Detecting referral and selection bias by the anonymous linkage of population, practice, hospital and clinic data using SAPREL (Secure and Private Record Linkage)

Tom Chan
Senior Lecturer in Biomedical Informatics
tchan@sgul.ac.uk

Simon de Lusignan
Reader in General Practice and Informatics
slusigna@sgul.ac.uk

Rob Navarro
Hon Research Assistant & Director Sapior
robert.navarro@sapior.com

Senior Investigators from the IAPT evaluation team to be invited to contribute:

Glenys Parry,
Professor of Applied Psychological Therapies
Email: G.D.Parry@sheffield.ac.uk

Kim Dent-Brown
Research Fellow
Email: K.Dent-Brown@sheffield.ac.uk

Tony Kendrick
Professor of Primary Care and Dean.
Hull York Medical School
tony.kendrick@hyms.ac.uk

Division of Population Health Sciences and Education
St. George’s – University of London
LONDON SW17 ORE
Tel: 020 8725 5661
Fax: 020 8725 3584
Sapior Ltd
8 Cheyne Avenue
London
E18 2DR
United Kingdom
Phone: +44 (0) 871 218 2580
Fax: +44 (0)20 7748 0970

Centre for Psychological Services Research
ScHARR, University of Sheffield
Regent Court, Regent Street
SHEFFIELD S1 4DA
Tel +44 (0)114 222 0753
Fax +44 (0)114 222 0785

Hull York Medical School,
University of Hull,
Hull HU6 7RX
Tel +44 (0)1482 464701
Fax +44(0)1482464705
Detecting referral and selection bias by the anonymous linkage of population, practice, hospital and clinic data using SAPREL (Secure and Private Record Linkage)

Abstract:
Background: The evaluation of pilot demonstration sites set up to provide improved access to psychological therapies (IAPT) involved the study of all people identified as having common mental health problems, those referred to the service, and a sample of service attenders studied in-depth. Information technology makes it feasible to link population, practice, hospital and clinic data to evaluate the representativeness of those referred, and the sample studied in-depth. However, researchers do not have permission to browse and link these data without the patients’ consent.

Objective: To demonstrate the use of a novel technology – secure and private record linkage (SAPREL) – to describe selection bias in subjects chosen for in-depth evaluation.

Method: We extracted, pseudonymised and used fuzzy logic to link multiple health records without the researcher knowing details of the patient’s identity. Within the data provider safe haven we extracted: Demographic data, hospital utilisation and IAPT clinic data; converted post code to index of multiple deprivation (IMD); and identified people with common mental health problems (CMHP).

We contrasted the age, gender, ethnicity and IMD for the population who had in-depth evaluation with people referred to IAPT, using hospital services and of the population as a whole.

Results: The in IAPT-in-depth group had a mean age of 43.1years; CI: 41.0 - 45.2 (n=166); the IAPT-referred 40.2years; CI: 39.4 - 40.9 (n=1118); and those with CMHP 43.6years SEM 0.15. (n=12210). Whilst around 67% of those with a CMHP were women, compared to 70% of those referred to IAPT, and 75% of those subject to in-depth evaluation (Chi square p<0.001).

The mean IMD score for the in-depth evaluation group was 36.6; CI: 34.2 - 38.9; (n=166); of those referred to IAPT 38.7; CI: 37.9 - 39.6; (n=1117); and of people with CMHP 37.6; CI 37.3-37.9; (n=12143).

Conclusions:
The sample studied in-depth were older, more likely to be female, and less deprived than the total population of people with common mental health problems, and fewer had recorded ethnic minority status. Anonymous linkage using SAPREL provides unique insight about the representativeness of a study population and possible adjustment for selection bias.

