Assessing the Legality of Using the Category of Race and Ethnicity in Clinical Algorithms - the EU Anti-discrimination Law Perspective

Malwina Anna Wójcik [1,2]

[1] University of Bologna (Italy),
[2] University of Luxembourg (Luxembourg)

Abstract

The growing use of machine learning (ML) in medical prognostics, diagnostics, and treatment recommendations offers new powerful tools for addressing pressing health challenges. However, similarly to their knowledge-based predecessors, ML algorithms are not neutral and can mirror prevailing patterns of power and disadvantage in healthcare, entrenching racial disparities. The developers of clinical algorithms often face a paradox. Ignoring race can introduce racial bias through proxies. However, explicitly taking race into account can also replicate harmful stereotypes embedded in the category of race. This paper explores the problem of race and ethnicity in clinical algorithms from the European legal perspective, offering three main contributions. Firstly, it tackles the importance of the correct operationalization of race, arguing that since race is a social construct, developers of clinical algorithms should pay particular attention to the dimension of race that the data represents. In doing so, the paper addresses the challenges to race operationalization in the European context. Secondly, the paper analyses how race is used in the design of clinical algorithms, with a particular focus on fairness measures that take race into account at the prediction time, including race correction. Thirdly, the paper explores the legality of such measures under the Racial Equality Directive.

Keywords [1]

Clinical algorithms, race-correction algorithms, ML in healthcare, racial fairness, EU anti-discrimination law

1. Introduction

The growing availability of digital health data paves the way for the development of machine learning (ML) algorithms for medical diagnosis, prognosis, and treatment recommendation. These technologies are increasingly used to support clinical decision-making. For instance, ML solutions are expected to transform the practice of oncology, helping with early diagnosis of lung cancer and delivering tailored therapy [33]. Other recently discussed potential applications include predicting the need for blood transfusion during surgery [36] or triaging patients with acute chest pain syndrome [31].

However, there are several barriers to the effective implementation of ML models in clinical tasks. One of them is their propensity to reinforce harmful biases, including racial ones. Studies conducted in the US demonstrated that ML algorithms often result in discriminatory treatment of non-White patients, replicating and amplifying disadvantages that they have been suffering when accessing healthcare [43], [51]. Algorithmic bias can have dramatic consequences in the medical domain, endangering the health, or even life, of patients belonging to racial minorities, for instance through an incorrect diagnosis of disease [18] or deprioritization in organ allocation [34].

EWAF'23: European Workshop on Algorithmic Fairness, June 07–09, 2023, Winterthur, Switzerland
EMAIL: malwinaanna.wojcik2@unibo.it
ORCID: 0000-0002-9565-8811

While a growing body of literature focuses on the problem of machine bias in automated decisions, it should be underlined that the issue of racial discrimination by clinical algorithms predates the popularization of ML. For instance, a recent study of algorithms predicting stroke has demonstrated that novel ML techniques do not significantly improve predictive accuracy in comparison with the well-established pooled cohort equations [24]. Both the new and the old techniques of stroke prediction are less accurate in the case of Black patients [24]. This shows that traditional knowledge-based algorithms that have long been used in medical domain are not immune to bias against non-White patients. In particular, the practice of race correction, understood as adjusting the outcome of the prediction based on a patient's race, has come under increased criticism for reinforcing racial stereotypes in medicine [55]. The problem receives growing attention in the US, where the Department of Health and Human Services has introduced a proposal intended to combat discrimination by algorithms in healthcare decision-making [54]. However, the issue remains underexplored from the European perspective. This contribution aims to address this gap.

The European anti-racist identity results in the rejection of the category of race in the social, legal and political discourse. However, the idea of European color-blindness has not necessarily led to eradicating racism, including in the area of healthcare. Studies conducted by the Fundamental Rights Agency (FRA) [15] and the European Commission against Racism and Intolerance (ECRI) [11] indicate that patients belonging to racial and ethnic minorities often face discriminatory treatment and structural inequalities when accessing healthcare. One of the most graphic examples of discriminatory treatment of ethnic minorities is obstetric violence towards Roma women, encompassing forced sterilization and segregation in maternal wards [4].

