Data quality and autism: Issues and potential impacts

Johannes Heyl [a,b], Flavien Hardy [a], Katie Tucker [c], Adrian Hopper [a,d], Maria J. Marchã [e], Ashley Liew [f,g,h,i], Judith Reep [d], Kerry-Anne Harwood [d], Luke Roberts [d], Jeremy Yates [e,j], Jamie Day [a], Andrew Wheeler [a], Sue Eve-Jones [a], Tim W.R. Briggs [a,k], William K. Gray [a,*]

[a] Getting It Right First Time, NHS England and NHS Improvement, London, UK
[b] Department of Physics and Astronomy, University College London, London, UK
[c] Innovation and Intelligent Automation Unit, Royal Free London NHS Foundation Trust, London, UK
[d] Guy's and St Thomas' NHS Foundation Trust, London, UK
[e] Science and Technology Facilities Council Distributed Research Utilising Advanced Computing High Performance Computing Facility, London, UK
[f] National & Specialist CAMHS, South London and Maudsley NHS Foundation Trust, London, UK
[g] Evelina London Children's Hospital, Guys and St Thomas, NHS Foundation Trust, London, UK
[h] Centre for Educational Development, Appraisal and Research (CEDAR), University of Warwick, Coventry, UK
[i] Institute for Mental Health, University of Birmingham, Birmingham, UK
[j] Department of Computer Science, University College London, London, UK
[k] Royal National Orthopaedic Hospital NHS Trust, London, UK

ARTICLE INFO

ABSTRACT

Introduction: Large healthcare datasets can provide insight that has the potential to improve outcomes for patients. However, it is important to understand the strengths and limitations of such datasets so that the insights they provide are accurate and useful. The aim of this study was to identify data inconsistencies within the Hospital Episodes Statistics (HES) dataset for autistic patients and assess potential biases introduced through these inconsistencies and their impact on patient outcomes. The study can only identify inconsistencies in recording of autism diagnosis and not whether the inclusion or exclusion of the autism diagnosis is the error.
Methods: Data were extracted from the HES database for the period 1st April 2013 to 31st March 2021 for patients with a diagnosis of autism. First spells in hospital during the study period were identified for each patient and these were linked to any subsequent spell in hospital for the same patient. Data inconsistencies were recorded where autism was not recorded as a diagnosis in a subsequent spell. Features associated with data inconsistencies were identified using a random forest classifiers and regression modelling.
Results: Data were available for 172,324 unique patients who had been recorded as having an autism diagnosis on first admission. In total, 43.7 % of subsequent spells were found to have inconsistencies. The features most strongly associated with inconsistencies included greater age, greater deprivation, longer time since the first spell, change in provider, shorter length of stay, being female and a change in the main specialty description. The random forest algorithm had an area under the receiver operating characteristic curve of 0.864 (95 % CI [0.862 – 0.866]) in predicting a data inconsistency. For patients who died in hospital, inconsistencies in their final spell were significantly associated with being 80 years and over, being female, greater deprivation and use of a palliative care code in the death spell.
Conclusions: Data inconsistencies in the HES database were relatively common in autistic patients and were associated a number of patient and hospital admission characteristics. Such inconsistencies have the potential to distort our understanding of service use in key demographic groups.

1. Introduction

Routinely collected, administrative healthcare data are increasingly being used to inform decisions taken by clinician and healthcare service manager.[1] Whilst the use of such data can be valuable, it is important to have a clear understanding of data quality and the strengths and

limitations of any dataset prior to analysis.

In England, the Getting It Right First Time (GIRFT) programme is an NHS England and NHS Improvement initiative with a remit to reduce unwarranted variation in clinical practice that negatively impacts on patient outcomes. The GIRFT programme is one of the largest users of administrative healthcare data in England, with the Hospital Episodes Statistics (HES) dataset being a key resource. Although there is usually no 'gold standard' reference dataset against which to validate administrative datasets, such as HES, missing and internally inconsistent data can be identified. The issue of data inconsistency is particularly pertinent when considering people with life-long conditions, such as autism, who may have regular hospital attendances linked to their condition or where their condition may need to be considered in the delivery of care. Autism can impact on a person's ability to communicate health problems and to engage with support and advice offered.[2] The recording of autism is mandatory in the HES dataset; all episodes of admitted hospital care for autistic people should include a diagnostic code for autism.[3]

The UK government's proposed national strategy for autistic children, young people and adults[4] has recognised the gap in autism data and identifies that a key enabler for the strategy is the need to improve data collection and quality of data on autism across the public services to drive system improvement. NHS Digital's Assuring Transformation initiative has improved record keeping of inpatient admissions for autism, but this does not include autistic patients in beds to address their physical health care.[5].

This was an exploratory analysis that aimed to identify the extent and pattern of inconsistent data for autistic patients within the HES administrative dataset and explore whether any inconsistencies were linked to outcomes. Our a-priori hypothesis was that poor recording of autism in medical notes would suggest a lack of focus on autism as a complicating condition that may impact of patient care and outcomes. In light of the LeDeR programme, a specific interest was in hospital spells where a death had occurred.

2. Methods

2.1. Ethics

Consent from individuals involved in this study was not required for this analysis of the HES administrative dataset. The analysis and presentation of data follows current NHS Digital guidance for the use of HES data for research purposes. Reported data are anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data.[6]

2.2. Study design and data collection

This was a retrospective exploratory analysis of HES data. HES data are collected by NHS Digital for all NHS-funded patients admitted to hospitals in England. Hospital trusts run all NHS hospitals in England. A hospital trust is an administrative unit typically operating between one and four secondary or tertiary care hospitals for a geographically defined catchment population. Data collection and reporting is mandatory, and data are entered by clinical coders at each trust.

2.3. Timing, case ascertainment, inclusion and exclusion criteria

Data were taken from HES for all patient discharges during the period 1st April 2013 to 31st March 2021. This period was chosen to reflect a period where clinical coding practice in the NHS in England was relatively stable. The time period defined the sample size.

Once a diagnosis of autism has been made, autism should be recorded in HES for all episodes of hospital care for that patient.[7] Autistic patients were identified in HES where the International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10) codes F84.0 (childhood autism), F84.1 (atypical autism), F84.5

(Asperger's syndrome) was listed as a diagnosis in any position in the diagnostic record.

The first use of the specified code in any position of the diagnostic record for a hospital spell during the study period was identified and data for all subsequent spells for the same person extracted.

2.4. Identification of clinical coding inconsistencies

Clinical coding inconsistencies are reported at the spell level; a spell in hospital being made up of individual episodes of care. In HES a hospital spell is defined as a continuous period in hospital from admission to discharge. A spell can include multiple smaller episodes of care in various hospital settings and under different consultants. As an example, following an emergency department attendance, a patient may initially be under the care of acute medicine (episode one), then transferred to a critical care setting (episode two) and then to a care of the elderly ward (episode three) prior to discharge.

The data were extracted at an episode level and aggregated to a spell level for analysis. Therefore, a subsequent spell was considered to contain an inconsistency if none of its constituent episodes mentioned the ICD-10 codes of interest. An inconsistency was recorded where none of the three autism ICD-10 codes appeared in the diagnostic record of a subsequent spell. We recognised that the precise diagnostic code used may change for a patient, but the broad diagnosis of autism should be present during each spell. This approach was felt to be a fairer and more pragmatic approach than the stricter definition requiring all episodes in a spell to contained one of the three autism codes.

2.5. Data features (variables)

Patient characteristics: sex, age in years, ethnicity (White, Black or Black British, Asian or Asian British, Mixed, other), frailty (Hospital Frailty Risk Score HFRS[8], and deprivation (Index of Multiple Deprivation (IMD scores)[9]. All ICD-10 codes present in the diagnostic record for each subsequent spell were investigated for their potential to improve the model performance. In modelling, the feature age was the age at the subsequent admission where the inconsistency was recorded.

Features of hospital stay: Emergency readmission within 30 days, spell length of stay, hospital trust, admission method (emergency or elective), number of days since the first spell with the diagnosis was recorded (difference between the discharge date of the first spell and the admission date of the subsequent spell), clinical specialty of admission episode. All categorical features were one-hot encoded.

2.6. Outcome (target) variable

The target was described by a binary flag indicating whether a clinical coding inconsistency was recorded in the subsequent spell.

2.7. Data analysis

Data were extracted onto a secure encrypted server controlled by NHS England and NHS Improvement. Analysis within this secure environment took place using Alteryx 2019.3 (Alteryx Inc., Irvine, CA, USA), Python 3.9.6 and the scikit-learn machine learning library 0.24.2 (Python Software Foundation, Beaverton, OR, USA).[10]

All machine learning models were developed using the scikit-learn library. Random forest classifiers were used to identify key co-variates associated with data inconsistencies. Random forest classifiers are ensemble classifiers that fit decision trees to portions of the data and average over all decision trees. The advantage of using these classifiers as opposed to deep neural networks is the interpretability and transparency of the results, particularly feature importance. This is of particular importance if a machine learning model is to provide useful information about the relationship between the features and outcome variable to clinicians. Machine learning has shown significant benefit in

analysing healthcare data and providing insight.[11–14] The task considered in this work is a binary classification task.

The dataset was randomly split in the ratio 70:15:15 to create a training set, validation and test set, respectively. Machine learning algorithms require the data to be split into a training set, from which the algorithm learns the relationships between features, and a testing subset where it applies that learning to an unseen part of the dataset. The performance on this unseen test set is used to assess the generalisability of the model, to check for overfitting and evaluate how well the features have been learned. The machine learning algorithm was trained on the training set and its performance evaluated based on how well it could predict inconsistencies in the test set. The validation set was used to tune the hyperparameters of the random forest. To ensure that the model did not simply classify according to the majority outcome (no inconsistency), the training set was reduced by random under-sampling from the majority class to ensure that there were an equal number of patients with and without inconsistencies in the training set. This eliminated the effect of the class imbalance on the model performance and ensured that the model had sufficient exposure to patients in the minority class. Random (unstratified) under-sampling was felt appropriate given the exploratory nature of the analysis. However, the test set on which the trained model was evaluated was not balanced, increasing the model's external validity.

There exist hyperparameters specific to the random forest classifier that can be tuned. The hyperparameters were determined by performing a nested k-fold cross-validation based grid search. The combination of hyperparameters with the highest Area under the Receiver Operator Curve (AUROC) for the validation set were selected. The hyperparameters that were tuned and their values tested were: the number of trees (100, 200, 300), the minimum samples per split (2, 4, 8) and the maximum depth (10, 100, 1000). The final hyperparameter values used in the model were the number of trees (300), the minimum samples per split (4) and the maximum depth ("100"). Data splitting was performed before any pre-processing steps.

Model performance was reported in terms of accuracy, balanced accuracy, specificity, sensitivity, and area under the receiver operator characteristic (AUROC) curve. The 95 % confidence intervals for these were calculated using 10-fold cross-validation. The 10-fold cross-validation data splitting procedure occurred before any pre-processing to avoid data-leakage. Given the size of the dataset, the sample size for model training and validation was considered adequate.

Shapley values were used to assess feature importance.[15] The Shapley value is the average marginal contribution of a feature value across all possible feature combinations. Shapley values assess the contribution of a given feature to the overall prediction. Model accuracy is reported in terms of the area under the receiver operating characteristic (AUROC) curve, overall accuracy and the true positive rate.

A sensitivity analysis was conducted to assess the impact of removing regular day attendances for dialysis (ICD-10 code N185 (chronic kidney disease, stage 5) and Z491 (extracorporeal dialysis)). On visual inspection of the data, spells containing these codes tended to have a very high proportion of data inconsistencies and they could add significant bias to any model. Their impact on the model was assessed by removing them and re-running the analysis. Coders for regular attendances for other reasons (e.g., cancer treatment) were not identified as significant outliers on visual inspection.

A sub-analysis was performed to identify features associated with data inconsistencies within spells where death is recorded. To do this, a mixed-effects multilevel logistic regression was used. A regression model was preferred over a random forest classifier for this analysis due to the much smaller size of the dataset.

Where data were missing the numbers are given. Missing data were rare for all features except ethnicity where 15–20 % of values were missing. When modelling ethnicity, a separate category for missing data was created. In all other cases missing data were excluded from modelling.

3. Results

Data were available for 172,324 unique patients with an autism diagnosis on first admission. The characteristics of these patients in their first spell during the study period are summarised in Table 1 together with the number of subsequent spells and the number of data inconsistencies. The median age of patients was 15 with a lower quartile of 9 and an upper quartile of 26. The majority of patients were aged younger than 18 years and over two-thirds were male. The Asperger's syndrome ICD-10 code was used in 36,625 (21.3 %) of first spells. There were 390,220 subsequent spells of which 170,447 (43.7 %) had inconsistencies. Data inconsistencies were more common in older age groups, females, patients from more deprived areas and patients of White ethnicity.

Table 2 summarises data for the subsequent spells according to aspects of hospital care. More than half of the spells did not involve an overnight stay and these spells had a noticeably higher rate of inconsistencies than where patients stayed overnight. Elective patients had a slightly higher inconsistency rate than emergency admissions. Spells under the care of a paediatrician had a comparatively low

Table 1

Number of first admission, number of subsequent spells and number of data inconsistences categorised by patient demographic characteristic on first admission.

| Variable	Number of patients, total = 172324	Number of subsequent spells (%), total = 390220	Number of subsequent spells with data inconsistencies (error rate %), total = 170447
Age band			
0–17	98,591 (57.2 %)	162,696 (41.7 %)	57,317 (35.2 %)
18–39	50,682 (29.4 %)	152,082 (39.0 %)	71,936 (47.3 %)
40–59	16,060 (9.3 %)	50,421 (12.9 %)	27,441 (54.4 %)
60–79	6171 (3.6 %)	22,371 (5.7 %)	12,025 (53.8 %)
80 years and over	820 (0.5 %)	2650 (0.7 %)	1728 (65.2 %)
Sex			
Female	49,414 (28.7 %)	138,417 (35.5 %)	64,650 (46.7 %)
Male	122,616 (71.2 %)	251,803 (64.5 %)	105,797 (42.0 %)
No sex recorded	294 (0.1 %)	0 (0.0 %)	
Deprivation quintile			
1 (most deprived)	48,539 (28.2 %)	112,451 (28.8 %)	50,739 (45.1 %)
2	38,254 (22.2 %)	89,978 (23.1 %)	40,305 (44.8 %)
3	31,311 (18.2 %)	73,208 (18.7 %)	30,392 (41.5 %)
4	26,332 (15.3 %)	61,298 (15.7 %)	26,390 (43.1 %)
5 (least deprived)	22,275 (12.8 %)	49,005 (12.6 %)	20,478 (41.8 %)
No deprivation recorded	5613 (3.3 %)	4280 (1.1 %)	2143 (50.1 %)
Ethnicity			
White	113,146 (65.4 %)	256,806 (65.8 %)	114,838 (42.9 %)
Asian or Asian British	6916 (4.0 %)	13,041 (3.3 %)	4733 (36.3 %)
Black or Black British	4964 (2.9 %)	10,154 (2.6 %)	4195 (41.3 %)
Mixed	3695 (2.1 %)	6508 (1.7 %)	2362 (36.3 %)
Other ethnicity	16,537 (9.6 %)	24,264 (6.2 %)	10,155 (41.8 %)
No ethnicity recorded	27066 (15.7 %)	79,447 (20.4 %)	34,064 (42.9 %)

Table 2
Numbers of data inconsistencies and clinical management profile.

Characteristic	Number of subsequent spells (total = 390220)	Data inconsistencies (error rate %), total = 170447
Length of stay		
0 days	220,100 (56.4 %)	104,196 (47.3 %)
1 day	66,525 (17.1 %)	26,557 (39.9 %)
2 days	27,412 (7.0 %)	10,774 (39.3 %)
≥ 3 days	76,179 (19.5 %)	28,917 (38.0 %)
Missing	4 (0.0 %)	3 (75.0 %)
Admission method		
Elective	190,948 (48.9 %)	84,845 (44.4 %)
Emergency	198,748 (50.9 %)	85,380 (43.0 %)
Missing	524 (0.1 %)	222 (42.4 %)
Speciality		
Paediatrics	77,794 (19.9 %)	27,388 (35.2 %)
General Medicine	34,869 (8.9 %)	16,834 (48.3 %)
General Surgery	24,015 (6.2 %)	11,348 (47.3 %)
Accident and Emergency	19,028 (4.9 %)	10,446 (54.9 %)
Gastroenterology	18,727 (4.8 %)	9412 (50.3 %)
Trauma and Orthopaedics	14,985 (3.8 %)	5575 (37.2 %)
Nephrology	14,025 (3.6 %)	11,327 (80.8 %)
Clinical Haematology	12,018 (3.1 %)	4755 (39.6 %)
Learning Disability	10,821 (2.8 %)	4529 (41.9 %)
Urology	10,782 (2.8 %)	4815 (44.7 %)
Other	153,156 (39.2 %)	64,018 (41.8 %)
Trust		
Same trust as first spell	252,048 (64.6 %)	97,724 (38.8 %)
Different trust to first spell	138,172 (35.4 %)	72,723 (52.6 %)

inconsistency rate and spells under the care of nephrology had a very high inconsistency rate. On inspection of the 14,025 nephrology spells, 12,547 (89.5 %) were single day admissions and 12,628 (90.0 %) were elective. Higher rates of data inconsistencies were also seen for subsequent spells where the hospital trust was not the same as the hospital trust for the first admission spell.

The random forest algorithm was found to have an accuracy of 77.7 % (95 % CI: [77.5 %, 78.0 %]), a balanced accuracy of 77.4 % (95 % CI: [77.2 % − 77.7 %]), sensitivity of 75.5 % (95 % CI: [75.0 % − 75.8 %]), a specificity of 79.4 % (95 % CI: [79.0 % − 79.8 %]) and an AUROC of 0.864 (95 % CI [0.862 – 0.866]) in predicting whether or not a given subsequent spell would have an inconsistency. Fig. 1 is a plot of the 20 most important features identified by the random forest model as well as the corresponding Shapley value for that feature in each sample. The features most strongly associated with inconsistencies included greater age, greater deprivation, longer time since the first spell, change in provider, shorter length of stay, being female and a change in the main specialty description. A small number of ICD-10 codes were identified as adding predictive value to the random forest classifier. Inconsistencies were less common where there was a diagnosis of at least one other learning difficulty/disability ('developmental disorder of scholastic skills' (ICD-10 code: F819) and 'attention-deficit hyperactivity disorder, predominantly inattentive type' (ICD-10 code: F900) or anxiety ('anxiety disorder, unspecified' (ICD-10 code: F419)). Inconsistencies were more common where 'chronic kidney disease, stage 5' (ICD-10 code: N185) and 'care involving dialysis' (ICD-10 code: Z491) were coded together. The latter code is used when a patient attends solely for renal dialysis preparation or treatment.[16] Tables 4 and 5 provide descriptions of the most important continuous and nominal variables.

Supplementary material Figures 1 and 2 show the random forest classifier's prediction of the probability of an inconsistency as a function
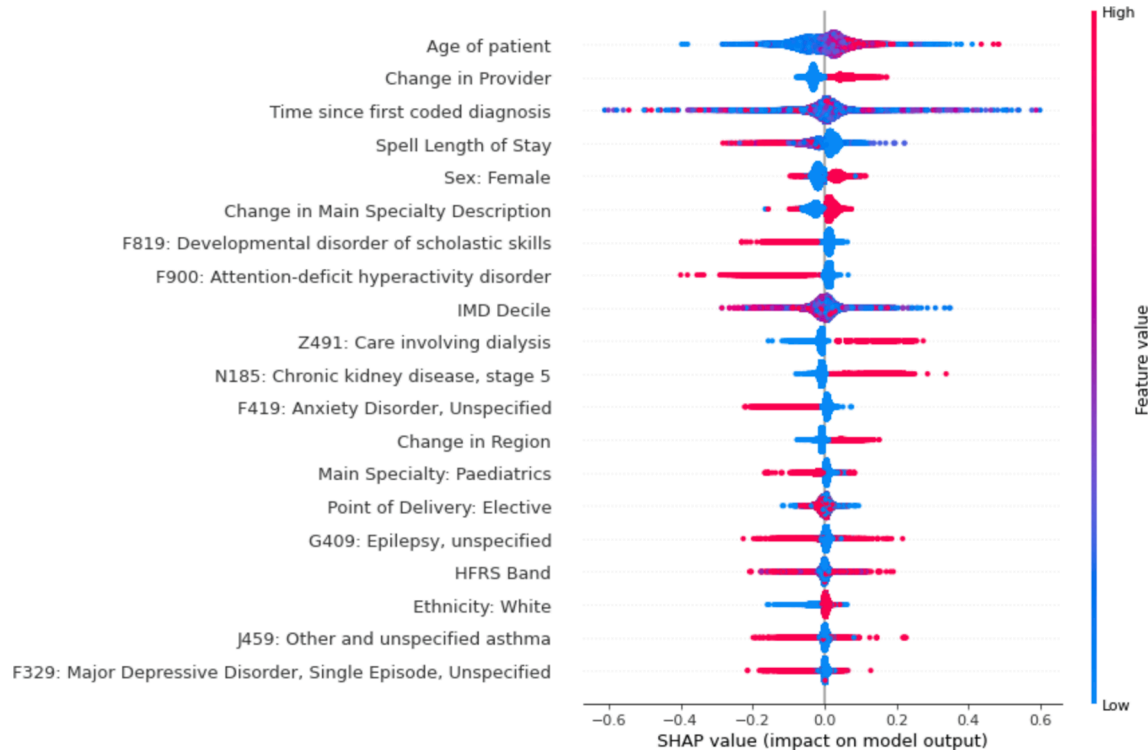
Fig. 1. Shapley Value Dot Summary Plot Each dot in the plot represents a subsequent spell. The x-axis indicates whether there is a positive or negative correlation between the value of the feature and its contribution to the model prediction of a spell containing an error. The colour of the dot represents the size of the feature relative to the range of values that feature can take, with red representing large feature values and blue low feature values. The horizontal axis represents the impact of the feature value on the output (an error in a subsequent spell). A positive SHAP value means the feature contributes to the spell having an error. A negative SHAP value means the feature contributes to the spell not having an error. The features are ranked by the mean of the absolute value of the SHAP values. IMD = Index of multiple deprivation.

Table 3

Population breakdown for deceased patients in the dataset, with the number of inconsistencies and the odds ratio for each feature based on multilevel logistic regression modelling.

Age band	Number of patients	Number of inconsistencies	Odds ratio
0–17	166 (6.3 %)	52 (31.3 %)	1 (reference)
18–39	444 (16.8 %)	116 (26.1 %)	0.896 (0.511 – 1.573), p = 0.702
40–59	759 (28.7 %)	227 (29.9 %)	1.104 (0.634 – 1.924), p = 0.726
60–79	992 (37.4 %)	305 (30.7 %)	1.188 (0.682 – 2.070), p = 0.543
80 years and over	288 (10.9 %)	130 (45.1 %)	2.270 (1.251 – 4.117), p = 0.007
Sex	(2 missing)		
Male	1950 (73.6 %)	589 (30.2 %)	
Female	697 (26.3 %)	241 (34.6 %)	1.234 (1.022 – 1.489), p = 0.029
Deprivation quintile			
1 (most deprived)	653 (24.7 %)	231 (35.4 %)	1 (reference)
2	615 (23.2 %)	182 (29.6 %)	0.734 (0.576 – 0.935), p = 0.012
3	545 (20.5 %)	167 (30.6 %)	0.783 (0.611 – 1.004), p = 0.054
4	468 (17.7 %)	131 (28.0 %)	0.672 (0.516 – 0.876), p = 0.003
5 (least deprived)	341 (12.9 %)	116 (34.0 %)	0.868 (0.653 – 1.154), p = 0.330
Missing	27 (1.0 %)	3 (11.1 %)	0.254 (0.075 – 0.865), p = 0.028
Ethnicity			
White	1730 (65.3 %)	536 (31.0 %)	1 (reference)
Asian or Asian British	62 (2.3 %)	19 (30.6 %)	1.002 (0.561 – 1.790), p = 0.994
Black or Black British	29 (1.1 %)	11 (37.9 %)	1.545 (0.711 – 3.354), p = 0.272
Mixed	15 (0.6 %)	5 (33.3 %)	1.267 (0.423 – 3.796), p = 0.672
Other ethnicity	155 (5.9 %)	47 (30.3 %)	1.013 (0.701 – 1.462), p = 0.946
No ethnicity recorded	658 (24.8 %)	212 (32.2 %)	1.037 (0.816 – 1.320), p = 0.762
HFRS Band			
None	2017 (76.1 %)	632 (31.3 %)	1 (reference)
Mild	271 (10.2 %)	81 (29.9 %)	0.947 (0.714 – 1.255), p = 0.704
Moderate	246 (9.3 %)	82 (33.3 %)	1.114 (0.837 – 1.484), p = 0.459
Severe	115 (4.3 %)	35 (30.4 %)	0.921 (0.608 – 1.396), p = 0.700
Palliative Care			
No palliative care code	1933 (73.0 %)	571 (29.5 %)	1 (reference)
Palliative care code present	716 (27.0 %)	239 (33.4 %)	1.322 (1.089 – 1.606), p = 0.005
Main Specialty of Final Spell			
General Medicine	696 (26.3 %)	200 (28.7 %)	1 (reference)
Geriatric Medicine	306 (11.6 %)	100 (32.7 %)	1.020 (0.751 – 1.384), p = 0.901
Respiratory Medicine	281 (10.6 %)	93 (33.1 %)	1.334 (0.750 – 2.834), p = 0.326
General Internal Medicine	171 (6.5 %)	52 (30.4 %)	1.053 (0.684 – 1.623), p = 0.814
General Surgery	155 (5.9 %)	49 (31.6 %)	1.198 (0.813 – 1.766), p = 0.361
Gastroenterology	126 (4.8 %)	40 (31.7 %)	1.170 (0.770 – 1.779), p = 0.461
Cardiology	98 (3.7 %)	34 (34.7 %)	1.375 (0.870 – 2.173), p = 0.172
Paediatrics	94 (3.6 %)	31 (33.0 %)	1.403 (0.695 – 2.834), p = 0.345

Table 3 (*continued*)

Age band	Number of patients	Number of inconsistencies	Odds ratio
Critical Care Medicine	93 (3.5 %)	28 (30.1 %)	1.226 (0.741 – 2.028), p = 0.427
Anaesthetics	66 (2.5 %)	14 (21.1 %)	0.771 (0.411 – 1.445), p = 0.416
Other	563 (21.3 %)	189 (33.5 %)	1.314 (1.021 – 1.691), p = 0.034

For the main specialties of the final spell, the reference category was general medicine.

Table 4

Descriptions of the most predictive continuous features used in the model.

Name of Feature	Description	Unit of Measure	Range	Mean (Standard Deviation)
Age of patient	Patient age	Years	(0, 101)	20.2 (16.5)
Time since first coded diagnosis	Time since first spell with autism	Days	(0, 2917)	742.8 (662)
Spell Length of Stay	Length of Stay of Hospital Spell	Days	(0, 2353)	3.9 (27.0)

Table 5

Descriptions of the most predictive nominal features used in the model.

Name of Feature	Description	Distribution of Values
Change in Provider	Indicates if there was (1) or was not (0) a change in provider relative to the first spell	0: 64.6 %, 1:35.4 %
Sex: Female	Indicates if patient was female (1) or male (0)	0: 64.5 %, 1: 35.5 %
Change in Main Specialty Description	Indicates if there was (1) or was not (0) a change in main specialty description relative to the first spell	0: 40.6 %, 1:59.4 %
F819: Developmental disorder of scholastic skills	Indicates if the code F819 was (1) or was not (0) used in the spell	0: 84.5 %, 1: 15.5 %
F900: Attention-deficit hyperactivity disorder	Indicates if the code F900 was (1) or was not (0) used in the spell	0: 92.1 %, 1: 7.9 %
Z491: Care involving dialysis	Indicates if the code Z491 was (1) or was not (0) used in the spell	0: 95.4 %, 1: 4.6 %
N185: Chronic kidney disease, stage 5	Indicates if the code N185 was (1) or was not (0) used in the spell	0: 94.9 %, 1: 5.1 %
F419: Anxiety Disorder, Unspecified	Indicates if the code F419 was (1) or was not (0) used in the spell	0: 91.8 %, 1: 8.2 %
Change in Region	Indicates if there was (1) or was not (0) a change in region relative to the first spell	0: 88.6 %, 1: 11.4 %
Main Specialty: Paediatrics	Indicates if the paediatrics was (1) or was not (0) the main specialty of the spell	0: 75.7 %, 1: 24.3 %
Point of Delivery: Elective	Indicates if the spell was elective (1) or emergency (0)	0: 51.1 %, 1: 48.9 %
G409: Epilepsy, unspecified	Indicates if the code G409 was (1) or was not (0) used in the spell	0: 88.4 %, 1: 11.6 %
Ethnicity: White	Indicates if the patient's ethnicity was white (1) or not (0)	0: 34.2 %, 1: 65.8 %
J459: Other and unspecified asthma	Indicates if the code J459 was (1) or was not (0) used in the spell	0: 87.8 %, 1:12.2 %
F329: Major Depressive Disorder: Single Episode, Unspecified	Indicates if the code F329 was (1) or was not (0) used in the spell	0: 91.0 %, 1: 9.0 %

of the patient age and time in days since first coded diagnosis, respectively. The plot for patient's age shows that inconsistencies are relatively common in very young children (aged < 1 year) but are much lower in older children before rising sharply in teenage years, with steadier increases from 20 years and above. The plot for the time since the first coded diagnosis indicates that the probability of an inconsistency is at its lowest five days after first spell where the diagnosis is recorded, but steadily increases after this before remaining relatively constant.

In sensitivity analysis, spells that were regular day attendances for kidney dialysis were removed. The performance of the model was similar to that of the main analysis with an accuracy of 76.1 % (95 % CI: [77.5 %, 78.0 %], a balanced accuracy of 77.6 % (95 % CI: [77.1 % − 77.8 %]), sensitivity of 75.6 % (95 % CI: [75.2 % − 75.8 %]), a specificity of 79.6 % (95 % CI: [79.0 % − 79.8 %]) and an AUROC of 0.847 (95 % CI [0.842 – 0.849]). The 20 most important features in the model are shown in **Supplementary material Figure S3** and were similar to the main analysis.

In total, 2649 patients died in hospital during the study period. Their characteristics are summarised in Table 3. The patients who did not have an autism code in the spell in which they died were more likely to be older and more deprived than those who did have an autism code. The patients with inconsistencies were also more likely to have a palliative care code in their final spell. Table 3 also shows the results of the multilevel multivariable logistic regression; inconsistencies in the final spell in those who died were significantly associated with being 80 years and over, being female, greater deprivation and use of a palliative care code in the death spell.

4. Discussion

Our study reveals the use of mandatory codes in the Hospital Episodes Statistics (HES) dataset to be inconsistent across spells in a large number of cases. Since our dataset reflects information recorded in the medical record during a hospital stay, it suggests that in some cases care may be sub-optimal and not take into account that a patient has a diagnosis of autism. Key factors associated with data inconsistencies included greater age, being female, greater deprivation, change of speciality or provider from the first spell, length of time from the first spell and day-case admission. The higher inconsistency rate in females, older age groups and people living in more deprived areas is perhaps the most finding striking and suggests a degree of under-recording of autism in these groups. This suggests that as people get older and may have multiple co-morbidities, clinicians may be less focussed on an underlying autism diagnosis. The higher rate of data inconsistences in females correlates with emerging evidence of late and misdiagnosis of autism in females due in part to the greater prevalence of camouflaging or masking behaviours.[17] Uncertainty of an autism diagnosis or a prolonged diagnostic process may also be possible factors impacting the inconsistency rate, particularly in younger patients. Such inconsistencies have the potential to distort our understanding of service use in key demographic groups.

Higher rates of inconsistencies were noted where there was a change of speciality or provider. The lack of joined up data systems has been acknowledged with the planned implementation of the NHS Digital and NHS England reasonable adjustment flag on the NHS Spine to improve sharing across systems.[18] Improved intra-trust (between departments) and inter-trusts communication could substantially reduce the number of data inconsistencies.

Inconsistencies were more common where the admission was associated with a routine day attendance for kidney dialysis. Although the autism codes are mandatory, this association with dialysis is not unexpected and consistent with coding practice in many trusts, where limited data is provided to clinical coders for routine day attendances and large numbers of regular day attendances may preclude extensive coding depth. Inconsistencies were less common where there was another learning disability code used in the same spell. This may be due to

clinicians recognising a learning disability more easily than autism and that autism is a common comorbidity with learning disability.

Through initiatives such as the GIRFT programme, the Model Healthcare system and National Consultant Information Programme (NCIP), the use of administrative data to inform clinical and management decision making in healthcare in England is increasing and this change is to be encouraged.[19–20] The COVID-19 pandemic has also shed light on the potential of such data sources.[21] However, if attention is not paid to the quality of the underlying data, then there is potential to distort our understanding. A greater ability to link datasets across settings may help to identify and address such inconsistencies. The recommendations of the recently published Goldacre review on the use of health data may help facilitate moves to improved data quality through allowing better linkage across datasets within trusted research environments.[1] SNOMED CT is a clinical terminology set to play an important role in improving data quality and interoperability.[22] Its design and structure enable it to support scalable and meaningful health data capture, storage, retrieval and communication.[23]

Since most interactions of autistic people with health services will be through primary care, comparison of primary and secondary care data at a patient level is likely to provide much greater insight than secondary care data alone. Improved intra-dataset linkage may help avoid issues around data consistency when patients change healthcare specialty or healthcare provider. In England, initiatives such as the Population Health Management Programme, developed across integrated care systems, have the potential to improve data quality.[24]

Autism can impact on a person's ability to communicate health problems and to engage with support and advice offered.[2] In England, it is a legal duty under the Equality Act (2010) to ensure that reasonable adjustments, such as staff being trained in autism awareness, longer appointment times and reducing noise and distractions, are made so that services are accessible.[18] This will vary for each autistic person and not all will need them. However, clinicians need to be aware that a person has a diagnosis of autism in order to make reasonable adjustments in the provision of care. Since the end of 2021 The Learning from lives and deaths (LeDeR) programme has extended from reviewing the deaths of people with learning disabilities to include the deaths of all autistic adults within its scope.[25] This programme aims to improve services for autistic people and reducing premature mortality. The programme relies on these deaths to be identified and reported.

The need for reliable data on the health inequalities faced by autistic people has been highlighted by NHS England and NHS Improvement Autism Programme and has resulted in the inclusion of autistic people who have died in LeDeR from 2022.[25] Our study highlights the issue of identifying deaths in people with autism. We also note a higher rate of inconsistencies in patients with a palliative care code and this is perhaps unsurprising given the care setting. However, this could result in a lack of referrals to the ongoing LeDeR programme and the potential for subsequent learning and improvement recommendations to be biased. Initiatives to allow easier identification of co-morbidities within HES will help improve recording of autism diagnoses and minimise the risks of diagnostic overshadowing, particularly later in life when age-related comorbidity may come to dominate the health recorded.[26]

Beyond autism, the potential impacts of data inaccuracy in people with learning disabilities have been previously identified. The 2013 confidential inquiry into premature deaths of people with learning disabilities found that the lack of reasonable adjustments was a contributory factor in a number of deaths and that hospital systems to identify people with learning disabilities who needed reasonable adjustments were limited.[27] The inquiry identified a need for clear identification of people with learning disabilities on NHS databases and for this information to be made available to care professionals in healthcare record systems, including a record of reasonable adjustment require. The issue of under-recording of learning disabilities on death certificates was identified in a recent systematic review.[28]

The reasons for under recording of autism in HES are likely to be

complex and specific to each trust entering data. Health professionals may not always feel able or confident enough, or feel it is necessary, to add an autism diagnosis to the clinical record if not directly relevant to the current spell of care. Autistic people themselves may not want to discuss their diagnosis and clinicians are advised by NHS England that terminology for autism is not universally accepted and that, "it is best to use the words they use to describe themselves and their loved ones".[29] Even when autism is recorded in the medical notes for a spell, out-dated terminology used by the patient may be used or an abbreviation.[29] The diagnostic term 'Asperger's syndrome' was used in over one-fifth of first spells, but is now considered an outdated term. Clinical coders can only code a diagnosis if it appears in a certain format and may be unable to code autism even if they strongly suspect it to be present from the information given. Within the specific area of learning disability, coding is also complex with slight differences in how information is recorded in medical notes leading to different ICD-10 codes being used. Therefore, to fulfil the aims of the National Autism Plan in the UK, particularly to reduce health inequalities and to ensure improved access to relevant supports and treatments, the findings of our analysis should be shared with relevant clinicians to ensure more consistent data entry, and to develop systems to optimise data quality.

Our study has several strengths including the size of the dataset and the ability to track patient data across multiple hospital trusts within England and over time. The usefulness of the retrospective approach taken in this work is in highlighting data inconsistencies and identifying particular patient or spell characteristics that relate to poor data quality. These could go on to help inform policy change.

Our study also has some limitations, the most important of which have already been acknowledged in relation to coding practice. No gold standard exists against which to externally validate the HES dataset. As such, we were limited to assessing internal consistency. Inconsistencies were assessed in relation to the first spell in which the autism code was recorded during the study period. We recognise that in some cases this first recorded use of an autism code may be an error and that later absence of an autism code may represent an accurate clinical record. In our analysis we do not attempt to identify the source of the error, only that the record is inconsistent with a previous record. A potential limitation is our use of random under-sampling, which has the potential to introduce bias to the analysis. However, given the relatively large sample size, substantial bias is unlikely.

In summary, our study identifies a large number of data inconsistencies in the recording of autism in the HES dataset. Improving data quality for people with autism admitted to hospital is important in ensuring they receive appropriate care. Identifying factors associated with data inconsistencies supports discussions regarding possible reasons for these inconsistencies; a first step in improving data quality. As the UK government's national strategy for autistic children, young people and adults (2021 to 2026)[4] acknowledges, good quality data is a fundamental requirement if care for people with autism is to be improved.

Summary Table

What was already known on the topic	What this study added to our knowledge
Once diagnosed, it is a mandatory requirement to record autism in the Hospital Episodes Statistics (HES) database for England.	Specific patient features were identified as relating strongly with inconsistencies in autism coding. These included greater age and being female.
The extent to which autism is consistently recorded in HES is not known.	Features relating to the hospital stay were also identified as being related to inconsistencies in autism coding. These included change of speciality or provider from the first spell in which autism was recorded, length of time from the first spell and day-case admission.
Not recording autism as present during hospital stay has potential to result in	

(continued on next column)

(continued)

What was already known on the topic	What this study added to our knowledge
sub-optimal care for patients and to bias data for people with autism.	Nephrology spells and spells that involved dialysis had very high rates of inconsistencies. Inconsistencies in the final spell for patients with autism who died in hospital were associated with being 80 years and over, being female, greater deprivation and the use of a palliative care code in the death spell.

CRediT authorship contribution statement

Johannes Heyl: Conceptualization, Data curation, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **Flavien Hardy:** Conceptualization, Data curation, Formal analysis, Writing – review & editing. **Katie Tucker:** Conceptualization, Formal analysis, Project administration, Writing – review & editing. **Adrian Hopper:** Formal analysis, Writing – review & editing. **Maria J. Marchã:** Conceptualization, Formal analysis, Writing – review & editing. **Ashley Liew:** Formal analysis, Writing – review & editing. **Judith Reep:** Formal analysis, Writing – review & editing. **Kerry-Anne Harwood:** Formal analysis, Writing – review & editing. **Luke Roberts:** Formal analysis, Writing – review & editing. **Jeremy Yates:** Writing – review & editing. **Jamie Day:** Writing – review & editing. **Andrew Wheeler:** Writing – review & editing. **Sue Eve-Jones:** Writing – review & editing. **Tim W.R. Briggs:** Writing – review & editing. **William K. Gray:** Conceptualization, Data curation, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

This report does not contain patient identifiable data. Consent from individuals involved in this study was not required. Requests for any underlying data cannot be granted by the authors because the data were acquired from data under licence/data sharing agreement from NHS Digital, for which conditions of use (and further use) apply. Individuals and organisations wishing to access HES data can make a request directly to NHS Digital.