Word count 345
Detecting referral and selection bias by the anonymous linkage of population, practice, hospital and clinic data using SAPREL (Secure and Private Record Linkage)

Introduction:
Selection bias may distort the results about the effectiveness of a new service\(^1\). In the NHS nearly all the population are registered with a single family practitioner; and have a single unique identifier (NHS number), which can be linked to health services utilisation, making it possible in theory to quantify selection bias and if needed adjust for it\(^2\). Although computerised records make it technically straightforward to link population, practice, hospital and clinic data it is not possible to extract a patient’s records without their consent. For a large population based study this is not feasible; and obtaining this consent may result in further bias\(^3\). Methods are needed which allow selective mining of key variables from individual patients’ records to enable researchers to know the extent of any selection bias and the likely representativeness of their sample. Such methods should allow anonymous extraction and linkage of data with only the data needed to make comparisons extracted; and the privacy of the patient in maintained through the researcher not having access to any strong identifiers (i.e. no access to name, date of birth and so on)\(^4,5,6,7\).

The Improving Access to Psychological Therapies (IAPT) programme is a quality improvement initiative\(^8\); and its evaluation involved in-depth study of a group of attendees and a case-control study of those referred compared with age-sex practice matched controls\(^9\). IAPT offers a series of stepped interventions, including the use of cognitive behavioural therapy (CBT) which aims to reduce the disease burden to the individual and economic burden to society of psychological illness and enable people to cope better with their mental health problems. The target population for the IAPT programme is people with common mental health problems (CMHP) in primary care, specifically people suffering from depression and/or anxiety disorders. The thresholds of severity of depression and/or anxiety disorders for referral to IAPT were not always strictly adhered to by referrers. Access to IAPT is further complicated as patients also have direct access to the service without seeing their GP. We linked practice, hospital and clinic data to conduct this evaluation – linking anonymised data (strictly pseudonymised as each patient was given an arbitrary identity which can only be re-identified in their originating practice) using privacy enhanced fuzzy matching maximise join quality. We called this process SAPREL – secure and private record linkage. The resulting merged data table enabled the tracking of health utilisation of individual patients of the participating GP practices across primary care, hospital and within the IAPT clinics.

We decided to use these linked data to explore any bias in terms of referral, and in terms of the selection of the sample of service attenders studied in depth. This paper compares the characteristics of the populations linked using the SAPREL process: (1) In-depth evaluation group; (2) Those referred to IAPT; (3) People with CMHP; (4) Uses of hospital services; and (5) The practice registered population. We contrasted their age, gender, ethnicity and level of deprivation.
Method:

Data sources:

We extracted data from 20 general practice electronic patient record (EPR) systems, 10 each in two localities which piloted the first IAPT services in England. One was within London in an area with ethnic diversity; the other a northern city with a predominantly white population. The primary care data were extracted using MIQUEST (Morbidity Information Query and Export Syntax – a Department of Health sponsored application which allows the same data extraction query to be run on different branded EPR systems). These data were extracted, processed and cleaned using well established processes.\textsuperscript{10,11}

The hospital and IAPT clinic data were exported using their standard data export methods. The hospital data were the standard reports of hospital episode statistics (actually called SUS – Secondary Uses Services data) and customised output from the IAPT clinic which was exported from two different bespoke applications developed specifically to support IAPT clinics. These data were de-identified within the premises where the data were held or accessed, and then subsequently linked using the SAPREL method. This method means that no person identifiable data left the premises where such data were held, and at no time did the researchers hold strong identifiers.

The in-depth evaluation data were collected from a cohort of patients recruited from GP practices, who volunteered to be involved in a research project and who consented to providing the researchers with their personal details as well as a number of demographic and clinical measures beyond what would be available from routine clinical practice.

The data set:

The GP practice data set included: personal identifier for data linkage – forename, surname, date of birth, NHS number and postcode; demographic information: gender, ethnicity, registered date; and postcode to link to the Index of Multiple Deprivation (IMD).\textsuperscript{12} Additionally we extracted clinical information which enabled us to report whether a patient had a CMHP, namely a diagnosis or medications for depression or anxiety. Where we compare the people with CMHP with the other groups we make the comparison with the adult (≥16 years) population.

