

Leading article

Ethnicity recording in primary care computerised medical record systems: an ontological approach

Z. Tippu

Academic Foundation Doctor, Section of Clinical Medicine and Aging, University of Surrey, Guildford

A. Correa

Primary Care Scientist, Section of Clinical Medicine and Aging, University of Surrey, Guildford

H. Liyanage and J. Van Vlymen

Research Fellow, Section of Clinical Medicine and Aging, University of Surrey, Guildford

D. Burleigh

SQL Developer, Section of Clinical Medicine and Aging, University of Surrey, Guildford

A. McGovern

Clinical Research Fellow, Section of Clinical Medicine and Aging, University of Surrey, Guildford

S. Jones

Research Professor, Department of Population Health, Division of Healthcare Delivery Science, NYU School of Medicine, New York, USA

S. de Lusignan

Professor of Primary Care and Clinical Informatics, Section of Clinical Medicine and Aging, University of Surrey, Guildford

Cite this article: Tippu Z, Correa A, Liyanage H, Van Vlymen J, Burleigh D, McGovern A, Jones S, de Lusignan S. Ethnicity recording in primary care computerised medical record systems: an ontological approach. *J Innov Health Inform.* 2017;23(4):799–806.

<http://dx.doi.org/10.14236/jhi.v23i4.920>

Copyright © 2017 The Author(s). Published by BCS, The Chartered Institute for IT under Creative Commons license <http://creativecommons.org/licenses/by/4.0/>

Author address for correspondence:

Z. Tippu,
Section of Clinical Medicine and Aging,
Department of Clinical and Experimental Medicine,
University of Surrey,
Guildford GU2 7XH, UK.
Email: s.lusignan@surrey.ac.uk

Accepted January 2017

ABSTRACT

Background Ethnicity recording within primary care computerised medical record (CMR) systems is suboptimal, exacerbated by tangled taxonomies within current coding systems.

Objective To develop a method for extending ethnicity identification using routinely collected data.

Methods We used an ontological method to maximise the reliability and prevalence of ethnicity information in the Royal College of General Practitioner's Research and Surveillance database. Clinical codes were either directly mapped to ethnicity group or utilised as proxy markers (such as language spoken) from which ethnicity could be inferred. We compared the performance of our method with the recording rates that would be identified by code lists utilised by the UK pay for the performance system, with the help of the Quality and Outcomes Framework (QOF).

Results Data from 2,059,453 patients across 110 practices were included. The overall categorisable ethnicity using QOF codes was 36.26% (95% confidence interval (CI): 36.20%–36.33%). This rose to 48.57% (CI: 48.50%–48.64%) using the described ethnicity mapping process. Mapping increased across all ethnic groups. The largest increase was seen in the white ethnicity category (30.61%; CI: 30.55%–30.67% to 40.24%; CI: 40.17%–40.30%). The highest relative increase was in the ethnic group categorised as the other (0.04%; CI: 0.03%–0.04% to 0.92%; CI: 0.91%–0.93%).

Conclusions This mapping method substantially increases the prevalence of known ethnicity in CMR data and may aid future epidemiological research based on routine data.

Keywords: epidemiology, ethnicity, primary care

INTRODUCTION

Ethnicity is an important determinant of healthcare inequality worldwide. Disease pattern and presentation,¹ alongside uptake and utilisation of healthcare services,² varies across ethnic groups.

Ethnicity recording within primary care health records was introduced in 1991, followed by the introduction into Hospital Episode Statistics in England in 1995.³ Despite its prolonged inclusion time ethnicity recording within primary care data remains suboptimal.⁴ This has been attributed to an oversimplification of classifying hierarchies,⁵ practical and administrative difficulties in data collection,^{6–8} alongside inaccuracies in the data recorded. The use of surname recognition software alongside estimations of ethnic makeup according to census data, as a means of inferring ethnicity, has been attributed as the reason for these inaccuracies.^{9,10} These methods have limitations and run significant risks of misclassifying individuals compared to the gold standard measure of self-reported ethnicity.^{11,12}

The Quality and Outcomes Framework (QOF) is a pay-for-performance system introduced in 2004 to improve chronic disease management in primary care. It has included incentivising the recording of ethnicity since 2006, resulting in more comprehensive recordings.³ Although QOF payments may distort coding,¹³ overall they have probably improved data quality. However, inconsistencies persist, with low concordance of multiple ethnicities recorded in some minority groups.¹¹

The UK coding system for primary care is the Read Classification.¹⁴ Diagnostic codes obtained using differing coding structures can vary in their inference and interpretation.^{8,15} The taxonomies for recording ethnicity within this classification are complex and tangled;¹⁶ different ethnicity classifications, language, country or origin codes and religion all convey some information about ethnicity with differing levels of granularity. Published literature has predominantly focused on two principal groups of ethnicity codes, '9i... Ethnic Category-2001 census' and '9S...ethnic group' codes, within the UK Read code system,¹⁴ which have been mapped onto the Office for National Statistics (ONS) Census.

Ontologies allow the description of key concepts and their relationships within a domain,¹⁷ with a view to explicitly defining data when utilising it for chronic disease management¹⁸ and integrated care.¹⁹ An ontological approach, allowing identification and classification of a number of varying coding hierarchies within primary care data, would allow identification of ethnicity of patients from a number of differing inferences.

We carried out this study to explore whether we could increase the number of appropriately coded ethnicities recorded in primary care data while taking a more

ontological approach to ethnicity data. Developing this ontological approach should allow an extended and more defined mapping of ethnicity within healthcare datasets and improve epidemiological studies of disease progression and service utilisation within ethnic groups.

METHODOLOGY

The objective of this study is to maximise the number of cases for which we can reliably report ethnicity by mapping relevant clinical codes onto the 2011 census categories through use of language and interpreter codes. This approach is an extension of an existing mapping method.²⁰ In addition, we have looked at whether other codes could be used as a proxy measure of ethnicity.

The ontological process takes place in three steps: (1) the conceptual or ontological layer that groups concepts and is coding system independent so could be applied to any computerised medical record (CMR) systems that had one or more of the data categories; (2) the coding layer (specific to the coding system); and (3) the logical data extract model.¹⁴

This analysis was conducted using data from the Royal College of General Practitioners' Research and Surveillance database, comprising 110 practices across England and Wales.

Ontological process

Ontological and coding layer

The ontological layer requires the definition of the relevant concepts, in this case of ethnicity, and their relationships. A number of ethnicity related concepts, such as ethnicity, nationality and language, were reviewed in a number of clinical terminologies including Read2, Clinical Terms version 3 (CTv3) and Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) to develop an overarching semantic information model (Figure 1).

NHS Browser derived Read 2 codes, CTv3 and SNOMED-CT were used to derive relevant coding hierarchies. Relevant coding hierarchies are highlighted (Table 1). Codes were divided into definite (coding hierarchies directly coding an identifiable ethnicity), probable (coding hierarchies which have a high probability of inferring an ethnicity), and possible (coding hierarchies which were deemed possible to infer an ethnicity from) layers.

Rates of code accuracy, other than census derived ethnicity codes, were measured in an attempt to determine whether they could be effective proxies for ethnicity. Utilising ontologically defined valid read codes, data extraction queries were developed, allowing the identification of the proportion of patients within the dataset that had ethnicity coded, as per

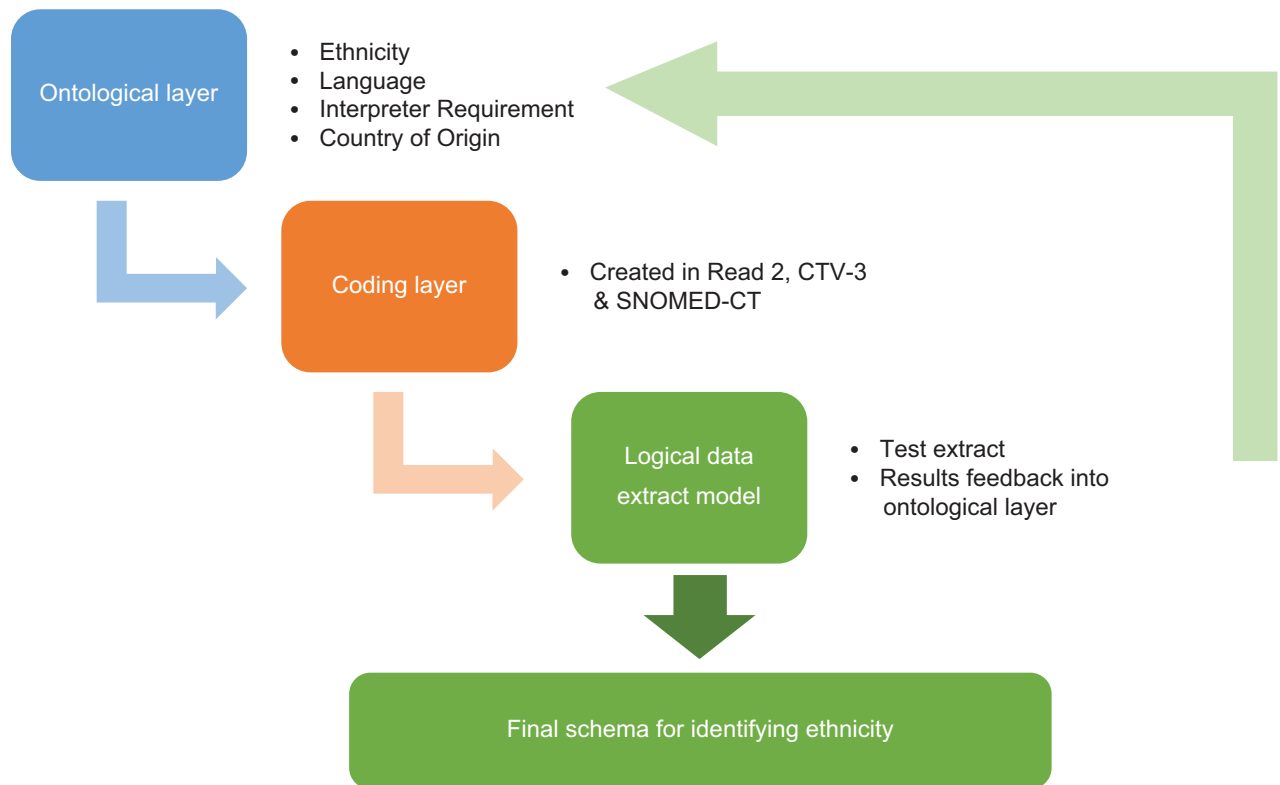


Figure 1 A three-step ontological process identifying ethnicity from CMRs¹⁷

Table 1 Iterated coding hierarchies in the Read 2, CTV3 and SNOMED-CT coding structures categorised as definite, probable and possible when inferring ethnicity

| Coding system | Definite coding hierarchy | Probable coding hierarchy | Possible coding hierarchy |
|-----------------------------|--|--|---|
| Read 2 CTv3 SNOMED-CT | Ethnic groups (census) 9S... XUVqQ S-0001B | Main spoken language 13l.. XUUSn F-02B21 | Language read 13n.. XUUSm F-94930 |
| Read 2 CTv3 SNOMED-CT | Ethnic category – 2001 census 9i... - - | Additional main spoken language 13u.. - - | Need for interpreter 9NU.. Xal8W F-025FB |
| Read 2 CTv3 SNOMED-CT | Ethnicity and other related national data 9T. - F-001A9 | Supplemental main language spoken 13w.. UaoHK F-04502 | |
| Read 2 CTv3 SNOMED-CT | Country of Origin 134.. XSCj5 F-03C9A | Language Spoken 13Z6 - F-028DC | |
| Read 2 CTv3 SNOMED-CT | | World languages 13b.. - - | |

each of the coding layers of the ontology (Table 1). The validity of probable or possible coding hierarchies as proxy markers for ethnicity was determined by a mapping of greater or equal than 80% to definite codes. If less than 80% of patients with a probable or possible code did not also have a corresponding definite code, then the probable or possible code was discarded as a not suitable proxy for ethnicity. For example, use of specific language and interpreter codes was seen to have a greater than 80% mapping to codes that definitively inferred a Bangladeshi ethnicity and could be used as proxy measures (Table 2).

Logical data extraction model

An algorithm was developed for the classification of patients into a 2011 Census ethnic group (Figure 2). Patients with a single ethnicity were easily classified, with preference given to definite codes followed by the proxy codes (probable and possible), if available. Those patients with multiple ethnicity codes recorded over time were classified according to the most common ethnicity recorded, using the same hierarchy

of preferred codes described above. In those patients for whom there was no majority ethnicity recorded, the most recent coded ethnicity was utilised. Some patients had multiple ethnicities recorded on their latest consultation; if there was no clear majority on the latest coded encounter, patients were deemed to have no identifiable usable ethnicity.

We report the inter-practice variability and geographic variation in ethnicity identification using this ethnicity classification method using descriptive statistics (mean and quartiles).

Rates of recording were assessed across both age and gender, allowing a determination of demographic trends in ethnicity recording within primary care. Further data extraction for demographic variable determines any inconsistencies in the ontological approach employed.

As a final assessment of the granularity of the extended coding hierarchies utilised, the ontologically derived codes were compared against the national code list recommended for primary care in the QOF,²¹ for a definitive determination of the extended data capture enabled utilising an ontological approach.

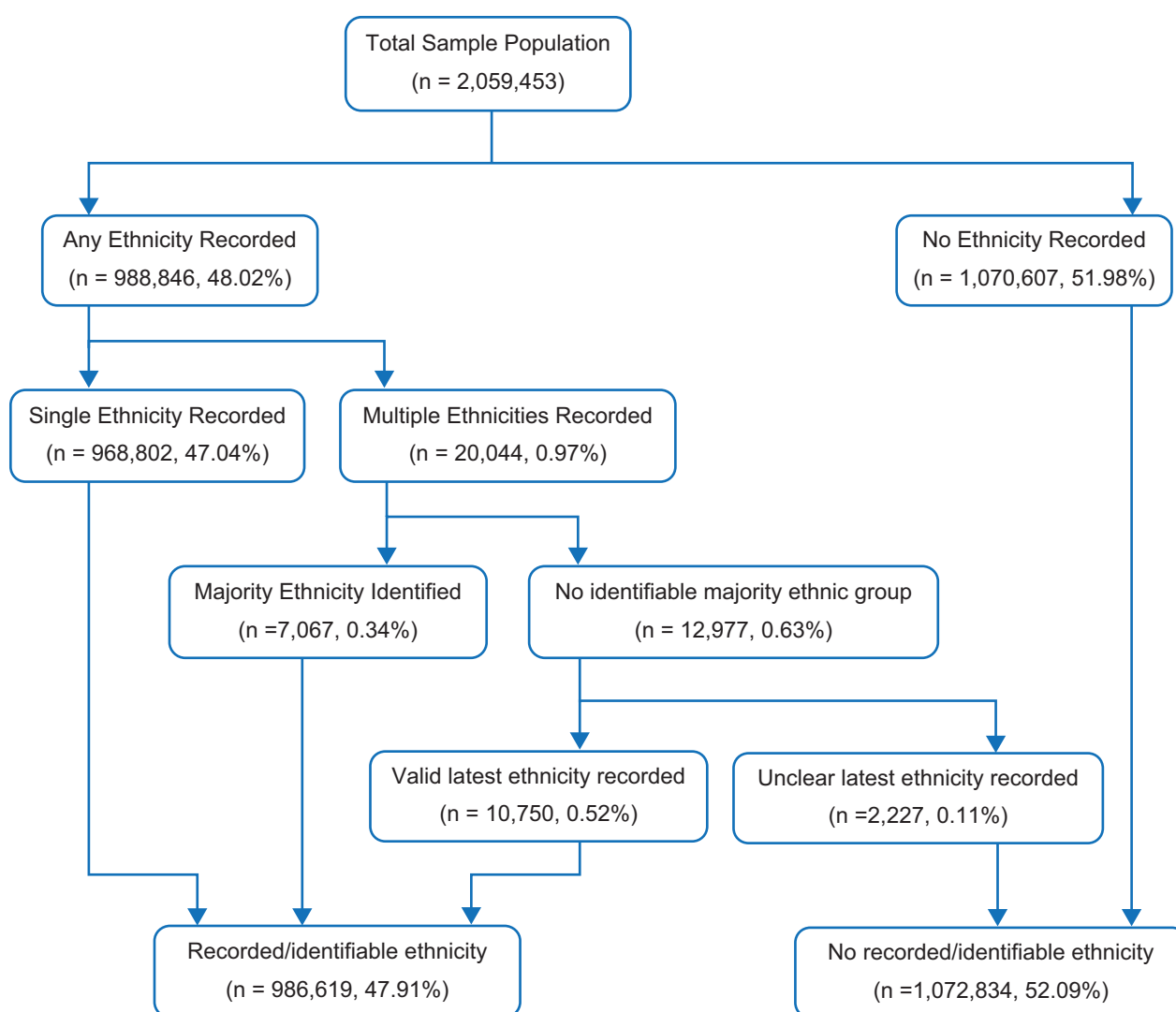


Figure 2 Final schema for ethnicity classification

Table 2 Iterated Read 2 codes in CTv3 and SNOMED-CT inferring a Bangladeshi ethnicity

| Coding system | Definite codes | Probable codes (%= 94.3) (matched = 10,808, n = 11,457) | Possible codes (%= 93.9) (matched = 1059, n = 1,128) |
|---------------|-------------------------------------|---|--|
| | Bangladeshi | Main spoken language Bengali | Reads Bengali |
| Read 2 | 9S8.. | XaG5q | XaIMQ |
| CTv3 | 9S8.. | 13I1. | 13n8. |
| SNOMED-CT | F-00188 | F-025BB | F-00E64 |
| | Bangladeshi/ Brit Bangladeshi | Language Bengali | Interpreter needed – Bengali |
| Read 2 | 9i9.. | 13Z61 | 9NU6. |
| CTv3 | 9i9.. | - | - |
| SNOMED-CT | - | R-41ECC | - |
| | RACE: Bangladeshi | Sylhety | Interpreter needed – Sylheti |
| Read 2 | - | Xa6fc | 9NUI. |
| CTv3 | 134I. | 13b2. | - |
| SNOMED-CT | - | R-4149A | - |
| | | Main spoken language Sylheti | |
| Read 2 | | 13IJ. | |
| CTv3 | | CaG68 | |
| SNOMED-CT | | F-025F6 | |

Data extraction was undertaken in SQL server management studio, with statistical analysis subsequently performed in the statistical software R.

RESULTS

Data was available for 2,059,453 patients across 110 practices in England and Wales, utilising the Read 2 classification.

Mean valid ethnicity recording was 48.97% ($n = 1,008,667$) across the sample population. Of those identified with a valid ethnicity recording, 96.14% ($n = 969,740$) were recorded to have a definite code. 3.71% ($n = 37,443$) and 0.15% ($n = 1,487$) had possible and probable codes, respectively.

Just under 1% (0.97%) were noted to have multiple ethnicities recorded over time. Of those with multiple ethnicity codes, 35.26% had a clearly identifiable majority ethnic group recorded and were classified accordingly. Of those who had no clear determined majority ethnicity code recorded, 82.83% were classified according to a valid single most recent ethnicity code within the dataset. 0.11% of the population had an underterminable ethnicity coded and were thus excluded from the dataset. Of the patient population, 52.09% had no definitive ethnicity code mapped and were excluded from the analysis. Figure 2 describes the logical data model used for ethnicity classification.

The median ethnicity identification rate across the dataset was 55.20% (inter quartile range: 35.6%). Considerable variations were noted between practices, with minimum and maximum ethnicity identification rates being 1.5% and 91.6%, respectively (Figure 3). Six practices were seen to have ethnicity identification proportions greater than 80%,

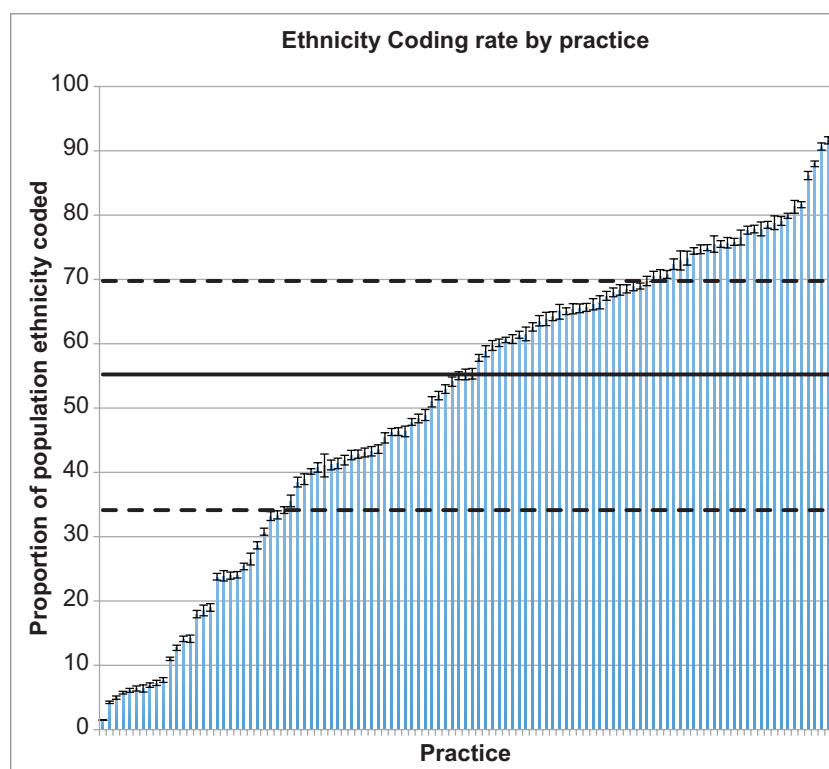


Figure 3 Ethnicity identification rate by practice (median 55.2%: solid line; interquartile range 35.7%: dashed lines).

two of which had a greater than 90% recording. Ten practices had less than 10% of the patient population coded with a recognisable ethnicity identifier.

Geographical variations exist within the dataset. Ethnicity identification was greatest in London-based practices ($n = 30$), mean 58.59%, with the identification lowest in those practices based in the South of the country ($n = 25$), mean 39.14%. Ethnicity identification was highest in practices with the highest proportions of non-white people; a 0.16% increase in the total ethnicity identified per percent increase in the proportion of non-white people. However, the correlation was poor ($R^2 = 0.016$).

Ethnicity identification was highest in children aged between 5 and 9 years ($n = 64,432/99,232$, 64.93%). In the adult population, ethnicity recording was consistent between age bands, with higher recording levels in young adults, 54.67% ($n = 97,587/178,493$) for 25–29 years and 53.53% ($n = 106,335/198,654$) for 30–34 years. For those greater than 40 years, ethnicity recording was 46.51% ($n = 463,203/996,013$), and for those greater than 65 years, this was 46.33% ($n = 175,286/378,323$). Ethnicity recording in females was greater than that observed in males, 50.81% ($n = 537,293/1,057,559$) and 47.05% ($n = 471,374/1,001,874$) respectively.

The QOF highlights the use of the '9i...' coding hierarchy for ethnicity within primary care datasets. Utilising iterated proxy markers for ethnicity, inclusive of language spoken and read, and requirement for an interpreter, identification was increased across each respective ethnic group (Figure 4), compared to using the QOF codes. Across the whole database, the detection of white ethnicity increased from 30.61% (CI:30.55%–30.67%) to 40.24% (CI:40.17%–40.30%); for black, from 1.65% (CI:1.63%–1.67%) to 2.29% (CI:2.27%–2.32%); for Asian, from 3.08% (CI:3.06%–3.11%) to 4.11% (CI:4.08%–4.14%); for those of a mixed ethnicity, from 0.89%

(CI:0.88%–0.90%) to 1.01% (CI: 0.99%–1.02%); and those from other ethnic groups from 0.04% (CI:0.03%–0.04%) to 0.92% (0.91%–0.93%).

DISCUSSION

Main findings of this study

Utilising an ontological approach, we have created updated method for mapping ethnicity related codes to the 2011 Office for National Statistics (ONS) Census major ethnicity categories. This mapping allows the use of iterated Read codes, which do not directly infer an ethnicity to act as proxy markers and substantially improved ethnicity identification in the tested primary care dataset.

In providing an update to the paper by Kumarapeli *et al*,²⁰ our study demonstrates that utilising iterated proxy markers of ethnicity improved the identification of ethnicity recording within primary care CMR systems (46.85% to 55.20%). Using limited coding hierarchies can result in an underrepresentation of individuals with ethnicity recorded within the dataset. In comparison with the QOF-endorsed '9i...' hierarchy, the ontology-based method developed in this paper has resulted in increased ethnicity identification across all major ONS ethnic groups (Figure 4). Utilising an algorithmic approach in those that have conflicting ethnicity codes, a further 0.87% ($n = 17,817$) were mapped to one of the ONS defined ethnic groups, allowing for clear identification and classification of these individuals.

There is considerable inter-practice variability in ethnicity recording, with higher mean ethnicity recordings within London-based practices (58.59%). Practices with a higher proportion of minority ethnicities had higher ethnicity coding rates. This suggests that a greater ethnic diversity amongst registered patients at these practices results in a greater tendency for clinicians to code ethnicity; the correlation was, however, poor.

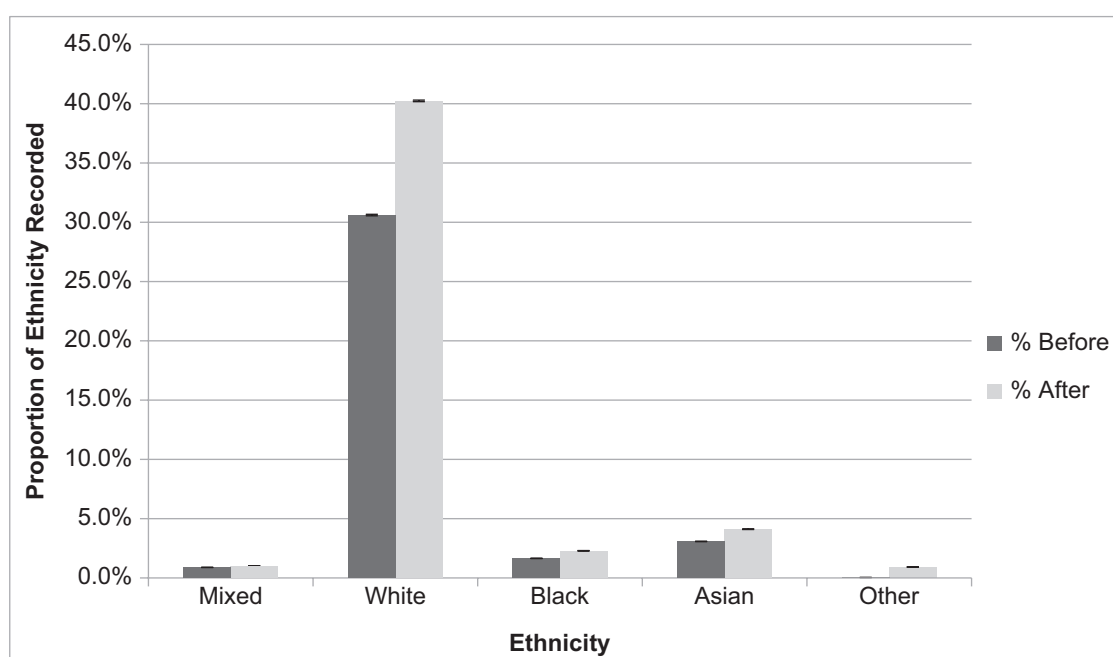


Figure 4 Proportion of ethnicity records mapped to a 2011 census ethnic group using an ontological approach versus QOF codes

What is already known on this topic

The recording of ethnicity utilising traditional coding hierarchies within primary care datasets is known to be sub-optimal. This has resulted in restrictions in the use of ethnicity in monitoring disease presentation, prevalence, pattern, and management amongst differing ethnic sub-groups.²² Incentivisation of coding ethnicity through the QOF from 2006 to 2007 has resulted in significant improvements in the completeness of ethnicity data within primary care, but this financial incentivisation has been removed from the business rules as of 2011/2012. The existence of multiple ethnicity coding hierarchies, with no clear inter-relationships or inferences, has resulted in inconsistencies in recording ethnicity within primary care datasets.

What this study adds

This paper provides an updated method to ethnicity identification from records within a large primary care dataset. This study is the first to utilise an ontological approach to maximise data capture for ethnicity in primary care data. Utilising iterated proxy markers, ethnicity recording was seen to increase across all ONS defined ethnic groups. Published literature has focused on two traditional coding hierarchies '9i...' and '9S...' codes.^{3,20} The use of these limited codes can result in poor capture of ethnicity data,⁵ further propagating its underutilisation in analysing trends in healthcare outcomes amongst differing ethnic groups. The use of limited coding hierarchies poorly reflect the numerous variables that infer a patient's ethnicity; multiple inferences affords greater flexibility in determining an individual's ethnicity.⁵ The use of a mapping structure allowed identification of codes which definitively infer an ethnicity, expanding on previous ethnicity codes utilised. The process of iteration of non-directly mapped codes justifies the use of proxy markers.

Since defining this ontological process we have since applied the method to ethnicity identification in a population with diabetes.²³ In this group the data quality was sufficiently higher and ethnicity identification using the ontological process set out here was 82.1%. Accurate ethnicity identification from real world data is of particular importance in this group given the high diabetes prevalence in those of Asian and black ethnicities.²⁴

This paper provides a clear ontology for classification of those patients with conflicting ethnicity codes within the

dataset. Similar to previous studies, we use the most commonly recorded ethnicity group as means of classification into an ethnic group,³ preferentially using those codes which directly infer an ethnicity, followed by the proxy codes. Further, in those patients with no clear majority, the use of the latest recorded ethnicity provides a mechanism with which ethnicity can be classified.

Limitations of this study

This paper was unable to establish the underlying reasons as to why individuals within the dataset did not have a recognisable ethnicity recorded. The authors are aware that those without an ethnicity recorded may be an important negative, as a result of underutilisation of healthcare services or difficulties in communication.

Due to the nature of our dataset, there is a reliance on the accuracy of coding. In those patients with multiple differing ethnicity codes, inferences were made in regard to ethnicity (based on most common or latest code), which run the risk of misclassification. The use of an algorithmic approach for these patients, who represent a small proportion of our dataset, minimises the risk of this misclassification.

As the UK coding system for primary care uses the Read classification, CTv3 and SNOMED-CT coding hierarchies could not be utilised in the logical data extract model. Extraction in the coding layer demonstrates that these coding hierarchies exist, and inclusion within both the ontological and coding layers means that this ontology remains viable on the introduction of these codes.

CONCLUSION

An ontological approach to ethnicity codes allows for more accurate and extensive identification of ethnicity of people, and hence, we obtained a substantially higher proportion of ethnicities.

Widespread inter-practice variations in ethnicity coding still persist; with the removal of incentives from the QOF frameworks in 2011/2012, these disparities may widen. We encourage clinicians and healthcare researchers to actively code ethnicity within encounters.

REFERENCES

1. Boyd AE, Murad S, O'Shea S, de Ruiter A, Watson C and Easterbrook PJ. Ethnic differences in stage of presentation of adults newly diagnosed with HIV-1 infection in south London. *HIV Medicine* 2005;6(2):59–65. Epub 2005/04/06. <https://doi.org/10.1111/j.1468-1293.2005.00267.x>. PMID:15807711.
2. Griffiths C, Cooke S and Toon P. Registration health checks: inverse care in the inner city? *British Journal of General Practice* 1994;44(382):201–4. Epub 1994/05/01. PMID:8204332; PMCID:PMC1238866
3. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *Journal of Public Health* 2014;36(4):684–92. Epub 2013/12/11. <https://doi.org/10.1093/pubmed/fdt116>. PMID:24323951; PMCID:PMC4245896.
4. Sultana K and Sheikh A. Most UK datasets of routinely collected health statistics fail to collect information on ethnicity and religion. *Journal of the Royal Society of Medicine* 2008;101(9):463–5. Epub 2008/09/10. <https://doi.org/10.1258/jrsm.2008.080007>. PMID:18779248; PMCID:PMC2587383.
5. Aspinall PJ. The operationalisation of race and ethnicity concepts in medical classification systems: issues of

