DOI: 10.1111/1471-0528.13170
www.bjog.org

Commentary

Race and ethnicity in biomedical publications

PJ Steer

Academic Department of Obstetrics and Gynaecology, Chelsea and Westminster Hospital, London, UK
Correspondence: Prof PJ Steer, Academic Department of Obstetrics and Gynaecology, Chelsea and Westminster Hospital, 369 Fulham Road, London SW10 9NH, UK. Email p.steer@imperial.ac.uk

Accepted 24 September 2014. Published Online 21 November 2014.

Please cite this paper as: Steer PJ. Race and ethnicity in biomedical publications. BJOG 2014; DOI: 10.1111/1471-0528.13170.

The etymological origins of the word 'race' are obscure, and its meaning and implications have been a cause of great controversy over the last 100 years. The concept that *Homo sapiens* could be divided into racial groupings has been poisoned by its association with eugenics. Although associated in popular consciousness with the heinous policies of the Nazis in the 1930s and 40s, eugenics was part of mainstream science before that time. Many papers refer to 'Fisher's exact test' but most authors are unaware that as well as being a fellow of the Royal Society, Sir Ronald Aylmer Fisher was a founder member of the University of Cambridge Eugenics Society, together with John Maynard Keynes and Horace Darwin (son of Charles), and was linked with the Eugenics Education Society of London, founded by Francis Galton. Fisher worked with JBS Haldane on population genetics, and in 1933 he became professor of eugenics at University College London. Over time eugenics became associated with forced sterilisation of the 'genetically unfit', even in such otherwise enlightened countries such as Sweden and the USA. Eugenics was supported by political activists such as George Bernard Shaw, Sidney and Beatrice Webb, and even by William Beveridge, the architect of the British welfare state. Because of universal revulsion at the policies of the Nazis, after 1945 eugenics was rejected, and it became very difficult to discuss genetic differences between groups of people with different geographic origins without being accused of racism, even in scientific publications.

Gradually, however, that situation is changing. Progress in molecular biology led to the development of tests for variations of haemoglobin, such as sickle cell and thalassaemia. This progressed inevitably to the concept of antenatal diagnosis, with the option of not proceeding with pregnancies where the child would be severely affected. It was indisputable that sickle cell affected mostly people with West African ancestry, and thalassaemia affected mostly people with Mediterranean ancestry. Soon cystic fibrosis, more common in white Europeans, was added to the list of genetic conditions with a different distribution between races. Clearly, such differences could not be ignored, and so it became acceptable once again to ask families about their racial origins.

Furthermore, the dramatic advances we have seen over the last 50 years in our understanding of the origins of *Homo sapiens* have led to a much deeper understanding of the similarities and differences between races. It has become clear that *Homo sapiens* developed in Africa about 120 000 years ago, and for half the time since then has shared the planet with other hominin species, such as *Homo erectus* and *Homo neanderthalensis*.[1–4] Anatomically modern humans (AMHs) first migrated out of Africa about 70 000 years ago, and by about 40 000 years ago had already reached Australasia (Figure 1). The Americas were invaded as recently as 16 000 years ago, although remarkably the evidence suggests that within 1000 years of reaching the continent they had spread out from the top to the bottom. Despite the extent of these migrations into different environments, the majority of human genetic variation (perhaps 80%) remains within Africa. This is because the relatively small number of people leaving Africa (perhaps only a few thousand),[5] created a genetic bottleneck that resulted in a comparatively homogeneous genetic inheritance in the subsequently expanding populations. *Homo neanderthalensis* became extinct less than 30 000 years ago, and the discovery of bones with surviving DNA has led to their genome being completely sequenced.[6] Remarkably, about 30% of their genome persists in AMHs, although with a maximum of about 4% in any particular individual. This legacy is confined to humans outside of Africa, a few of whom must have interbred with Neanderthals during their migrations across Asia and Europe.[7] The Neanderthal genetic legacy persists, particularly in relation to lipid metabolism,[8] and has been implicated in the high incidence of type–II diabetes in South Asians and native Americans.[9] My own group has shown the importance of racial origin in relation to the incidence of gestational diabetes,[10] and there are potential implications in relation to the duration of pregnancy and the timing of elective caesarean section.[11]

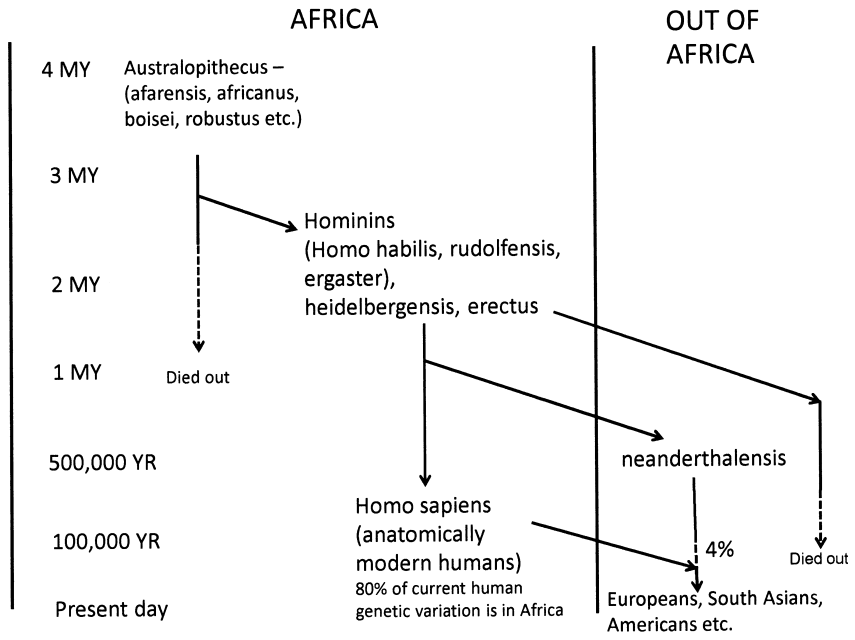

Figure 1. An evolutionary outline of human ancestry; MY, million years; YR, years.

The word 'ethnicity' is derived from the Greek 'ethnos', and is commonly used as an alternative to 'race'; however, it actually means people joined by practising similar customs or a common culture (for example, dietary rules or habits and marriage arrangements commonly vary between cultures, and are clearly not genetically determined), and so ethnicity does not cross-correlate directly with genetic ancestry. Ethnicity commonly parallels race, but can change when people migrate, whereas their genetic structure is relatively fixed. In evolutionary terms, when people live in a particular environment, genetic variations that suit the climate and diet are likely to increase in frequency; however, when a population moves suddenly to a different culture, this can result in important medical consequences, such as an increased incidence of diabetes. For example, when Latinos (Latin-American) immigrated into the USA, many changed from a 'Mediterranean' diet to a high-fat diet, resulting in high rates of obesity.[12] There is therefore scientific value in distinguishing between race and ethnicity.[13]

It is a fundamental axiom of medicine that we should treat people as of equal worth, irrespective of race, religion, or politics. Although it is vital that we maintain the principle that we are all born equal from an ethical point of view, ideal management must take account of differences in our genetic background. We are gradually moving towards personalised medicine, in which therapies can be tailored to produce optimal benefit, and ultimately knowing our own personal genetic structure will enable completely individualised treatment. This will be particularly important in global cities such as London, where there is already considerable racial mixing, which is likely to increase, making racial classification more difficult and less useful. In the meantime, however, knowing our racial (geographical) ancestry can still give important information about our genetic inheritance, and hence how best each individual should be treated. The recent rapid growth of online genealogy resources has made it easier for individuals to determine their origins and therefore discover if they may be at increased risk of genetically influenced diseases: for example, about one in 30 Ashkenazy Jews are carriers of the gene for Tay–Sachs disease, and about one in five of the Irish (Celtic) population carry the gene for haemochromatosis.

Given the sensitivity around this topic, it is understandable that many scientists prefer to avoid the word 'race' and instead use 'ethnicity'. It is ironic that despite this sensitivity, it is still common for people to use the word 'Caucasian', a word that is closely linked with the eugenics movement. It is self-evident that most white people living in the UK are not Caucasian, highlighted by the fact that a significant number of people from the Caucasus have immigrated into the UK in recent years! In fact, Caucasian is poorly defined and non-specific, and originally included peoples from North India, the Middle East, North Africa, and Europe. In 2003, a 'special communication' in *JAMA* commented that 'At a minimum, journals should follow the guidance of the Council of Science Editors in prohibiting use of the term "Caucasian", which, like "Caucasoid", "Mongoloid", and "Negroid", is based on an outmoded theory of racial distinction, and require use of the term "Asian" instead of "Oriental" or "Asiatic".'[14] In 2004, the National Library of Medicine in the USA changed its medi-

cal subject headings (MeSH) for racial groups (retitled 'continental population groups') with new terms to emphasise that they should refer to geographic origins rather than eugenic theory. So, for example, 'Caucasoid race' was changed to 'European continental ancestry group' (www.nlm.nih.gov/pubs/techbull/nd03/nd03_med_data_changes.html). This is somewhat cumbersome, however, and the most common replacement term for Caucasian is white European, although ideally the description should be as specific as possible (e.g. white British).