Additional data were extracted for our evaluation, but are not reported in this paper. The ethnicity codes are mapped using an established method to the National Statistics “5+1” categories.\textsuperscript{13} The categories are: Not stated, white, mixed, Asian or Asian British, Black or Black British, Chinese or other ethnic group). Full postcode of each of the GP practice list patients was mapped to IMD using Geographical Information System (GIS) methods. IMD is divided into deciles of equal sizes, where the first decile (IMD ≤5.63) is the least deprived and decile ten (IMD≥45.33) the most deprived. Each decile is assigned a specific numeric range (e.g. 5th decile is 13.74 to 17.08).
We always used NHS number (a unique personal ID) to link primary and secondary care data; NHS number was pseudonymised and encrypted and not directly browsed by the researcher. There were a large number of people with a temporary, duplicate and missing NHS numbers (table 1), as well as errors with missing name or gender (n=609); though this is less than 0.5% of records (N=150,000). Where a unique NHS number was not available in the primary or secondary care data, that record was discarded. Missing post codes (n=129) are less surprising as additional post codes are created with building developments. Although for the evaluation we separated types of attendance for this paper we have combined any hospital attendance during the study period. The GP practice and IAPT clinic data only had forename, surname, date of birth and postcode in common. As part of the SAPREL process, various functions of these fields were paired in various combinations. The link strategy with the lowest estimated false positive and false negative count was ultimately selected. Where an IAPT forename, surname, date of birth or postcode was missing – that record was discarded (n=98; 91 missing postcodes, 7 missing dates of birth).

Table 1: Missing, temporary or duplicate unique identifiers (NHS number) by practice

We used a single variable to encompass all usage of hospital services. This variable included out-patient clinics, accident departments, day treatment units and in-patient episodes.

The purpose of this study is to explore any selection bias in the people referred to IAPT and those who underwent in-depth evaluation. The IAPT clinic data contained information about all those who were referred to the clinics. The IAPT programme records an extensive data set to evaluate their services; though it does not include NHS number. For the evaluation we extracted an extensive dataset, but for the purposes of this study we just characterise the population referred; as our evaluation of these clinics is on an ‘intention to treat’ basis. Additionally, a cohort of people who attended the IAPT clinic was followed-up in depth by the investigation team, and we analysed this group’s linked data separately.

In this paper we contrast the characteristics of this in depth group with those referred to IAPT, people with CMHT, and the populations they were drawn from. Finally, we compared referral rates from practices.

Statistical methods:
We used descriptive statistics to compare the samples. We quoted 95% confidence intervals (CI) and standard error of the mean (SEM) to allow comparison between age groups; and used a t-test to give the probability that age-groups were significantly different. We always used a 2-tailed t-test as one population may be older or younger than the other. We used Chi square to compare proportions of categorical variables. We used the Wilcoxon non-parametric test to compare the distribution of 5-year age-bands between different populations.

Ethical considerations:
A national research ethics committee (reference No: 08/H0715/101) provided ethical review and the research office of the local healthcare organizations involved provided local site approval. Specific approval was obtained from the former Patient Information Advisory Group (PIAG) for Section 60 exemption for the transient holding of patient identifiable information while it is pseudonymised and encrypted on health service premises (Reference number: PIAG 6-06(h)/2008). The SAPREL process complies with both research and information governance frameworks in England in the use and protection of patient information; and was commended by PIAG as an example of best practice.

**Results:**

*Age and gender*

The in-depth study cohort was roughly three years older than the people referred to IAPT (One sample t-test \( p<0.000 \)) and about 0.5 years younger than people with CMHP (One sample t-test \( p=0.001 \)). The mean for the in-depth cohort was 43.1; CI: 41.0 - 45.2 (n=166) for people referred to IAPT: mean=40.2; CI: 39.4 - 40.9 (n=1118); and for people with CMHP 43.6 years; CI: 43.3 – 43.9 (n=12210). The practice population was younger (mean age 35.3years; CI: 35.1 – 35.4 (n=152302)) though the population attending hospital were a similar age (41.6; Std D=22.2; CI: 41.4 - 41.8 (n=60143)). Similarly the age-sex profiles of the practices were not significantly different from the locality populations they were drawn from; Wilcoxon non-parametric test \( p=0.133 \) (non-significant), though there was a slight excess of people between 25 and 34 years old. However, although the people who use hospital services and IAPT clients have similar mean ages, this the majority of people referred to IAPT are in the age bands between 20 and 54; reflecting how initially this service was targeted at people of working age. Nobody below the age of 16 and few people aged 65 and over referred to IAPT. The in-depth evaluation group was over-represented in the over 35 years age group; and increased use of hospital starts after age 45 (Graph 1).

