Full title: Using data linkage to electronic patient records to assess the validity of selected mental health diagnoses in English Hospital Episode Statistics (HES)

Short title: The validity of mental health diagnoses in Hospital Episode Statistics

Katrina A.S. Davis¹,², Oliver Bashford²,³, Amelia Jewell², Hitesh Shetty², Robert J Stewart¹,², Cathie L.M. Sudlow⁴, Matthew Hotopf¹,² *

¹ King's College London, Department of Psychological Medicine, Institute of Psychiatry Psychology and Neuroscience, London, UK

² South London and Maudsley NHS Foundation Trust, National Institute for Health Research Biomedical Research Centre, De Crespigny Park, Denmark Hill, London, UK

³ Surrey and Borders Partnership NHS Foundation Trust, Surrey, UK

⁴ UK Biobank and Centre for Clinical Brain Sciences, University of Edinburgh, UK

* Corresponding author: Matthew Hotopf, Department of Psychological Medicine PO62, Weston Education Centre, Cutcombe Road, London SE5 9RJ, UK. (Email: matthew.hotopf@kcl.ac.uk)
Background: Administrative data can be used to support research, such as in the UK Biobank. Hospital Episode Statistics (HES) are national data for England that include ICD-10 diagnoses for inpatient mental healthcare episodes, but the validity of these diagnoses for research purposes has not been assessed.

Methods: 250 people's HES records were selected based on a HES recorded inpatient stay at the South London and Maudsley NHS Foundation Trust with a diagnosis of schizophrenia, a wider schizophrenia spectrum disorder, bipolar affective disorder or unipolar depression. A gold-standard research diagnosis was made using Clinical Records Interactive Search pseudonymised electronic patient records using, and the OPCRIT+ algorithm.

Results: Positive predictive value at the level of lifetime psychiatric disorder was 100%, and at the level of lifetime diagnosis in the four categories of schizophrenia, wider schizophrenia spectrum, bipolar or unipolar depression was 73% (68-79). Agreement varied by diagnosis, with schizophrenia having the highest PPV at 90% (80-96). Each person had an average of five psychiatric HES records. An algorithm that looked at the last recorded psychiatric diagnosis led to greatest overall agreement with the research diagnosis.

Discussion: For people who have a HES record from a psychiatric admission with a diagnosis of schizophrenia spectrum disorder, bipolar affective disorder or unipolar depression, HES records appear to be a good indicator of a mental disorder, and can provide a diagnostic category with reasonable certainty. For these diagnoses, HES records can be an effective way of ascertaining psychiatric diagnosis.
Mental health research using data derived from clinical records and administrative data can be highly informative.[1-4] There are benefits of using records to enrich research cohorts with variables such as hospital admissions, diagnoses and medication use.[5] The use of administrative data to ascertain diagnoses is an efficient means to follow participants in cohort studies. However, in a recent systematic review on the validity of psychiatric diagnoses in administrative data we showed that there were large differences in the validity of diagnoses between data sources.[6] We concluded that researchers should conduct validation studies on the datasets they proposed using to guide interpretation.

We present a validation of English National Health Service (NHS) Hospital Episode Statistics' (HES) diagnostic codes for psychiatric admissions. HES is an administrative data resource which provides records of hospital admissions, outpatient and accident and emergency department visits for individuals receiving NHS hospital treatment in England. Similar systems are available in Scotland and Wales. HES data are widely used in research, including mental health research.[7-9] Mental health providers have contributed HES inpatient data since 1996. Figure 1 shows how HES inpatient data are assembled from full text patient records via a coded aggregate dataset that represents the activity per hospital, and then combined to capture all hospital activity in England.[10] NHS Digital manage access to HES, which includes publishing regular aggregate data and managing access to individual patient-level data, including linkages to research datasets.[11]

Figure 1: Inpatient Hospital Episode Statistics pathway from patient to researcher.

Audits of inpatient HES from general hospitals have demonstrated that diagnosis can be unreliable [12] but no audit of HES diagnosis has been conducted for mental health providers. The motivation for this validation study was primarily to inform research on mental health disorders conducted
using data from UK Biobank. This research resource recruited 500,000 people aged between 40-69 years in 2006-2010 from across the UK [13] who agreed to have their health followed through linkages to health-related records, which include HES. We aimed to investigate the accuracy and reliability of mental health diagnoses in HES inpatient records from a mental health provider, to produce an external validation for schizophrenia spectrum and affective disorders diagnoses. Since it has been shown that the choice of algorithm for extracting psychiatric diagnosis from administrative datasets can have substantial impact on the accuracy of the diagnoses derived [14-16] we aimed to provide a range of accuracy statistics for possible algorithms, to enable future researchers to evaluate and choose the algorithms most suited to their purposes.

/H1/ Materials and Methods /H1/

/H2/ Data Source /H2/

The South London and Maudsley (SLaM) NHS Foundation Trust provides comprehensive NHS mental health services for a defined geographic catchment of around 1.2 million residents in South London, along with a number of smaller specialist tertiary referral centres. SLaM introduced an electronic records system across all its services from 2006. The Clinical Record Interactive Search (CRIS) system and its associated governance structures were developed to allow approved researchers to interrogate fully de-identified electronic health records, and has been used to provide a SLaM case register.[17, 18] A CRIS records includes the entirety of the electronic patient record from the NHS Trust. This includes structured fields such as age and ethnicity, forms such as for care planning and any detentions under mental health law, free text such as clinical notes and clerkings, and attachments of correspondence such as letters to primary care physicians and onward referrals. Changes and additions in the patient record are updated to CRIS on a daily basis to maintain it as a contemporary source.
Linkage of the CRIS/SLaM register with other databases is managed through the SLaM Clinical Data Linkage Service (CDLS). Linkage to HES is carried out by NHS Digital using NHS numbers, which are unique patient identifiers. A record in the CRIS/SLaM register will have linked HES records that include admissions to SlaM, to other mental health providers in England, and to general hospitals.

