPT dataset used for analysis

does not include Person Identifiable Data (PID) such as NHS number and instead uses IAPT\_PERSON\_ID as a pseudonym. The IAPT Person ID is attributed to each unique record using a process of matching. When a person record is received during processing the following combinations are used to try to match the record to one which has been submitted in a previous period:

1. NHS number, (date of birth - 2 out of three element match)
2. Local patient ID and full DOB
3. Postcode, date of birth

If the record cannot be matched using any of these combinations a new IAPT person ID is created for the record. In addition, if a record has no NHS Number and the postcode is invalid, default or missing then this system is bypassed and instead the record is assigned an "E Code", which is an incremental 20 digit number starting at E0000000000000000001. These bypassed records will never be matched from month-to-month so long as the NHS number is missing and the postcode is invalid default or missing.

## Recreating constructions

Once users have recreated the relevant derivations they are then in a position to attempt to calculate the activity measures in our reports using the constructions in the excel sheets provided. In the examples that follow this document explains how these constructions can be applied to calculate three key measures: access, recovery and waiting times. This should help data users to understand how to apply constructions to their data extracts to replicate NHS Digital reported numbers however we are also happy to provide advice on carrying out data reconciliation via [enquiries@nhsdigital.nhs.uk](mailto:enquiries@nhsdigital.nhs.uk).

## Calculating access figures

One of the most requested measures is access, (i.e. number of people entering treatment,) particularly due to the government target of reaching a 15% access rate by the end of Q4 2014/15. The number of people who had a first treatment and therefore entered treatment in the period can be found in the monthly csv file. The construction for this measure is as follows:

**COUNT** of distinct IC\_PATHWAY\_ID

**WHERE** IC\_USE\_QTR\_REFERRAL\_FLAG = 'Y'

**AND** IC\_DATE\_FIRST\_THERAPEUTIC\_SESSION is between the period start and end dates or is null

**Note:** The number entering treatment is just the numerator to calculate the access rate. The denominator is the prevalence for the area, an estimate based on the Adult Psychiatric Morbidity Survey, 2000. This information is not held at Provider or CCG level by NHS Digital and so is not included in our reports, however it is available from NHS England at CCG level.

This construction relies on the derived field of IC\_DATE\_FIRST\_THERAPEUTIC\_SESSION, which is the earliest appointment in a referral which had an appointment type of 02 Treatment, 03 Assessment and Treatment or 05 Review and Treatment and that was attended. It is key that the appointment must have occurred within the referral. Appointments occurring before the referral received date or after the referral end date are not

included. As mentioned before, the construction of this measure does not look at the second treatment appointment and when this occurred, it is concerned only with the date of the first treatment appointment and whether it occurred in the period.

Data users should also be aware that the definition of a treatment appointment as used in this measure has recently changed. In version 1.0 of the dataset the Appointment Type field suffered from poor data quality and so could not be used to identify treatment appointments. Instead an appointment was classed as a treatment appointment if there was a Therapy Type submitted, indicating that therapy had been provided and so making that appointment a treatment appointment. When version 1.5 of the dataset was implemented in July 2014, Appointment Type became a mandatory field meaning that data could not be submitted unless each appointment had a valid Appointment Type provided. As a result the methodology was updated so that treatment appointments would now be classified as above.

## Calculating waiting times

Waiting time figures are related to the access rate as they are defined as the time in days waited between referral and first treatment. NHS England has recently issued updated guidance<sup>6</sup> on waiting times calculations as the definition of these has been updated since the KPI collections ended. Currently in NHS Digital reports, waiting times are calculated for all those referrals that entered treatment in the period as follows:

(IC\_Date\_First\_Therapeutic\_Session – REFRECDATE)

In the monthly reports, these waiting times are split into groups of 28 days or less, 29 to 56 days, 57 to 90 days, more than 90 days, less than 6 weeks and less than 18 weeks.

**Waiting Time Pauses:** In version 1.5 of the dataset an additional table was added which allows providers to identify a waiting time pause, which is a period of time when the patient was unavailable for an appointment. The intention behind these pauses is that they would be removed from the overall waiting time, however this has not yet been incorporated into central waiting times calculations.

