

Announcement of methodological change

Impact of improving our Data Processing
Services on Mental Health Services
Monthly Statistics

Version 1: Published November 2019



Information and technology
for better health and care

Contents

Summary	3
What has changed?	3
What is the effect of this change?	3
Background	4
Improving our Data Processing Services (DPS)	4
Mental Health Services Dataset (MHSDS)	4
Mental Health Services Monthly Statistics	4
Description of changes	5
Strategic Data Collection Service in the cloud (SDCS Cloud)	5
Faster data processing	5
Master Person Service (MPS)	6
Impact	7
SDCS Cloud and data processing	7
MPS	7
Further information	10
Appendices	11
Appendix A: More information about the MPS	11
Appendix B: The previous MHSDS Person Index	17

Summary

This section briefly outlines key points about the changes that have been made to the collection of the Mental Health Services Dataset and the effect they have had on the Mental Health Services Monthly Statistics publication series.

What has changed?

We are improving how we collect and process the Mental Health Services Dataset, beginning with the collection of data from April 2019. This will give us the power to process larger volumes of data, faster than ever before, whilst providing the tools to manage incoming data to ensure it is accurate, useful and secure. It is also an internet-facing service that does not require a specialist health and social care data connection, making it easier for more providers to submit data.

We are changing how we identify people within national datasets which use this new technology, including the Mental Health Services Dataset. Using one method for doing this across datasets helps us increase the amount of usable, better quality, linkable data available to support research and planning.

What is the effect of this change?

We had operational issues putting the new technology in place

There were some initial operational issues in implementing this new technology which may have affected the quality of statistics in the **Mental Health Services Monthly Statistics** publication series. Several providers have been contacted to understand the background to specific issues, and findings are highlighted in the data quality note which accompanies the **Final April 2019** publication. Any further information we receive about issues providers have experienced submitting MHSDS via SDCS Cloud will be made available in future editions of this publication series. All editions of this series can be found at:

<https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-services-monthly-statistics>

We are now more likely to match records for people

In general, records submitted to the Mental Health Services Dataset are more likely to result in a match using the new method than using the previous one. This increased matching means that:

- Statistics which are a count of people will tend to be lower using the new method.
- Statistics which use historically submitted statuses will tend to be higher using the new method.

The effect on England-level statistics is small. We estimate that the new method reduces the number of people thought to be in contact with NHS funded secondary mental health, learning disabilities and autism services nationally by less than 0.1% on any given day. The effect is greater when people are counted across multiple months, but still small.

Background

This section provides information on the work we are doing to transform how we collect and process health and social care data, what we collect in the Mental Health Services Dataset, and what we make available in the Mental Health Services Monthly Statistics publication series.

Improving our Data Processing Services (DPS)

We are implementing modern technologies and processes to enable us to perform our statutory role as the safe-haven for health and care information. This requires us to collect and process data needed to run the health service.

These secure technologies and processes will enable us to collect, process and access data in a smarter, more efficient way. This will lead to faster access to better linked data giving commissioners and researchers a clearer picture of health and care.

Most importantly, our data processing services will improve patient care by empowering the health and care system to use information more effectively for research into the prevention and treatment of diseases and planning services essential to the sustainability of the NHS.

Mental Health Services Dataset (MHSDS)

These improvements have changed the way we collect and process mental health services data. The MHSDS is a patient level, secondary uses dataset delivering robust, comprehensive and comparable person-based information for children, young people and adults who are in contact with mental health services. As a secondary uses data set it intends to re-use clinical and operational data for purposes other than direct patient care.

The MHSDS covers not only services provided in hospitals, but also in outpatient clinics and in the community. MHSDS brings together key information from Adult and Children's mental health, learning disabilities or autism spectrum disorder, Children and Young People's Improving Access to Psychological Services (CYP-IAPT) and early intervention care pathway that has been captured on clinical systems as part of patient care.

Mental Health Services Monthly Statistics

The MHSDS is the source of the **Mental Health Services Monthly Statistics** publication series, which provides the most timely picture available of people using NHS funded secondary mental health, learning disabilities and autism services in England. This information is of use to people needing access to information quickly for operational decision making and other purposes.

All editions of this publication series and related annual publication series can be found on the NHS Digital website at:

<https://digital.nhs.uk/data-and-information/data-collections-and-datasets/data-sets/mental-health-services-data-set/statistics-about-mentalhealth-learning-disabilities-and-autism-services>

The effect of the changes included in this announcement on annual statistics will be discussed alongside the **Mental Health Bulletin, Annual Report 2019-20** and the **Mental Health Act Statistics, Annual Figures 2019-20**. Some early understanding of the potential impact on annual statistics can be found in the discussion of MHS69: Children and young people receiving second contact with services, April 2018 to March 2019 in this document.

The MHSDS is the source of some statistics published in other NHS Digital publication series, such as the **Learning Disability Services Monthly Statistics** series. The discussion in this document can be used to gain an understanding of the likely impact on similar MHSDS-sourced statistics in these publications.