Unfortunately, data concerning racial and ethnic discrimination and disparities in medicine is scarce, as the European ambivalence towards the concept of race prevents the establishment of coherent measurement metrics. This has crucial consequences for the development of accurate and fair clinical algorithms, exacerbating the problems concerning the operationalization of race. General consensus persists regarding the fact that race is not a natural attribute, but rather a social construct. Thus, developing successful racial fairness interventions in a clinical algorithm requires understanding what race means in a given context. This problem is often overlooked by the developers of algorithms, who rarely pay attention to the purpose for which the racial data was collected and their measurement methods. Further complications arise from the ongoing debate about the merits of race-based medicine and the misconceptions of race as a biological concept.

This paper focuses on the use of race and ethnicity in clinical algorithms and its compatibility with European anti-discrimination law. In particular, it tackles the importance of correct operationalization of race and the issue of using race at the time of prediction. As these problems are universal, and not limited to AI technologies, this paper adopts a broad definition of clinical algorithms, encompassing both knowledge-based and ML models that are deployed to diagnose patients, prognose the development of diseases, and offer treatment recommendations.

This contribution starts by outlining the origins of race as a biological concept and describes the subsequent rejection of this idea in favor of a social definition of race. It then proceeds to explore the European approach to the concepts of race and ethnicity, pointing out definitional struggles and inconsistencies in the legal field. Next, the paper notes an unexpected development in the regulation of pharmaceuticals in Europe – the increasing inclusion of racial and ethnic differences in drug effect in the summaries of product characteristics. It argues that this practice could have an effect on the popularization of race-aware medicine in Europe, which is problematic coupled with a lack of regulatory guidelines concerning the operationalization of race. The paper continues to explore the latter issue in the context of clinical algorithms, arguing that an incorrect and incoherent operationalization of race can prevent developers of clinical algorithms from successfully addressing healthcare inequalities. Then, the contribution proceeds to explore how race is used in the design of clinical algorithms, focusing, in particular, on the practice of race correction. It argues that, while using the category of race is often necessary to ensure the fairness of the algorithm, race correction practices that are not evidence-based can in fact constitute discrimination. Hence, the final part of the paper concerns the legality of using race and ethnicity in fairness interventions under the Racial Equality Directive

(RED).[2] It argues that in cases where race plays a decisive role in algorithmic decision-making, a *prima facie* case of direct discrimination might be established. It then explores whether such a practice can be justified under the positive action doctrine.

2. Race and ethnicity in healthcare - the European perspective

2.1. Race as a biological concept - the origins of racial categorization and race-based medicine

The origins of racial categorization can be traced back to the 19th century scientific theories on race, which sought to justify colonialism and slavery based on the existence of biologically distinct races, identified with reference to physical attributes, including skin color and facial appearance [3]. These theories drew from the eugenics movement in order to establish a scientific basis for the physical and intellectual superiority of the White European race over other peoples. Scientific race theories influenced the development of medicine, contributing to the belief that different races have different blood types, which subsequently led to the racialization of certain diseases which were believed to be more prevalent in non-Caucasian populations, such as Sickle Cell Anemia or Tay Sachs Disease [26]. In turn, the "purity of blood" and disease prevention concerns were used to legitimate practices compromising the autonomy of racial minorities, including sterilization, marriage control, and birth control. The entrenched belief in biological differences between races produced pseudo-scientific conclusions about the Black body, which was perceived as more "hard" and "durable" than the White body [26].

After the 2nd World War, the atrocities of the Holocaust, whose ideological basis largely stemmed from the scientific race theory and eugenics, accelerated the spread of emerging criticism of race as a biological concept. In 1978, Art. 1 of the UNESCO Declaration on Race and Racial Prejudice clarified: "The differences between the achievements of the different peoples are entirely attributable to geographic, historical, political, economic, social, and cultural factors. Such differences can in no case serve as a pretext for any rank-ordered classification of nations or peoples."