Research using the CRIS system, based on SLaM electronic records and the associated linkages has the endorsement of a patient-led oversight group. CRIS/SLaM has approval for analysis as a source of secondary data from the Oxford Research Ethics Committee C (reference 06/H0606/71+5) with access to restricted to researchers holding an honorary or substantive contract with SLaM.[17]

/

We studied HES records generated by SLaM in 2008-2013 where the primary diagnosis was schizophrenia (F20), wider schizophrenia spectrum disorder (F21-29), bipolar affective disorder (F30-31) or unipolar depression (F32-33), and the patient was age 30 years or over – this is the youngest a UK Biobank participant could have been when HES began in 1996. The selection was independent of participation in UK Biobank. A dataset was produced with 100 cases with a HES diagnosis of schizophrenia, 100 with bipolar affective disorder, 100 with unipolar depression, and 50 with a wider schizophrenia spectrum disorder, with no replacement, such that each record belonged to a different person. For each of the 350 cases identified by a HES record from SLaM, all available HES records from any hospital were also extracted to a file, with their source, the primary diagnosis, and all secondary diagnoses. The researcher was able to access this file only after the validation procedure was complete, in order to extract all HES records with an F-chapter ICD-10 diagnosis, which would include admissions to SLaM hospitals, other mental health trusts, and general hospitals where a mental disorder had been included in the record.
Once the procedures below were set, it was decided that assessment of 250 cases (out of the 350) would be sufficient to draw conclusions on overall validity. These cases were chosen at random, and the clinical assessors had no access to the file of HES diagnoses, in order to ensure blinding to the HES diagnosis during the research diagnosis procedure. The main clinical assessor received only the CRIS ID for each of the cases, which enabled them to access the pseudonymised version of the entire electronic patient record generated by South London and Maudsley as described above. This record would include all of the contacts of the selected cases between the years of 2006, when electronic records began, and October 2015, when the validation procedure began.

/H2/ Validation Procedure /H2/

We used a comprehensive approach to extracting diagnostic information from the record, following the “gold standard” Longitudinal, Expert, All Data (LEAD) diagnostic system defined by Spitzer [19] and guidelines for reporting of validation studies for routine data.[20] This involved extracting longitudinal psychopathology and diagnostic information from the full notes, and then using the available information to determine the most likely clinical diagnosis as an ICD-10 code for the index event (the admission leading to the HES record) and as a hierarchical lifetime diagnosis. The hierarchy that we use aligns with most diagnostic manuals in prioritising schizophrenia-like disorders over affective disorders, on the basis that the former are life-long and pervasive, and may explain symptoms of illnesses further down the hierarchy.[16, 21] In our case we specified that schizophrenia had priority, with wider schizophrenia spectrum disorders next, followed by bipolar affective disorder and unipolar depression.

A psychiatrist assessor (KD) extracted data and used a semi-structured process to explore each patient’s entire CRIS/SLaM record, with a view to gain sufficient detail about the presentation of the patient to complete an operational criteria checklist. The checklist used was the enhanced OPCRIT (or OPCRIT+),[22] which is a structured clinical and research tool consisting of a form that enquires
about psychopathology and other diagnostic criteria in order to give an algorithmic guide to the likely diagnostic code in ICD and DSM coding manuals. The process involved first extracting data from structured fields then reading free-text records, in particular, the assessor paid careful attention to the period before and during the admission that generated the HES record, the first and last assessments available in the notes, standardised care planning forms (the Care Programme Approach (CPA) record), and where structured fields showed a change in the diagnosis. Finally, searches were run on the full-text record with probes to identify mention of the following: diagnostic terms (schizo*, bipolar, mania, depress*); symptom terms (manic, euphoric, hallucination, delusion*, voices, thought disorder, FTD (formal thought disorder)); and important comorbidities (alcohol, cannabis, personality). OPCRIT+ was then completed and run twice – once for the “current” (i.e. index) and again for “lifetime” diagnosis. OPCRIT+ can also guide the formation of a structured abstract of the case,[23] which was assembled for 80 cases. These abstracts avoided diagnostic terms to allow a second psychiatrist to assign diagnoses without knowledge of the opinion of the treating team, and also avoided mentioning ethnicity. An example of a structured abstract and OPCRIT+ output for a hypothetical case is provided in the appendix.

Both psychiatrist assessors used the OPCRIT+ output along with other details to make research-standard diagnoses of one primary disorder and unlimited secondary disorders for index and lifetime formulations. The primary diagnosis for the index admission was allocated as the diagnosis that was the main reason for admission at that time. The lifetime primary diagnosis was allocated by hierarchy of all diagnoses considered to be met at any point by an individual. The primary diagnosis was required to be a specific code or “no psychiatric disorder”, and could not be left blank or vague. If the assessor felt there was real ambiguity, the assigned diagnosis could be flagged as uncertain. Assessor KD’s formulations are used as the ‘gold standard’ research diagnoses presented in the results section.
All data were entered into MS Excel, which was used for statistical analyses, graphics and random number generation. Confidence intervals around proportions were estimated using Wilson’s method [24] at 95% confidence levels. For comparing agreement between sets of diagnoses the primary measure used was the positive predictive value (PPV) – the proportion of cases identified by HES considered to be true cases of that diagnosis according to the research diagnosis. To assess the validity of single episode diagnoses for indicating the true reason for admission and/or the most serious mental disorder diagnosis for each individual we compared the index HES (the record by which the case was selected into the research) against the index and lifetime research diagnoses, and at multiple levels of detail in the diagnosis – from presence of any mental disorder, through broad diagnostic groups, to exact agreement, as shown in Figure 2.

Figure 2: Hierarchy diagram showing the diagnoses considered in this study, with different levels of detail.