Description of changes

This section describes changes to how we collect and process MHSDS data which could change how you interpret statistics in the Mental Health Services Monthly Statistics publication series.

Strategic Data Collection Service in the cloud (SDCS Cloud)

MHSDS submissions were previously made by data providers to NHS Digital via the Bureau Service Portal. From the implementation of MHSDS v4.0 onwards in April 2019 submissions have been made to NHS Digital via the Strategic Data Collection Service in the cloud (SDCS Cloud).

SDCS Cloud will provide:

- improved user experience and faster data quality feedback
- a secure solution using the cloud technology of the future which will integrate with our improved data processing services

SDCS Cloud is a completely new tool, and different from other existing NHS Digital collection tools, as it is being designed from the ground up to optimise use in the cloud. It uses a two-factor authentication as the secure method of confirming user identity. It is also an internet-facing service that does not require an N3 or HSCN connection, making it easier for more providers to submit data.

Faster data processing

We are using modern data processing services to streamline and automate existing ways of collecting, processing and accessing data. This will give us the power to process larger volumes of data, faster than ever before, whilst providing the tools to manage incoming data to ensure it is accurate, useful and secure.

Master Person Service (MPS)

MPS helps us increase the amount of usable, better quality data available to support research and planning. It does this by verifying the demographic information contained in a person's health and care records and matching it to their unique NHS number to confirm their identity.

Patients can visit multiple places where they register to receive care or treatment. They move, get married (or divorced) and go on holiday. At any given time, we store the health and care records of individuals as recorded in various systems around the country. This can create challenges, but MPS aims to match the right person with the right record with a 99% accuracy rate.

MPS uses a four-stage algorithm to provide a single best result for each record against Patient Demographics Service (PDS). DPS core processing assesses the data quality of submitted data before passing to the MPS. MPS then attempts to match all the records to a single NHS Number held in the Personal Demographic Service (PDS). MPS also checks demographic details supplied in the submitted data file, such as age, gender and postcode, for their 'closeness' to the data held in PDS and produces an associated match confidence score. Any results matched under the confidence threshold will not result in a match. The value will be more precisely determined as part of a future planned development which will introduce probabilistic scoring. Currently the lowest value score for Mental Health is 77.6% (which is an exact match on postcode and gender and partial match on Date of Birth year).

As we develop MPS we will expand coverage to more datasets and continue to make improvements. More detailed information about MPS can be found in **appendix A**. Information about the person index applied to the MHSDS prior to April 2019 can be found in **appendix B**.

We will issue a new version of this notice if MPS developments result in changes to how you can interpret statistics in the **Mental Health Services Monthly Statistics** publication series.

Impact

This section summarises the impact of these changes on how you can interpret statistics within the Mental Health Services Monthly Statistics publication series.

SDCS Cloud and data processing

The **Mental Health Services Monthly Statistics: Final April 2019** publication was the first edition of this series that uses data collected through SDCS Cloud. There were some operational issues in implementing this new system.

Extensive time series analysis has been undertaken at national and provider level to understand trends following these changes compared to previous patterns. From the results analysed, the proportion of providers and Clinical Commissioning Groups (CCGs) for each measure with large changes between March and April 2019 is similar to that between March and April 2018. These are both higher than the proportion of providers and CCGs exhibiting with large changes between months within the same financial year (including between April and May 2019). There is likely to be a number of factors leading to greater changes between months in different financial years.

Several providers have been contacted to understand the background to specific issues and findings are highlighted in the data quality note which accompanies the **Final April 2019** publication. Additional Data Quality information for the month can be found in the Provider comments csv available at:

<https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-services-monthly-statistics/final-april-provisional-may-2019>

Any further information we receive about issues providers have experienced submitting MHSDS via SDCS Cloud will be made available in future editions of this publication series. All editions of this series can be found at:

<https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-services-monthly-statistics>

MPS

The implementation of MPS results in different records being identified as belonging to a specific person than would have been identified using the person index applied to the MHSDS prior to April 2019. The main scenarios where the results of MPS differ from the previous person index can be found in table 1. Detailed analysis of the effect of using MPS on selected England and provider-level statistics from the **Mental Health Services Monthly Statistics** publication series can be found in the **reference tables** which accompany this notice. These statistics have selected for this assessment due to the availability of data items used in their construction within the new data processing system. They cover a range of different types of statistics produced in this publication series and can be used to gain an understanding of the impact of this change on the publication series as a whole. If you need to understand the impact on any specific statistics not included in this assessment then

please contact us by emailing enquiries@nhsdigital.nhs.uk quoting 'MHSDS MPS Impact Assessment'.

