

Race and ethnicity in biomedical literature: A narrative review

Rossella Ferrari

Independent medical writer
Milan, Italy

Correspondence to:

rossella_ferrari@virgilio.it

Abstract

Race and ethnicity are not clearly defined in biomedical literature and misaligned with genomics and epigenomic findings; the guidelines for consistent reporting in publications and regulations from health authorities are lacking.

Minority populations are underrepresented in clinical studies; this limits the identification of risk profiles for diseases (which is the main objective of precision medicine) and fuels false beliefs and implicit bias of clinical decisions. This setting hinders interpreting, generalisation of findings, and prevention planning, and increases socio-economic disparities in healthcare access.

This review outlines recent studies on race and ethnicity and criteria for the proper use of terminology according to evidence, clarity, transparency, and ethics in biomedical documents.

Introduction

For hundreds of years, there has been a vigorous debate about dealing with race and ethnicity categories in studies regarding human health. In particular, how defining and describing race and ethnicity concepts in biomedical literature.¹ In the last century, humans had been classified in the distinct anthropological groups of Caucasoid, Congoloid, Mongoloid, Capoid, and Australoid;² however, these terms have been discarded as new genetic information came to light. In 1994, the Italian geneticist Luigi Luca Cavalli Sforza published the book *The History and Geography of Human Genes*,³ which summarised and evaluated genetic information and the data of genetic diversity of that time. Stemming from the idea that DNA helps track

human origins and history, this book documented the genetic similarities among humans and misleadingly classifying humans according to any “race” concept to explain phenotypic differences. The human genetic studies and genomics research that followed were more advanced and confirmed overwhelming DNA similarities and the negligible DNA differences among humans.^{4,5} However, genetic variations cannot account entirely for the phenotypic differences among humans. To date, as the picture grows in complexity, the debate about race and ethnicity as diversity measures persists in biomedical literature.

The disparity in healthcare access due to race and ethnicity are crucial confounding factors leading to severe consequences. Underrepresented subpopulations in clinical studies can mask the epidemiology of several diseases.^{6–9} Missing data have led to bias; therefore, it is difficult identifying possible connections between socio-demographic and genetic determinants and clinical variables.⁴

This literature review provides an overview of the concepts and terminology of “race” and “ethnicity” and how they have been applied in the biomedical literature and their implications therein.

Definitions of race and ethnicity

In general dictionaries, “race” is defined as a group of people sharing a common origin, and physical features.¹⁰ “Ancestry” or “ethnicity” refers to categories as having a common descent or national and cultural traditions.¹⁰ The designations of race and ethnicity or ancestry in the biomedical literature are highly heterogeneous and inconsistent across countries, clinical studies, and clinical genetics practice.^{2,4} Race and ethnicity data in clinical databases and algorithms are often absent, inconsistent, incomplete, or contradictory, which leads to unreliable interpretation of results.^{11,12} For example, discrepant comparison of lung function

between Blacks and Whites still points to the questionable “race correction” of spirometric measurement in the US since the 19th century and has not yet been updated using the scientific approach and modern methods.¹ Race and ethnicity assessed by different criteria fuel the debate around determinants of diseases. These criteria are often US- or EU-centric and should be evaluated with caution. In the present article, the terminology for race and geographic origins are those used by the original authors of the

articles. Ethnicity and ancestry designations are used interchangeably in the literature, although ancestry usually includes cultural and behavioural features relevant for healthcare.

American Medical Association (AMA) Style suggestions for harmonised designation of race and ethnicity are reported in the paragraph “Because words matter” of this article.

Genetic variants

Natural selection has contributed to genetic variation of individuals or populations.⁵ The sequence of DNA bases of one gene or a

group of genes can permanently change; DNA modification not linked to disease is named a “genetic variant”. Research has highlighted the existence of more genetic variants than socio-cultural categories such as race. Genome-wide association studies were performed to stratify populations according to clusters of gene expression and not geographic origin.² Indeed, the percentage of genetic variation between two subpopulations is low by increasing the number of *loci* analysed, and most genetic variations are tracked among subjects belonging to a single population.⁵

Although genetic variants are not typically linked to diseases, some genetic variants may be associated with the risk of some diseases. For example, the incidence of end stage kidney disease (ESKD) is much higher in African Americans than Whites.¹³ ESKD has been associated with polymorphisms at the *APOLI*