Secondly, we test possible algorithms for extracting diagnosis or selecting cohorts from datasets where multiple HES records exist per person, and report the performance in terms of sensitivity, specificity, PPV, negative predictive value (NPV) and Cohen’s kappa, so that future researchers can select the algorithm that is optimised for their purpose (eg maximum sensitivity or highest NPV). These algorithms are largely based upon the work for Sara et al. [16]:

- "Ever": Any inpatient HES record (from mental health or not) with ICD-10 diagnosis in that diagnostic category – some individuals will have multiple diagnoses
- "More than once": Two or more HES records with a diagnosis in the range – some individuals will have multiple diagnoses
- "Last": Allocated the most recent F-code HES diagnosis, excluding F99 (a non-specific code)
• "Hierarchy": Allocated the diagnostic category received highest in the sequence above (F20 > F21-29 > F30-31 > F32-33)
• "Hierarchy > 1": Allocated the diagnostic category highest in the sequence (F20 > F21-29 > F30-31 > F32-33) of diagnoses received more than once
• "Most": Allocated the category of diagnosis that they have received most often. Where there is tie, the hierarchy rule is used

Kappa was included as a measure for overall agreement that reduces the effect of the prevalence of the disorder under study,[25] and 95% confidence intervals were calculated using the RealStatistics plug-in.[26]. For these analyses, we considered three of the individual disorder categories - schizophrenia, bipolar and depression – and the compound groups schizophrenia spectrum (schizophrenia plus wSS) and “severe mental illness” (schizophrenia spectrum plus bipolar). Schizophrenia spectrum was included instead of wSS due to the poor predictive value of wSS found when looking at the index HES.

All statistics reflect the performance of the HES diagnosis in this sample, which is people who have had an admission to a mental health provider – i.e. PPV is the proportion of people of those who had an admission to a mental health provider that resulted in a HES diagnosis of ‘condition x’ who truly had ‘condition x’; sensitivity is proportion of people with an admission to a mental health provider due to ‘condition x’ who received a HES diagnosis of ‘condition x’.

/H2/ Data Availability /H2/

All aggregate data used to draw the conclusions in this paper are shown. Patient level data is restricted for patient privacy in accordance with CRIS governance procedures,[17, 27] informed by NHS Digital requirements where linkage data is used. Access would usually be subject to gaining an honorary contract with the NHS Trust. To enquire about accessing patient-level data, please contact the authors and/or the CRIS administrators for SLaM (cris.administrator@kcl.ac.uk).
Results

Of the 250 individual cases, one was no longer in the database and three were transferred out of trust before discharge, for all of whom there was insufficient detail on which to make a diagnosis. Four cases had no SLaM admission corresponding to the HES record, but in all of these cases there were previous admissions upon which to base a lifetime research diagnosis. This meant there were 242 assessor-made “index” diagnoses and 246 “lifetime” diagnoses, as shown in figure 3. Of the 242 index research diagnoses, 119 (49%) were marked as uncertain. In one case, there was uncertainty at the border with normality (i.e. whether the person had a mental disorder or not); in all other cases, the uncertainty concerned the boundary between disorders.

Figure 3: Flow chart of cases used to derive gold-standard diagnoses

Schizophrenia (ICD-10 F20); wSS = wider schizophrenia spectrum disorder (ICD-10 F21-29); BPAD = bipolar affective disorder (ICD-10 F30-31); depression = unipolar depression (ICD-10 F32-33)

Table 1 shows the patient characteristics for the 249 cases. There were some differences between the groups, especially between the group with a HES schizophrenia diagnosis compared with a depression diagnosis on the items of race and duration of contact with services. Characteristics of those with wider schizophrenia spectrum (wSS) and bipolar diagnoses were between those of schizophrenia and depression. There is potential for some of these factors to influence the validity of the clinical (HES) diagnosis, meaning they may affect comparisons of validity between different diagnoses.

Table 1: Patient characteristics based on diagnosis in HES record by which they were selected.

<table>
<thead>
<tr>
<th>Index HES record diagnosis:</th>
<th>F20 Schizophrenia</th>
<th>F21-29 wSS</th>
<th>F30-31 Bipolar affective disorder</th>
<th>F32-33 Unipolar Depression</th>
<th>Total</th>
</tr>
</thead>
</table>


Table 2 shows agreement between the index HES diagnosis and the primary research diagnoses. The strictest comparison is with index research diagnoses at three figures of the ICD-10 code (see figure 2), where only 21% of records agreed. However, at the level of diagnostic group (i.e. schizophrenia, wSS, bipolar and unipolar depression) there was a 66% agreement in index diagnosis. Lifetime research diagnostic group agreed with index HES in 73% of cases. Agreement was highest for schizophrenia diagnosis, and lowest for wSS. Merging the schizophrenia and wSS category to make a schizophrenia spectrum category gave agreement of 86% (95%CI: 77-92). The proportion of cases for which the assessor was uncertain was highest for wSS and lowest for schizophrenia.

**Table 2: Agreement rates for index HES diagnosis and research primary diagnosis (index and lifetime) by index HES diagnosis.** Index (strict) refers to exact diagnosis precipitating admission (3 figure ICD-10 code), Index (category) refers to diagnostic group precipitating admission, Lifetime (category) refers to diagnostic group lifetime occurrence. “Uncertain” refers to the research assessor marking the lifetime diagnosis as being uncertain.
The diagnoses in this study can be considered to lie on a psychosis-affective spectrum from F20 schizophrenia, through F21-29 wSS and F30-31 bipolar affective disorder, to F32-33 unipolar depression. The most common disagreement between index HES and research diagnoses was a HES diagnosis of wSS and a research diagnosis of a type of schizophrenia. The documentation of a case of schizophrenia as wSS represents a shift in the administrative record away from the psychosis end of the spectrum. Considering all discordant diagnoses (both index and lifetime) 91/145 (63%) represented a shift in the HES coded record away from psychosis diagnoses relative to the research diagnosis and 23/145 (16%) represented a shift towards psychosis. In thirty cases, the primary research diagnosis was found to be outside the diagnoses of reference, with 22/145 (16%) diagnoses of functional disorders (ICD-10 F34-F69), and 8/145 (6%) organic, neurodevelopmental or personality disorder diagnoses.

