

# Ontologies to improve the identification of ethnicity in people with Type 2 Diabetes

Zayd Tippu, Harshana Liyanage, Ana Correa, David Burleigh, Andrew McGovern, Simon Jones, Simon de Lusignan

Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK



## Aim

Ethnicity as an independent risk factor for Type 2 Diabetes is well established. We present a description of an ontological approach to increase ethnicity identification in primary care data. Allowing for improved epidemiological study of disease progression and service utilisation within ethnic groups.

## Background

Prevalence and risk of Type 2 Diabetes vary with ethnicity, studies using routine data are limited by missing ethnicity data. The recording of ethnicity data within primary care datasets remains suboptimal (Figure 3).

Ontologies formally define concepts and relationships in a given domain. Ontologies improve case finding and using formal methods enables their consistent application across datasets.

## Methods

We reviewed ethnicity and its related concepts (nationality, language etc.) in a number of clinical terminologies in routine use and developed an overarching semantic information model.

We graded the associated concepts to indicate the extent to which they act as proxies for ethnicity; and tested this on the >1million patient records in the Royal College of General Practitioners Research and Surveillance Centre database, comprising 110 practices across England and Wales.

We compared ethnicity identification across all 2011 English census groups using our ontology with the national code list recommended for primary care in the Quality and Outcome Framework (QOF).

## Results

Utilising iterated proxy markers for ethnicity, language spoken and interpreter requirements, we increased identification of ethnic group.

Data from 2,059,454 patients across 110 practices were included. The overall categorisable ethnicity using QOF codes was 36.26% (CI: 36.20%-36.33%). This rose to 48.57% (CI: 48.50%- 48.64%) using the described ethnicity mapping process (Figure 1). Mapping increased across all ethnic groups (Figure 2). The largest increase was seen in the white ethnic 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 other (0.04%; CI 0.03%-0.04% to 0.92%; CI 0.91%- 0.93%).

## Conclusion

An ontological approach to ethnicity codes allows for more accurate identification of ethnicity and a substantially higher proportion of people's ethnicities identified.

We encourage clinicians and healthcare researchers to actively code ethnicity within encounters.

## Key findings

- An ontological approach to ethnicity identification allows for a truer epidemiological prevalence of Type 2 Diabetes in primary care.