Race and ethnicity data in clinical databases and algorithms are often absent, inconsistent, incomplete, or contradictory, which leads to unreliable interpretation of results.



locus in non-diabetic people with West African ancestry. This genetic variant was selected because it conferred protection against sleeping sickness common in West Africa due to *Trypanosoma brucei*.⁶ However, polymorphisms do not account for the increased risk of kidney disease and no mechanistic relation has been demonstrated until now.^{6,13}

The phenotypes observed among populations have other sources of diversity; the risk of diseases may be linked to external factors that impact the epigenome; various chromatin and RNA modifications have consequences on health from gestation to death.^{6,14} For example, the low mortality rate of SARS-CoV-2 virus infection in Africa compared to Europe, the US, and Asia can

be explained by the differences in environment instead of race. Moreover, clinical laboratories may apply different classifications of genetic data and other parameters suitable for clinical evaluation.⁴

Studies involving South Asian populations living in the US or EU countries (with a high prevalence of type 2 diabetes) lack well-characterised genetic and epigenetic profiles.¹⁵ Conversely, the inclusion of subjects with African ancestry identified novel *loci* in obesity, metabolic syndrome, or immune diseases such as multiple sclerosis.⁷

The genetic and epigenetic profiles can help identify subpopulations at risk of syndromes and diseases, which may be fundamental for

prevention strategies.

Therefore, limiting inclusion of subjects of various subpopulations prevents targeting the objective of precision medicine. Conversely, including diverse subpopulations in genome-wide association studies may strengthen the research, cast light on genotype-phenotype interactions in diseases, and identify new drug targets.⁶

Race and ethnicity in clinical studies

The persistence of false beliefs in race and ethnicity categories in randomised clinical trials and observational studies may impact clinical decisions.¹⁶ Standardised data in registries favour measuring disparities in healthcare access among

different subpopulations. Also, it provides comprehensive epidemiology and prevention strategies in many medical fields.¹⁷ For example, African Americans and Whites with newly diagnosed nonmetastatic prostate cancer and treated with standard healthcare access, after adjusting for demographics, cancer, treatment-related baseline differences, and inverse probability weighting, displayed comparable stage-to-stage prostate-cancer mortality.¹⁸

The Platelet Oriented Inhibition in New TIA and Minor Ischemic Stroke (POINT) trial compared aspirin plus clopidogrel versus aspirin alone at 90-day follow-up in 4044 US subjects.¹⁹ The subgroups of Black participants (918/4044, 22.7%) had a higher cumulative risk of stroke than White patients. The adjustment for covariates (demographic data, comorbidities, and adherence to aspirin plus clopidogrel treatment) confirmed the higher risk of early recurrence of stroke after minor ischemic stroke or TIA of Black participants.¹⁹ The Reasons for Geographic and Racial Differences in Stroke (REGARDS) study involved 9416 Blacks and 13,091 Whites without a history of CV diseases.²⁰ At 6.1-year follow-up, compared to Whites, Blacks showed a significantly higher risk of sudden cardiac death (SCD), confirmed after adjustments for socio-demographic, comorbidities, health behaviour changes, intervening CV events, and risks of non-SCD mortality causes. However, these variables did not account for the higher incidence of SCD in Black patients.²⁰

The perception of pain is complex and may be influenced by cultural differences. Still, implicit bias about race and ethnicity in pain can increase the burden of pain, blur the assessment and mislead recommendations.¹⁶ In 2017, the National Institute of Health (NIH) supported the OPPERA cohort study on orofacial pain enrolling White, Black/African American, Hispanic, Asian, Native Hawaiian/Pacific Islander, American Indian/Alaskan Native, and other populations subjects. The study did not find any racial differences among the populations in tissue characteristics and nociceptive sensitivity by 34 pressure, mechanical, and thermal pressures.²¹ The NIH study of Reynolds Losin et al. (2020) analysed the pain perception pathways in three ethnic groups by functional



Magnetic Resonance Imaging. This study highlighted similar nociceptive pain processing among the groups, which overturns the influence of races, ethnicity or culture on the complexity of pain perception.²²

Underreporting race and ethnicity data or inconsistent reporting without standardised methods in orthopaedics,²³ surgery,²⁴ real-world data of medical devices,²⁵ anaesthesia,²⁶ or other specialties hinders the possibility to identify differences in treatments, post-intervention outcomes,²³ and real-world evidence.²⁵ Most US and UK healthcare systems usually collect data on race and ethnicity,¹⁷ but most EU countries do not.⁸ Study designs and statistical protocols are essential to highlight, deepen, or confirm clinical similarities or differences among subpopulations. It also eases to evaluate the contribution of socio-demographic variables and comorbidities.

Implications

Limited studies on the health of ethnic minorities can have a negative impact on healthcare expenditure for diseases like diabetes, mental health, or infectious diseases.²⁷ Race and ethnicity (alongside other determinants) seem to account for differences in insulin regulation and glycaemic response to carbohydrates; however, given the scarcity of studies, recommendations on insulin dosing and formulations in more diverse populations is still lacking.²⁸

The *New England Journal of Medicine* editors' team has recently marked the value of inclusion

of various subpopulations in research studies for the generalisability of the findings and the extension of new treatments.⁹ Subgroup analyses of clinical studies can highlight risk factors or diseases determinants of the diverse subpopulations; they also can increase the equity of the access and provision of healthcare. Moreover, transparency favours the decision of the reviewers and publishers on publishing manuscripts. From January 2022, authors who intend to publish in NEJM will be asked to provide supplementary information tables about the representativeness of the patient populations enrolled in the studies.⁹

Because words matter

The guidelines for specifying the reasons to use race and ethnicity terms in biomedical publications (e.g., generalisability, disparities in healthcare and expenditure) were published in 2003 when the Human Genome Project was completed.⁴ However, the original five-group anthropologic classification is still used, yet reduced to the three major NIH population ancestries (European/ Caucasians, African, and Asian).² Numerous medical documents may include race and ethnicity terms:

- Regulatory documents such as protocols, case report forms, Summary of Product Characteristics, and leaflets
- In medical communication such as manuscripts for publication in peer-reviewed journals
- Project descriptions for grants or funding proposals

Table 1 summarises the current evaluations and suggestions of the editorial associations for reporting race and ethnicity in biomedical literature.

Recently, the JAMA editorial team has published practical guidelines to improve the quality of reporting race and ethnicity data in regulatory documents or clinical studies.¹⁰ Race and ethnicity designations must always be consistent and justified. As social constructs, the utility of race and ethnicity in biomedical research and practice is limited; however, its pretextual use can help highlight disparities and pitfalls. In this view, the solutions proposed by the AMA Manual of Style committee are

continuously under revision according to cultural and social evolution and open to feedback of authors, editors, and readers to enhance the correctness of reporting terms.¹⁰

- **Methods:** should explain how race and ethnicity or ancestry have been identified (e.g., self-reported or by the investigator or database or other modalities). Data collection on race and ethnicity must be motivated and contextualised according to socio-economic settings relevant as health determinants.
- **Results:** the ethnic categories can be listed in alphabetic order instead of numerical majority and specified as the “others” group.
- **Discussion:** structural racism or disparities in healthcare can be highlighted and contextualised. Discussion or Conclusions sections should suggest appropriate studies to identify variables and determinants of health. The terms for defining race and ethnicity have to be specific. For example, “African American”

or “Black” can be substituted by “African descendant” as this term underlines not only the origin but also culture and traditions. However, the “African descendant” designation is questionable if culture and traditions are not practised.

The AMA committee suggests capitalising the name of races or ethnicities, e.g., White, Hispanic, Latino, or Asian. They suggest avoiding categories like “Asians” or “Blacks”; instead, adjectival nouns would be more appropriate (e.g., “Asian women” and “African American patients”). Adding the geographic origin to race and ethnicity definition can be relevant. It can, however, be challenging. The term “Caucasian” refers to the region of Eurasia. Therefore, “Caucasian” should be used only for people from that region and not as a synonym for White people.¹⁰

Since the inclusion criteria of various

populations entail a standard designation of race and ethnicity, different protocol measures and suitable calculations methods for sample sizes are required for study designs and proper reporting.¹⁰

Based on the existing International Committee of Medical Journal Editors (ICMJE), the editorial guidelines of the publishers should focus on reporting race and ethnicity data with clear clinical motivations related to the research questions for the biomedical studies. The editorial teams should harmonise recommendations and suggestions in collecting and reporting data. Statements and designations of race and ethnicity should be applied not only by authors but also by publishers and reviewers.

Legislative framework

The primary part of the EU legislation is the treaties and the secondary are laws (Directives). Discrimination based on race and ethnicity is

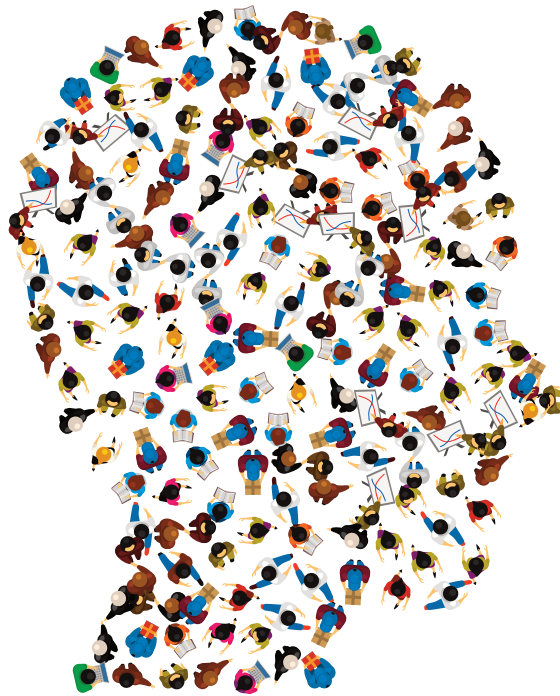
Table 1. Current evaluations and suggestions of editorial associations for reporting race and ethnicity in biomedical literature

2019 ICMJE ²⁹	Because the relevance of such variables as age, sex, or ethnicity is not always known at the time of study design, researchers should aim for inclusion of representative populations into all study types and at a minimum provide descriptive data for these and other relevant demographic variables and recommend “Authors should define how they determined race and ethnicity and justify their relevance. Authors should use neutral, precise, and respectful language to describe study participants and avoid the use of terminology that might stigmatise participants.” ²⁹
EQUATOR (Enhancing the Quality and Transparency Of health research) network ³⁰	The use of the “race and ethnicity” terms is only partially addressed.
COSORT 2010 ³¹	The “ethnicity” variable has been quoted in the Item 21 paragraph about the generalisability of trials findings in some examples (i.e., in Table 4), but no designation and suggestion about reporting race and ethnicity have been provided.
STROBE ³²	No mention
European Association of Science Editors ³³	“Race” was mentioned only as a variable to be disaggregated.
COPE (Committee on Publication Ethics) ³⁴	Does not provide any specific core practice for reporting in research studies.
CDC (Centers for Disease Control and Prevention) ³⁵	Acknowledged the problem regarding race and ethnicity and highlighted general principles with different expressions in health communication, such as: <ul style="list-style-type: none"> ● Instead of “high risk group” prefer “disproportionately affected groups” ● Instead of “racial or ethnic groups” prefer “people from racial or ethnic groups” ● Instead of “minority” prefer “(people from) racial and ethnic minority groups”

banned and is explicitly stated in the Directive 2000/43/EC and treaties. However, the EU legislation lacks non-discrimination laws on access to healthcare. The responsibility of non-discrimination is held by the national regulations of each EU country that have variably weak legal platforms regarding race and ethnicity (article 168 of the Treaty on the Functioning of the EU Union).³⁶ As reported in cases studies (e.g., the antihypertensive BiDil³⁷), the racialisation of drug regulation has been rising in the US and EU.³⁷ The concept of “racialisation of pharmaceutical regulation” refers to how race and ethnicity have become important to drug testing and evaluation.³⁷ A recent comparison of 397 new drugs approved in the US and Europe has highlighted the uninterrupted lack of concordance between the pharmaceutical legislations by specific tools like the International Conferences on Harmonisation.³⁷ This comparison has revealed inconsistent designations of race and ethnicity in the labels or “Summary of Product Characteristics” of pharmaceutical products. FDA emphasises the inclusion of race and ethnicity subgroups in the labels more than EMA, but this inclusion and the reported differences are less frequent in clinical trials.³⁷

Pharmaceutical regulations on drug approvals lack data on the different effects of pharmaceutical products in various populations because most of the registrational clinical trials performed during drug development include mostly White patients. For example, given the difference in genetic variants, the algorithm for the dose of warfarin may differ in Whites and African descendants.⁶ To note, health authorities do not require pharmaceutical industries to enrol subjects belonging to minorities in clinical studies, nor in numbers that enable proper analyses and conclusions on drug effectiveness.³⁸ This shortcoming lowers the robustness of meta-analyses,³⁹ limits having a complete pharmacovigilance system of a drug’s adverse events, and consequently, risks of knowledge gaps in drugs profiles.

An amendment to the EU legislative framework should be considered essential. In particular, pharmaceutical regulations should require an equitable enrolment of patients in clinical studies.



Conclusions

“Race” and “ethnicity” or “ancestry” are complex terms that need increased knowledge and in-depth analysis in biomedical literature. The support of the legislation at EU and local levels could ease the advancement of the scientific evidence with positive implications in healthcare access.

Evidence and evaluations of all the stakeholders can lead to the consistent and specific use of race and ethnicity concepts in regulatory documents and publications and pinpoint their relevance in clinical practice.

Acknowledgements

The author thanks Dr Komla Sena Amouzou, Surgery Department of Sylvanus Olympio Teaching Hospital of Lomé (Togo), and Dr Opeyemi Eseyin, Department of Paediatrics, General Hospital, Ebute-metta, Lagos (Nigeria) for the revision of the manuscript and their critical feedback.

The author thanks Clare Chang for her help in editing this manuscript.

Disclaimers

The opinions expressed in this article are the author’s own and not necessarily shared by her employer or EMWA.

Disclosures and conflicts of interest

The author declares no conflict of interest.

References

- Braun L. Race correction and spirometry: why history matters. *Chest*. 2021;159(4):1670–75.
- Zhang F, Finkelstein J. Inconsistency in race and ethnic classification in pharmacogenetics studies and its potential clinical implication. *Pharmacogenomics Pers Med*. 2019;12:107–23.
- Cavalli-Sforza LL, Menozzi P, Piazza A. The history and geography of human genes. Princeton, NJ: Princeton University Press, 1996.
- Bonham VL, Green ED, Perez-Stable EJ. Examining how race, ethnicity, and ancestry data are used in biomedical research. *JAMA*. 2018;320(15):1533–4.
- Rosemberg NA. A population-genetic perspective on the similarities and differences among worldwide human populations. *Hum Biol*. 2021;92(3):135–52.
- Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell*. 2019;177(1):26–31.
- Bentley AR, Callier SL, Rotimi CN. Evaluating the promise of inclusion of African ancestry populations in genomics. *NPJ Genom Med*. 2020;5:5.
- Shaaban AN, Peleteiro B, Martins MRO. The writing’s on the wall: on health inequalities, migrants, and Coronavirus. *Front Public Health*. 2020;8:505.
- The Editors. Striving for diversity in research studies. *N Engl J Med*. 2021;385(15):1429–30. doi: 10.1056/NEJMe2114651.
- Flanagin A, Frey T, Christiansen SL for the AMA Manual of Style Committee. Updated guidance on the reporting of race and ethnicity in medical and science journals. *JAMA*. 2021;326(7):621–7.
- Hartzler BM, Snyder A. Caring by numbers: evaluation of inconsistencies and incompleteness in the reporting of racial and ethnic data. *J Racial Ethn Health Disparities*. 2017;4(6):1092–99.
- Singh S, Steeves V. The contested meanings of race and ethnicity in medical research: a case study of the DynaMed Point of Care tool. *Soc Sci Med*. 2020;265:113112.
- Umeukeje EM, Young BA. Genetics and end stage kidney disease disparities in African Americans. *Am J Kidney Dis*.

- 2019;74(6):811–21.
14. Torano EG, Garcia MG, Fernandez-Morera JL et al. The impact of external factors on the epigenome: *in utero* and over lifetime. *Biomed Res Int*. 2016;2016:2568635. doi: 10.1155/2016/2568635.
 15. Canchizos-Achiricha M, Fedeli U, Sattar N et al. Epidemiology, risk factors, and opportunities for prevention of cardiovascular disease in individuals of South Asian ethnicity living in Europe. *Atherosclerosis*. 2019;286:105–13.
 16. Hoffman KM, Trawalter S, Axt JR et al. Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proc Natl Acad Sci*. 2016;113(16):4296–301.
 17. Rodriguez-Lianz A, McDonald M, Fonseca-Ford M et al. Collection of data on race, ethnicity, language, and nativity by US public health surveillance and monitoring systems: gaps and opportunities. *Public Health Reports*. 2018;133(1):45–54.
 18. Dess RT, Hartman HE, Brandon MS et al. Association of black race with prostate cancer specific and other-cause mortality. *JAMA Oncol*. 2019;5(7):975–83.
 19. Kamel H, Zhang C, Kleindorf DO et al. Association of black race with early recurrence after minor ischemic stroke or transient ischemic attack. *JAMA Neurol*. 2020;77(5):601–5.
 20. Deo R, Safford MM, Khodneva YA et al. Differences in risk of sudden cardiac death between Blacks and Whites. *Am Coll Cardiol*. 2018;72:243X-2439.
 21. Ostrom C, Bair E, Maixner W et al. Demographic predictors of pain sensitivity: results from the OPPERA study. *J Pain*. 2017;18(3):295–307.
 22. Reynolds Losin EA, Woo CW, Medina NA et al. Neural and sociocultural mediators of ethnic differences in pain. *Nat Hum Behav*. 2020;4(5):517–39.
 23. Paul RW, Lee D, Brutico J et al. Reporting and analyzing race and ethnicity in orthopaedics clinical trials: a systematic review. *J Am Acad Orthop Surg Glob Res Rev*. 2021;5(5):e21.00027.
 24. Maduka RC, Broderick M, White EM et al. The reporting of race and ethnicity in surgery literature. *JAMA Surg*. 2021;e212752.
 25. Tarver ME. Race and ethnicity in real-world data sources: considerations for medical device regulatory efforts. *J Prim Care Community Health*. 2021;12:2150132721994040.
 26. Begic D, Janda-Martinac C, Vrdoljak M et al. Reporting and analyses of sex/gender and race/ethnicity in randomized controlled trials of interventions published in the highest-ranking anesthesiology journals. *J Comp Eff Res*. 2019;8(16);1417–23.
 27. Snijder MB, Gelenkamp H, Prins M et al. Cohort profile: the healthy life in an urban setting (HELIUS) study in Amsterdam, The Netherlands. *BMJ Open*. 2017;7:e017873.
 28. do Vale Moreira NC, Ceriello A, Basit A et al. Race/ethnicity and challenges for optimal insulin therapy. *Diabetes Res Clin Pract*. 2021;175:108823
 29. ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated December 2019.
 30. Enhancing the QUALity and Transparency of health research (cited 2021 Dec) Available from: <https://www.equator-network.org>
 31. Moher D, Hopewell S, Schultz K et al. CONSORT 2010 Explanation and elaboration: updated guidelines for reporting parallel group randomized trials. *BMJ*. 2010;340:c869.
 32. van Elm E, Altman DG, Egger et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12:1495–9.
 33. EASE Guidelines for Authors and Translators of Scientific articles to be published in English. November 2018.
 34. COPE (Committee on Publication Ethics) (cited 2021 Sept 17). Available from: <https://publicationethics.org/core-practices>
 35. CDC (cited 2021 Sept 17). Available from: https://www.cdc.gov/healthcommunication/Key_Principles.html
 36. Orzechowski M, Nowak M, Bielinska K et al. Social diversity and access to healthcare in Europe: how does European Union's legislation prevent discrimination in healthcare? *BMC Public Health*. 2020;20:1399.
 37. Mulinari S, Vilhelmsson A, Ozieranski P et al. Is there evidence for the racialization of pharmaceutical regulation? Systematic comparison of new drugs approved over five years in the USA and the EU. *Soc Sci Med*. 2021;280:114049.
 38. Nayak A, Hicks AJ, Morris AA. Understanding the complexity of heart failure risk and treatment in African Americans. *Circ Heart Fail*. 2020;13(8):e007264.
 39. Liu P, Ross JS, Ioannidis JPA, Dhruva SS et al. Prevalence and significance of race and ethnicity subgroup analyses in Cochrane interventions reviews. *Clin Trials*. 2020;17(2):231–4.



Author information

Rossella Ferrari has a degree in biology. After working in the marketing departments of multinational pharmaceutical companies, she was responsible for medical communications and editorial projects at specialised agencies and publishers from 2003 to 2011. Currently, Rossella works as a freelance medical writer.