Regarding secondary diagnoses, only 15 (6%) index HES records had a secondary diagnosis. The research diagnoses had least one, often several, secondary diagnosis in 113 cases (47%), including substance use disorder in 70 (28%), a further functional diagnosis (including personality disorder) in 40 (16%), and an organic or neurodevelopmental disorder in 25 (10%). Table 3 shows that the prevalence of secondary diagnoses was fairly even across different primary diagnoses.
Among the eighty cases reviewed by two assessors for index and lifetime diagnoses (160 formulations), the agreement between the two raters for primary diagnosis at the category level was 81%, with kappa 0.75 (95%CI 0.61-0.84). The degree of agreement of the assessor diagnosis with the HES primary diagnosis at the level of diagnostic group for those cases that were double reviewed was virtually identical (assessor 1: 69% kappa 0.57, assessor 2: 68% kappa 0.58). As table 4 shows, agreement between assessors was fairly even over the different primary diagnoses, while agreement of the two assessors with HES diagnosis showed a similar pattern of agreement between diagnostic groups.

Table 4 Inter-rater reliability for primary diagnoses of index and lifetime formulations in 80 cases (160 formulations)

<table>
<thead>
<tr>
<th>Primary diagnosis (categories)</th>
<th>F20 Schizophrenia</th>
<th>F21-29 wSS</th>
<th>F30-31 Bipolar disorder</th>
<th>F32-33 Unipolar Depression</th>
<th>Fx Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessor 1 diagnoses (index and lifetime)</td>
<td>63</td>
<td>12</td>
<td>38</td>
<td>27</td>
<td>20</td>
<td>160</td>
</tr>
<tr>
<td>Assessor 1 agreement (index and lifetime)</td>
<td>51/63 81%</td>
<td>10/12 83%</td>
<td>33/38 87%</td>
<td>22/27 81%</td>
<td>14/20 70%</td>
<td>130/160 81%</td>
</tr>
<tr>
<td>HES index diagnoses</td>
<td>58</td>
<td>10</td>
<td>48</td>
<td>44</td>
<td>Na</td>
<td>160</td>
</tr>
<tr>
<td>Assessor 1 agreement (index and lifetime)</td>
<td>49/58 84%</td>
<td>2/10 20%</td>
<td>36/48 75%</td>
<td>23/44 52%</td>
<td>Na</td>
<td>110/160 69%</td>
</tr>
<tr>
<td>Assessor 2 agreement (index and lifetime)</td>
<td>46/58 79%</td>
<td>4/10 40%</td>
<td>33/48 69%</td>
<td>26/44 59%</td>
<td>Na</td>
<td>109/160 68%</td>
</tr>
</tbody>
</table>

Abbreviations: wSS = wider schizophrenia spectrum disorder, Na = not applicable
For the 250 individuals in the sample, all inpatient HES records were examined. There were 1015 HES records from SLaM (including the 250 index), 99 HES records (in 48 cases) from other mental health trusts and 139 HES records (in 72 cases) from general hospitals that mentioned an ICD-10 F-code diagnosis in primary or secondary diagnosis, a mean of 5.0 HES records per person. Sixty-nine (28%) people had HES records indicating a diagnosis from more than one diagnostic category. We explored algorithms that could be used with multiple records, as defined in methods.

The performance of the algorithms by diagnostic category are shown in table 5. No single algorithm clearly performed best. Sensitivity was highest with the "Ever" algorithm, up to 93% for severe mental illness. Specificity for severe mental illness was highest in the "More than one" algorithm, at 83%, but the Hierarchy algorithms were better for specificity in depression. Agreement as assessed by kappa was maximised overall using the “last” algorithm, at kappa between 0.67 and 0.74.

Table 5: Accuracy statistics for algorithms to predict lifetime research diagnosis from multiple HES records, based on review of 246 cases. Numbers of cases given for each diagnostic group predicted by stated algorithm (explained in methods). PPV = positive predictive value, NPV = negative predictive value

<table>
<thead>
<tr>
<th>Table 5</th>
<th>F20 Schizophrenia</th>
<th>F30-31 Bipolar Disorder</th>
<th>F32-33 Unipolar Depression</th>
<th>F20-29 Schizophrenia Spectrum</th>
<th>F20-31 Severe Mental Illness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifetime research primary diagnosis (reference)</td>
<td>86</td>
<td>51</td>
<td>49</td>
<td>117</td>
<td>168</td>
</tr>
<tr>
<td>Ever</td>
<td>96</td>
<td>79</td>
<td>75</td>
<td>139</td>
<td>188</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>78% (69-86)</td>
<td>91% (81-96)</td>
<td>84% (72-92)</td>
<td>90% (84-94)</td>
<td>93% (88-96)</td>
</tr>
<tr>
<td>Specificity</td>
<td>84% (77-89)</td>
<td>85% (80-90)</td>
<td>85% (79-89)</td>
<td>77% (69-84)</td>
<td>69% (57-79)</td>
</tr>
<tr>
<td>PPV</td>
<td>74% (64-82)</td>
<td>65% (54-75)</td>
<td>61% (50-71)</td>
<td>80% (73-86)</td>
<td>89% (84-93)</td>
</tr>
<tr>
<td>NPV</td>
<td>87% (81-91)</td>
<td>97% (93-99)</td>
<td>95% (91-97)</td>
<td>89% (82-94)</td>
<td>78% (66-87)</td>
</tr>
<tr>
<td>Agree</td>
<td>82% (77-86)</td>
<td>87% (82-90)</td>
<td>85% (80-89)</td>
<td>84% (79-88)</td>
<td>86% (82-90)</td>
</tr>
<tr>
<td>Cohen’s kappa</td>
<td>0.61</td>
<td>0.67</td>
<td>0.60</td>
<td>0.68</td>
<td>0.65</td>
</tr>
<tr>
<td>More than one</td>
<td>73</td>
<td>59</td>
<td>42</td>
<td>95</td>
<td>147</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>PPV</td>
<td>NPV</td>
<td>Agree</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>66% (56-75)</td>
<td>77% (64-86)</td>
<td>44% (32-57)</td>
<td>66% (57-73)</td>
<td>76% (69-81)</td>
</tr>
<tr>
<td></td>
<td>92% (86-95)</td>
<td>92% (87-95)</td>
<td>91% (86-94)</td>
<td>88% (82-93)</td>
<td>83% (72-90)</td>
</tr>
<tr>
<td></td>
<td>82% (72-89)</td>
<td>74% (62-83)</td>
<td>57% (42-71)</td>
<td>85% (76-91)</td>
<td>92% (86-95)</td>
</tr>
<tr>
<td></td>
<td>82% (76-87)</td>
<td>93% (88-96)</td>
<td>85% (79-89)</td>
<td>72% (64-79)</td>
<td>56% (46-65)</td>
</tr>
<tr>
<td></td>
<td>82% (77-86)</td>
<td>88% (84-92)</td>
<td>80% (75-85)</td>
<td>77% (71-82)</td>
<td>78% (72-82)</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.68</td>
<td>0.38</td>
<td>0.54</td>
<td>0.51</td>
</tr>
<tr>
<td>Last</td>
<td>71</td>
<td>66</td>
<td>64</td>
<td>111</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>71% (61-79)</td>
<td>84% (72-91)</td>
<td>83% (71-91)</td>
<td>82% (75-88)</td>
<td>91% (86-95)</td>
</tr>
<tr>
<td></td>
<td>96% (91-98)</td>
<td>90% (85-94)</td>
<td>90% (85-93)</td>
<td>92% (86-96)</td>
<td>81% (70-89)</td>
</tr>
<tr>
<td></td>
<td>91% (82-96)</td>
<td>72% (60-81)</td>
<td>70% (58-80)</td>
<td>91% (84-95)</td>
<td>93% (88-96)</td>
</tr>
<tr>
<td></td>
<td>85% (79-90)</td>
<td>95% (91-97)</td>
<td>95% (91-97)</td>
<td>84% (77-89)</td>
<td>77% (66-85)</td>
</tr>
<tr>
<td></td>
<td>87% (82-90)</td>
<td>89% (84-92)</td>
<td>88% (84-92)</td>
<td>87% (82-91)</td>
<td>89% (84-92)</td>
</tr>
<tr>
<td></td>
<td>0.71</td>
<td>0.70</td>
<td>0.67</td>
<td>0.74</td>
<td>0.71</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>96</td>
<td>49</td>
<td>53</td>
<td>139</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>78% (69-86)</td>
<td>73% (60-82)</td>
<td>71% (58-82)</td>
<td>90% (84-94)</td>
<td>93% (88-96)</td>
</tr>
<tr>
<td></td>
<td>84% (77-89)</td>
<td>96% (92-98)</td>
<td>93% (88-95)</td>
<td>77% (69-84)</td>
<td>69% (57-79)</td>
</tr>
<tr>
<td></td>
<td>74% (64-82)</td>
<td>85% (72-92)</td>
<td>73% (60-83)</td>
<td>80% (73-86)</td>
<td>89% (84-93)</td>
</tr>
<tr>
<td></td>
<td>87% (81-91)</td>
<td>92% (87-95)</td>
<td>92% (87-95)</td>
<td>89% (82-94)</td>
<td>78% (66-87)</td>
</tr>
<tr>
<td></td>
<td>82% (77-86)</td>
<td>91% (86-94)</td>
<td>88% (83-91)</td>
<td>84% (79-88)</td>
<td>86% (82-90)</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.73</td>
<td>0.64</td>
<td>0.68</td>
<td>0.65</td>
</tr>
<tr>
<td>Hierarchy &gt;1</td>
<td>73</td>
<td>53</td>
<td>30</td>
<td>92</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>66% (56-75)</td>
<td>73% (60-83)</td>
<td>40% (28-54)</td>
<td>64% (55-72)</td>
<td>75% (68-80)</td>
</tr>
<tr>
<td></td>
<td>92% (86-95)</td>
<td>94% (90-97)</td>
<td>96% (92-98)</td>
<td>89% (83-94)</td>
<td>83% (72-90)</td>
</tr>
<tr>
<td></td>
<td>82% (72-89)</td>
<td>79% (66-88)</td>
<td>73% (55-86)</td>
<td>86% (77-92)</td>
<td>92% (86-95)</td>
</tr>
<tr>
<td></td>
<td>82% (76-87)</td>
<td>92% (87-95)</td>
<td>85% (80-89)</td>
<td>71% (63-78)</td>
<td>55% (45-64)</td>
</tr>
<tr>
<td></td>
<td>82% (77-86)</td>
<td>89% (85-92)</td>
<td>84% (78-88)</td>
<td>77% (71-81)</td>
<td>77% (71-82)</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.69</td>
<td>0.43</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>Most (hier)</td>
<td>78</td>
<td>67</td>
<td>58</td>
<td>116</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td>72% (62-80)</td>
<td>87% (76-94)</td>
<td>76% (63-85)</td>
<td>81% (73-87)</td>
<td>92% (87-95)</td>
</tr>
<tr>
<td></td>
<td>92% (86-95)</td>
<td>91% (86-94)</td>
<td>92% (87-95)</td>
<td>88% (81-92)</td>
<td>75% (63-84)</td>
</tr>
<tr>
<td></td>
<td>83% (73-90)</td>
<td>74% (62-83)</td>
<td>72% (59-82)</td>
<td>87% (80-92)</td>
<td>91% (86-94)</td>
</tr>
<tr>
<td></td>
<td>85% (79-90)</td>
<td>96% (92-98)</td>
<td>93% (88-96)</td>
<td>82% (75-88)</td>
<td>77% (65-86)</td>
</tr>
<tr>
<td></td>
<td>84% (79-88)</td>
<td>90% (86-93)</td>
<td>88% (83-92)</td>
<td>84% (79-88)</td>
<td>87% (83-91)</td>
</tr>
</tbody>
</table>
We set out to determine the accuracy of certain psychiatric diagnoses as recorded in national statistics (Hospital Episode Statistics or HES) of episodes of inpatient mental healthcare. The SLaM/CRIS database of cases with electronic patient notes and associated data linkage allowed us to validate HES against a research diagnosis without needing to recontact patients. An analysis of 246 cases diagnosed with schizophrenia, a wider schizophrenia spectrum disorder, bipolar affective disorder or unipolar depression showed a perfect agreement with the presence of any mental disorder and good agreement with the presence of the stated diagnostic group of disorder (73% 68-79). When considering multiple HES records with mental disorder diagnoses, a good overall approach was to take the most recent, which showed the positive predictive value of a diagnosis was 91% for schizophrenia spectrum, 72% for bipolar affective disorder and 70% for unipolar depression. This puts the accuracy of HES records to identify the presence of a mental disorder in the range of some “objective” biomedical tests and discharge diagnoses of physical illness from general hospitals.[28] This implies that HES and similar records can be used with some confidence to indicate the likely presence of these three categories of mental disorder serious enough to require hospitalisation.