Nowadays, race is no longer considered a scientific variable, but a dynamic social construct, shaped by geographic, cultural, and socio-political forces [17]. However, deeply rooted stereotypes about the biological origin of race persist in bio-medical practice and research. For instance, a study of UpToDate, a widely used medical knowledge support tool, revealed that 93.3% of documents in its database biologized race, often inappropriately linking the Black race with genetics or clinical phenotype [6]. An analysis of how racial and ethnic categories are used in genetic and genomic research highlights that genetic essentialism, understood as the belief that most differences between humans can be explained and analyzed through genetic variance, often leads to false biologization of race, which should not be considered an adequate proxy for genetic ancestry [37]. While genetic and phenotypic differences between humans are a scientific fact, they "do not create biologically understood races, but merely provide the basis for the emergence of socio-cultural racial classifications" [39]. This is well illustrated by the example of kidney transplant criteria. As described by Lebret, in the US, being African American had been long considered a risk factor in kidney donation [34]. However, the new criteria which replaced race with genotype have led to more accurate and fair results, proving that race and genetic ancestry are not synonyms [27].

2.2. The European silence about race

In the modern socio-political and legal discourse the concepts of race and ethnic origin in Europe remain definitional puzzles. The strong European identification with anti-racism, stemming from decolonization efforts and the legacy of the Holocaust, translates into the "silence about race" tradition based on the presumption that defining the concept of race "has become increasingly unacceptable in modern societies".[3] In his monograph, Möschel identifies four reasons for the exclusion of "race" from

[2] Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin.

[3] Opinion of Advocate General Wahl delivered on 1 December 2016 in case C-668/15 *Jyske Finans AS v Ligebehandlingsnaevnet*, para 31.

European legal and political discourse [41]. Firstly, the use of racial categories is feared to perpetuate the existence of biologically defined races, and thus, to promote racism. Secondly, the European legacy of Marxism and Socialism entrenched the belief that other categories, in particular class and gender, are more relevant for studying the patterns of discrimination and disadvantage. Thirdly, the European experience of the Holocaust makes the very act of collecting data based on racial classifications unconscionable. Fourthly, there is a fear that such sensitive data, even if collected without racist intentions, could be misused by the governments.

The European ambivalence about race is reflected in the EU anti-discrimination law, and, in particular, the Racial Equality Directive (RED), which prohibits direct and indirect discrimination based on race and ethnic origin, without defining them. The rationale for the absence of definition is provided by Recital 6 which states: "The European Union rejects theories which attempt to determine the existence of separate human races. The use of the term 'racial origin' in this Directive does not imply an acceptance of such theories."

In a similar manner, none of the EU Member States provides an explicit definition of race or ethnic origin in their domestic law [13]. Likewise, no definition of these terms exists in the European Convention on Human Rights (ECHR). Moreover, neither of the regional anti-discrimination regimes defines the term "racial discrimination" either, leaving it open to judicial interpretation. In this matter, the European Court of Justice (ECJ) and the European Court of Human Rights (ECtHR) often evoke Art. 1 of the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), which states that racial discrimination "shall mean any distinction, exclusion, restriction or preference based on race, color, descent, or national or ethnic origin."