**Graph 1: Comparing the 5-year age bands of people in the in-depth evaluation, referred to IAPT and who use hospital services**

Women represented just under half of the population; yet female gender is associated with greater use of hospital, common mental health problems, referral to IAPT, and with three-quarters of the in-depth study cohort being female. The gender distribution in the five comparator populations are shown in table 2, with all the differences statistically significant. The proportion of women referred to IAPT is similar to the greater proportion with CMHP, but there is a step up in proportion in the in-depth study cohort. The people referred to IAPT consisted almost entirely of adults of working age (15 – 64), two thirds of whom are women.

**Table 2: Comparison by gender of practice list, using hospital services, common mental health problems (CMHP), referred to IAPT and part of the in-depth evaluation cohort**

*Ethnicity*
The level of ethnicity recording in the in-depth evaluation group was about half that recorded for the rest of the population (table 3). Around two-thirds (64.8%) of the population referred to IAPT and over half (53.1%) of the subgroup with CMHP had their ethnicity recorded; while generally ethnicity was recorded for 60% of the practice population and those referred to hospital. The level of recording of ethnicity codes varied significantly between the two demonstration sites (table not shown); the London site had a higher level of recorded ethnicity (81.4%) compared with the northern site (35.1%).

Table 3: Rates of recording of ethnicity by patient group

Around 60% of people in the different study populations have their ethnicity recorded; except in the in-depth study cohort where it fell to just over 30%. White ethnicity is recorded in 32.3% of the GP practices’ population, in 40.5% of those with CMHPs and 47.9% of people referred to IAPT clinics. However, only 28.9% of people in the in-depth study cohort were recorded as having white ethnicity; though this represents 90% of the ethnic recording in this group. There were no members of the in-depth evaluation group with recorded Mixed, Chinese or Black ethnicity; Asian, black and Chinese ethnicity recording was under-represented in the group referred to IAPT (Graph 2).

Graph 2: Proportion of each ethnic group in four study populations (In-depth evaluation, Referred to IAPT, Use hospital services, Registered with a study practice)

Index of multiple deprivation (IMD)

Over 99% of people were mapped to an appropriate IMD decile; less than a thousand out of the 152302 patients (about 0.5%) could not be mapped due to invalid postcodes. The mean IMD scores for the four populations are similar. The mean IMD score for the sample selected for the in-depth IAPT study is 36.6. The equivalent data for those referred to IAPT, people with CMHP, those who use hospital services, and for the population were: 38.7, 37.6, 38.1 and 38.0 respectively (Table 4).

Table 4: Mean deprivation (IMD) score for each study population

The majority (30.5%+32%=62.5%) of the patients of the GP registered population are in the most deprived 20% (i.e. 9th and the 10th decile). The hospital, CMHP, and the IAPT clinic populations broadly reflected the IMD scores of the GP list population from which they were drawn, with perhaps a small rise in the 10th decile and a small reduction in the 8th decile for IAPT clients compared with the GP list population. The in-depth IAPT study cohort were slightly over-represented (compared with the other three study populations) in the 3rd, 5th and the 6th decile (Graph 3).

Graph 3: Distribution of study populations by index of multiple deprivation (IMD) decile

People with common mental health problems (CMHP):
We had expected that those referred to IAPT would be drawn from people with diagnosed CMHP. However, not everyone referred to IAPT or in the in-depth evaluation group was in the CMHP population. We therefore compared those with and without CMHP in each of our four groups: in-depth evaluation, referred to IAPT, received hospitalisation, and for the registered population as a whole (Table 5).