Table 1: Scenarios where the result of MPS is different to the previous person index

Scenario	MPS	Previous Person Index
New record from same provider, incoming record missing Date of Birth (DoB)	Not matched	Matched
New record from same provider, incoming record missing DoB & Gender	Not matched	Matched
New record from same provider, incoming record missing DoB & Postcode	Not matched	Matched
New record from same provider, incoming record missing DoB, Postcode & Gender	Not matched	Matched
New record from same provider, incoming record missing NHS Number	Not matched	Matched
New record from different provider, incoming record missing DoB	Not matched	Matched
New record from different provider, incoming record missing DoB & Gender	Not matched	Matched
New record from different provider, incoming record missing DoB & Postcode	Not matched	Matched
New record from different provider, incoming record missing DoB, Postcode & Gender	Not matched	Matched
New record from different provider, incoming record missing NHS Number	Matched	Not matched

The effect of these scenarios on published statistics will depend on the number of records submitted to the MHSDS that have these characteristics. In general, records submitted to the MHSDS are more likely to result in a match using MPS than using the previous person index.

There were 1,546,944 MHSDS person records within submissions for March 2019. The previous person index identified 1,495,473 unique people across these records. In comparison to a simple record count this reduces the number of people in the dataset for this month by 51,471 (3.3%). When MPS is applied to the same month the number of unique people across these records falls to 1,494,033. Compared to a simple record count this reduces the number of people in the dataset for this month by 52,911 (3.4%), and compared to the previous person index it reduces the number of people by 1,440 (0.1%).

This increased matching has two main effects on published statistics:

- Statistics which are a count of people will tend to be lower using MPS. This is caused by the higher matching rate reducing the overall number of distinct people within a single submission period. An example of a measure affected by this is MHS01: People in contact with services at the end of the reporting period.
- Statistics which use historically submitted statuses will tend to be higher using MPS. This is caused by the higher matching rate increasing the number of statuses that can be found for a particular person in previous submissions. An example of a measure affected by this is MHS13: People in contact with services at the end of the reporting period with accommodation status recorded.

Some measures are affected by both of these effects. An example of this is the measure MHS20: People in contact with services at the end of the reporting period with a diagnosis recorded. In these cases the two effects may balance and result in a very small change overall.

In all cases the effect on England-level statistics has been found to be small. We estimate that the new method reduces the number of people thought to be in contact with NHS funded secondary mental health, learning disabilities and autism services nationally by less than 0.1% on any given day. From this we estimate that if we had used this method for the previous reporting period that the number of people in contact with these services on 31 March 2019 would have been 1,358,786 instead of the original published statistic of 1,359,992.

The effect is greater when people are counted across multiple months, but still small. If we had used MPS to derive MHS69: Children and young people receiving second contact with services, April 2018 to March 2019 from the MHSDS the published value would have been 1,056 (0.35%) lower at 296,776 children and young people¹.

Provider-level impact

For the majority of providers the effect of the change on published statistics is small. The one exception to this is Wiltshire Council. Using the previous person index we published that 420 people were in contact with mental health, learning disabilities and autism services provided by Wiltshire Council on 31 March 2019. If we would have used MPS to derive this the number of people would have been 150. Records submitted by Wiltshire Council for this month had low levels of NHS Number recording. In the absence of NHS Number, MPS was able to match a larger number of these records than the previous person index. In the comparison of September 2018 statistics, the number of people in contact with Wiltshire Council is largely unaffected by using MPS, as the level of NHS Number recording in this month was much higher for this provider.

The only other providers with statistics which would have been more than 20% different in March 2019 if MPS would have been used to derived them were Derbyshire County Council, YMCA Downlink Group, and Young People Cornwall. For these providers the only measure which changed by this amount was MHS69: Children and young people receiving second contact with services, April 2018 to March 2019. This measure counts people across a twelve month period, which would indicate that provider-level differences may be larger for annual statistics. The effect of these on annual statistics will be explored further in the **Mental Health Bulletin, Annual Report 2019-20** and the **Mental Health Act Statistics, Annual Figures 2019-20**.

¹ The MHSDS derived value for this statistic has been superseded on the basis of a data validation exercise for 2018-19. Our best estimate for the number of individual children and young people aged under 18 accessing treatment by NHS funded community services in England, in 2018-19 is 377,866.

Further information

This section signposts to places where you can find more information about the changes discussed in this document.

Improving our Data Processing Services

Further details about DPS can be found here:

<https://digital.nhs.uk/data-and-information/data-insights-and-statistics/improving-our-data-processing-services>

SDCS Cloud

More information about SDCS Cloud, including support for submitters, can be found at:

<https://digital.nhs.uk/services/strategic-data-collection-service-in-the-cloud-sdcs-cloud>

MPS

Please see **appendix A** from more information about MPS. More information may be provided in future at:

<https://digital.nhs.uk/services/master-person-service>

MHSDS v4.0

Information on the latest version of the MHSDS can be found at:

<https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/mental-health-services-data-set/mental-health-services-data-set-specifications-and-guidance>

Changes to the data collected, validated and derived in MHSDS v4.0 can be found in the **Summary of Changes** tab of the **MHSDS v4.0 Technical Output Specification**, available at the above link.

Mental Health Services Monthly Statistics

All editions of this publication series include information on the quality of data submitted to the MHSDS, and metadata describing how MHSDS data is used to produce these statistics. These can be found at:

<https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-services-monthly-statistics>

Other mental health statistics

Various sources of mental health statistics can be found at:

<https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/mental-health-data-hub>

Appendices

These appendices provide more detailed information about the changes discussed in this document.