Inter-rater reliability is strong (at kappa 0.75), but considering the inter-rater reliability involved one psychiatrist researcher providing a second opinion based on the facts of the case put forward by another psychiatrist of similar background, this is lower than might be expected. There was also a high level of uncertainty in the research diagnoses (49%). This leads to a sense that some the differences between clinical and research diagnoses reflect not 'error' by the clinician or in the administrative record, but uncertainty in the true diagnosis. It is widely acknowledged that current mental disorder classifications are an imperfect representation of human psychopathology.[21, 29,
Out of the three main diagnoses, unipolar depression was the most uncertain, and this may be because the types of ‘depression’ that present to inpatient mental healthcare are atypical compared to the standard presentation in the community,[31] and allied with a high level of comorbidity in our sample. For bipolar affective disorder, the uncertainty seemed to be with alternative diagnoses on the schizophrenia spectrum, although there were some cases with complex personality traits and other morbidities. Schizophrenia diagnoses were accurate and stable, which was not the case with diagnoses in the wider spectrum. wSS diagnoses were often assigned to more complex cases of psychosis after some time in the service, but the validation process showed that in many cases criteria were in fact met for schizophrenia. This leads to the concept of “schizophrenia spectrum” performing well, and we would not recommend using HES records where the distinction between schizophrenia and the other schizophrenia spectrum disorders was required.

Secondary diagnoses were generally not documented in HES when present. We suspect this reflects a wide-spread tendency not to document of code for secondary diagnoses in mental health inpatients. Great caution is thus required using HES or other administrative data in evaluating mental disorders which are more likely to be coded as secondary or ‘co-morbid’ – such as substance misuse, learning disability and personality disorder – as the data is likely to be unrepresentative. Specific registers may be more appropriate.[32]

Our recent review of the validity of administrative diagnoses found that there was a wide range of validity between the studies.[6] Aggregating the studies showed that the diagnoses of schizophrenia spectrum, bipolar affective disorder and unipolar depression performed similarly on PPV, with a median of 75% and a wide spread of results. Schizophrenia spectrum as a category of diagnosis performed better than schizophrenia and schizoaffective disorders separately (a large proportion of the wSS cases in this study were schizoaffective diagnoses). Our results are entirely consistent with the results of the review. With the studies ranked from lowest to highest PPV, this study is within the
25-50\textsuperscript{th} centile for bipolar affective disorder and unipolar depression diagnoses and within the 50-75\textsuperscript{th} centile for schizophrenia and schizophrenia spectrum. Our overall PPV of 73\% (CI 95\% 68-79) can be compared to the 13 results from studies from the review carried out using inpatient diagnoses, in which the average PPV was 77\%, which is not significantly different.

\textit{Strengths and weaknesses}\n
Our study used a set of electronic patient records to explore patient histories in depth with patient anonymity protected through the CRIS system. We did not interview the cases under study, but it has been shown to be possible to make gold-standard psychiatric diagnoses without re-interviewing patients, where the procedure is consistent with LEAD diagnosis.\cite{19, 33, 34} We explored index HES diagnoses from a defined service and geographic catchment in South-East London that is considered to be a centre of excellence in psychiatry, which could lead to concerns over generalisability. However, HES records came from numerous wards across four locality hospitals, providing care in a highly pressurised healthcare system for patients from various settings, including urban areas of high deprivation; therefore not untypical of mental health services elsewhere in the UK. A drawback of concentrating on a mental health trust was that we were unable to look in detail at diagnoses made in the general hospital, outpatient departments or A&E departments – which all go in to making up the entirety of HES output.

We studied three of the most prominent diagnoses for general psychiatry - schizophrenia, bipolar affective disorder and unipolar depression - covering also the whole range of what is sometimes termed “severe mental illness” by our inclusion of wider schizophrenia spectrum. However, we did not study HES diagnoses important in some psychiatric specialties, such as eating disorder and dementia, and so cannot inform questions of validity for these. We have provided a variety of outcomes (PPV, sensitivity, kappa, etc.) for a variety of algorithms to help future studies choose
algorithms that are well suited to their objectives. However, the results are based on one assessor, and the inter-rater reliability showed there was some variation between assessors.