**Table 5:** Comparing those with and without common mental health problems within the four study populations

In the in-depth cohort those with CMHP were around seven years younger and more likely to be female compared with those with no CMHP; there are no significant differences in multiple deprivation scores. In the group referred to IAPT those with CMHP came from slightly less deprived areas, and more likely to be female compared with those with no CMHP; there were no significant difference in age. In the adult population who had attended hospital the people with CMHP were approximately two years younger, came from slightly more deprived areas, and more likely to be female. Taking the practice population as a whole people with CMHP tended to be a little older and more likely to be female.

**Inter-practice and inter-locality variation:**
All the practices except one referred patients to IAPT; and two other practices were low referrers. All the practices referred between 0.5% and 2.5% of their adult registered list, which is also between 3.5% and 12.5% of the number of people with CMHP (though not all those referred came from this group). The age distributions in the two localities was not significantly different from the population in the sample practices (Wilcoxon non-parametric tests comparing 5-year age bands) p=0.58 and p=0.145 for the northern and southern localities respectively. The northern was almost a perfect match, there was an excess of young adults in the southern locality.

**Discussion:**

**Principal findings:**
By linking three differently structured health databases we were able to characterise the population referred and the group who were part of the in-depth evaluation. The practice populations were not significantly different from the locality from which they were drawn. The people referred to IAPT were not exclusively drawn from those recorded as having CMHP; and the group studies in-depth for the service evaluation were relatively older, more likely to be women, and included fewer with a recorded ethnic minority status.

This approach has allowed databases designed to serve a different purposes and using different coding systems to be linked. The SAPREL process also meets with stringent research and information governance requirements for the ethical use and protection of patient information.
The finding also shows that IAPT services can engage with people from the lowest IMD deciles; over 60% of the patients in the study practices’ populations lived in the 20% most deprived areas as measured by the Index of Multiple Deprivation. However, men, older people and some ethnic groups were apparently less likely to be referred.

**Implications of the findings:**
This method of data linkage has the potential to make more data available for research and for quality improvement.

The SAPREL technique has the capability to be used prospectively in future studies to identify cases that are more representative. These data were linked so that we could conduct a case controlled, before and after study of hospital utilisation in the group referred to IAPT, and by the time that permissions were obtained and this linkage completed this study was well underway. However, the SAPREL technology would allow in future studies a sample for in-depth study to be purposively sampled and therefore to be more representative of the population or group under study. This approach has the potential to quantify any selection bias and allow researchers to avoid or adjust for it.

Linking data at the individual personal level across care boundaries offers the opportunity of evaluating system impact of policy initiatives which cannot be effectively measured separately in different sectors of the healthcare provision. The resulting file tracks a cohort of individual patients through the system across primary and secondary healthcare organizational boundaries, and avoids the potential biased conclusion through the analysis of different cohorts of patients from separate cross-sectional databases.

**Comparison with the literature:**
Analysis in this paper found that compared with the 2001 census, the study practices’ populations had more people of working age (20 to 50 for men, and 20 to 35 for women), and a larger proportion of children aged under 5. Other studies and household surveys conducted in the UK report a higher prevalence of common mental health problems in females. Deprivation is associated with poor outcomes including in mental health, so the accessibility of the IAPT service to people of low socioeconomic status is important.

Strategies such as placing researchers with honorary contracts into practices have been suggested as an alternative method of accessing records, methods have also been piloted to use “agents” (software to flag eligible patients) to meet this need. However, all of these methods, including SAPREL rely on the clinicians responsible for patient’s health data having trust in the professionalism of the person extracting these data.

**Limitations of the method:**
Our reporting of ethnicity could have been more complete. We could have also increased the ethnicity recording using data from hospital information but did not retain this data field.

Recruitment of patients for research projects is challenging; and it is possible that older patients had more time available to participate. The fuzzy logic linkage used in SAPREL was not independently checked for accuracy. We did not find any
contradictions in these data, however this does not mean there was a perfect match. We feel that this linkage is sufficiently good for a research study but because there will be errors (false positive and false negative) it is not recommended as a method for identifying the care in practice of individual patients.

Call for further research:
Further research is needed to assess the acceptability of this approach to patients and to test its reliability – using a dataset where we can do open matching as well as using the SAPREL fuzzy logic.

