

## Data Profile: Hospital Episode Statistics Ethnicity Data Product –Hospital admission for Global Burden of Disease conditions by ethnic group

### Introduction

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 in Hospital Episode Statistics (HES) from more than half of records of patients who received inpatient care in England.

To address the gap in the completeness of ethnicity data, CDRC worked with NHS Digital to enhance HES data collected between 1999/00 and 2013/14 with name-based ethnicity imputation using the CDRC Ethnicity Estimator software (EE).

Details about the completeness of ethnicity records and the prediction success of the EE software for different ethnic groups over time and across regions can be found in the accompanying documentation for the CDRC HES Ethnicity data products. Two data products were produced in conjunction with this project: 1) Preventable hospitalisation admission by ethnic group; and 2) Hospital admission for major disease categories as defined in the Global Burden of Disease project.

This document concerns Hospital admission for major disease categories as defined in the Global Burden of Disease project.

Reducing inequalities in health has explicitly been part of the government agenda in the United Kingdom since 1997. 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. NHS-commissioned 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 1999/00-2003/04 and 2009/10-2013/14.

### Scale and Extent

Field	Value
Data Provider	NHS Digital
Analytical Units	Local authority districts
Data Format	Csv
Temporal Extent	1999/00-2003/04, 2009/10-2013/14
Geographical Extent	England
Variables	1999/00-2003/04: 14 variables 2009/10-2013/14: 14 variables
Observations	326

### Citation Information

The following statement should be included when citing the use of this dataset:

“The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC [Project Number], ES/L011840/1; ES/L011891/1”

### Data Classification and Access Summary

These data are classified as Safeguarded and are available only upon approved application. To make an initial application, please visit: <https://www.cdrc.ac.uk/data-services/using-our-data/>

### Content

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.

## Data Profile: Hospital Episode Statistics Ethnicity Data Product –Hospital admission for Global Burden of Disease conditions by ethnic group

---

### **Novelty**

The dataset on hospital admissions for major disease categories by ethnic group is novel in a few different aspects. First, it breaks down ethnicity by different ethnicity classifications using a combination of NHS-recorded ethnicity and names-based ethnicity. Second, it includes ethnic variation across a wide spectrum of disease categories. Third, it provides admission statistics by local authority district.

### **Quality, Representation and Bias**

HES has near-complete coverage of NHS commissioned hospital admissions in England. Coding of diagnoses may vary in consistency but has been validated for research and auditing purposes in an earlier study.

As a limitation, it should be noted that ethnicity is a complex concept encompassing biological, cultural, and subjective aspects. 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.

We used denominator data from Census 2011 as the most complete dataset on the ethnicity of the residential population in England. The census contains self-reported ethnicity. HES draws on the central NHS patient register with self-reported ethnicity. The proportion of patients without ethnicity record was 38.8% in 1999/00-2003/04 and 9.7% in 2009/10-2013/14. The dataset contain admission estimates by ethnic group as recorded by NHS as well as three different ethnicity classifications enhanced with names-based ethnicity.

The prediction success of the EE software is reported in the accompanying documentation.

### **Related Datasets**

Hospital Episode Statistics Ethnicity Data Product – Preventable hospitalisation.

Data Profile: Hospital Episode Statistics Ethnicity Data Product –Hospital admission for Global Burden of Disease conditions by ethnic group

**Data Triangulation: data sources used to establish provenance**

Source	Variable	Spatial granularity of comparator	Temporal granularity of comparator	Note(s)
ONS	Oslaua	Local authority district (2016)	2016	Local Authority district code (2016)
HES, Census 2001 for 1999/00-2003/04, Census 2011 for 2009/10-2013/14	r_[ethnic group]	Local authority district (2016)	1999/00-2003/04 or 2009/10-2013/14	Age- and sex-standardised preventable hospitalisation incidence per 100,000 pop
HES, Census 2001 for 1999/00-2003/04, Census 2011 for 2009/10-2013/14	r_ll_[ethnic group]	Local authority district (2016)	1999/00-2003/04 or 2009/10-2013/14	Standardised incidence 95% CI lower
HES, Census 2001 for 1999/00-2003/04, Census 2011 for 2009/10-2013/14	r_ul_[ethnic group]	Local authority district (2016)	1999/00-2003/04 or 2009/10-2013/14	Standardised incidence 95% CI upper
HES, Census 2001 for 1999/00-2003/04, Census 2011 for 2009/10-2013/14	rzcat_[ethnic group]	Local authority district (2016)	1999/00-2003/04 or 2009/10-2013/14	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)
HES, Census 2001 for 1999/00-2003/04, Census 2011 for 2009/10-2013/14	Ethcat	Local authority district (2016)	1999/00-2003/04 or 2009/10-2013/14	1) NHS-recorded; 2) NHS-recorded with surname imputation; 3) Surname prediction; 4) Prediction with forename and surname agreement

Data Profile: Hospital Episode Statistics Ethnicity Data Product –Hospital admission for Global Burden of Disease conditions by ethnic group

HES, Census 2001 for 1999/00-2003/04, Census 2011 for 2009/10-2013/14	gbd1a	Local authority district (2016)	1999/00-2003/04 or 2009/10-2013/14	GBD1 disease code
HES, Census 2001 for 1999/00-2003/04, Census 2011 for 2009/10-2013/14	gbd_label	Local authority district (2016)	1999/00-2003/04 or 2009/10-2013/14	GBD1 disease label
HES, Census 2001 for 1999/00-2003/04, Census 2011 for 2009/10-2013/14	Census	Local authority district (2016)	1999/00-2003/04 or 2009/10-2013/14	"01" for 1999-2003 and "11" for 2009-2013