Informed consent

Informed consent was not sought for the present study because it was an analysis of routine administrative data.

Ethical approval

Ethical approval was not sought for the present study because it did not directly involve human participants. This study was completed in accordance with the Helsinki Declaration as revised in 2013.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2022.104938.

References

[1] B. Goldacre, J. Morley, Better, broader, safer: using health data for research and analysis, 2022; https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1067053/goldacre-review-using-health-data-for-research-and-analysis.pdf (accessed 3rd May, 2022).

[2] B. Muskat, P. Burnham Riosa, D.B. Nicholas, W. Roberts, K.P. Stoddart, L. Zwaigenbaum, Autism comes to the hospital: the experiences of patients with autism spectrum disorder, their parents and health-care providers at two Canadian paediatric hospitals, Autism 19 (4) (2015) 482–490.

[3] NHS Digital, *Coding clinic (vesion 9.1)*. London, UK, 2021.

[4] Department of Health and Social Care and Department for Education. Policy paper: National strategy for autistic children, young people and adults: 2021 to 2026, 2021; https://www.gov.uk/government/publications/national-strategy-for-autistic-children-young-people-and-adults-2021-to-2026. Accessed 24th May, 2022.

[5] Digital N. Assuring Transformation. 2022; https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/assuring-transformation. Accessed 24th May, 2022.

[6] M. Oswald, Anonymisation Standard for Publishing Health and Social Care Data Specification (Process Standard). *Leeds, UK, Information Standards Board for Health and Social Care,* 2013.

[7] N.H.S. Digital, *National Clinical Coding Standards ICD-10 5th Edition: Accurate data for quality information*, Terminology and Classifications Delivery Service, Leeds, UK, 2021.

[8] T. Gilbert, J. Neuburger, J. Kraindler, et al., Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study, Lancet 391 (10132) (2018) 1775–1782.

[9] Ministry of Housing and Communities and Local Government. English indices of deprivation, 2019; https://www.gov.uk/government/collections/english-indices-of-deprivation. Accessed 2021-08-25.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., Scikit-learn: Machine learning in Python, J. Machine Learning Res. 12 (2011) 2825–2830.

[11] S. Subudhi, A. Verma, A.B. Patel, et al., Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19, NPJ Digit Med. 4 (1) (2021) 87.

[12] P. Baqui, V. Marra, A.M. Alaa, I. Bica, A. Ercole, M. van der Schaar, Comparing COVID-19 risk factors in Brazil using machine learning: the importance of socioeconomic, demographic and structural factors, Sci. Rep. 11 (1) (2021) 15591.

[13] H. Estiri, Z.H. Strasser, J.G. Klann, P. Naseri, K.B. Wagholikar, S.N. Murphy, Predicting COVID-19 mortality with electronic medical records, NPJ Digit Med. 4 (1) (2021) 15.

[14] F. Cabitza, A. Campagner, The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies, Int. J. Med. Inf. 153 (2021), 104510.

[15] L.S. Shapley, A value for n-person games, RAND Corporation, Santa Monica, CA, USA, 1952.

[16] N.H.S. Digital, National Clinical Coding Standards - ICD-10, 5th ed., The Stationary Office, London, UK, 2017.

[17] M.C. Lai, M.V. Lombardo, B. Auyeung, B. Chakrabarti, S. Baron-Cohen, Sex/gender differences and autism: setting the scene for future research, J. Am. Acad. Child Adolesc. Psychiatry 54 (1) (2015) 11–24.

[18] NHS Digital. Reasonable adjustments flag, 2021; https://digital.nhs.uk/services/reasonable-adjustment-flag#reasonable-adjustments-an-overview. Accessed 24th May, 2022.

[19] National Consultant Information Programme. NCIP. 2022; https://www.gettingitrightfirsttime.co.uk/ncip/. Accessed 16th January, 2022.

[20] Model Health System. Model Health System: Supporting NHS teams to provide high quality patient care and continuous improvement. 2022; https://model.nhs.uk/. Accessed 16th January, 2022.

[21] A.V. Navaratnam, W.K. Gray, J. Day, J. Wendon, T.W.R. Briggs, Patient factors and temporal trends associated with COVID-19 in-hospital mortality in England: an observational study using administrative data, Lancet Respir. Med. 9 (4) (2021) 397–406.

[22] NHS Digital. *SCCI0034: SNOMED CT.* London, UK, 2020.

[23] NHS Digital. *An introduction to SNOMED.* London, UK, 2022.

[24] NHS England. Population Health and the Population Health Management Programme. 2022; https://www.england.nhs.uk/integratedcare/what-is-integrated-care/phm/. Accessed 3rd May, 2022.

[25] NHS England and NHS Improvement. *Learning from lives and deaths – People with a learning disability and autistic people (LeDeR) policy.* London, UK, 2021.

[26] R. Molloy, G. Brand, I. Munro, N. Pope, Seeing the complete picture: A systematic review of mental health consumer and health professional experiences of diagnostic overshadowing, J. Clin. Nurs. (2021).

[27] P. Heslop, P. Blair, P. Fleming, M. Hoghton, A. Marriott, L. Russ, Confidential Inquiry into premature deaths of people with learning disabilities (CIPOLD), University of Bristol, Bristol, UK, 2013.

[28] F. Dunwoodie Stirton, P. Heslop, Medical Certificates of Cause of Death for people with intellectual disabilities: A systematic literature review, J. Appl. Res. Intellect. Disabil. 31 (5) (2018) 659–668.

[29] NHS England and NHS Improvement. Making information and the words we use accessible. 2022; https://www.england.nhs.uk/learning-disabilities/about/get-involved/involving-people/making-information-and-the-words-we-use-accessible/. Accessed 24th May, 2022.