Conclusions:

Patients referred to the IAPT are predominantly of working age, i.e. aged between 20 and 64 and recording of white ethnicity is overrepresented in the IAPT referred group. The sample who volunteered for the in-depth IAPT study were not entirely representative of the total population referred to the IAPT programme. They tended to be drawn from less deprived areas, were even more likely to be female, and older. These biases must be borne in mind when attempting to extrapolate the findings of the study to other patient populations. Linking data using SAPREL is technically feasible, ethically acceptable and has provided insight into selection bias.
Acknowledgements:

Participating practices and colleagues in the participating NHS Trusts and IAPT clinics. This project was funded by Department of Health (IAPT team) and NIHR SDO project funding (Principal Investigator Professor Glenys Parry, Sheffield University).

Contributions:

TC: Developed the detailed protocol, managed the ethics and PIAG application, and made major contributions to writing the paper.

SdeL: Principal investigator for the data linkage project; overall design and responsibility for the project. Developed the analysis plan with TC, drafted the paper outline and made major contributions to subsequent drafts. Liaised with RN over the detail of the study.

RN: Honorary Research Assistant at St. George’s and Technical Director of Sapior. Worked with SdeL to explore to how methods designed for private linkage of large datasets could be used in the research context. Used Sapior privacy enhanced fuzzy logic system to link.

GP: PI of NIHR SDO evaluation of IAPT demonstration sites, input into protocol and contributed to the paper.

KD-B: Input into the detailed organisation and running of all stages of the project and contributed to the paper.

TK: Led the write up of the GP data within the NIHR SDO final report and contributed to the paper.

Conflict of interest:

TC: None

SdeL: None

RN: Technical Director of Sapior Ltd – who supply privacy enhanced solutions for the collection, de-identification and linking of sensitive data. The algorithms and fuzzy logic used are the intellectual property of Sapior.

GP: None

KD-B: None

TK: None
Tables, Graphs and Figures:

Table 1: Missing, temporary or duplicate unique identifiers (NHS number) by practice
Table 2: Comparison by gender of practice list, using hospital services, common mental health problems (CMHP), referred to IAPT and part of the in-depth evaluation cohort
Table 3: Rates of recording of ethnicity by patient group
Table 4: Mean deprivation (IMD) score for each study population
Table 5: Comparing those with and without common mental health problems within the four study populations

Graph 1: Comparing the 5-year age bands of people in the in-depth evaluation, referred to IAPT and who use hospital services
Graph 2: Proportion of each ethnic group in four study populations (In-depth evaluation, Referred to IAPT, Use hospital services, Registered with a study practice)
Graph 3: Distribution of study populations by index of multiple deprivation (IMD) decile
<table>
<thead>
<tr>
<th>Practice number</th>
<th>List size</th>
<th>Duplicate or default NHS number</th>
<th>Missing NHS Number problems</th>
<th>Total of NHS Number problems</th>
<th>Missing Forename or Surname</th>
<th>Missing Postcode</th>
<th>Demographic data errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>2</td>
<td>13552</td>
<td>0</td>
<td>145</td>
<td>145</td>
<td>1.07</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>14</td>
<td>7190</td>
<td>68</td>
<td>0</td>
<td>68</td>
<td>0.95</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>10354</td>
<td>61</td>
<td>0</td>
<td>61</td>
<td>0.59</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>5189</td>
<td>49</td>
<td>0</td>
<td>49</td>
<td>0.94</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>7611</td>
<td>52</td>
<td>0</td>
<td>52</td>
<td>0.68</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>11828</td>
<td>45</td>
<td>0</td>
<td>45</td>
<td>0.38</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>6875</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>0.1</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>19</td>
<td>4715</td>
<td>38</td>
<td>0</td>
<td>38</td>
<td>0.81</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>8951</td>
<td>36</td>
<td>0</td>
<td>36</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>5971</td>
<td>31</td>
<td>0</td>
<td>31</td>
<td>0.52</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>6661</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0.03</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>17</td>
<td>7338</td>
<td>23</td>
<td>0</td>
<td>23</td>
<td>0.31</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>9508</td>
<td>0</td>
<td>14</td>
<td>14</td>
<td>0.15</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>7951</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>0.06</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>18</td>
<td>3740</td>
<td>11</td>
<td>0</td>
<td>11</td>
<td>0.29</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>4466</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>14708</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>10049</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2092</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3614</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>152363</td>
<td>436</td>
<td>170</td>
<td>606</td>
<td>0.4</td>
<td>3</td>
<td>129</td>
</tr>
</tbody>
</table>
Table 2: Comparison by gender of practice list, using hospital services, common mental health problems (CMHP), referred to IAPT and part of the in-depth evaluation cohort