---

<sup>6</sup> with <https://www.england.nhs.uk/mentalhealth/adults/iapt/>

## Calculating recovery figures

Recovery is a key figure in IAPT publications and generates a lot of interest politically and in the media. It is also one of the more complicated measures in IAPT reports. In simple terms, the number of referrals which have recovered is the number of referrals which completed a course of treatment and which moved from caseness to not being at caseness by the end of the referral. It is measured through using the assessment scores that are collected at each IAPT session and evaluating the change in these scores over the case of the referral.

### Finished course of treatment

In order to be considered for recovery a referral must first have finished a course of treatment the construction for which is as follows:

**COUNT** of distinct IC\_PATHWAY\_ID

**WHERE** IC\_USE\_QTR\_REFERRAL\_FLAG = 'Y'

**AND** ENDDATE between the start and end dates of the reporting period

**AND** have attended at least two appointments (i.e. the ATTENDANCE field must be '5' or '6'), that were held between the REFRECDATE and the ENDDATE

**AND** where appointment type is recorded as 02, 03 or 05 for at least two of those appointments

This means that the referral must have ended within the period having had at least two attended treatment appointments in the course of the referral.

The construction for recovery is then as follows:

**COUNT** of distinct referrals that ended in the quarter having finished a course of treatment

**WHERE** the referral has non-NULL paired PHQ and anxiety measure or GAD7 scores appropriate to the problem descriptor (note that for calculating recovery, GAD7 is currently used for problem descriptors of 'F41.', rather than the PANIC\_DISORDER\_SEVERITY\_SCALE score)

**AND** the first score is above caseness on either the PHQ9 or anxiety measure /GAD7 (or both)

**AND** both last scores are below caseness.

### First and Last Scores

This construction makes heavy use of the first and last score derivation from the IC\_Derivations tab and it is useful to calculate these first. First and last scores should be calculated for the PHQ9, GAD7 and all other Anxiety Specific Disorder Measures. Please note that the last score cannot be taken from the same appointment as the first score. Additionally a common misconception is that both scores being evaluated (e.g. the PHQ9 and GAD7) must be taken from the same appointments but this is not the case. When looking for the first and last scores for each measure you should evaluate each individually. For example, a referral may have the following set of appointments in the course of its referral:

Appointment 1 – PHQ9 and OCI recorded

Appointment 2 – PHQ9 and GAD7 recorded

Appointment 3 – PHQ9, OCI and GAD7 recorded

Appointment 4 – PHQ9 and GAD7 recorded

In this case the first PHQ9 score is taken from appointment 1 and the last PHQ9 score is Appointment 4. The first GAD7 score is appointment 2 however and the last GAD7 is appointment 4 whilst the first OCI is appointment 1 and the last OCI score is appointment 3.

### **Anxiety measure selection**

Once all first and last scores are calculated and before any further calculations can occur it is important to determine whether GAD7 or an anxiety measure should be used in the recovery analysis. NHS Guidance is that a problem descriptor should be provided for each referral and that the measure used in recovery should be the appropriate anxiety measure for that measure. However, in cases where there is no problem descriptor the anxiety measure is not present, (with at least two scores recorded,) or does not show a value that is above the caseness threshold on the first score, GAD7 should be used instead. Going back to the example above therefore, if the referral had no problem descriptor then GAD7 would be used in the recovery calculation. If there was a problem descriptor of OCD however the OCI would be used instead, (as long as the score recorded in appointment 1 was above the caseness threshold for the OCI.) If another problem descriptor was recorded then GAD7 would be used as no other anxiety measures have been recorded.

PHQ9 is used in all recovery calculations alongside the GAD7 or relevant anxiety measure.

### **Caseness**

Once the correct anxiety measure has been determined, referrals then need to be assessed for caseness at the start of the referral. A referral can only recover if it was at caseness (i.e. classed as a clinical case) at the beginning of treatment. Each measure has a caseness threshold and to be at caseness at least one of the two first scores being evaluated for the referral must be above the caseness threshold.

The following thresholds are used for caseness:

- PHQ  $\geq 10$
- GAD7  $\geq 8$
- Agoraphobia Mobility Inventory  $\geq 2.3$
- Health Anxiety Inventory (Short Week)  $\geq 18$
- Obsessive Compulsive Inventory  $\geq 40$
- Impact of Events Scale  $\geq 33$
- Social Phobia Inventory  $\geq 19$

Referrals which are not at caseness on either score are excluded from the calculation and are removed from the denominator (referrals with a finished course of treatment) when calculating recovery rates.

## Paired Scores

In order to assess recovery all referrals must have two scores for the measures being considered so that the change in these scores can be evaluated. Once those referrals that are not at caseness have been removed those who do not have a first and last PHQ9 and a first and last GAD7 or other anxiety measure must also be removed. They are still included in the denominator but cannot contribute to the numerator.

## Recovery

Once all the above steps are completed the data user can assess the final scores. Whilst only one score is needed to be above caseness on the first score to be included in the recovery calculation, **both** final scores must be below the caseness threshold in order for the referral to have recovered. The caseness thresholds to be used are the same as above but it is important to note that recovery requires a score below the threshold. A score that is at the threshold does not meet this requirement and would prevent the referral from demonstrating recovery.

## Common issues and misconceptions

This section of the document is designed to address a number of commonly asked questions and also to address issues that frequently cause problems for those who are trying to replicate NHS Digital figures.

### Can I replicate analysis from my live system?

This is one of the most fundamental issues that affects reconciliation of figures. The only way to replicate NHS Digital data is to use the same source data and so data users must use post-deadline extracts. While constructions can be used on live systems to calculate measures for local monitoring, they would not be expected to match NHS Digital figures.

### Do my first and last scores need to coincide with the first and last appointments?

No, the first and last scores do not need to be taken from the first and last appointments. A first score is just the first time the score was recorded and the last score is the latest time the score was recorded. Providers should bear in mind however that NHS Guidance<sup>7</sup> indicates that scores should be recorded at each appointment.

In addition to this, please note that the first score of one measure does not need to be taken from the same appointment as the first score of the second measure. If a PHQ9 was recorded at appointment 1 but the anxiety measure /GAD7 was not recorded until appointment 7, appointment 1 is the first score for the PHQ9 but appointment 7 is the first score for the anxiety measure /GAD7.

### What is the definition of a treatment appointment?

A treatment appointment in version 1.5 of the dataset (since July 2014) is any appointment with an Appointment Type of 02, 03 or 05. In most cases these appointments need to be within the referral, (between the REFRECDATE and ENDDATE, if not ended then end of reporting period,) and need to be attended, (with an attendance code of 5 or 6.)

In version 1.0 of the dataset Appointment Type was not used due to data quality issues and so a treatment appointment was any appointment where one of the Therapy Type fields were not null and appointment type was not used except to exclude follow up appointments, (appointment type of 06.)

### Why are end dates missing from my submission?

A common issue that can affect numbers of people who finish a course of treatment, (and therefore affect recovery and other outcomes,) is end codes not flowing to NHS Digital. The most common reason for this is end dates not being submitted in the submission window for the month that they occurred.

For example, if a referral ends in March that referral must flow to NHS Digital as part of the March submission. If the March submission is missed and the record instead flows with a March end date in another period such as April, the system will remove the end date during processing and the referral will never be ended on NHS Digital systems.

It is important therefore to ensure timely reporting of data to ensure end dates are recorded on the system and are included in the last good submission for the month.

## **Why are some appointments missing from my submission?**

Just like the end dates for a referral, appointments must flow in the submission they occurred in, for example May appointments must flow in May's submission. If they flow in another month the system will remove them and they will never be received by NHS Digital. This can have a very large impact on all measures that deal with appointments, particularly access, finished course of treatments and recovery.

In some cases providers send effectively empty appointment tables due to all appointments submitted being outside the period. It is believed that this is an issue that occurs when the data is extracted from clinical systems and the wrong period is selected for export. Please note that if this occurs during the refresh submission the empty table will overwrite any legitimate appointments from the primary submission and they will not be used in analysis. There is no way to refresh data after the final submission so it is important to check diagnostic information available on submission to ensure appointments have been submitted.

## **What causes a broken pathway?**

In order to track a referral across multiple submissions the Service ID and IAPT Person ID must remain the same. The Service ID is locally provided and identifies the referral number across periods. The same ID must continue to be submitted for the referral in each submission. The Person ID is a field generated when processing the data on submission and identifies a unique person across periods by matching on data such as NHS number, date of birth and postcode.

More information about the construction of the Person ID can be found in the "Replicating our figures" section of this document. One of the most common reasons for a broken pathway is the provision of inconsistent person identifiers which leads to a new Person ID being created and prevents the Pathway ID from matching previous periods.

## **What are Bypass Patients and how do they affect my data?**

Bypass patients are patients for whom no valid NHS Number and no valid date of birth are provided. If this occurs no attempt is made to match the person to an existing Person ID and they are assigned a bypass number as their IAPT Person ID. As a new IAPT Person ID is generated for these records with each submission so long as the NHS number and date of birth remain missing the referral will never be matched across submissions.

No derivations are calculated for bypass patients and so they are excluded from most activity measures although they will be counted when the referral is received for the first time and when the referral ends. Access and recovery will not be calculated for these referrals.

## How should I record referrals that have been stepped up or down to a different provider?

If a referral has been stepped to another provider it should be ended in the current provider and a new referral (with a new Service ID) should be started in the new provider. The referral should be ended with an end code of '40 - Stepped up from low intensity Improving Access to Psychological Therapies Service' or '41 - Stepped down from high intensity Improving Access to Psychological Therapies Service', and the new organisation should be recorded in the 'Organisation Code (IAPT Stepped To Provider)' field.

The new provider should then start the new referral with a referral received date on or later than the previous end date and record a source of referral of either N1 (stepped up from low intensity IAPT service – where the previous end code was 40) or N2 (stepped down from high intensity IAPT service – where the previous end code was 41).

This should only be used when stepping to another IAPT provider. If the referral is being stepped down to primary care then the discharge code of '42 – Completed scheduled treatment' should be used. If a patient is stepped up to secondary care then the code '44 - Referred to non IAPT service' should be used.

## How should I record referrals that have been stepped up or down within the same provider?

If a referral is stepped up or down within the same provider, (i.e. it will be included under the same Provider code in your submission,) then the referral should **not** be ended. Instead the step of care can be tracked at each appointment through using the 'Stepped Care Intensity Delivered' field. This allows providers to code whether the appointment was high or low intensity and whether it is the first step of care or a later step.

Please note when using this field, although the codes available are 01, 02, 03 and 04, these DO NOT represent steps of care 1, 2, 3 and 4. Instead codes 01 and 03 identify low intensity (either the first step or a later step in the pathway), whilst 02 and 04 identify high intensity (again, either the first step or a later step in the pathway).

## How are records attributed to commissioners?

The primary tool used to determine the commissioner for a referral is the Organisation Code of Commissioner as recorded by the Provider. Commissioners will only receive extracts containing those records that have their organisation code recorded in this field. In some cases invalid commissioner codes are provided, (such as old PCT codes,) or this field is not completed. In order to attempt to match these referrals to a CCG in published figures we will attempt to derive the CCG from the GP Practice code. If the GP Practice code is also invalid or unavailable we will instead derive a CCG from the postcode of residence. If none of these methods produces a valid CCG for a record no CCG will be assigned.

Please bear in mind that this processing may mean that CCG level data derived from post-deadline extracts may not be the same as published figures as additional referrals that had an invalid or missing Organisation Code of Commissioner may be attributed to the CCG. This would lead to volumes in published figures being higher than figures derived from post-deadline commissioner extracts.

## **We've realised we have a data quality issue in our data. What can we do? Can we refresh our past submissions?**

We strongly encourage providers and commissioners to make us aware of data quality issues as soon as these are identified. If the issue pertains to a Primary submission and the Refresh window as not yet closed you may be able to make a further submission to rectify the problem. If you have identified the issue after the Refresh submission window for that month has closed you will not be able to make a Refresh submission. We do not have the capability to carry out further refreshes nor to perform an annual refresh of data and so the Refresh window for a given month is the last opportunity to ensure the data is correct.

If you become aware of an issue once the Refresh window has ended we would still encourage you to [contact the IAPT team](#) as we may be able to include a notice regarding the problem in the data quality reports which accompany our publication. Again however, this is dependent on receiving information in a timely manner in order that we can incorporate notes relating to the issue with the relevant publication. If information is provided regarding an already published period there may be limits to what we can do to make data users aware of the issue.

The need for timely feedback and the lack of an ability to perform an annual refresh emphasises the importance of the data quality information that is available to providers at submission and in data quality notices. It is of paramount importance that providers use and act on these notices to ensure data quality issues are addressed before the refresh window closes for a submission period.