In Europe the terms "race" and "ethnic origin" are often used interchangeably, generating confusion about the relationship between them. Both the ECJ[4] and the ECtHR[5] held that discrimination based on ethnic origin can constitute a form of racial discrimination, embracing the definition provided by ICERD. However, jurisprudence also notes differences between the two concepts. The delineation leads to inconsistencies – while ethnic origin falls under the scope of race, the two of them are defined as very different phenomena. In its seminal judgment, the ECtHR explained:

"Ethnicity and race are related and overlapping concepts. Whereas the notion of race is rooted in the idea of the biological classification of human beings into subspecies according to morphological features such as skin color or facial characteristics, ethnicity has its origin in the idea of societal groups marked by common nationality, tribal affiliation, religious faith, shared language, or cultural and traditional origins and backgrounds."[6]

Notably, although the judgment makes an explicit reference to ICERD, the Strasbourg Court does not use the language of Art. 1 (for instance "color and descent") to define race. Instead, it reduces race to a biological concept of human "subspecies". This is problematic for two main reasons. Firstly, the definition adopted by the Court can be read as an affirmation of the existence of biological races. Secondly, it fails to capture "race" as a social phenomenon, entrenching the narrow understanding of racism which does not encompass cultural or systemic racism.

2.3. Towards race-aware medicine? The European paradox of racial labeling of pharmaceuticals

In spite of the controversies surrounding ethno-racial categories in medicine, many regulatory regimes institutionalize their use. A key example is the US, where the medical practice has long supported the development of racially tailored treatments and medicines. For instance, in 2005, the Food and Drug Administration (FDA) approved BiDil, the first heart failure drug recommended exclusively for Black patients [28]. Although BiDil remains the only race-specific drug to date, the federal law mandates the FDA to assess the inclusion of race and ethnic minorities in clinical trials of drugs, biologics, and medical devices, as well as to ensure the presence of safety and effectiveness data

[4] Case C-83/14 *CHEZ Razpreldelenie Bulgaria AD v Komisia za Zastita ot Diskriminatsia* [2015] EU:C:2015:480.

[5] *Timishev v Russia*, Applications nos. 55762/00 and 55974/00, judgment of 13 December 2005.

[6] Ibid, para 55.

by race and ethnicity.[7] Moreover, robust FDA guidelines exist on the operationalization of racial and ethnic census data in medical research [9].

Similar institutional commitment is lacking in the case of the European Medicines Agency (EMA), which did not issue any guidance pertaining to the inclusion of racial and ethnic categories in the regulation of pharmaceuticals and medical devices. The only official guidelines on this point come from "Guidance on Ethnic Factors in the Acceptability of Foreign Clinical Data" introduced by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH-E5 Guidance) and adopted by the EMA. It suggests that the registration of pharmaceuticals in the ICH regions (USA, EU, and Japan) requires data on three major races: Asians, Blacks, and Caucasians. Moreover, there is evidence that when approving US pharmaceuticals, the EMA shapes its reporting of ethnic and racial demographics on the US model [42].

The EMA's hesitancy to explicitly regulate the matter could be attributed to the fact that race remains a deeply problematic concept in Europe because of its association with biological racism. On the other hand, the interest in clinical differences between races seems growing. A recent comparative analysis of summaries of product characteristics of medicines approved by the EMA and the FDA shows that race labeling is increasingly common in Europe, with almost half of the examined summaries containing ethno-racial demographic information [42]. Moreover, while US summaries contained more information on the demographics of clinical trial participants, EU summaries contained more statements about actual racial and ethnic differences in drug effects. This is highly surprising given the lack of EMA's guidelines on the matter and the European "silence about race" tradition.

Although the EMA has not recommended any drug for a specific race or ethnicity, the increasingly common inclusion of possible racial differences in summaries of product characteristics is not trivial, as it can influence the practice of medical professionals, leading to the popularization of race-conscious medicine.

It must be noted that the concept of race-based medicine has been highly contentious. US scholars have long criticized the use of racial categories in healthcare for being based on racial stereotypes instead of sound medical practice [28]. Apart from inheriting this criticism, transplanting the concept of racial medicine to Europe risks further challenges. Unlike the US where census data contain self-reported race, the vast majority of Member States do not collect race data in a systemic manner. This is problematic because the availability of local demographic data is the prerequisite for developing race-conscious interventions which serve equity purposes. The three-race classification of the ICH-E5 Guidance might be inaccurate in the European context, as it is based on a shaky compromise regarding the definitions of race and ethnicity [32]. Likewise, caution should be advised when transplanting the race categories used in the US to the European practice of medicine.

3. Race in clinical algorithms

The ambiguity surrounding the definition of race and ethnicity and their potential clinical relevance invite careful consideration of how race is conceptualized and used in clinical algorithms, especially in light of the potentially growing interest of the European regulator in race-aware medicine. The lack of coherent European standards concerning the operationalization of race in clinical practice and research affects the quality of medical data concerning racial minorities, creating obstacles to the development of fair and accurate clinical algorithms.

3.1. How is race operationalized in clinical algorithms?

The concept of race, as opposed to being a fixed feature, consists of different dimensions. These include:

- self-identified race (the race that a person identifies with);
- self-classified race (racial self-classification based on available criteria);
- observed race (the race that others ascribe to a person based on appearance and interaction);
- reflected race (the race that a person believes others assume them to be);

[7] Section 907 of the Food and Drug Administration Safety and Innovation Act.

- phenotype (racial appearance judged by predetermined criteria) [49].

These dimensions are measured in a different manner and are appropriate for studying differing phenomena, including various contexts of discrimination. Therefore, inconsistencies can exist between different dimensions of race. For instance, in the context of healthcare, self-identified race often differs from race ascribed in medical records or interviewer classifications [50].

Unfortunately, the multidimensionality of race is often overlooked in the process of developing clinical algorithms. While the main sources of health data, such as electronic health records, may contain explicit racial labels, the developers of the algorithm rarely focus on the purpose for which the data were collected. Moreover, little attention is paid to how racial categories are measured, leading to the lack of method and consistency in using different dimensions of race [37]. As a result, the category of race is often "flattened" and falsely reduced to a single attribute. Since race is essentially a social construct, it cannot be treated as an intrinsic and immutable feature. Its incorrect or inconsistent operationalization can lead to the misrepresentation of racial disparities and can jeopardize the success of an anti-racist algorithmic project.

Let us consider an ML algorithm trained to detect skin cancer. A common fairness problem associated with these algorithms is that they underperform on darker-skinned individuals because the database on which they are trained is often non-representative [1]. In this case, phenotypic labeling appears to be the most appropriate to address the problem. Self-identified race would not constitute a reliable label, since individuals identifying as Black can, in fact, represent a variety of skin tones. This is illustrated in the seminal study of Buolamwini and Gebru who used phenotype instead of demographic data on race and ethnicity to benchmark facial recognition algorithms [6].

Conversely, self-identified race can be an appropriate label in an algorithm to predict the development of certain mental disorders [38]. For instance, it has been suggested that stress associated with experiences of racism and exclusion puts individuals belonging to racial and ethnic minorities at an increased risk of first-episode psychosis, regardless of their socio-economic status and living conditions [30]. Therefore, the use of phenotype or observed race risks not capturing the individual discrimination experience, which appears to be the cause of possible mental health problems.

Observed race, on the other hand, can be an appropriate label in algorithms deployed in clinical tasks whose outcomes are known to be influenced by historic patterns of discrimination [58]. For instance, there is evidence that doctors often suspect patients of minority backgrounds of exaggerating their health problems to claim benefits [15]. This, along with other prejudices, can cause doctors to administer inadequate doses of painkillers to racial or ethnic minority patients [40]. Mitigating these inequalities in an algorithm recommending medicine doses in pain therapy is thus best achieved through the use of observed race in the labels, which captures the doctor's perception of the patient's race.

The race measurement problem is exacerbated by the European culture of silence about race. The choice of correct race measurement in clinical algorithms requires, first and foremost, an identification of the nature of disparities relevant in a given context. These are best captured by reviewing data on inequality. The lack of definition of race and ethnic origin at European and Member States levels contributes to the limited and inadequate collection of disaggregated data on minorities, especially using the observed race dimension, which is crucial for detecting inequalities arising from discriminatory practices in healthcare. Thus, racial and ethnic minorities are often categorized in binary opposition to the majority population of Member States, using general terms, such as "migrant background" [13]. Unfortunately, treating various minorities as a homogeneous group does not capture the unique nature of disparities that some of these groups might suffer. Another problem arises in the context of self-identified race since minorities in continental Europe tend not to identify with race at all, as evidenced by the fact that "claims for recognition as a racial minority are sporadic, peripheral or non-existent" [13].

Unreliable data or lack thereof increase the probability of choosing an incorrect dimension of race in the algorithmic design process. Therefore, the lack of a European standard for collecting racial and ethnic minority data has been increasingly criticized. For instance, in its latest annual report, ECRI urges that establishing a standard for collecting data on patient's race, ethnicity and language is necessary to identify and address health disparities [11].

3.2. How is race used in the design of clinical algorithms?

The complexities associated with the use of race in medicine, and clinical algorithms in particular, have caused some commentators to argue for alternative solutions. The first is to completely erase the category of race from algorithmic inputs and outputs, creating a race-blind algorithm. However, the absence of race as a feature does not guarantee that an algorithm does not replicate racial prejudices through the use of proxies. For instance, in the well-known study by Obermeyer *et al*., the algorithm that used healthcare spending as a proxy for illness falsely attributed a lower risk of serious disease to Black patients, reflecting unequal access to healthcare [43]. This case illustrates that even a race-blind algorithm can be discriminatory and therefore needs to be accompanied by appropriate fairness interventions that rely on the most suitable operationalization of race.

The second solution focuses on substituting the category of race with a different one, with an aim to focus on racial disparities. For instance, public health discourse gradually departs from studying the effects of race on medical outcomes, focusing on the effect of racism instead [59]. This marks a shift from race-based to post-racial medicine that discards the category of race as useful in biomedical research [45]. Similarly, noting the problem of bias embedded in socially constructed racial categories, Benthall and Haynes propose to use unsupervised machine learning to replace the concept of "race" with "race-like categories", aimed at detecting patterns of racial disadvantage [5].

These solutions appear to mirror the European approach of silence about race, and thus they risk falling under the same paradox of impossibility to effectively address racism by rejecting the category of race. As noted by Malinowska and Żuradzki, "as long as the folk category of race influences the construction of social reality, we should not get rid of the concept of race (understood as a social construct) from medicine" [38]. Race remains indispensable for understanding certain disparities which stem from being a representative of a given minority group, and cannot be fully explained with reference to socio-economic and environmental factors. These include the psychological effects of race that, as mentioned above, can have serious implications on human health and can even be passed through generations [48].

Moreover, race, understood as a social phenomenon, is often the key intersecting factor in discrimination suffered by other protected groups, including women, persons with disabilities, or members of the LGBTQ community. Therefore, the category of race is crucial to understanding the intersectional nature of oppression that these groups suffer. The concept of intersectionality was first brought to the legal discourse by Crenshaw who highlighted the experiences of Black women in the US, describing how distinctive patterns of disadvantage can arise based on multiple identities, such as gender and race [10]. In the 1990s, intersectionality entered the European fundamental rights discourse thanks to feminist advocates who drew attention to discrimination faced by racial and ethnic minority women in the European labor market [60]. This sensitivity to intersectional disadvantage was also reflected in the drafting of the RED. Recital 14 of the Directive acknowledges that women are especially prone to become victims of multiple discrimination. Nowadays, intersectionality is increasingly recognized as a cross-cutting principle in addressing equality concerns, including in the context of health [44].