<table>
<thead>
<tr>
<th></th>
<th>In-depth study cohort</th>
<th>IAPT referrals</th>
<th>CMHP</th>
<th>Use hospital services</th>
<th>Practice list</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>123</td>
<td>734</td>
<td>8229</td>
<td>33785</td>
<td>75523</td>
</tr>
<tr>
<td></td>
<td>74.1%</td>
<td>65.7%</td>
<td>67.4%</td>
<td>56.0%</td>
<td>49.6%</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>384</td>
<td>3981</td>
<td>26457</td>
<td>76779</td>
</tr>
<tr>
<td></td>
<td>25.9%</td>
<td>34.3%</td>
<td>32.6%</td>
<td>44.0%</td>
<td>50.4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>166</td>
<td>1118</td>
<td>12210</td>
<td>60143</td>
<td>152302</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Pearson $X^2$ $p<0.001$  
Pearson $X^2$ $p=0.023$  
Pearson $X^2$ $p<0.001$  
Pearson $X^2$ $p<0.001$  
NPar $X^2$ $p=0.016$
### Table 3: Rates of recording of ethnicity by patient group

<table>
<thead>
<tr>
<th></th>
<th>In-depth study cohort</th>
<th>IAPT clinic referral</th>
<th>CMHP</th>
<th>Use hospital services</th>
<th>GP list population</th>
</tr>
</thead>
<tbody>
<tr>
<td>No ethnicity code /not stated</td>
<td>113 (68.1%)</td>
<td>394 (35.2%)</td>
<td>5732 (46.9%)</td>
<td>38.6%</td>
<td>62389 (41.0%)</td>
</tr>
<tr>
<td>White</td>
<td>48 (28.9%)</td>
<td>536 (47.9%)</td>
<td>4940 (40.5%)</td>
<td>21318 (35.4%)</td>
<td>49127 (32.3%)</td>
</tr>
<tr>
<td>Mixed</td>
<td>0 (0.0%)</td>
<td>10 (0.9%)</td>
<td>78 (0.6%)</td>
<td>605 (1.0%)</td>
<td>1432 (0.9%)</td>
</tr>
<tr>
<td>Asian or Asian British</td>
<td>5 (3.0%)</td>
<td>91 (8.1%)</td>
<td>710 (5.8%)</td>
<td>7470 (12.4%)</td>
<td>19550 (12.8%)</td>
</tr>
<tr>
<td>Black or Black British</td>
<td>0 (0%)</td>
<td>78 (7.0%)</td>
<td>639 (5.2%)</td>
<td>6345 (10.5%)</td>
<td>15957 (10.5%)</td>
</tr>
<tr>
<td>Chinese or other ethnic group</td>
<td>0 (0%)</td>
<td>9 (0.8%)</td>
<td>111 (0.9%)</td>
<td>1200 (2.0%)</td>
<td>3847 (2.5%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>166 (100%)</strong></td>
<td><strong>1118 (100%)</strong></td>
<td><strong>12210 (100%)</strong></td>
<td><strong>60143 (100%)</strong></td>
<td><strong>152302 (100%)</strong></td>
</tr>
</tbody>
</table>

Over two-thirds of the in-depth study cohort do not have ethnicity recorded in their GP record.
Table 4: Mean deprivation (IMD) score for each study population