Appendix A: More information about the MPS

This appendix includes details of the function of MPS as at the publication date of this document. This document will be revised following any major changes to how MPS functions in future.

Scope

DPS core processing assesses the data quality (DQ) of submitted data before passing to the MPS. MPS then attempts to match all the records to a single NHS Number held in the Personal Demographic Service (PDS). MPS also checks demographic details supplied in the submitted data file, such as Age, Gender and Postcode, for their 'closeness' to the data held in PDS and produces an associated match confidence score.

DPS Core processing

For records that successfully pass initial DQ checks in DPS, the next step is to identify what demographic data is available on the record.

Where supplied, NHS Number is then cross-checked with the associated demographic data items with the Personal Demographic Service (PDS) in DPS Core. PDS is the electronic national database of person demographic details for the NHS in England.

- Where NHS Number is verified, the record is not passed to MPS.
- Where NHS Number cannot be verified or has not been provided, the record is sent to MPS for further processing.
- Less than 5% of new submissions and less than 10% of historic Mental Health and Maternity Services data records require onward processing within MPS.
- PDS is updated in DPS daily.

Outcomes of cross-check matching:

- **Successful Match**

Where a match is successful, and the NHS Number is verified, MPS Return Code – 00 SUCCESS – is assigned to the record. NHS Number is then used as the Person ID to enable linkage to different records within a data set and across different data sets. NB: In most cases, Person ID is presented in a pseudonymised/tokenised form.

- **Unsuccessful Match**

Where matching is unsuccessful, the record is passed to MPS for further processing using more complex algorithmic tracing logic.

Master Person Service (MPS) complex algorithmic processing

On receiving an unmatched record following DPS cross-check processing, MPS tries to match the record with PDS using NHS Number and other demographic data submitted on the record.

If a record contains NHS Number, MPS will validate this against PDS and, along with other supplied demographic details, attempt to find a unique match.

If the record does not contain an NHS Number, or the submitted NHS Number cannot be validated, the supplied demographic details are used to find a unique match in PDS.

Where a 'sufficiently close match' is found, the NHS Number is verified and MPS adds the corresponding PDS values to the demographic data items. The MPS Return Code – 00 SUCCESS – is assigned to the record. NHS Number is used as the Person ID to enable linkage to different records within a data set and across different data sets.

If no match is found, i.e. it appears the person does not have a PDS record or NHS number, a unique MPS ID is created. This ID is reusable and linkable i.e. records with the same demographic details are assigned the same MPS ID, unless and until an NHS Number can be identified. An MPS Return Code – 98 NO MATCH FOUND – is assigned to the record. The MPS ID is used as the Person ID to enable linkage to different records within a data set and across different data sets.

If multiple potential matches are found, a unique MPS ID is not assigned. Instead, a unique DPS ID is created. An MPS Return Code – 97 NOT ENOUGH DATA – is assigned to the record. The DPS ID is used as the Person ID but this ID is non-reusable i.e. records with the same details are assigned a different DPS ID each time and are non-linkable. Records with a DPS ID cannot be linked within the same data set or across different data sets.

NB: for most viewers of the data the Person ID and other Person Confidential Data (PCD) will be presented in a pseudonymised, tokenised or derived form.

Unmatched Record Identifiers - MPS ID (UPRI 1) and UPRI 2

Spine generates the MPS ID, a unique identifier (known to DPS Core and the programme as UPRI 1) and in some scenarios DPS Core will create their own unique ID known as UPRI 2.

The MPS ID (UPRI 1) is linkable and reusable – where the same details are received in a different record then the same MPS ID will be assigned to the record.

The DPS Core generated unique ID (UPRI 2) is not linkable and is non-reusable - where the same details are received in a different record a new UPRI 2 Identifier will be created and assigned to the record.

Here is the logic:

An NHS Number is provided:

- Code=00 (Success) **and** NHS# is neither 0000000000 nor 9999999999
- Code=90 (Success Superseded)
- Code=92 (Sensitive)

MPS Spine generates (or reuses) a reusable and linkable MPS ID:

- Code=00 (Success) **and** NHS#=0000000000 (no match) **and** MATCHED_ALGORITHM_INDICATOR=4
- Code=98 (Not Found)

DPS Core generates a non-reusable and non-linkable unique id aka UPRI 2 ID:

- Code=00 **and** NHS#=0000000000 (no match) **and** MATCHED_ALGORITHM_INDICATOR != 4
- Code=00 **and** NHS#=9999999999 (multiple matches)
- Code=15 (No Trace Performed)
- Code=91 (Invalid)
- Code=96 (Not Enough Data)
- Code=97 (Multiple Close Matches)

PHASE 4 MPS TRACE STEPS

STEP 1: CROSS CHECK TRACE

This step is always run.

Please note there is currently no scoring of matches in this step.

The mandatory parameters used in Cross Check Trace are **NHS NUMBER** and **DATE OF BIRTH (YYYYMMDD)**

To start with, the Cross Check Trace will check for:

1) An exact match on **DOB** and **NHS Number**:

- NHS Number matches exactly
- DOB matches exactly

Outcome: If details match, processing is complete, Steps 2 and 3 are not run so no further checks performed on supplied demographic data attributes or scores calculated. Person ID is NHS Number.

If no match and NHS Number and Date of Birth are the only parameters available then, MPS will check for:

2) An exact match on **NHS number** and partial match on **DOB** and name or postcode (ELSE 3):

- NHS number matches exactly
- Any 2 of the 3 DOB components (day, month, year) match. NOTE each component is only compared to its corresponding item i.e. day is only compared to day, month to month and year to year OR DD/MM are swapped e.g. 06/12 becomes 12/06 or DD or YY are flipped

- Post code is limited to type = home and includes both current and historic postcode. If no match and Given Name and Family Name are available in the request file then the Cross Check Trace will check
- 1st character of forename and first 3 characters of surname match and Outcode in postcode match. NB: Outcode is all the characters before the space e.g. LS1 or BD20

Outcome: If details match processing is complete, Steps 2 and 3 are not run so no further checks performed on supplied demographic data attributes or scores calculated. Person ID is NHS Number.

3) not a match

Note: Where partial date of birth match, if forename or surname are given (provided in the request) but don't match (i.e. the initials of forename and first 3 characters of surname don't match) then return "not a match". In this scenario, even if the Outcode of postcode matches, it should return "not a match".

STEP 2: SIMPLE TRACE

Deleted as all parts are now included in Step 3.

STEP 3: ALPHANUMERIC TRACE

Can only be run if Family Name data is provided, therefore is not applicable to MHSDS where this information is not captured.

STEP 4: ALGORITHMIC TRACE

If Alphanumeric Trace is unsuccessful and the minimum required data set is provided, then perform an Algorithmic Trace. (note that wildcard is not supported)

Please note currently scoring only occurs on this trace step.

Minimum set of Parameters is one of:

- Family Name, Given Name, Date of Birth
 - Family Name, Gender, Date of Birth, Postcode
 - Given Name, Gender, Date of Birth, Postcode
- Note: MPT Algorithmic Trace does not match on Family Name, PAF Address Key and Other Given Name
- D.O.B, Postcode, Gender

Following 4 steps are performed:

- **BLOCK1:** Find all matches using FAMILYNAME(SOUNDEX), GIVENNAME(SOUNDEX), DATEOFBIRTH (YYYY/ YYYYMM/ YYYYMMDD)
- **BLOCK2:** Find all matches using FAMILYNAME(SOUNDEX), GENDER, DATEOFBIRTH (YYYY/ YYYYMM/ YYYYMMDD), POSTCODE
- **BLOCK3:** Find all matches using GIVENNAME(SOUNDEX), GENDER, DATEOFBIRTH (YYYY/ YYYYMM/ YYYYMMDD), POSTCODE

- **BLOCK4:** Find all matches using GENDER, DATEOFBIRTH (YYYY/ YYYYMM/ YYYYMMDD), POSTCODE

Phase 4 MPS is been used for the Mental Health and Maternity datasets, where no name data will be supplied. Therefore 3 out of 4 Blocks listed above will not be performed.

Block 4 is looking for **exact** match on the first two following items and partial on DATEOFBIRTH:

- GENDER
- POSTCODE
- DATEOFBIRTH (includes partial match on – YYYY/YYYYMM and full match on YYYYMMDD)

BLOCK 4 scoring:

Scored fields	Provided in dataset (MH/Maternity)	Possible score values in the returned response file
Date of birth YYYY – partial YYYYMM – partial YYYYMMDD – full	Provided	YYYY- 33 YYYYMM – 66 YYYYMMDD – 100
Gender	Provided	100
Postcode	Provided	100

Note '0' values = no match, so the result will not be returned.

The total probability score for the interim, (for results where there is no exact match on NHS# and DOB) can only be an aggregate total of the above e.g.:

<i>Algorithmic_Trace_DoB_score_%</i>	66 (partial match on YYYYMM)
<i>Algorithmic_Trace_GENDER_score_%</i>	100 (exact match)
<i>Algorithmic_Trace_POSTCODE_score_%</i>	100 (exact match)
= MATCHED_CONFIDENCE_PERCENTAGE	$266/300 \times 100 = 88.66\%$

If multiple matches as a result of Algorithmic Trace/ Block 4, then the difference between the first and the second record must be greater than 5 percent. If so, then MPS will take the

record with the highest score. Otherwise, we return MULTIPLE_CLOSE_MATCHES (Outcome Code: 97).

CONFIDENCE RATING FILTER

MPS also features a confidence rating filter feature which can be set to a configurable value. Any results matched under the confidence threshold will not be returned in the response file. The value will be more precisely determined as part of a future planned development which will introduce probabilistic scoring.

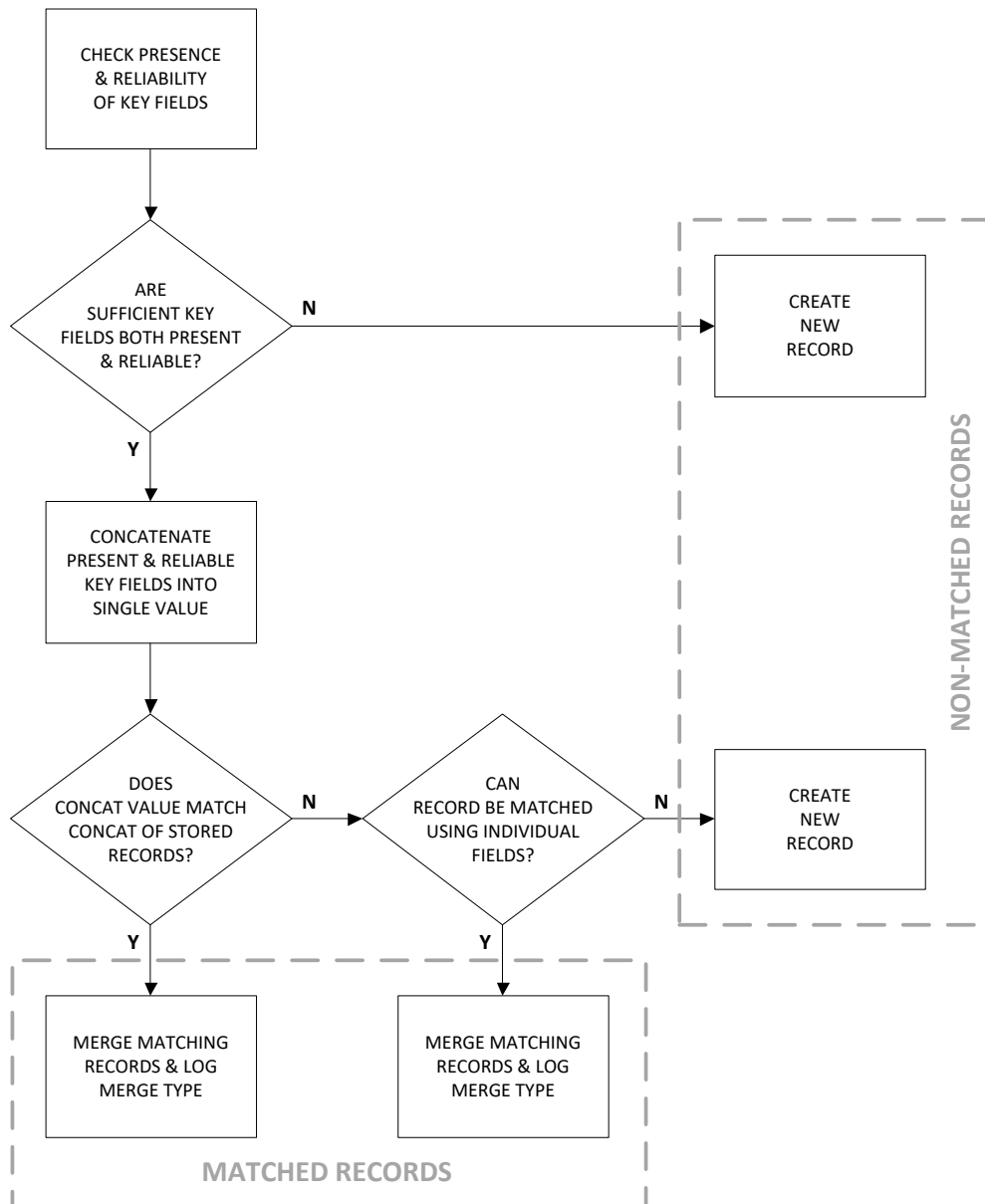
Currently the lowest value score for Mental Health will be 77.6% (which is an exact match on postcode and gender and partial match on DoB YYYY).

Appendix B: The previous MHSDS Person Index

This appendix includes details of the function of the person index applied to MHSDS prior to the implementation of MPS in April 2019.

Overview

The process map below demonstrates the basic principles of the person index:



Key aspects of the process to note are:

- Before attempting to match the record the system checks that the fields required for matching are populated and if so that they are populated with reliable data. If the record does not contain sufficient reliable data in these fields then matching is not attempted. This avoids wasting processing time on matching attempts that cannot succeed.
- When attempting to match the record the system begins by concatenating all the key fields together into a single value and attempting a single one-to-one match. This is a more efficient way of matching and most records within MHSDS can be matched in this way. Only if the concatenated value cannot be matched will the system then move on to the more processing intensive field by field matching process.

Defining the Priority Order of Key Fields

The system uses six fields for matching purposes. Initially all the values from all six are concatenated together into a single value and a deterministic match is attempted. If this is unsuccessful however, then the system will attempt deterministic matching on some of these individual fields in priority order instead.