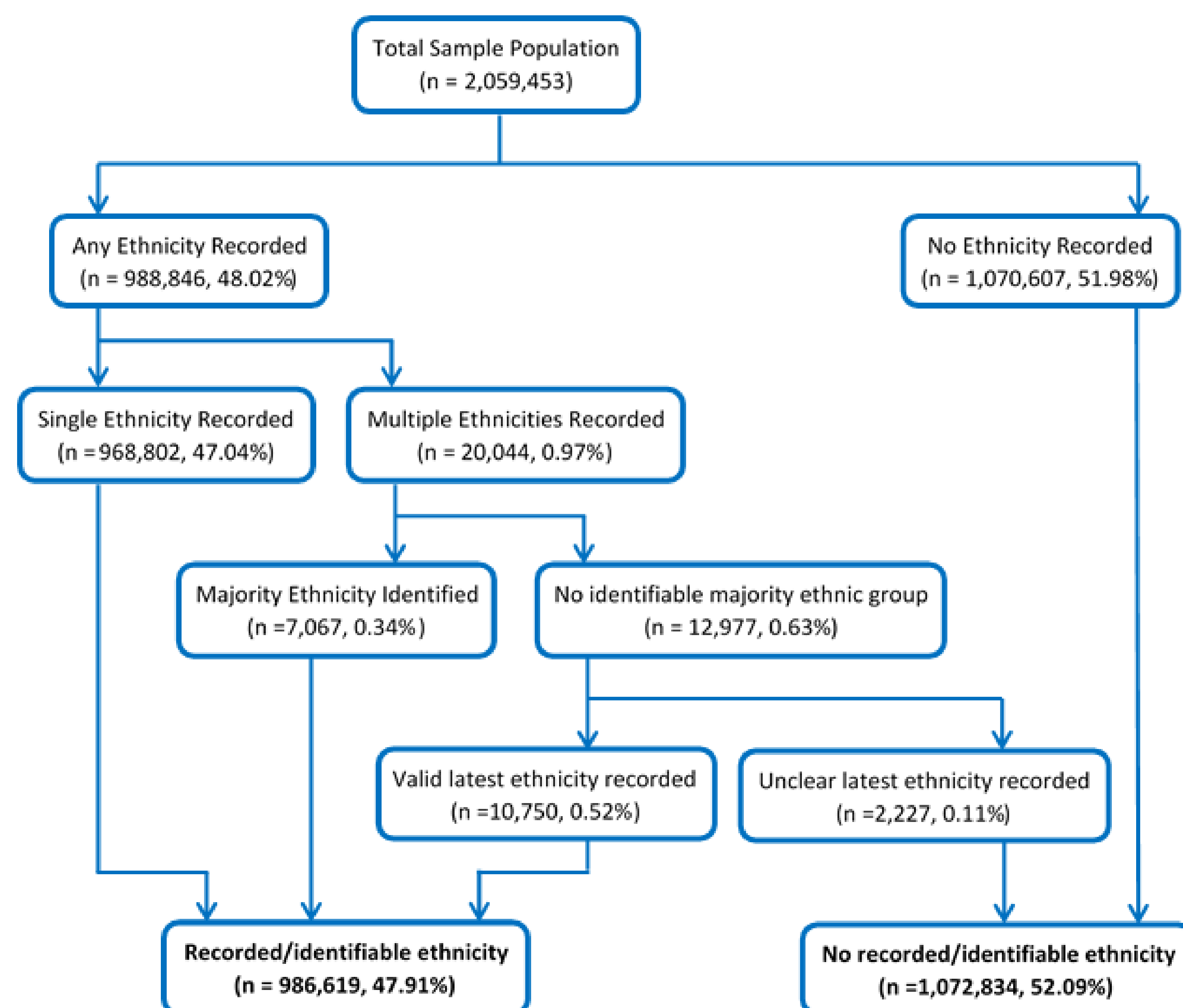


Figure 1. Algorithm for ethnicity classification.

## References

Available online with an electronic copy of this poster at:

<http://clininf.eu/>



Research sponsored by:  
Eli-Lilly Pharmaceuticals

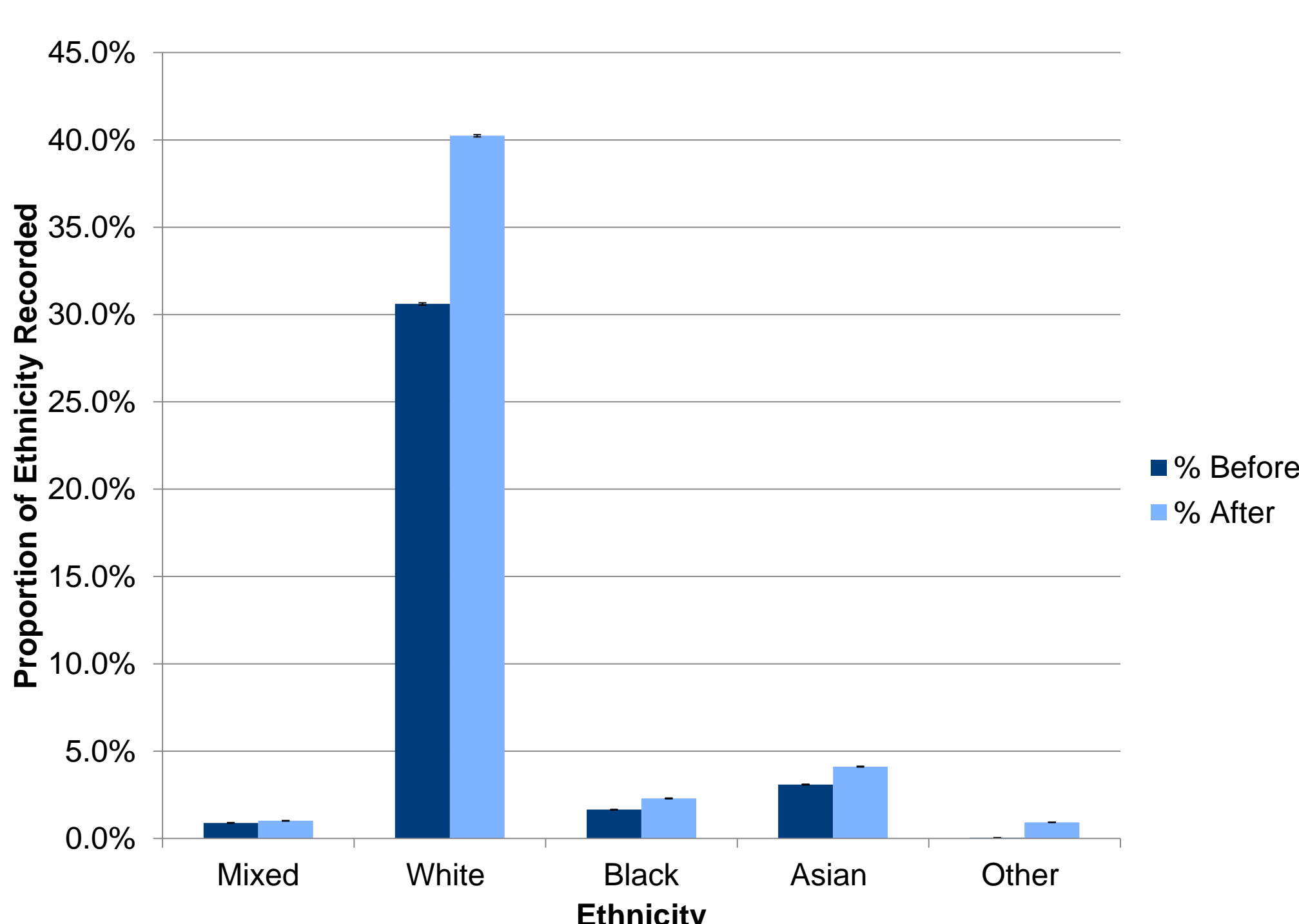


Figure 2. Proportion of ethnicity records mapped to 2011 census.

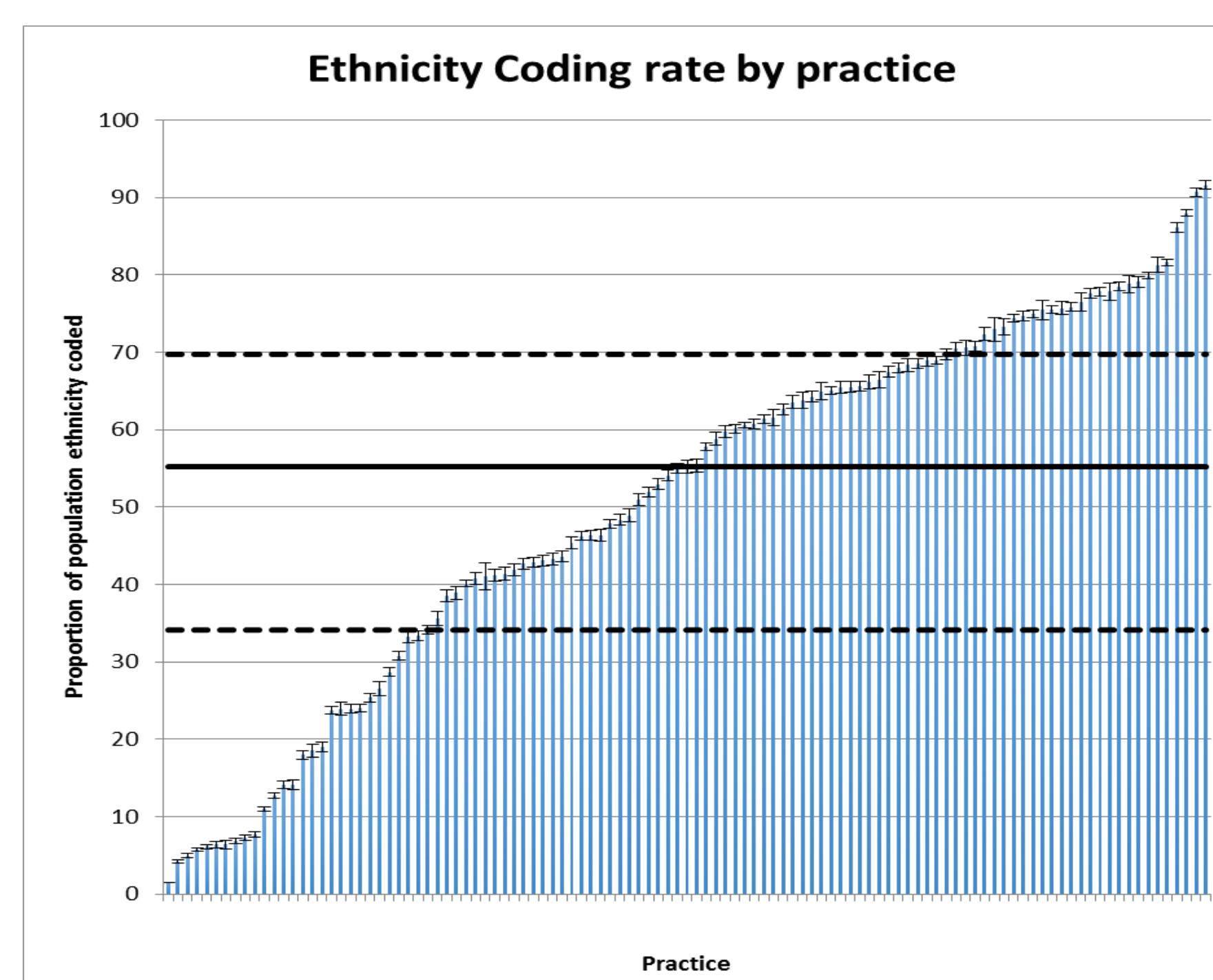


Figure 3. Ethnicity identification rate by practice.