<table>
<thead>
<tr>
<th></th>
<th>In-depth study cohort</th>
<th>IAPT clinic referral</th>
<th>CMHP Use hospital services</th>
<th>GP list population</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>166</td>
<td>1,117</td>
<td>12,143</td>
<td>59,903</td>
</tr>
<tr>
<td>Mean IMD score</td>
<td>36.6</td>
<td>38.7</td>
<td>37.6</td>
<td>38.1</td>
</tr>
<tr>
<td>95% Confidence Intervals</td>
<td>34.2-38.9</td>
<td>37.9-39.6</td>
<td>37.3-37.9</td>
<td>38.0–38.2</td>
</tr>
</tbody>
</table>


Table 5: Comparing those with and without common mental health problems within the four study populations

<table>
<thead>
<tr>
<th></th>
<th>In-depth study population</th>
<th>Referred to IAPT</th>
<th>Hospital adult population</th>
<th>GP listed adult population</th>
</tr>
</thead>
<tbody>
<tr>
<td>People with NO record of Common Mental Health Problem (CMHP):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of population</td>
<td>10.80%</td>
<td>30.80%</td>
<td>85.20%</td>
<td>89.90%</td>
</tr>
<tr>
<td>Age (SEM)</td>
<td>49.3 (3.4)</td>
<td>40.2 (0.74)</td>
<td>46.5 (0.05)</td>
<td>42.3 (0.05)</td>
</tr>
<tr>
<td>Gender (F% : M%)</td>
<td>61.1% : 38.9%</td>
<td>55.8% : 44.2%</td>
<td>55.0% : 45.0%</td>
<td>47.7% : 53.3%</td>
</tr>
<tr>
<td>IMD (SEM)</td>
<td>36.8% (3.2)</td>
<td>40.4 (0.72)</td>
<td>37.6 (0.07)</td>
<td>37.7 (0.04)</td>
</tr>
<tr>
<td>People with a Common Mental Health Problem (CMHP):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of population</td>
<td>89.20%</td>
<td>69.20%</td>
<td>14.80%</td>
<td>10.10%</td>
</tr>
<tr>
<td>Age (SEM)</td>
<td>42.4 (1.1)</td>
<td>40.2 (0.46)</td>
<td>44.7 (0.19)</td>
<td>43.6 (0.15)</td>
</tr>
<tr>
<td>Gender (F% : M%)</td>
<td>75.7% : 24.3%</td>
<td>70.0% : 30.0%</td>
<td>69.7% : 30.5%</td>
<td>67.4% : 32.6%</td>
</tr>
<tr>
<td>IMD (SEM)</td>
<td>35.5 (1.3)</td>
<td>38.0 (0.53)</td>
<td>38.1 (0.17)</td>
<td>37.6 (0.13)</td>
</tr>
<tr>
<td>Total population size and differences between people with and without a CMHP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total population size</td>
<td>100% (n=166)</td>
<td>100% (n=1,118)</td>
<td>100% (n=53,318)</td>
<td>100% (n=121,199)</td>
</tr>
<tr>
<td>Age difference (Significance t-test)</td>
<td>-6.9 years (p&lt;0.05)</td>
<td>None (n.s.)</td>
<td>-1.8 years (p&lt;0.001)</td>
<td>+ 1.5 years (p&lt;0.001)</td>
</tr>
<tr>
<td>Gender differences (Pearson Chi-square)</td>
<td>14.60% (p&lt;0.001)</td>
<td>14.20% (p&lt;0.001)</td>
<td>14.70% (p&lt;0.001)</td>
<td>19.70% (p&lt;0.001)</td>
</tr>
<tr>
<td>IMD difference (Significance t-test)</td>
<td>-1.3 (n.s)</td>
<td>-2.4 (p&lt;0.05)</td>
<td>0.5 (p&lt;0.01)</td>
<td>-0.1 (n.s.)</td>
</tr>
</tbody>
</table>
Graph 1: Comparing the 5-year age bands of people in the in-depth evaluation, referred to IAPT and who use hospital services
Graph 2: Proportion of each ethnic group in four study populations (In-depth evaluation, Referred to IAPT, Use hospital services, Registered with a study practice)
Graph 3: Distribution of study populations by index of multiple deprivation (IMD) decile

- GP list population
- Patients of Hosp services
- Referred to IAPT
- In-depth study cohort
References: