Hospital Episode Statistics – Ethnicity data products

Scope

To enhance hospital admission records in England with names-based ethnicity coding and produce admission statistics for ambulatory care sensitive conditions and major disease categories by ethnic group and regions, April 1999 – March 2014. The report is divided into chapters as follows:

1. Names-based ethnicity enhancement of hospital admission records in England
2. Ethnic inequalities in preventable hospital admissions
3. Ethnic inequalities in hospital admissions by major disease groups

CONTENTS

/

Contributors

Jakob Petersen[1], Jens Kandt[2], Paul A. Longley[1]

1 Consumer Data Research Centre (CDRC), Department of Geography, University College London (UCL), Gower Street, London, WC1E 6BT 2 The Bartlett Centre for Advanced Spatial Analysis (CASA), Gower Street, UCL, London, WC1E 6BT.

Funding

The UK Economic and Social Research Council is acknowledged for its support for the UCL Consumer Data Research Centre (CDRC) enabling this research (Grant ES/L011840/1).

1 NAMES-BASED ETHNICITY ENHANCEMENT OF HOSPITAL ADMISSION RECORDS IN ENGLAND

1.1 ABSTRACT

Background

Accurate recording of ethnicity in electronic healthcare records is important for the monitoring of health inequalities. Yet until the late 1990s, ethnicity information was absent from more than half of records of patients who received inpatient care in England. In this study, we report on the usefulness of names-based ethnicity classification, Ethnicity Estimator (EE), for addressing this gap in the hospital records.

Methods

Data on impatient hospital admission were obtained from Hospital Episode Statistics (HES) between April 1999 and March 2014. The data were enhanced with ethnicity coding of participants' surnames using the EE software. Only data on the first episode for each patient each year were included.

Results

A total of 111,231,653 patient-years were recorded between April 1999 and March 2014. NHS recording of ethnicity improved from 59.5% in 1999 to 90.5% in 2013. Biggest improvement was seen in the White British group, which increased from 55.4% in 1999 to 73.9% in 2013. The correct prediction of NHS-reported ethnicity varied by ethnic group (2013/14 figures): White British (89.8%), Pakistani (81.7%), Indian (74.6%), Chinese (72.9%), Bangladeshi (63.4%), Black African (57.3%), White Other (50.5%), White Irish (45.0%). For other ethnic groups the prediction success was low to none. Prediction success was above 70% in most areas outside London but fell below 40% in parts of London.

Conclusion

Studies of ethnic inequalities in hospital inpatient care in England are limited by incomplete data on patient ethnicity collected in the 1990s and 2000s. The prediction success of a names-based ethnicity classification tool has been quantified in HES for the first time and the results can be used to inform decisions around the optimal analysis of ethnic groups using this data source.

1.2 INTRODUCTION

Ethnicity is defined as a sensitive personal characteristic under European Union (2016) General Data Protection Regulation (GDPR) [1]. It is often considered to be inherently subjective [2] and may not always be collected for reasons of statute [3,4]. This can handicap the conduct of equality audits, analysis of corporate governance [5] and, most recently, monitoring of hospital admissions and outcomes during the COVID-19 pandemic [6,7].

Provision has been made for Hospital Episode Statistics (HES) to include patient-reported ethnicity since 1995 by drawing on a central NHS patient register [5]. Yet until the late 1990s, ethnicity information was absent from more than half of records of patients who received inpatient care. General practitioners were financially incentivised to record patient ethnicity through the Quality Outcomes Framework (QOF) between 2006-2012 with a resultant increase in completeness of inpatient ethnicity data to more than 80% during this time [5].

The problem of missing ethnicity data in NHS datasets has previously been studied [8,9]; although not in the full range of ethnic groups in a national study over several years. Personal names are commonly used to impute ethnicity information when self-reported ethnicity data are not collected systematically or available through linkage [10,11].

In this paper we report on the use of names-based ethnicity classifications to address incomplete ethnicity information in inpatient hospital records. It is a national study covering the whole of England over fifteen years (1999/00-2013/14). The study quantifies the prediction success of the complete range of ethnic groups – nationally and regionally – against self-reported, NHS-recorded ethnicity. A freely available software, Ethnicity Estimator, was used [10]. EE was developed by the Consumer Data Research Centre (CDRC: cdrc.ac.uk) in partnership with the Office for National Statistics (ONS) and using enhanced algorithmic procedures [10,12]. The results of this study can be used to inform decisions around analysis of ethnicity in HES.

1.3 METHODS

Hospital inpatient admission records were obtained from NHS England HES for the period April 1999-March 2014 (financial years referred by the first year only from here onwards). The ethnicity information was coded on patient forename and surname separately using an enhanced version of the Ethnicity Estimator (EE) software [10]. Where a patient changed surname, e.g. due to marriage, the ethnicity category of the earliest name was used. To retain full anonymity, the coding was carried out in an air-gapped, secure data facility by NHS Digital linking name information in the Patient Demographic Service to HES. The main 12 EE categories map onto the Census 2011 ethnicity categories except for mixed ethnicity, which is not predicted by EE. All NHS-recorded mixed ethnicities were combined into a single mixed category. Not-Stated and Missing were combined, and Black Other was combined with Other. It should be noted that the NHS used a simpler coding frame in 1999-2001, which did not include categories for mixed, White Irish or Asian Other.

Using the EE software, we developed three different ways of estimating ethnicity; each of which we compared to the benchmark of self-reported ethnicity as recorded by the NHS.

1. NHS-recorded ethnicity with additional ethnicity estimation based on patient surname where data were missing (in the following: supplementary estimation)
2. Ethnicity estimation based on patient surname alone (surname-based estimation)
3. Ethnicity estimation based on patient forename and surname; selecting only those records where the estimated ethnic groups were identical for forename and surname (full name-based estimation), e.g. a record would only be classified as Pakistani where both forename and surname were estimated as Pakistani by the EE software, etc.

Annual estimates of the prediction success or sensitivity (proportion of true positives among the sum of true positives and false negatives) were calculated as the percentage of correctly predicted records based on surname – nationally and regionally. Only data on the first episode of care for each patient in each financial year were used.

The geography of correctly predicted ethnicities based on patient surname was mapped at Local Authority level. The specificity (proportion of true negatives among the sum of true negatives and false positives) of the estimator was also calculated.

Ethical approval was obtained from Bromley REC (Reference: 13/LO/1355) for analyses of patient-level HES data. The HES data licence reference was DARS-NIC-28051-Q3K7L.

1.4 RESULTS

A total of 111,231,653 patient-years were recorded between 1999 and 2013. NHS recording of ethnicity improved from 59.5% in 1999 to 89.2% in 2009 and peaked at 90.5% in 2013 (Figure 1). The biggest absolute improvement was seen in the White British group, which increased from 55.4% in 1999 to 73.9% in 2013. Figure 2 shows increased representation for other ethnic groups.

The sensitivity analysis comparing EE estimates with NHS-recorded ethnic group, in 2013, suggested that the accuracy of prediction was highest for White British individuals (89.8%) followed by those of Pakistani (81.7%), Indian (74.6%), Chinese (72.9%) and Bangladeshi (63.4%) extraction. Lower levels of success were recorded for Black African (57.3%), White Other (50.5%) and White Irish (45.0%) groups (Table S1, Appendix A, supplementary materials). For other ethnic groups the sensitivity was very low and none at all for mixed ethnic groups. The sensitivity increased for the White Other group from 10.5% in 1999 to 50.5% in 2013, whereas it remained more stable for other ethnic groups over time (Figure 3). The confusion matrix for NHS-recorded ethnicity against surname predicted ethnicity can be found in Table 1. The sensitivity and specificity of the EE prediction by ethnic group each year can be found in Table S1 (Appendix A, supplementary materials).

The prediction success within ethnic groups were similar for males and females (Figure 3). The prediction success was however higher for females than males among Bangladeshis. The prediction success of the full name-based classification was consistently lower than when using patient surname alone for all ethnic groups (Figure S1, Appendix A, supplementary materials). The prediction success of the surname-based estimation for the different ethnic groups across regions, in 2013, were relatively similar except for Asian others, White Other, White Irish, and Indian (Figure 4).

The proportion of patient ethnicities predicted by the software, in 2013, was calculated and mapped for English Local Authorities (Figure 5). Prediction success was above 70% in most areas outside London but fell below 40% in parts of London.

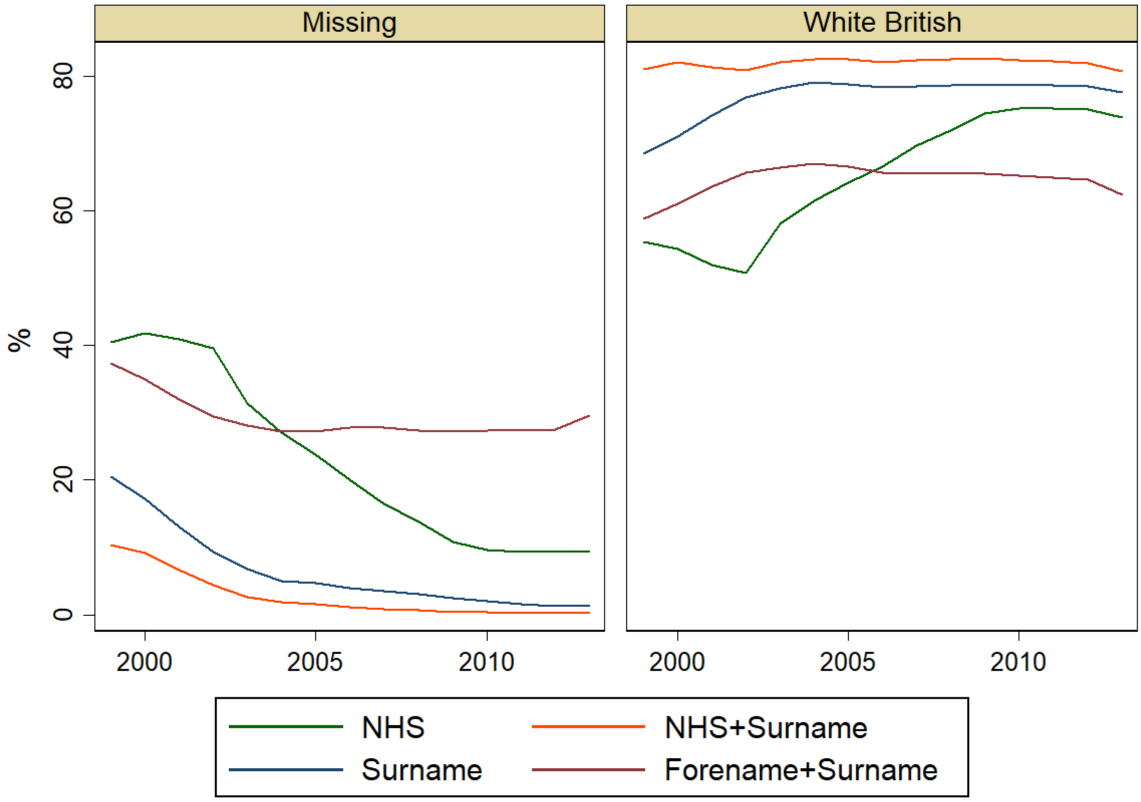

Figure 1   Proportion of patients for missing and White British ethnicity over time.

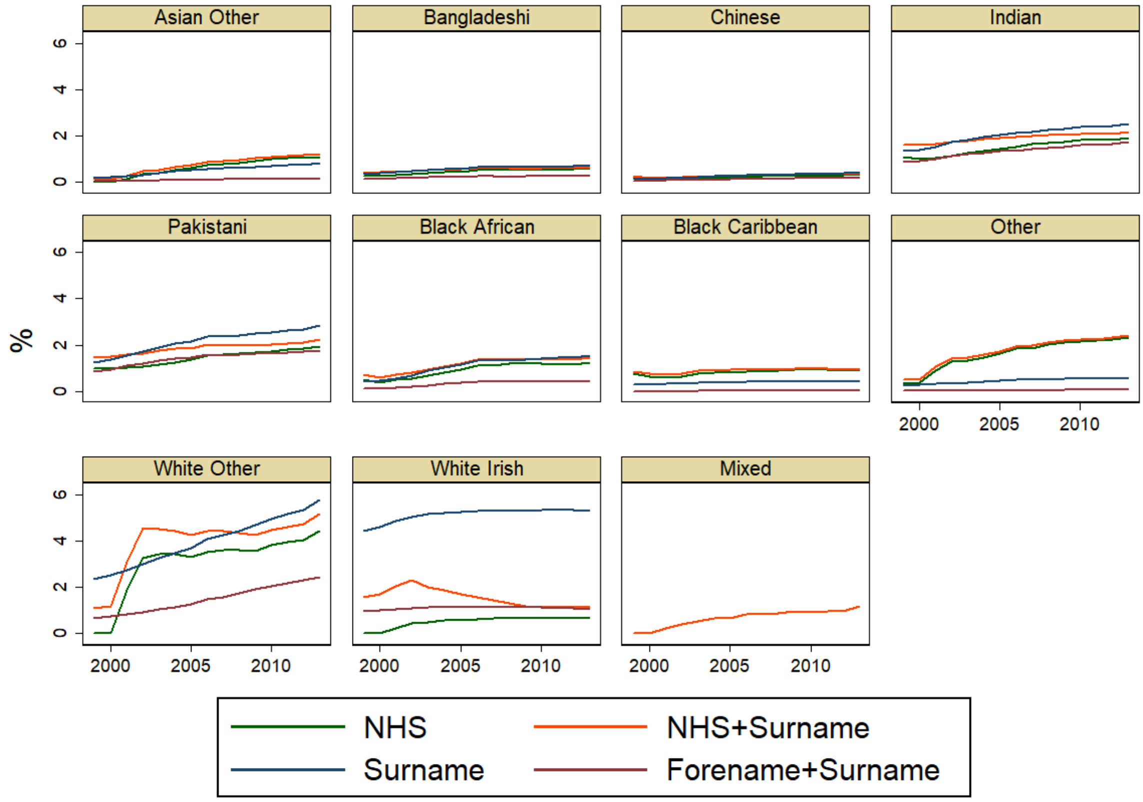

Figure 2   Proportion of patients for each ethnic minority group over time.

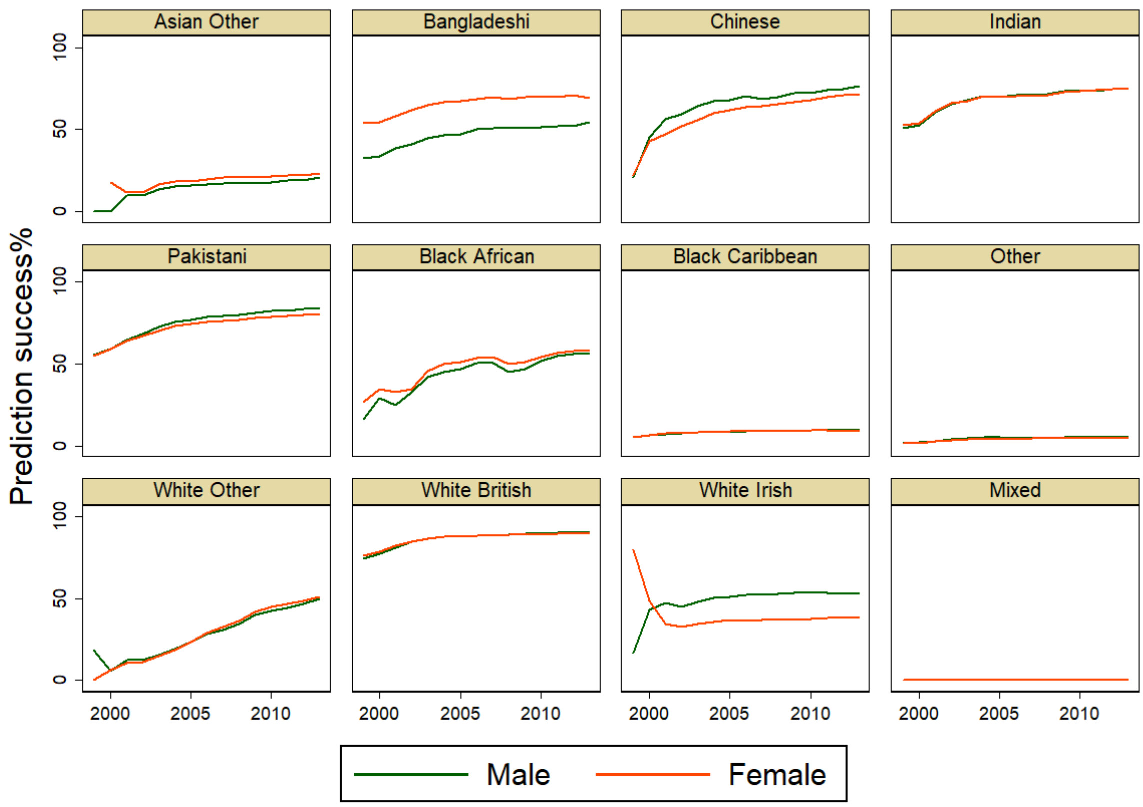

Figure 3 Sensitivity% of Ethnicity Estimator (EE) software (Kandt & Longley, 2018) in predicting NHS-recorded ethnicity by ethnic group and gender.

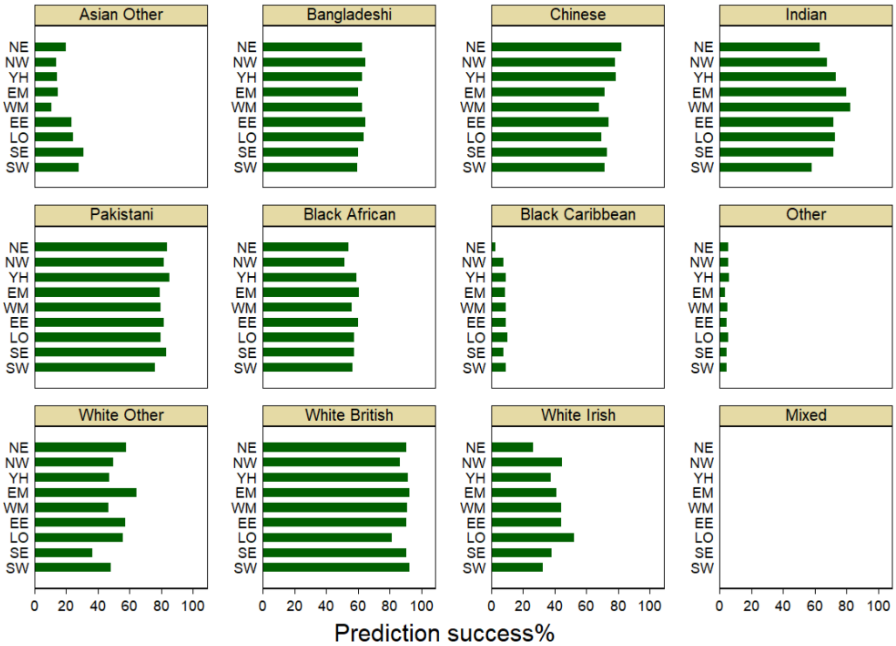

Figure 4 Prediction success (Sensitivity) of Ethnicity Estimator (EE) software (Kandt & Longley, 2018) in predicting NHS ethnic group from patient surname in 2013 by region. Abbreviations: North East (NE), North West (NW), Yorkshire & The Humber (YH), East Midlands (EM), West Midlands (WM), East of England (EE), Greater London (LO), South East (SE), South West (SW)

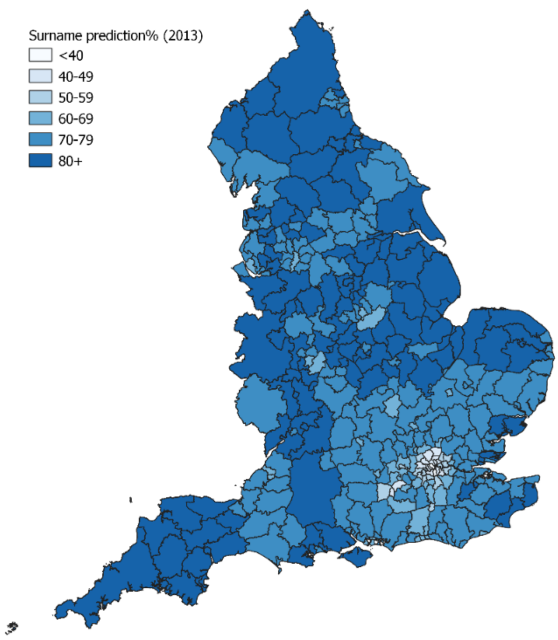

Figure 5 Proportion of patient ethnicities predicted on surname (%) in England, 2013, by Local Authority, using the EE software (Kandt & Longley, 2018).

Table 1 Confusion matrix of NHS-recorded versus EE surname predicted ethnicity (Kandt & Longley 2018) of HES patients in 2013.

	Surname prediction (EE)												
NHS-recorded	Asian other	Bangladeshi	Chinese	Indian	Pakistani	Black African	Black Caribbean	Missing	Other	White Other	White British	White Irish	Total
Asian Other	21,200	4,119	2,068	18,513	18,750	3,142	266	5,803	4,242	6,897	10,495	409	95,904
Bangladeshi	685	32,388	21	1,486	12,805	600	17	1,272	392	482	901	51	51,100
Chinese	1,621	21	18,623	171	104	106	20	1,017	115	584	3,016	139	25,537
Indian	5,399	2,418	125	122,042	12,799	2,198	212	3,116	1,620	3,658	9,772	320	163,679
Pakistani	1,770	11,213	29	8,171	137,807	2,137	46	2,224	1,329	1,084	2,825	96	168,731
Black African	1,736	331	153	1,786	4,945	62,144	661	6,887	3,982	6,249	19,019	568	108,461
Black Caribbean	303	59	153	1,078	334	2,939	7,655	1,916	1,154	2,570	58,790	1,858	78,809
Missing	10,478	5,753	4,744	24,459	24,363	16,206	3,679	21,563	7,073	63,538	596,372	39,381	817,609
Other	13,803	2,518	2,397	9,680	14,441	24,353	3,184	15,336	10,183	42,466	60,324	3,711	202,396
White Other	4,981	721	790	4,073	3,573	4,876	906	21,113	6,200	192,404	131,211	10,050	380,898
White British	5,946	2,509	3,067	21,425	10,530	9,010	17,365	33,579	12,200	163,615	5,715,973	372,119	6,367,338
White Irish	72	29	25	353	105	101	214	749	128	1,390	27,928	25,403	56,497
Mixed	3,387	977	1,351	5,606	5,967	7,253	2,346	4,160	3,215	11,823	50,965	3,512	100,562
Total	71,381	63,056	33,546	218,843	246,523	135,065	36,571	118,735	51,833	496,760	6,687,591	457,617	8,617,521

1.5 DISCUSSION

The coronavirus disease 2019 pandemic (COVID-19) has re-emphasised the importance of a better understanding of the many factors causing ethnic inequalities such as poorer living and working conditions as well as co-morbidities exacerbating infection and survival [13]. It is therefore timely to reassess the completeness and quality of ethnicity information in electronic healthcare records.

Rates of NHS recording of ethnicity in HES have improved over the period of this study, especially for the White British group between 1999 and 2009. The gaps in the ethnicity records in the 1990s and early 2000s, are however likely to limit studies of ethnic inequalities. The availability of patient names was more complete than NHS-recorded ethnicity during the entire study period. There are therefore good reasons to consider alternative ways to enhance the ethnicity records either by linkage or by using name-based ethnicity classification softwares. We report on the latter approach.

NHS-recorded ethnicity supplemented with surname-based ethnicity yielded the highest completeness across the years with a few exceptions. The main exception was that using surnames alone would assign eight times more patients to White Irish background than recorded by the NHS. An Irish surname alone is in other words not a very strong predictor of individuals perceiving themselves as Irish. This is likely due to the long migration history of people from Ireland to Great Britain. The regional data showed that the prediction success for the White Irish group was higher in London than other regions, which may reflect that London has more first generation Irish migrants who still perceive themselves as Irish [10].

We would expect that the full name-based estimation would lead to higher sensitivity in the prediction. Empirically, however, we found the opposite, surname-based estimation would outperform full name-based estimation in predicting self-reported, NHS-recorded ethnicity. In the subsequent analyses, we therefore focused on the surname-based estimation. Kandt and Longley (2018) came to similar conclusions finding that forenames added little to ethnicity estimations that were based only on surnames [10]. It should in this context be mentioned that classifications based on groups of closely associated forenames and surnames are also available, i.e. the methodology used for creating the related Onomap software [10,14]. Onomap was validated against the Scottish birth registration database for 2004-2008, with slightly higher sensitivity for White, South Asian, and Chinese names than found in this study [14]. The reported sensitivity for Black African names was however as low as 25% (compared with nearly 50% in this study). It should be noted that is not possible to make a direct comparison in this case as the name-based classification methodology (surname vs forename-surname groups), study population (England HES vs Scottish birth register), and study period (1999-2011 vs. 2004-2008) were not identical.

The sensitivity of the EE software in correctly predicting NHS-recorded ethnicity was stable over time, except for the White Other group where success rates improved over time, possibly following successive EU enlargements. The sensitivity, in 2013, was nearly 90% for the White British group followed by those of Pakistani, Indian, Chinese, or Bangladeshi extraction (63%-82%) and was close to 50% for the Black African, White Other, and White Irish groups. For other ethnic groups the sensitivity was very low and none at all for mixed ethnic groups. The surname prediction results are broadly comparable with those reported in an analysis of Census 2011 microdata for England and Wales using the same name-based classification [10]. For the comparison, it should be noted that the HES only covers England and that the HES patient population is skewed towards older individuals, whereas the Census is designed to cover the entire residential population. The geographical coverage (England vs. England and Wales) and time period (1999-2013 vs. 2011) of the two sources were not identical but overlapping.

In many cultures women change surname upon marriage and this can lead to lower prediction success. The prediction success within ethnic groups were however similar in HES for males and females (Figure 3). This may in part be because we used the earliest name for name-based coding if there had been any changes over time. The prediction success was notably higher for females than males among Bangladeshis. This was also found in a study of 2011 Census microdata [10].

The problem of imputing ethnicity in NHS databases has previously been considered by other authors. Ryan et al. (2012) who used Onomap and Nam Pehchan to impute the ethnicity of White, South Asian, Black and Other groups in the UK's West Midlands [8]. Nam Pehchan is based on distinctive surnames and Onomap is based on clusters of closely associated forenames and surnames. Ryan et al. 2012 used a multiple imputation strategy with characteristics of the individual patients, their care, and the ethnic composition of their neighbourhoods: they reported that the sensitivity of the multiple imputation was above 90% for White and South Asian ethnicities but was very low for other groups. Smith et al. 2017 used the Onomap software to assign children and young people with cancers to either White, South Asian, or Other groups in a Yorkshire study, concluding that combining different data sources including names-based ones increased the representation of ethnic minorities, albeit with some ambiguity [9]. Both studies concluded that there is no perfect substitute for more complete self-reported ethnicity data.

The Scottish NHS presents a parallel case with many similarities [15]. Knox et al. (2019) imputed missingness in the Scottish hospital admission database in two ways [15]. First, by assigning last recorded ethnic group to previous records for each patient. Second, by assigning remaining patients to an ethnic group based on the distribution of the different ethnic groups by sex and 5-year age band under a missing-at-random assumption. Knox et al. were in this way able to increase the completeness of the ethnicity information from 76% to 100%. The unevenness of ethnicity recording for different groups over time in both England and Scotland does not support the missing-at-random assumption [5,15], which indicate that further work is required to either collect more accurate on ethnicity or develop more sophisticated methods of imputation.

When the sensitivity of EE was mapped it also showed that successful prediction was greatest outside London. This is likely to be a compositional effect with a significant presence of groups with relatively low prediction success. This is supported by the fact that the prediction success in London within most ethnic groups was among the highest in the country.

There is currently no good surname distinction of, e.g. patients of Black Caribbean background. Future work may consider involving geographical and demographic data to improve the prediction for these groups. This may however be challenging, especially, for rarer or more geographically dispersed ethnic groups [8,15].

In summary, studies of ethnicity in HES, 1999-2013, are compounded by a number of caveats. The completeness of ethnicity data was below 60% in 1999. It improved in the 2000s and reached a plateau of 89-90% in 2009-2013. The completeness of patient surnames improved from 79% in 1999 to 93% in 2002; then gradually improved to 99% in 2013. If patient names are used for ethnicity estimation, it should be noted that the sensitivity (prediction success) varies by ethnic group, e.g. it is close to 90% for White British and approximately 50% for Black African. For White Other, the sensitivity notably increased from 10.5% in 1999 to 50.5% in 2013. We also found that the surname estimation inflated the White Irish group considerably relative to individuals reporting themselves as White Irish. The representation of different ethnic groups in HES could potentially be improved by retrospective linkage to the patient register or other data sources with better quality data. Names-based classification can however be a method for estimating ethnicity in studies where linkage is not feasible.

Limitations

As a limitation, it should be noted that ethnicity is a complex concept encompassing biological, cultural, and subjective aspects. Which aspect matters most depends on the kind of inequalities that are the object of the study and the related assumptions about disease aetiology. Variation in prediction success of name-based ethnicity classification can therefore arise for different reasons including individuals' sense of belonging and resulting choice of ethnic group, socio-cultural naming and name-change practices, distinctiveness of names across ethnic groups, and the extent to which the name-based classification covers different origins at a given time point, e.g. when later waves of immigration have widened the range of diasporic names in the host country since the creation of the software. More detailed analysis of non-matching ethnicity predictions can help to disentangle these different aspects but were outside the scope of the current study.

1.6 Conclusion

Studies of ethnic inequalities in hospital inpatient care in England are limited by incomplete data on patient ethnicity in the 1990s and 2000s. Financial incentives for general practitioners to collect and report ethnicity to the central patient register between 2006 and 2012 have greatly improved completeness during this period. Personal names of patients remain an untapped source for closing this gap for the earlier years. As demonstrated in this - and other studies - name-based ethnicity classifications have merit for the predictions of many ethnic minorities. The case for name-based ethnicity classification is naturally stronger for databases where ethnicity is not collected systematically, e.g. accident and emergency department data [16] or more recently COVID-19 admissions in the Welsh hospital admission database [7]. The current work also highlights areas where name-based ethnicity classifications can be improved. There is currently no good surname distinction of, e.g. patients of Black Caribbean background. Future work may consider involving geographical and demographic data to improve the prediction for these groups.

What is already known on this subject

- Studies of ethnic inequalities in hospital inpatient care in England are limited by incomplete data on patient ethnicity collected in the 1990s and 2000s.
- Name-based ethnicity classifications have merit for the predictions of many ethnic minorities, yet there has to date not been any studies of the completeness of personal names in HES or the prediction success of name-based classifications over time.

What this study added to our knowledge

- The prediction success of a names-based ethnicity classification tool has been quantified in HES for the first time and the results can be used to inform decisions around the optimal analysis of ethnic groups using this data source.
- The work also highlights areas where name-based ethnicity classifications can be improved, e.g. for patients of Black Caribbean background.
- Future work may consider involving geographical and demographic data to improve the prediction for these groups.

2 ETHNIC INEQUALITIES IN PREVENTABLE HOSPITAL ADMISSIONS

2.1 INTRODUCTION

Emergency hospital admissions are distressing for patients, associated with poorer long-term outcomes, and are costly to the healthcare system. Many healthcare systems are therefore undergoing reforms to reduce emergency admissions by improving early detection, treatment and monitoring of a range of conditions in less intensive settings, i.e. primary and community care services [17,18]. These conditions are known as ambulatory care sensitive conditions (ACSC). ACSC include acute, chronic, and vaccine-preventable conditions such as urinary tract infections, chronic obstructive pulmonary disease (COPD), and pneumonia. ACSC admissions have been associated with patients under the age 5 years, the elderly, deprivation, and ethnicity [17].

The English NHS saw a 40% rise in ACSC admissions in 2001-2011 [17] and a 42% rise in emergency admissions between 2006 and 2017 making this a policy area of urgency [19]. ACSC indicators were introduced into the NHS Commissioning Outcome Framework in 2012 to monitor this area for quality of care improvements for the general population [17]. Whilst ACSC has been studied before in England, there has to our knowledge not been a study of ethnic inequalities in ACSC in England nor of its geographical distribution for these groups. A study of ACSC is particularly pertinent for the understanding of ethnic inequalities, because they are indicative of how patients from different minorities access and navigate the healthcare system. Studies in US, New Zealand, and Scotland have found higher risk of ACSC admission for many ethnic minorities compared to the White majority populations [20–22]. A recent Scottish study found that South Asian groups had higher risk of ACSC admission compared to the White majority group [20].

For this study we gathered data on hospital admission from Hospital Episode Statistics (HES) for different ethnic groups over a five-year period, 2009-2013, and linked them to the 2011 Census population estimates.

2.2 METHODS

Inpatient hospital admission records with an emergency admission route were obtained from NHS England's Hospital Episode Statistics (HES), April 2009-March 2014. Diagnoses in HES are coded to the International Classification of Diseases (ICD10) system [23]. The primary diagnosis of the first episode in spells with emergency admission were coded with definitions for acute ambulatory care sensitive conditions (ACSC), chronic ACSC, and vaccine-preventable diseases [17]. HES-recorded ethnic group was used but supplemented with surname-coded ethnicity information where missing using the Ethnicity Estimator (EE) software [10,12]. To retain full anonymity, the surname coding was carried out in an air-gapped, secure data facility by NHS Digital linking name information in the Patient Demographic Service to HES. HES-recorded ethnicity categories for Not-Stated and Missing were combined, and Black Other was combined with Other. For Census base population data, Arabic and Black Other were combined with Other to harmonise the two data sources.

We validated the surname-based ethnicity against HES-recorded ethnicity by calculating the proportion of ethnicities correctly predicted by the EE software for each group, aka diagnostic sensitivity. We found that surname imputation overestimates the White Irish group and the results for this group were omitted as it would not be possible to create accurate population estimates based on surnames.

Incidence per 100,000 population standardised by age and sex according to the European Standard Population was calculated by ethnic group and local authority district for a combined ACSC outcome (acute, chronic, and vaccine-preventable). The results for areas with less than 20 cases were suppressed.

2.3 DATA PRODUCT

The data are available as a comma separated text file (file name: oslaua_acsc.csv).

Variable	Description
oslaua	Local Authority district code (2016)
r_[ethnic group]	Age- and sex-standardised preventable hospitalisation incidence per 100,000 pop
r_ll_[ethnic group]	Standardised incidence 95% CI lower
r_ul_[ethnic group]	Standardised incidence 95% CI upper
rz_[ethnic group]	z-score of r_[ethnic group] relative to national mean and SD of White British group
rzcat_[ethnic group]	z-score categories (label): -1000 (Less than -4 SD), -4 (-4 to -2.1 SD), -2 (-2 to -1.1 SD), -1 (-1 to -0.1 SD), 0 (0-0.9 SD), 1 (1-1.9 SD), 2 (2 to 3.9 SD), 4 (4 SD and more), -999 (Low count: <20 cases)

Ethnic groups included: Asian Other, Bangladeshi, Chinese, Indian, Pakistani, Black African, Black Caribbean, Other, White Other, White British, Mixed. Missing value code for suppression of cells based on <20 cases: "-999".

2.4 METADATA

Field	Value
Title	Ethnic inequalities in preventable hospital admissions
Dataset description	Age- and sex-standardised hospital admission incidence per 100,000 population by ethnic group and local authority district for acute ambulatory care sensitive conditions (acute, chronic, and vaccine-preventable combined). The hospital admission data were enhanced with ethnicity coding of participants' names using the Ethnicity Estimator (EE) software.
Access status	Safeguarded
Attributes	oslaua: Local Authority district code (2016) r_[ethnic group]: Age- and sex-standardised preventable hospitalisation incidence per 100,000 pop r_ll_[ethnic group]: Standardised incidence 95% CI lower r_ul_[ethnic group]: Standardised incidence 95% CI upper rz_[ethnic group]: Z-score of r_[ethnic group] relative to national mean and SD of White British group rzcat_[ethnic group]: Z-score categories: z-score categories (label): -1000 (Less than -4 SD), -4 (-4 to -2.1 SD), -2 (-2 to -1.1 SD), -1 (-1 to -0.1 SD), 0 (0-0.9 SD), 1 (1-1.9 SD), 2 (2 to 3.9 SD), 4 (4 SD and more), -999 (Low count: <20 cases)
Controller	University College London (UCL)
Time period	April 2009 – March 2014
Source	Hospital Episode Statistics
Funder	The UK Economic and Social Research Council (Grant ES/L011840/1).
Data and resources	Ethnic inequalities in preventable hospital admission (oslaua_acsc.csv) CDRC Hospital Episode Statistics Ethnicity data products (cdrc_hes_ethnicity_data_documentation.pdf)
Modified	N/A
Release date	TBC
Homepage URL	TBC
Spatial/geographical coverage location	England
Granularity	Local authority district (2016)
Minimum embounding rectangle	Latitude (49.89198 to 55.79742)

	Longitude (-6.352647 to 1.760443)
Author	Jakob Petersen, Jens Kandt, Paul Longley
Contact name	Longley, Paul
Contact email	data@cdrc.ac.uk

3 Ethnic inequalities in hospital admissions by major disease groups

3.1 Introduction

Reducing inequalities in health has explicitly been part of the government agenda in the United Kingdom since 1997 [24,25]. Inequalities are associated with poverty and may be exacerbated for ethnic minorities due to discrimination, lack of health knowledge or other barriers in access to health services such as language [26]. National Health Service (NHS) hospitals monitor their use by ethnic group in the national database, Hospital Episode Statistics (HES). In this study we analyse hospital admission records by ethnic group across all major disease categories in the Global Burden of Disease (GBD) classification in 2009-2013 [27].

3.2 Methods

Hospital admission records were obtained from NHS England's Hospital Episode Statistics (HES), April 1999-March 2004 and April 2009-March 2014. Diagnoses in HES are coded to the International Classification of Diseases (ICD10) system [23]. Data on the primary diagnosis for each admission were coded with definitions for the GBD conditions (Level 1).

If a patient was re-admitted within two days, only the first admission was counted. NHS-recorded ethnic group was used but replaced with surname-coded ethnicity information where missing using the Ethnicity Estimator (EE) software [10,12]. To retain full anonymity, this step was carried out in an air-gapped, secure data facility by NHS Digital linking name information in the Patient Demographic Service to HES. The sensitivity of the EE software in predicting the self-reported ethnic group was calculated. Self-reported ethnicity categories for Not-Stated and Missing were combined, and Black Other was combined with Other. For Census base population data, Arabic and Black Other were combined with Other. Preliminary work showed that surname imputation tends to inflate the White Irish group relative to self-reported data and the admission results for White Irish are not shown.

The incidence of hospital admission was calculated for each GBD1 disease category at local authority level and standardised by age and sex using Census denominators and 2013 European Standard Population weights [28]. The incidence estimates are available for two time periods, 1999/00-2003/04 and 2009/10-2013/14 using the nearest Census population estimates as the base population. Counts below 20 cases were suppressed.

Table 1 Global Burden of Disease (Level 1) categories.

GBD Level 1 Disease categories
1A Infectious and parasitic diseases
1B Respiratory infections
1C Maternal conditions
1D Perinatal conditions
1E Nutritional deficiencies
2A Malignant neoplasms
2B Other neoplasms
2C Diabetes mellitus
2D Endocrine disorders
2E Neuro-psychiatric conditions
2F Sense organ diseases
2G Cardiovascular diseases
2H Respiratory diseases
2I Digestive diseases
2J Genito-urinary diseases
2K Skin diseases
2L Musculoskeletal diseases
2M Congenital anomalies
2N Oral conditions
30 Injuries
X102 Nonspecific chest pain
X176 Contraceptive and procreative management
X251 Abdominal pain
X257 Other aftercare
X259 Residual codes – unclassified
XR Symptoms signs and abnormal clinical and laboratory
XZ Factors influencing health status and contact with health

3.3 DATA PRODUCT

The data are available as a comma-separated file for 1999/00-2003/04 (file name: inci_oslaua_01_ethcat.csv) and 2009/10-20013/04 (file name: inci_oslaua_11_ethcat.csv), respectively.

Variable	Description
Oslaua	Local Authority district code (2016)
r_[ethnic group]	Age- and sex-standardised incidence per 100,000 pop
r_ul_[ethnic group]	95% CI lower limit
r_ll_[ethnic group]	95% CI upper limit
rzcat_[ethnic group]	z-score category of standardised all cause hospitalisation incidence per 100,000 pop:, -1000 (-4 SD and less), -4 (-4 to 2.1 SD), -2 (-2 to -1 SD), -1 (-1 to -0.1 SD), 0 (0-0.9 SD), 1 (1-1.9 SD), 2 (2 to 3.9 SD), 4 (4 SD and more), -999 (Low count: cases <20)
Ethcat	1) NHS-recorded; 2) NHS-recorded with surname imputation; 3) Surname prediction; 4) Prediction with forename and surname agreement
gbd1a	GBD1 disease code
gbd_label	GBD1 disease label
Census	“01” for 1999-2003 and “11” for 2009-2013

Ethnic groups included: Asian Other, Bangladeshi, Chinese, Indian, Pakistani, Black African, Black Caribbean, Other, White Other, White British, Mixed.

3.4 Metadata

Field	Value
Title	Ethnic inequalities in hospital admissions by major disease groups
Dataset description	Age- and sex-standardised hospital admission incidence per 100,000 population by ethnic group and local authority district for Global Burden of Disease Level 1 conditions (GBD). The hospital admission data were enhanced with ethnicity coding of participants' names using the Ethnicity Estimator (EE) software.
Access status	Safeguarded
Attributes	oslaua: Local Authority district code (2016) r_[ethnic group]: Age- and sex-standardised incidence per 100,000 pop r_ul_[ethnic group]: 95% CI lower limit r_ll_[ethnic group]: 95% CI upper limit rzcat_[ethnic group]: z-score category of standardised all cause hospitalisation incidence per 100,000 pop:, -1000 (-4 SD and less), -4 (-4 to 2.1 SD), -2 (-2 to -1 SD), -1 (-1 to -0.1 SD), 0 (0-0.9 SD), 1 (1-1.9 SD), 2 (2 to 3.9 SD), 4 (4 SD and more), -999 (Low count: cases <20) ethcat: 1) NHS-recorded; 2) NHS-recorded with surname imputation; 3) Surname prediction; 4) Prediction with forename and surname agreement gbd1a: GBD1 disease code gbd_label: GBD1 disease label Census: “01” for 1999-2003 and “11” for 2009-2013
Controller	University College London (UCL)
Time period	April 1999 – March 2004, April 2009 – March 2014
Source	Hospital Episode Statistics
Funder	The UK Economic and Social Research Council (Grant ES/L011840/1).
Data and resources	Ethnic inequalities in hospital admissions by major disease groups 1999/00-2003/04 (inci_oslaua_01_ethcat.csv) Ethnic inequalities in hospital admissions by major disease groups 2009/10-20013/04 (inci_oslaua_11_ethcat.csv) CDRC Hospital Episode Statistics Ethnicity data products (cdrc_hes_ethnicity_data_documentation.pdf)
Modified	N/A
Release date	TBC
Homepage URL	TBC
Spatial/geographical coverage location	England
Granularity	Local authority district (2016)
Minimum embounding rectangle	Latitude (49.89198 to 55.79742) Longitude (-6.352647 to 1.760443)
Author	Jakob Petersen, Jens Kandt, Paul Longley
Contact name	Longley, Paul
Contact email	data@cdrc.ac.uk

4 REFERENCES

[1] European Union, Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e40-1-1 (accessed August 20, 2020).

[2] B. Byrne, C. Alexander, O. Khan, J. Nazroo, W. Shankley, Ethnicity, Race and Inequality in the UK - State of the Nation, Policy Press, 2020. https://policy.bristoluniversitypress.co.uk/ethnicity-race-and-inequality-in-the-uk (accessed November 4, 2020).

[3] The Economist, American ideas about racism are influencing Europe, The Economist. (2020). https://www.economist.com/europe/2020/08/08/american-ideas-about-racism-are-influencing-europe (accessed August 19, 2020).

[4] The Economist, An edgy inquiry, The Economist. (2015). https://www.economist.com/europe/2015/04/04/an-edgy-inquiry (accessed August 19, 2020).

[5] R. Mathur, K. Bhaskaran, N. Chaturvedi, D.A. Leon, T. vanStaa, E. Grundy, L. Smeeth, Completeness and usability of ethnicity data in UK-based primary care and hospital databases, J. Public Health Oxf. Engl. 36 (2014) 684–692. https://doi.org/10.1093/pubmed/fdt116.

[6] R.W. Aldridge, D. Lewer, S.V. Katikireddi, R. Mathur, N. Pathak, R. Burns, E.B. Fragaszy, A.M. Johnson, D. Devakumar, I. Abubakar, A. Hayward, Black, Asian and Minority Ethnic groups in England are at increased risk of death from COVID-19: indirect standardisation of NHS mortality data, Wellcome Open Res. 5 (2020). https://doi.org/10.12688/wellcomeopenres.15922.2.

[7] D.R. Thomas, O. Orife, A. Plimmer, C. Williams, G. Karani, M.R. Evans, P.A. Longley, J. Janiec, R. Saltus, A.G. Shankar, Ethnic variation in outcome of people hospitalised with Covid-19 in Wales (UK): A rapid analysis of surveillance data using Onomap, a name-based ethnicity classification tool, (In prep.).

[8] R. Ryan, S. Vernon, G. Lawrence, S. Wilson, Use of name recognition software, census data and multiple imputation to predict missing data on ethnicity: application to cancer registry records, BMC Med. Inform. Decis. Mak. 12 (2012) 3. https://doi.org/10.1186/1472-6947-12-3.

[9] L. Smith, P. Norman, M. Kapetanstrataki, S. Fleming, L.K. Fraser, R.C. Parslow, R.G. Feltbower, Comparison of ethnic group classification using naming analysis and routinely collected data: application to cancer incidence trends in children and young people, BMJ Open. 7 (2017). https://doi.org/10.1136/bmjopen-2017-016332.

[10] J. Kandt, P.A. Longley, Ethnicity estimation using family naming practices, PLOS ONE. 13 (2018) e0201774. https://doi.org/10.1371/journal.pone.0201774.

[11] P. Mateos, P.A. Longley, D. O'Sullivan, Ethnicity and Population Structure in Personal Naming Networks, PLOS ONE. 6 (2011) e22943. https://doi.org/10.1371/journal.pone.0022943.

[12] J. Kandt, J. Van Dijk, P.A. Longley, Family name origins and inter-generational demographic change in Great Britain, Am. Geogr. Soc. 110 (2020) 1726–1742.

[13] N. Bhala, G. Curry, A.R. Martineau, C. Agyemang, R. Bhopal, Sharpening the global focus on ethnicity and race in the time of COVID-19, The Lancet. 395 (2020) 1673–1676. https://doi.org/10.1016/S0140-6736(20)31102-8.
[14] F. Lakha, D.R. Gorman, P. Mateos, Name analysis to classify populations by ethnicity in public health: Validation of Onomap in Scotland, Public Health. 125 (2011) 688–696. https://doi.org/10.1016/j.puhe.2011.05.003.
[15] S. Knox, R.S. Bhopal, C.S. Thomson, A. Millard, A. Fraser, L. Gruer, D. Buchanan, The challenge of using routinely collected data to compare hospital admission rates by ethnic group: a demonstration project in Scotland, J. Public Health. fdz175 (2019). https://doi.org/10.1093/pubmed/fdz175.
[16] J. Petersen, P. Longley, M. Gibin, P. Mateos, P. Atkinson, Names-based classification of accident and emergency department users, Health Place. 17 (2011) 1162–1169. https://doi.org/10.1016/j.healthplace.2010.09.010.
[17] M. Bardsley, I. Blunt, S. Davies, J. Dixon, Is secondary preventive care improving? Observational study of 10-year trends in emergency admissions for conditions amenable to ambulatory care, BMJ Open. 3 (2013). https://doi.org/10.1136/bmjopen-2012-002007.
[18] J. Busby, S. Purdy, W. Hollingworth, How do population, general practice and hospital factors influence ambulatory care sensitive admissions: a cross sectional study, BMC Fam. Pract. 18 (2017) 67. https://doi.org/10.1186/s12875-017-0638-9.
[19] K. Hodgson, S.R. Deeny, A. Steventon, Ambulatory care-sensitive conditions: their potential uses and limitations, BMJ Qual. Saf. (2019). https://doi.org/10.1136/bmjqs-2018-008820.
[20] S.V. Katikireddi, G. Cezard, R.S. Bhopal, L. Williams, A. Douglas, A. Millard, M. Steiner, D. Buchanan, A. Sheikh, L. Gruer, Assessment of health care, hospital admissions, and mortality by ethnicity: population-based cohort study of health-system performance in Scotland, Lancet Public Health. 3 (2018) e226–e236. https://doi.org/10.1016/S2468-2667(18)30068-9.
[21] I. Blunt, Focus on preventable admissions, Health Found. (2013). https://www.health.org.uk/publications/qualitywatch-focus-on-preventable-admissions (accessed July 16, 2020).
[22] T. Dalla Zuanna, T. Spadea, M. Milana, A. Petrelli, L. Cacciani, L. Simonato, C. Canova, Avoidable hospitalization among migrants and ethnic minority groups: a systematic review, Eur. J. Public Health. 27 (2017) 861–868. https://doi.org/10.1093/eurpub/ckx113.
[23] WHO, ICD-10 Version:2016, (2016). https://icd.who.int/browse10/2016/en (accessed January 17, 2020).
[24] J.P. Mackenbach, Can we reduce health inequalities? An analysis of the English strategy (1997–2010), J. Epidemiol. Community Health. 65 (2011) 568–575. https://doi.org/10.1136/jech.2010.128280.
[25] M. Marmot, J. Allen, T. Boyce, P. Goldblatt, J. Morrison, Health Equity in England: The Marmot Review 10 Years On, Health Found. (2020). https://www.health.org.uk/publications/reports/the-marmot-review-10-years-on (accessed July 10, 2020).
[26] R.S. Bhopal, Migration, Ethnicity, Race, and Health in Multicultural Societies, OUP Oxford, 2014.
[27] N. Steel, J.A. Ford, J.N. Newton, A.C.J. Davis, T. Vos, M. Naghavi, S. Glenn, A. Hughes, A.M. Dalton, D. Stockton, C. Humphreys, M. Dallat, J. Schmidt, J. Flowers, S. Fox, I. Abubakar, R.W. Aldridge, A. Baker, C. Brayne, T. Brugha, S. Capewell, J. Car, C. Cooper, M. Ezzati, J. Fitzpatrick, F. Greaves, R. Hay, S. Hay, F. Kee, H.J. Larson, R.A.

Lyons, A. Majeed, M. McKee, S. Rawaf, H. Rutter, S. Saxena, A. Sheikh, L. Smeeth, R.M. Viner, S.E. Vollset, H.C. Williams, C. Wolfe, A. Woolf, C.J.L. Murray, Changes in health in the countries of the UK and 150 English Local Authority areas 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016, The Lancet. 392 (2018) 1647–1661. https://doi.org/10.1016/S0140-6736(18)32207-4.

[28] Eurostat, Revision of the European Standard Population - Report of Eurostat's task force - 2013 edition, (2013). https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-13-028 (accessed May 21, 2019).

5 Appendix A. Supplementary materials

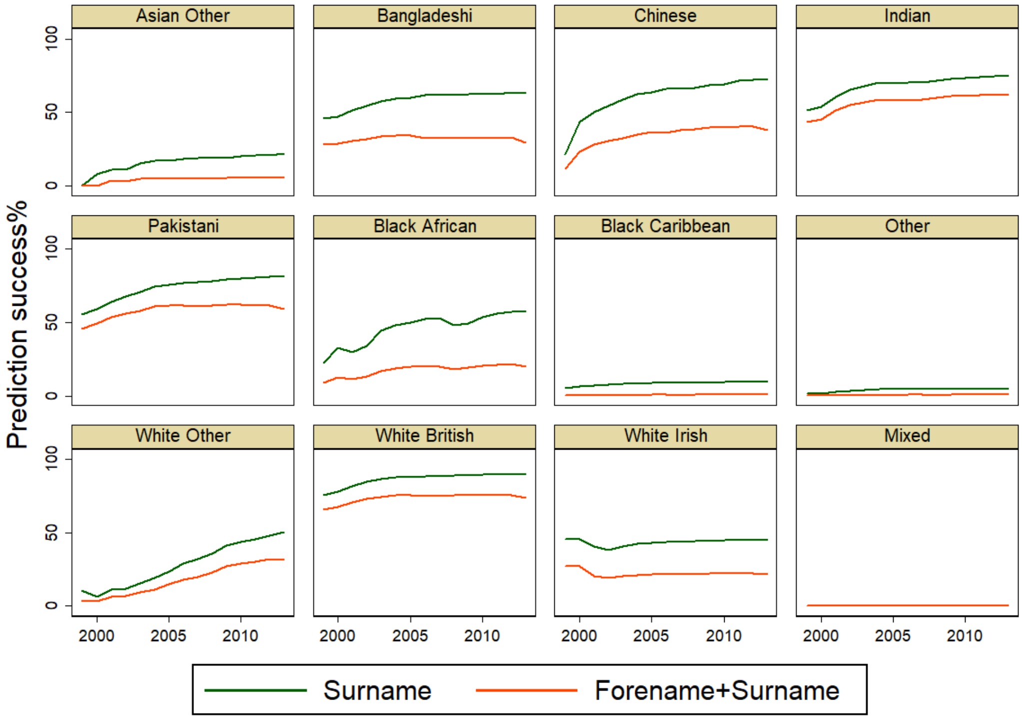

Figure S1 Sensitivity% of Ethnicity Estimator (EE) software (Kandt & Longley, 2018) in predicting NHS ethnic group from either patient surname or where both forename and surname point to the same ethnic group.

Table S1 Sensitivity and specificity for EE surname predicted ethnicity (Kandt & Longley, 2018) of HES patients in 1999-2013.

Financial year	Ethnic group	NHS-recorded	EE-predicted	Sensitivity	Specificity
1999	Asian Other	<5	13,156	-	-
	Bangladeshi	18,126	23,174	46.0	99.8
	Chinese	14,383	9,496	21.6	99.9
	Indian	72,266	89,169	51.7	99.2
	Pakistani	65,850	84,288	55.3	99.3
	Black African	34,870	27,898	22.8	99.7
	Black Caribbean	50,430	20,041	5.6	99.7
	Other	23,719	17,732	1.9	99.7
	White Other	77	157,001	10.4	97.7
	White British	3,712,919	4,591,702	75.6	40.4
	White Irish	11	297,003	45.5	95.6
	Mixed	<5	0	-	-
	Missing	2,715,319	1,377,316	-	-
	Total	6,707,976	6,707,976		
2000	Asian Other	36	14,435	8.3	99.8
	Bangladeshi	18,624	25,353	47.2	99.8

	Chinese	7,825	10,463	43.9	99.9
	Indian	67,538	91,851	53.7	99.2
	Pakistani	64,007	90,632	59.3	99.2
	Black African	24,955	30,646	33.1	99.7
	Black Caribbean	42,498	20,412	6.6	99.7
	Other	24,201	18,738	1.9	99.7
	White Other	1,093	165,602	6.3	97.5
	White British	3,601,869	4,702,322	78.2	37.6
	White Irish	92	305,388	45.7	95.4
	Mixed	57	-	-	-
	Missing	2,769,608	0	-	-
	Total	6,622,403	6,622,403		
2001	Asian Other	9,928	16,826	11.0	99.8
	Bangladeshi	19,800	27,738	51.4	99.7
	Chinese	7,610	11,355	50.6	99.9
	Indian	68,800	100,061	60.8	99.1
	Pakistani	67,369	101,609	64.4	99.1
	Black African	32,601	35,760	30.2	99.6
	Black Caribbean	39,252	21,625	7.5	99.7
	Other	59,545	20,588	2.9	99.7
	White Other	121,380	177,325	11.3	97.4
	White British	3,382,150	4,825,563	81.7	33.9
	White Irish	15,138	315,110	40.8	95.2
	Mixed	14,967	0	-	-
	Missing	2,655,192	850,172	-	-
	Total	6,503,732	6,503,732		
2002	Asian Other	23,826	21,550	11.0	99.7
	Bangladeshi	22,642	31,566	54.6	99.7
	Chinese	8,632	12,989	54.8	99.9
	Indian	73,857	114,204	65.8	99.0
	Pakistani	70,915	114,888	67.7	99.0
	Black African	37,964	45,600	34.3	99.5
	Black Caribbean	41,366	23,404	8.0	99.7
	Other	84,150	23,651	3.5	99.7
	White Other	216,932	197,110	11.9	97.3
	White British	3,365,488	5,088,451	84.7	31.3
	White Irish	27,644	332,201	38.3	95.1
	Mixed	24,695	0	-	-
	Missing	2,621,287	613,784	-	-
	Total	6,619,398	6,619,398		
2003	Asian Other	28,431	27,135	15.6	99.7
	Bangladeshi	27,104	36,773	57.9	99.7
	Chinese	10,913	15,594	58.7	99.9
	Indian	87,169	126,923	67.9	99.0
	Pakistani	79,385	131,610	71.1	98.9
	Black African	47,603	60,331	44.8	99.4