Three of the fields are used primarily only in the concatenated matching process as while they can be valuable for matching purposes when used in conjunction with other fields, they are not considered appropriate for matching alone as follows:

Organisation Code (the official ONS code of the service provider)

Every patient receiving treatment from the same provider will have the same ORG code and so this field is not considered sufficiently identifying to be worth using in individual field matching.

Local Patient Identifier (an internal patient ID assigned by service providers)

These IDs are issued locally within providers and are often simple numbers. While this is perfectly appropriate at a provider level, when considering patients at a national level there is a high probability that LPIDs will be duplicated across providers, especially if they use the same third party IT systems. Subsequently this field is not considered sufficiently identifying to be worth using in individual field matching.

Gender

The very small number of possible values for this field means that it is not considered sufficiently identifying to be worth using in individual field matching.

The remaining three fields can be used individually for matching purposes and the priority order in terms of reliability was defined as follows:

1. NHS Number
2. Date of Birth
3. Postcode
4. Organisation Code + Local Patient Identifier

This order was determined by considering two aspects of each field, namely the probability of duplicate values and the probability of values changing over time with lower values being better in both cases.

In these terms NHS Number is a strong first since it should always be unique and should never change meaning a probability of effectively zero for both aspects.

Date of Birth and Postcode are much closer in these terms however. On one hand there are only a maximum of 366 possible Dates of Birth per year, while there are hundreds of thousands of different postcodes meaning that the probability of duplication is higher for Date of Birth.

On the other hand, a patient's Date of Birth changes very rarely whereas postcodes would be expected to change numerous times during a patient's life. Subsequently, combining these two aspects results in Date of Birth being assigned a higher priority.

Finally, if the NHS Number of a record has been matched but the Date of Birth and Postcode have not, a concatenated value of organisation code and local patient identifier is used as a last resort.

Methodology

Preparatory Data Validation at the Point of Submission

When providers submit data to the MHSDS a series of validations are applied to the data. If any data within the submission breaks the specified data quality rules then, depending on the severity of the problem, the data will either be included in the submission but the provider will receive a warning of the problem, or the offending data will actually be removed from the submission and the provider informed.

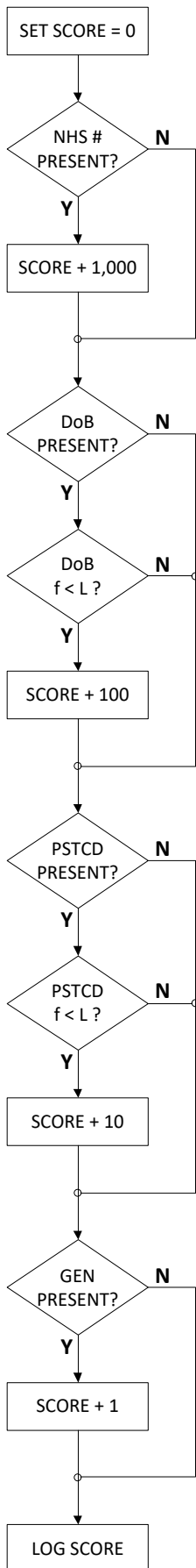
In this way these validation rules define the nature of the data the MHSDS Person Index has to be able to accommodate. The provider's organisation code and the local patient identifier can both be relied upon to always be present however the remaining fields used for record matching must be considered in terms of their validations as follows:

	NULL VALUE	INVALID VALUE	INVALID FORMAT
NHS NUMBER	accepted	changed to null	changed to null
DATE OF BIRTH	accepted	changed to null	changed to null
DATE OF BIRTH (babies in MSDS)	rejected	rejected	changed to null
POSTCODE	accepted	accepted	changed to null
GENDER	accepted	accepted	changed to null

(NB. Date of Birth for babies related to the Maternity Services Dataset and is not applicable to the MHSDS)

In this context the impact of these validations is that the system will only ever have to deal with either values of the correct format, or null values, and so has been designed accordingly.

Presence Checking and 'Score' Calculation



This process map demonstrates the initial 'presence checking' part of the system. As described above, within the MHSDS it can be safely assumed that the organisation code and local patient identifier fields will always be present and so they do not need to be included in this process.

For each record received, the system looks at each of the remaining matching fields in turn and checks whether a value is present. Non-null Date of Birth and Postcode values are also checked for reliability. The record is then assigned a 'score' between zero and 1,111.

This 'score' is actually just a form of shorthand notation that records which fields are populated and reliable and which are not. As the score is a numeric value it can then be used in subsequent record matching calculations.

The calculation of a score for a record is very simple: a field populated with a reliable value is represented by a '1' and a null field by a '0'. These values are then effectively multiplied by a different multiple of ten dependent on the priority of the field.

