Table 3 Validation of Onomap

Ethnicity	Ethnicity reported by participant	Ethnicity predicted by Onomap	Ethnicity correctly predicted	Sensitivity	Specificity
White British or Irish	1681	1811			
Other white	3235	3418			
Total white	4916	5229	4844	**98.5%**	**77.7%**
Indian	364	239			
Pakistani	313	348			
Bangladeshi	96	88			
Chinese	55	18			
Other Asian	<10	118			
Total Asian or Asian British	837	811	609	**72.8%**	**96.5%**
Black – African	344	142			
Black – Caribbean	10	<10			
Other black	23	<10			
Total black or black British	377	143	112	**29.7%**	**99.5%**
Arabic	39	279			
Other	<10	<10			
Other ethnic group	45	283	24	**53.3%**	**96.1%**
Mixed	234	<10	–		
Unclassified/unknown	231	174	<10	**3.5%**	**97.4%**
Total	6640	6640	**5589**	**87.4%**	**96.1%**

Estimated sensitivity and specificity of Onomap by ethnic group. Calculated by measuring the performance of Onomap to predict ethnicity in three clinical data sets* already containing self-reported or healthcare professional-reported ethnicity.
*Three data sets that included self-reported or healthcare professional-reported ethnicity were used to validate Onomap: a list of individuals attending a mosque in Wales who were offered screening for hepatitis C (n=189), a list of tuberculosis patients notified by doctors in Wales (n=3267) and a list of patients attending an infectious disease clinic in Poland (n=3184).

The absence of individual-level data on comorbidities is a limitation of this study. Further work is currently being carried out using linkage of COVID-19 notification data with other routine health records to further understand risks associated with hospitalisation. Also, we only had access to deaths that occurred in hospital. It is possible that there may have been ethnic differences in the proportion of people dying outside of hospital.

Onomap has been used widely as a tool in public health, for example, in studies investigating variation in influenza mortality,[17] hepatitis B infection[18] and human papillomavirus (HPV) vaccination uptake.[19] However, Onomap has limitations, and all findings should be interpreted in light of these. We previously validated the tool using data containing self-reported or healthcare professional-reported ethnicity. Onomap performs well for most ethnicities but has a low sensitivity for black or black British individuals. Risks identified for black and black British groups are therefore likely to be underestimated. Kandt and Longley[20] have published a comparison of Onomap with self-reported ethnicity in the 2011 Census. Notwithstanding apparent success in 30 reported studies in public health, healthcare and epidemiology (and wider application in equity audits in, inter alia, housing allocation, management science and social media), the reliability and limitations of such methods should be acknowledged and understood. In the absence of good ethnicity recording in routine health records, it does facilitate scientific investigation with margins of error that are understood. Moreover, many of the existing studies where individual person ethnicity is available have missing data and are not without their own classification bias. Anecdotally, members of minority ethnic groups are more likely to defer from reporting their ethnicities, and clinician-based classification is understood to be unreliable.

One of the recommendations of the Welsh Government Advisory Group is to improve recording of ethnicity in routine health data, and a data improvement plan is urgently required so ethnicity can be included in routine public health surveillance. There is an urgent need for all European countries carrying out COVID-19 surveillance to report trends by ethnicity, in order to inform local infection prevention and control policy and practice. Ethnic variation should also be considered in the design of interventions and in crisis communication.

Wales is less ethnically diverse than many other areas of the UK, but its BAME population has increased in