It is *BJOG* policy (along with many other biomedical journals) to encourage a more precise description of race based upon geographical (and hence genetic) origin (Box 1), and to differentiate this from ethnicity (culture). This is a policy that will continue to evolve and that hopefully at some point can be replaced by a simple statement of relevant genetic make–up. Indeed, given that self-reported racial origin may not always be reliable, some studies already employ key genetic markers of relevance when describing their populations.

In its MeSH terms, the National Library of Medicine recognises the divisions shown in Box 1; however, the classification relates largely to groups found in significant numbers in the Americas, whereas most human genetic variation occurs in Africa.[15] For example, in African populations at least 22 different alleles of the insulin minisatellite (located within the promoter of the human insulin gene) have been recognised, compared with just three in non-African populations.[16] North African populations have greater overlap with Mediterranean populations than with other African populations, from which they are separated by the Sahara. Some African groupings are very ancient and distinct, for example the Forest people ('pygmies') and the Khoi-San ('bushmen').

I have therefore added the major groupings that are commonly used in *BJOG* in italics, but the basic principle should be to describe the geographical genetic ancestry(ies) of the population studied as accurately as possible, while avoiding terminology associated with the discredited eugenics movement.

In June 2014 the US Department of Health and Human Services Food and Drug Administration recommended collecting both ethnicity and race separately in relation to participants in clinical trials (www.fda.gov/RegulatoryInformation/Guidances/ucm126340.htm), with the ethnicity question preceding the question about race. In the US, they recognise two main ethnic groupings, Hispanic/Latino and non-Hispanic/Latino, whereas they recommend that the following minimum choices be offered for race: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White.

Box 1. **Suggested racial groupings in the National Library of Medicine MeSH (plus groupings set in italics that are commonly used in BJOG)**

Continental ancestry group:

African
African Americans
North African
Central African
Horn of African
South-east African
Southern African
West African
African-Caribbean
American
Indians, Central American
Indians, North American
Indians, South American
Inuits
Asian
Asian Americans
South Asian
India,
Pakistan,
Bangladesh
East Asian
China
Japan
European
White European
Mediterranean
Oceanic
Polynesia

Disclosure of interests

PS has no conflicts of interest to declare. ■

References

1 Olson S. *Mapping Human History*. London: Bloomsbury, 2002.
2 Stringer C. Human evolution: out of Ethiopia. *Nature* 2003;423:692–3, 695.
3 Stringer C, Andrews P. *The Complete World of Human Evolution*. London: Thames and Hudson, 2005.
4 Robin D. *Human Evolution*. London: Pelican Books, 2014.
5 Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, et al. Proportionally more deleterious genetic variation in European than in African populations. *Nature* 2008;451:994–7.
6 Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 2014;505:43–9.
7 Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 2014;507:354–7.
8 Khrameeva EE, Bozek K, He L, Yan Z, Jiang X, Wei Y, et al. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nat Commun* 2014;5:3584.
9 Williams AL, Jacobs SB, Moreno-Macias H, Huerta-Chagoya A, Churchhouse C, Marquez-Luna C, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* 2014;506:97–101.

10 Makgoba M, Savvidou MD, Steer PJ. An analysis of the interrelationship between maternal age, body mass index and racial origin in the development of gestational diabetes mellitus. *BJOG* 2012;119:276–82.
11 Balchin I, Whittaker JC, Lamont RF, Steer PJ. Timing of planned cesarean delivery by racial group. *Obstet Gynecol* 2008;111:659–66.
12 Lara M, Gamboa C, Kahramanian MI, Morales LS, Bautista DE. Acculturation and Latino health in the United States: a review of the literature and its sociopolitical context. *Annu Rev Public Health* 2005;26:367–97.
13 Senior PA, Bhopal R. Ethnicity as a variable in epidemiological research. *BMJ* 1994;309:327–30.
14 Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication. *JAMA* 2003;289:2709–16.
15 Gomez F, Hirbo J, Tishkoff SA. Genetic variation and adaptation in Africa: implications for human evolution and disease. *Cold Spring Harb Perspect Biol* 2014;6:a008524. doi: 10.1101/cshperspect.a008524.
16 Stead JD, Hurles ME, Jeffreys AJ. Global haplotype diversity in the human insulin gene region. *Genome Res* 2003;13:2101–11.