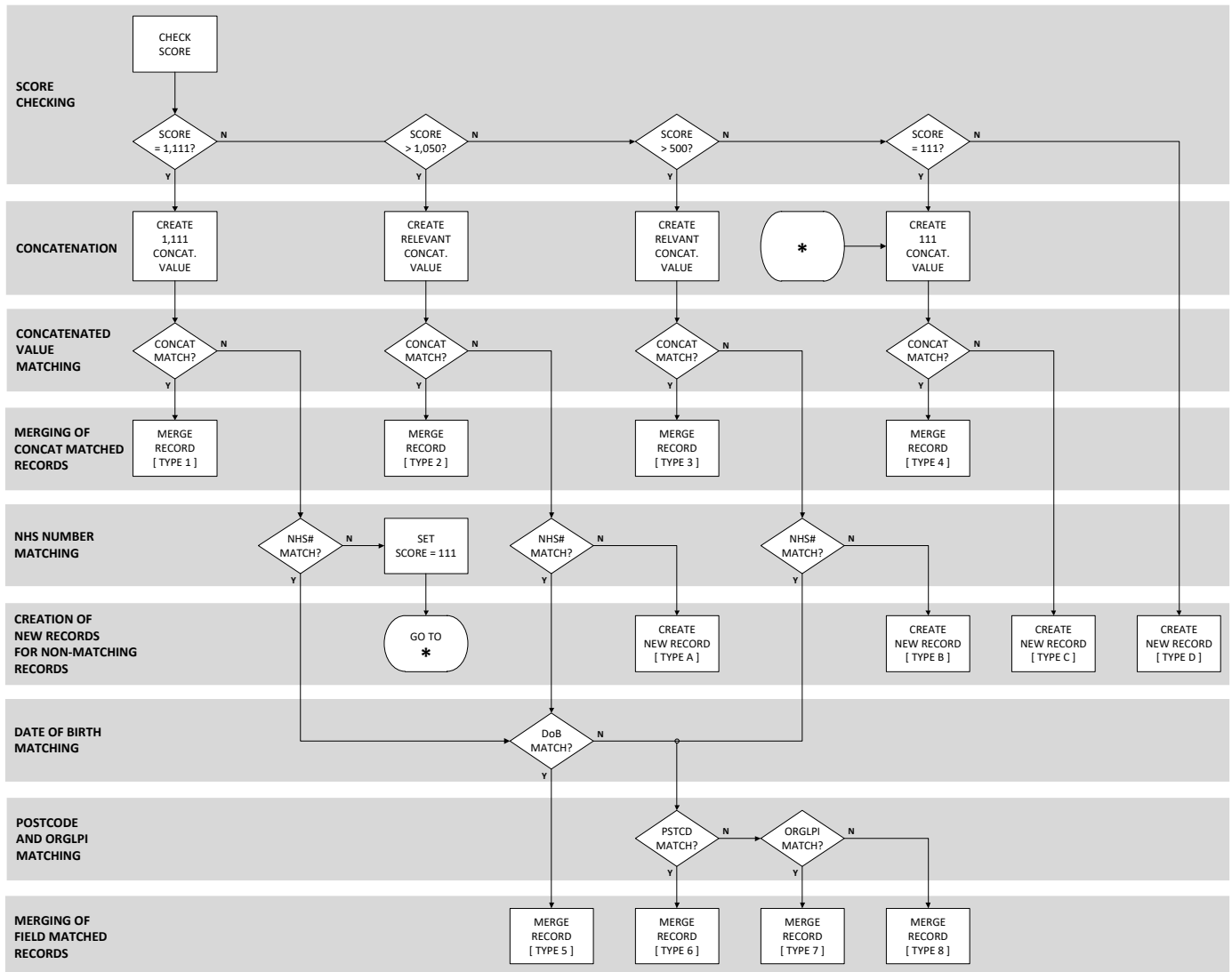
The following examples demonstrate the calculation of a variety of different scores:

Record	Field	Populated?	Reliable?	Score
A	NHS Number	Y	[N/A]	1,100
	Date of Birth	Y	Y	
	Postcode	Y	N	
	Gender	N	[N/A]	
B	NHS Number	N	[N/A]	101
	Date of Birth	Y	Y	
	Postcode	N	[N/A]	
	Gender	Y	[N/A]	
C	NHS Number	N	[N/A]	10
	Date of Birth	N	[N/A]	
	Postcode	Y	Y	
	Gender	Y	[N/A]	

(NB. reliability checks are only carried out on the Date of Birth and Postcode fields and only when a value is present,)

5.4 Record Matching Process Overview

The process map below demonstrates the primary record matching logic used by the MHSDS Person Index:



Once a record has been assigned a presence and reliability score, that score is then used to determine which record matching logic path the record should follow. This prevents the system attempting to match using values that are not present and thereby increases efficiency.

The first matching attempt involves concatenating all of the values present, including the organisation code, local patient identifier as well as Gender where it is present. If this form of matching is unsuccessful then the process moves down into individual field matching.

In most cases, if a record cannot be matched at the concatenation stage nor on NHS Number alone then it is stored as a new record. The exception to this is if the record enters the system with a score of 1,111. In this instance, before creating a new record the system checks to see if the record can be perfectly matched on all other fields with a previously submitted record which did not have an NHS Number.

Any records entering the system with a score of less than 111 are considered insufficiently populated for matching to even be attempted and so are immediately created as new records.