	Black Caribbean	52,838	25,983	8.5	99.7
	Other	91,886	27,387	4.3	99.7
	White Other	234,798	223,061	15.4	97.2
	White British	3,986,832	5,370,461	86.3	33.1
	White Irish	34,159	353,615	40.6	95.0
	Mixed	34,838	0	-	-
	Missing	2,153,680	470,763	-	-
	Total	6,869,636	6,869,636		
2004	Asian Other	35,623	32,034	17.1	99.6
	Bangladeshi	29,303	39,390	59.3	99.7
	Chinese	12,671	17,606	62.7	99.9
	Indian	93,863	136,211	70.4	99.0
	Pakistani	86,772	144,524	74.4	98.8
	Black African	56,258	72,616	48.5	99.3
	Black Caribbean	55,794	27,448	8.7	99.7
	Other	102,834	30,331	4.7	99.6
	White Other	238,616	241,650	19.0	97.1
	White British	4,263,928	5,479,611	87.7	34.7
	White Irish	38,073	362,008	42.5	95.0
	Mixed	44,125	0	-	-
	Missing	1,873,720	348,151	-	-
	Total	6,931,580	6,931,580		
2005	Asian Other	43,868	36,246	17.5	99.6
	Bangladeshi	32,765	41,822	59.9	99.7
	Chinese	14,166	19,429	64.0	99.9
	Indian	103,485	146,058	69.9	99.0
	Pakistani	97,667	155,534	75.5	98.8
	Black African	67,140	84,162	50.1	99.3
	Black Caribbean	59,796	28,542	8.9	99.7
	Other	115,873	32,858	4.7	99.6
	White Other	236,102	264,063	23.6	97.0
	White British	4,594,942	5,641,839	88.0	37.7
	White Irish	41,375	375,948	43.4	95.0
	Mixed	47,375	0	-	-
	Missing	1,708,425	336,478	-	-
	Total	7,162,979	7,162,979		
2006	Asian Other	54,888	43,566	18.5	99.6
	Bangladeshi	38,429	49,203	61.8	99.7
	Chinese	17,136	22,687	66.0	99.9
	Indian	115,159	159,409	70.5	98.9
	Pakistani	115,001	178,779	76.8	98.8
	Black African	83,569	100,947	52.6	99.2
	Black Caribbean	64,932	30,403	9.0	99.7
	Other	138,160	37,447	4.7	99.6
	White Other	264,774	304,575	28.8	96.8
	White British	4,988,936	5,872,779	88.2	41.1

	White Irish	45,084	394,932	43.6	95.0
	Mixed	60,844	0	-	-
	Missing	1,502,812	294,997	-	-
	Total	7,489,724	7,489,724		
2007	Asian Other	59,969	46,589	19.2	99.7
	Bangladeshi	39,705	49,626	62.4	99.7
	Chinese	19,382	24,136	66.0	99.9
	Indian	124,427	164,994	70.8	99.0
	Pakistani	119,918	181,086	77.4	98.8
	Black African	85,228	102,001	53.1	99.2
	Black Caribbean	66,372	30,961	9.3	99.7
	Other	141,980	38,275	4.8	99.6
	White Other	274,193	322,128	32.1	96.8
	White British	5,298,702	5,964,190	88.5	44.6
	White Irish	48,409	400,350	44.0	95.0
	Mixed	62,716	0	-	-
	Missing	1,254,369	271,034	-	-
	Total	7,595,370	7,595,370		
2008	Asian Other	66,465	49,467	19.3	99.5
	Bangladeshi	39,887	50,299	61.7	99.7
	Chinese	20,594	25,639	67.1	99.9
	Indian	131,051	174,872	71.6	98.9
	Pakistani	126,306	189,263	78.2	98.8
	Black African	92,490	103,164	48.4	99.2
	Black Caribbean	69,370	31,986	9.3	99.7
	Other	159,578	40,204	4.7	99.6
	White Other	280,085	345,940	35.7	96.7
	White British	5,615,453	6,137,686	88.9	47.7
	White Irish	50,896	413,067	44.3	95.0
	Mixed	66,510	0	-	-
	Missing	1,082,065	239,163	-	-
	Total	7,800,750	7,800,750		
2009	Asian Other	74,103	53,168	19.3	99.5
	Bangladeshi	41,416	52,052	62.2	99.7
	Chinese	22,387	27,263	68.8	99.9
	Indian	139,818	182,948	73.1	99.0
	Pakistani	134,900	198,030	79.5	98.8
	Black African	99,466	110,407	49.7	99.2
	Black Caribbean	73,251	32,968	9.3	99.7
	Other	167,224	42,445	4.9	99.6
	White Other	283,574	372,150	41.2	96.7
	White British	5,913,602	6,244,019	89.1	51.8
	White Irish	53,043	420,879	44.4	95.0
	Mixed	75,711	0	-	-
	Missing	854,286	196,452	-	-
	Total	7,932,781	7,932,781		

2010	Asian Other	80,849	58,071	20.0	99.5
	Bangladeshi	43,006	54,070	62.8	99.7
	Chinese	22,954	28,482	69.4	99.8
	Indian	147,886	192,774	73.4	98.9
	Pakistani	141,229	206,443	80.3	98.8
	Black African	98,678	115,411	53.7	99.2
	Black Caribbean	75,296	33,993	9.5	99.7
	Other	175,968	44,335	4.8	99.6
	White Other	308,908	399,330	44.0	96.6
	White British	6,091,457	6,369,906	89.5	54.3
	White Irish	54,738	430,983	45.0	95.0
	Mixed	75,297	0	-	-
	Missing	779,932	162,400	-	-
	Total	8,096,198	8,096,198		
2011	Asian Other	83,626	60,363	20.6	99.5
	Bangladeshi	43,842	54,746	62.8	99.7
	Chinese	23,169	29,370	71.5	99.8
	Indian	148,023	195,413	73.9	98.9
	Pakistani	147,017	213,961	80.7	98.8
	Black African	95,827	118,098	56.3	99.2
	Black Caribbean	75,438	34,176	9.6	99.7
	Other	177,800	45,170	4.9	99.5
	White Other	320,765	417,147	45.9	96.5
	White British	6,108,787	6,382,086	89.7	55.1
	White Irish	54,749	432,391	45.0	94.9
	Mixed	77,640	0	-	-
	Missing	757,530	131,292	-	-
	Total	8,114,213	8,114,213		
2012	Asian Other	85,956	62,284	21.0	99.5
	Bangladeshi	45,011	56,567	63.6	99.7
	Chinese	23,922	31,048	72.2	99.8
	Indian	149,170	199,871	74.6	98.9
	Pakistani	150,788	219,924	81.3	98.8
	Black African	95,648	121,242	57.2	99.2
	Black Caribbean	74,755	34,374	9.6	99.7
	Other	183,743	46,015	4.8	99.5
	White Other	329,775	436,016	47.9	96.5
	White British	6,129,165	6,413,441	89.9	55.7
	White Irish	54,843	436,048	44.8	94.9
	Mixed	79,890	0	-	-
	Missing	764,726	110,562	-	-
	Total	8,167,392	8,167,392		
2013	Asian Other	95,904	71,381	22.1	99.4
	Bangladeshi	51,100	63,056	63.4	99.6
	Chinese	25,537	33,546	72.9	99.8
	Indian	163,679	218,843	74.6	98.9

Pakistani	168,731	246,523	81.7	98.7
Black African	108,461	135,065	57.3	99.1
Black Caribbean	78,809	36,571	9.7	99.7
Other	202,396	51,833	5.0	99.5
White Other	380,898	496,760	50.5	96.3
White British	6,367,338	6,687,591	89.8	56.8
White Irish	56,497	457,617	45.0	95.0
Mixed	100,573	0	-	-
Missing	817,609	118,735	-	-
Total	8,617,521	8,617,521		