- validity and utility. *Health Informatics Journal* 2005;11(4):259–74. <https://doi.org/10.1177/1460458205055688>
6. McLeod D, Mansell J, Harris R, Bailey T, Dowell A, Robson B et al. The collection of patient ethnicity data: a challenge for general practice. *The New Zealand Family Physician* 2000;27(3):51–7. <https://doi.org/10.1136/bmj.312.7038.1080> PMID:8616418; PMCID:PMC2350904.
 7. Pringle M and Rothera I. Practicality of recording patient ethnicity in general practice: descriptive intervention study and attitude survey. *BMJ* 1996;312(7038):1080–2. Epub 1996/04/27.
 8. Morrison Z, Fernando B, Kalra D, Cresswell K, Robertson A, Sheikh A. The collection and utilisation of patient ethnicity data in general practices and hospitals in the United Kingdom: a qualitative case study. *Inform Prim Care*. 2014;21(3):118–31. Epub 2014/09/11. <https://doi.org/10.14236/jhi.v21i3.63>.
 9. Nitsch D, Kadalayil L, Mangtani P, Steenkamp R, Ansell D, Tomson C et al. Validation and utility of a computerized South Asian names and group recognition algorithm in ascertaining South Asian ethnicity in the national renal registry. *QJM* 2009;102(12):865–72. Epub 2009/10/16.
 10. Hull SA, Rivas C, Bobby J, Boomla K and Robson J. Hospital data may be more accurate than census data in estimating the ethnic composition of general practice populations. *Informatics in Primary Care* 2009;17(2):67–78. Epub 2009/10/08.
 11. Saunders CL, Abel GA, El Turabi A, Ahmed F and Lyratzopoulos G. Accuracy of routinely recorded ethnic group information compared with self-reported ethnicity: evidence from the English Cancer Patient Experience survey. *BMJ Open* 2013;3(6). Epub 2013/07/03.
 12. Aspinall PJ. The utility and validity for public health of ethnicity categorization in the 1991, 2001 and 2011 British Censuses. *Public Health* 2011;125(10):680–7. Epub 2011/09/13.
 13. de Lusignan S and Mimnagh C. Breaking the first law of informatics: the Quality and Outcomes Framework (QOF) in the dock. *Informatics in Primary Care* 2006;14(3):153–6. Epub 2007/02/10.
 14. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. *Informatics in Primary Care* 2005;13(1):65–70. Epub 2005/06/14.
 15. Rollason W, Khunti K and de Lusignan S. Variation in the recording of diabetes diagnostic data in primary care computer systems: implications for the quality of care. *Informatics in Primary Care* 2009;17(2):113–9. Epub 2009/10/08. <https://doi.org/10.14236/jhi.v17i2.723>.
 16. Barrett D, Liaw ST and de Lusignan S. Unravelling the tangled taxonomies of health informatics. *Informatics in Primary Care* 2014;21(3):152–5. Epub 2014/09/11. <https://doi.org/10.14236/jhi.v21i3.78>.
 17. de Lusignan S. In this issue: Ontologies a key concept in informatics and key for open definitions of cases, exposures, and outcome measures. *Journal of Innovation in Health Informatics* 2015;22(2):170. Epub 2015/08/08. <https://doi.org/10.14236/jhi.v22i2.170>. PMID:26245238.
 18. Liyanage H, Liaw ST, Kuziemsky C and de Lusignan S. Ontologies to improve chronic disease management research and quality improvement studies – a conceptual framework. *Studies in Health Technology and Informatics* 2013;192:180–4. Epub 2013/08/08. PMID:23920540.
 19. Liyanage H, Liaw ST, Kuziemsky C, Terry AL, Jones S, Soler JK et al. The Evidence base for using ontologies and semantic integration methodologies to support integrated chronic disease management in primary and ambulatory care: realist review. Contribution of the IMIA Primary Health Care Informatics WG. *Yearbook of Medical Informatics*. 2013;8:147–54. Epub 2013/08/27. PMID:23974562.
 20. Kumarapeli P, Stepaniuk R, de Lusignan S, Williams R and Rowlands G. Ethnicity recording in general practice computer systems. *Journal of Public Health* 2006;28(3):283–7. Epub 2006/07/15. <https://doi.org/10.1093/pubmed/fdl044>. PMID:16840765.
 21. Committee GP. *Ethnicity and first language recording-GPC guidance*. 2011. In Mathur R, Grundy E, Smeeth L (Eds.). Availability and use of UK based ethnicity data for health research 2013.
 22. British Medical Association. Ethnicity and first language recording. URL: <https://www.bma.org.uk/advice/employment/contracts/gp-partner-contracts/desethnicity> Accessed 30 January 2017. Mathur R, Grundy E and Smeeth L. Availability and use of UK based ethnicity data for health research. 2013. Available from: <http://www.ncrm.ac.uk/news/show.php?article=5353>. Accessed 30 January 2017.
 23. McGovern A, Hinton W, Correa A, Munro N, Whyte M and de Lusignan S. Real-world evidence studies into treatment adherence, thresholds for intervention and disparities in treatment in people with type 2 diabetes in the UK. *BMJ Open* 2016;6(11):e012801. Epub 2016/11/26. <https://doi.org/10.1136/bmjopen-2016-012801>. PMID:27884846.
 24. English S, Tippu Z, Chan T, Vlymen Jv, Burleigh D, Correa A et al. Type 2 diabetes prevalence among people of South Asian ethnicity in the UK. *Diabetes & Primary Care*. 2016;18(1):28–32.