Groups that already suffer intersectional disadvantage are particularly likely to be affected by algorithmic bias. In particular, ML techniques that are deployed to detect patterns and correlations in data, often result in profiling individuals into distinctive sub-groups. Thus, the output of ML algorithms is rarely based on a single characteristic, but rather on "a combination of characteristics and behavior that is unique to a particular person, or perhaps to a small group of persons" [19]. Because of AI's propensity to encode social injustices, these intersecting characteristics often overlap with groups protected under anti-discrimination law. For instance, the study by Boulamwini and Gebru has shown that leading commercial facial recognition algorithms underperform on Black women due to being trained on non-representative datasets [6]. A similarly biased outcome is likely to occur in the medical domain, as both women and Black patients are under-represented in medical datasets [1], [35]. Thus, when designing fairness interventions for various protected characteristics, the developers of clinical algorithms should also consider how these characteristics interact with race.

However, taking into regard the multi-layered character of racial categories and the distribution of healthcare disparities across ethno-racial lines, the inclusion of race in a clinical algorithm design is always a value choice, which can lead to an anti-racist or racist outcome. Therefore, it should be treated as a crucial sub-problem in the design of clinical algorithms, both those knowledge-based and those that rely on AI. When it comes to ML, algorithmic fairness techniques allow race and ethnic origin to be included in the model in a variety of ways, both at the time of training and prediction. The latter can turn out to be problematic, insofar as using race in the process of determining the outputs could go against the principle of formal equality.

3.2.1. Race correction in clinical algorithms - the road to hell is paved with good intentions?

Ruling out the instances of intentionally inserting racial bias into the algorithm, the motivations behind incorporating race in algorithmic outputs stem from health equity concerns. These, in turn, are predominantly based on two underlying assumptions. The first is that race can be a proxy for social, cultural, and economic determinants of health. Patients coming from minority racial and ethnic groups often face discrimination by healthcare professionals. Moreover, they also experience complex patterns of systemic disadvantage and divergences in access to good quality healthcare and social determinants of health. The second and, as already mentioned, the more problematic assumption is that racial differences in clinical outcomes are due to genetic differences. Both of these assumptions could justify the deployment of a race-aware algorithm in order to correct the inequalities.

In this sense, algorithms that take race or ethnic origin into account at prediction time can be perceived as pursuing "racial projects" [5], defined as "an interpretation, representation, or explanation of racial identities and meanings, and an effort to organize and distribute resources (economic, political, cultural) along particular racial lines" [45]. This section considers a particular type of such a racial project, namely race correction in clinical algorithms.

When it comes to the racialization of the algorithmic output, clinical algorithms which adjust the outcome of the prediction based on the patient's race or ethnicity, with an aim to offer individualized diagnosis, have long been a part of medical practice. These algorithms can vary from machine learning models through complex knowledge-based systems to simple equations. An example of the latter includes eGFR, an equation developed to measure kidney function by estimating the glomerular filtration rate from the level of creatinine. The algorithm is race-adjusted, providing higher scores for Black patients, associated with better kidney function. The rationale for race correction is that Black patients typically have higher levels of creatinine, allegedly due to being more muscular. However, the exact cause of differences in creatinine levels between Black and non-Black patients remains scientifically unexplained. Thus, opponents of race correction suggest that it could lead to delayed diagnosis and specialist referral for Black patients, exacerbating the disparities in kidney disease that they already suffer [12], [52]. Others, however, warn that caution is required in abolishing race adjustment algorithms, arguing that over-diagnosis of Black patients can also lead to detrimental outcomes [8]. Similar examples of controversial race adjustment algorithms exist in different areas of medicine including cardiology, obstetrics, urology, oncology, endocrinology, and pulmonology [55].

The problem with algorithms that adjust their output based on a patient's race is that medical professionals often misinterpret a correlation between race and clinical outcomes for biological causation. In other words, they tend to rely on the assumption that racial differences in clinical outcomes are due to genetic differences. However, a correlation between race and clinical outcomes "is insufficient to translate a data signal into a race adjustment without determining what race might represent in the particular context" [55]. Thus, in order to pursue an anti-racist project, clinical algorithms should exhibit a contextual understanding of race as a social phenomenon. Race correction is only appropriate if its deployment alleviates health disparities and it is based on robust evidence that plausibly explains a correlation between race (correctly operationalized) and a given clinical outcome. Conversely, race-adjusting algorithms which are based on prejudices and oversimplifications deriving from the deeply embedded belief in biological and genetic differences between races contribute to the reinforcement of racism in medicine.