/H2/ Conclusions /H2/
Administrative health data are being used in studies such as UK Biobank to collect information on health status without the need for face to face reassessment, recontact or (in some cases) reconsent of participants. Hospital Episode Statistics (HES) in England contain ICD-10 diagnostic information, including regarding psychiatric admissions. In our study, a clinical psychiatrist assessor agreed with the HES diagnosis 73% of the time, with level of agreement varying by diagnosis. All were felt likely to have a disorder in the F chapter (mental and behavioural disorders) of ICD-10. It should be remembered that administrative data will always under-represent those who do not, or cannot, access services. Even in highly developed countries, access to services for those with mental disorder is low, at around a third.[35, 36] Our study shows that HES inpatient psychiatric diagnoses can be used, with appropriate caution, to identify cases of severe mental disorder and distinguish between some common categories of diagnosis.

Acknowledgements:
This work was funded by a grant from UK Biobank
Cathie Sudlow is supported by UK Biobank and the Scottish Funding Council
This work was part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the UK Biobank, Scottish Funding Council, NHS, NIHR or Department of Health.
Thank you to other members of the BRC who responded to queries and contributed to discussion regarding this work.
Author Contributions: Conceptualisation by CS and MH, methodology by KD, RS, CS and MH, supervision by MH. Data curation by AJ and HS, with analysis by KD and validation by OB. Writing – original draft preparation by KD, review and editing by all authors.

Supporting information

Appendix: OPCRIT+ Output for hypothetical case (abstract and OPCRIT results)

7. RSA Open Public Services Network. Exploring how available NHS data can be used to show the inequality gap in mental healthcare. 2015.
Figure 1

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Description</th>
</tr>
</thead>
</table>
| Patient notes                                   | - Mainly unstructured data  
- Written by clinicians, increasingly electronic |
| Patient Administration System (PAS)             | - Mainly structured data  
- Entered by clerical and coding staff from information in notes |
| Hospital Episode Record                         | - 50 structured fields, including main and secondary diagnoses  
- One produced per patient per episode of care |
| Commissioning Data Set                          | - Aggregate set of hospital episode records  
- Submitted by the hospitals to commissioners |
<p>| Hospital Episode Statistics Database            | - Department of Health aggregates data sets for all hospitals in England, over 12 million records per year |
| Hospital Episode Statistics (HES)               | - NHS Digital publishes HES aggregate data and provides individual level HES data for approved scientific purposes |</p>
<table>
<thead>
<tr>
<th>High level diagnosis</th>
<th>Broad diagnostic group</th>
<th>Exact diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F0-99</strong></td>
<td><strong>F2-</strong></td>
<td></td>
</tr>
<tr>
<td>Mental Disorders</td>
<td><strong>Schizophrenia</strong></td>
<td><strong>F20: Schizophrenia</strong>*</td>
</tr>
<tr>
<td>(* = “severe mental illness”)</td>
<td><strong>Spectrum disorders</strong></td>
<td><strong>F20.0 Paranoid schizophrenia</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>F20.1 Hebephrenic schizophrenia</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>... F20.9 Schizophrenia, unsp.</strong></td>
</tr>
<tr>
<td></td>
<td><strong>F21-29: Wider</strong></td>
<td><strong>F21 Schizotypal disorder</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Schizophrenia</strong></td>
<td><strong>F22.0-9 Persistent delusional disorders</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Spectrum disorders</strong>*</td>
<td><strong>... F29 Unspecified nonorganic psychosis</strong></td>
</tr>
<tr>
<td><strong>F3-</strong></td>
<td><strong>F30-31: Bipolar</strong></td>
<td><strong>F30.0-9 Manic episode (by severity)</strong></td>
</tr>
<tr>
<td>Affective disorders</td>
<td><strong>Affective Disorder</strong>*</td>
<td><strong>F31.0-9 Bipolar affective disorder (by episode type/severity), eg F31.6 Current episode mixed</strong></td>
</tr>
<tr>
<td></td>
<td><strong>F32-33: Unipolar</strong></td>
<td><strong>F32.0-9 Depressive episode (by severity)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Depression</strong></td>
<td><strong>F33.0-9 Recurrent depressive disorder (by episode severity), eg F33.0 Current episode mild</strong></td>
</tr>
</tbody>
</table>
Figure 3

Pool of SLaM cases extracted via CRIS with linkage to HES records 2008-2013

Case selection based on HES diagnosis:
100 schizophrenia +
50 wSS +
100 BPAD +
100 depression

100 random cases not selected for further examination

250 cases chosen at random

1 case, record no longer in CRIS
3 cases, transferred out of trust during admission and information inadequate

246 gold-standard "lifetime" diagnoses

4 cases, no SLaM admission corresponding to index HES record, hence no research index diagnosis

242 gold-standard "index" diagnoses
Supporting information

Appendix S1: OPCRIT+ output for hypothetical case

Abstract:

DEMOGRAPHICS AND PRESENTING COMPLAINT

64 year old married man presents walking down the road semi-naked claiming to be sent from God.

Currently self-employed and living in owner occupied accommodation.

Strange behaviour in last week. Previous episodes of similar behaviour, but not for many years. Recently had complained of low mood, poor energy, poor sleep, poor concentration.

FAMILY HISTORY

There is a family history of mental illness, specifically his father had

- Major depressive disorder
- Drug or alcohol dependence

died of alcohol-related disease when pt was 25.

There is no known family history of physical illnesses.

PERSONAL HISTORY


Worked in a bank, but was recently made redundant, and now trying to set up business from home.

Practicing Christian evangelical.

One year ago had stroke and seriously ill in hospital. Made good physical recovery, but continued to be troubled by fatigue.

PAST PSYCHIATRIC HISTORY

Six previous admissions with elated, disinhibited behaviour and religious delusions, usually under section - responding quickly to treatment. Note religious at baseline, but different quality. Triggers have been problems at work and poor sleep. Two previous suicide attempts (poisoning). No episodes in last ten years, discharged from services. Stopped maintenance medications after stroke.
PAST MEDICAL HISTORY

Past medical history of-

- Hypertension
- Diabetes (type 2, tablet controlled)
- Hyperlipidaemia
- Cerebrovascular disease

Ischaemic CVA twelve months before this episode.

DRUG AND ALCOHOL HISTORY

Ex-cigarette smoker. 15 pack-years.

History of harmful use of-

- Alcohol
- Cannabis

ALCOHOL HISTORY-

Age of onset of regular alcohol use was 15.

Has not drunk for many years. Previous misuse, but no dependence.

CANNABIS HISTORY-

Age of onset of regular Cannabis use was 15. On an average week uses approximately £0-£9 worth of cannabis.

Smokes 1-2 spliffs 'herbal' cannabis per week - finds it relaxing. Increased in the week prior to presentation - exact amount.

OPIATE HISTORY-

No history of harmful misuse of opiates.

STIMULANT HISTORY-

No history of harmful misuse of stimulants.

MEDICATION HISTORY

Usually compliant. Had been without medication for 12 months prior to admission, as stopped when he had a stroke, as had been symptom free for 10 years.
No psychotropic medication is currently being prescribed. History of response to antipsychotics and mood stabilisers. History of 'switching' on antidepressants.

FORENSIC HISTORY
There is no known history of criminal or violent behaviour.
Contact with police only in the context of mental health.

SOCIAL HISTORY
Married with one child. Occupationally and socially capable between episodes.

PERSONALITY
There is an enduring pattern of inner experience and behaviour that deviates markedly from the expectations of the culture. This is manifested in...
...ways of perceiving and interpreting self, other people, and events
The pattern is inflexible and pervasive across a broad range of personal and social situations.
The onset of this pattern is in adolescence.
Tendency towards bizarre beliefs and experiences
Tendency towards preoccupation with perfection, order and routine

MENTAL STATE EXAMINATION

APPEARANCE AND BEHAVIOUR

The following abnormalities in appearance were noted:-
- gross lack of self care
- a mildly increased BMI

The following abnormalities in behaviour were noted:-
- Bizarre behaviour
- Excessive activity
- Reckless activity
- Distractibility
- Agitated activity

SPEECH AND FORM OF THOUGHT

The following abnormalities of speech and thought form were elicited:
- Positive formal thought disorder
- Pressured speech
- Thoughts racing

MOOD, AFFECT AND ASSOCIATED FEATURES

The following abnormalities of affect and mood were elicited:
- Inappropriate affect
- Irritable mood
- Dysphoria
- Poor concentration
- Increased sociability
- Increased self esteem

No complaints of suicidal ideation.

The following disturbances in sleep pattern were noted:
- Early morning waking

The following disturbances in appetite, eating behaviour and body image were noted:
- Increased appetite for 1 month or more.
- Mild weight gain recently.

ANXIETY, TRAUMA AND ASSOCIATED FEATURES

The following anxiety symptoms, were elicited, reaching time threshold for diagnostic significance:
- Autonomic arousal symptoms during times of anxiety
- Prominent, excessive, free-floating anxiety and suffers from... (always anxious. since stroke, worse)
- ...restlessness, feeling keyed-up or on-edge
- ...being easily fatigued
- ...poor concentration or 'mind going blank'
- ...irritability
- ...muscle tension
- ...sleep disturbance
- Obsessions and/or compulsions present... (cleanliness and cleaning rituals)
- ...that are repetitive and unpleasant
- ...that are resisted unsuccessfully by the subject
- ...that are acknowledged by the subject as originating in their own mind
- ...that are recognised as excessive or unreasonable

THOUGHT CONTENT
On this assessment affective symptoms predominate although psychotic symptom also sometimes occur.
Has been receiving messages through TV and other media that god wants him to go out and warn people about the end of the world, and cleanse them of their sin. Feels guilty that can't do this as well as he would like.
The following delusions, reaching time threshold for diagnostic significance, were elicited-
- Grandiose delusions
- Delusions of guilt

No passivity phenomena elicited on this assessment.

PERCEPTIONS
Hearing screaming and other noises that he thinks are damned souls.

COGNITION, INSIGHT AND CAPACITY
Cognition was not tested on this assessment.
Insight absent.
Capacity to consent to hospital admission was felt to be absent at this assessment.
Capacity to consent to medical treatment was felt to be impaired at this assessment.
Outcome

Responded to reinstatement of medication after around one week. Went home on leave, but felt overwhelmed and took an overdose. Returned back to the ward for another month. After discharge, felt able to continue work, and did this for a number of years before retiring. Discharged from services around the same time, with no contact in the last two years.

OPCRIT Results:

ICD10 criteria: F30.2 Mania with psychosis

ICD10 criteria: F12.1 Harmful use of cannabis

DSM3 criteria: Mania with psychosis

DSM3R criteria: Bipolar disorder

DSM4 criteria: Bipolar I disorder

Taylor & Abrams criteria: Mania

Research diagnostic criteria (RDC): Schizoaffective, bipolar

Carpenter criteria: Level 5 schizophrenia

Farmer criteria: Schizophrenia, hebephrenic-like subtype

Crow criteria: Schizophrenia, type 1 (positive symptoms)

Tsuang & Winokur criteria: Schizophrenia, undifferentiated

ICD10 criteria: F60.X personality disorder

Severe trait: Obsessive-compulsive (consider obsessive-compulsive (anankastic) PD)

For more information, see the OPCRIT+ webpages http://sgdp.iop.kcl.ac.uk/OPCRITplus/