4. Legality of using race and ethnicity in fairness interventions - clinical algorithms under the Racial Equality Directive

4.1. Algorithmic discrimination as direct discrimination?

As already mentioned, the RED enshrines the principle of substantive equal treatment, protecting against both direct and indirect discrimination based on racial and ethnic origin.[8] Direct discrimination takes place "where one person is treated less favorably than another is, has been or would be treated in a comparable situation on grounds of racial or ethnic origin."[9] Indirect discrimination occurs "where an apparently neutral provision, criterion or practice would put persons of a racial or ethnic origin at a particular disadvantage compared with other persons."[10] Discrimination by clinical algorithms falls under the scope of the Directive, as it covers discrimination in public and private sectors in relation to "social protection, including social security and healthcare."[11]

A wide body of literature explores algorithms' legality from the anti-discrimination law perspective [19], [21], [22], [23], [56]. One of the ongoing debates in European legal scholarship is on whether algorithmic discrimination is better captured by the doctrine of direct or indirect discrimination. This distinction is crucial from the point of view of possible justifications for discriminatory treatment. While direct discrimination, in principle, cannot be justified, a "provision, criterion or practice" that leads to indirect discrimination can be "objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary."[12] Therefore, in the latter case, an algorithm deployed with a legitimate aim whose overall predictive accuracy is similar to or higher than the accuracy of doctors could be deemed proportionate, even if it underperforms on certain racial or ethnic groups [21].

The majority of scholars favor the view that a biased algorithm will result in direct discrimination only in exceptional cases, where there is evidence that it aimed to disadvantage a specific group [19], [21], [56]. For instance, Schönberger [53] argues that discrimination by a healthcare algorithm is most likely to constitute indirect discrimination since such an algorithm will generally be considered an "apparently neutral provision, criterion or practice" that puts persons of a racial or ethnic origin at a particular disadvantage. This appears to be the case for race-blind algorithms that can encode harmful biases.

This prevailing view has been challenged by the recent contribution by Adams-Prassl *et al.* who convincingly argue that many forms of algorithmic bias, including discrimination through the use of proxies, fall under the scope of direct discrimination [2]. The authors claim that the propensity to analyze algorithmic discrimination through the lens of indirect discrimination comes from the influence of US legal scholarship. However, opposed to the US doctrine of disparate treatment, direct discrimination in EU law does not require proof of intention. Instead, it focuses solely on finding differential treatment based on a prohibited ground [14].

An analysis of the ECJ's case law on racial discrimination confirms that direct discrimination is broader in its scope than disparate treatment. For instance, in *Feryn*,[13] the Belgian company director who claimed that he will not hire immigrants argued that the statement was motivated by the expectations of his clients who wanted only White Belgians to perform the job. However, the Court considered this justification irrelevant, finding that the statements constitute direct discrimination based on race. It was not the intention or racist attitude of the director that mattered, but rather the effect that his statements had.

Another crucial judgment, *CHEZ*,[14] affirmed that it is not necessary for the practice to disadvantage only persons of racial or ethnic origin to be deemed direct discrimination under the RED. The case concerned discrimination against the Romani population in Bulgaria where the national electricity provider installed electricity meters at a height of six to seven meters in Romani-dominated districts. The alleged purpose was to prevent their inhabitants from tempering with the meters. Interestingly, the case was brought by one of the non-Romani residents of the district affected by the policy. Thus, as the provision appeared to apply equally to Romani and non-Romani residents, one of the questions before the ECJ was whether the case constitutes direct discrimination, or should be more adequately described as indirect discrimination. The Court found that such a practice is capable of constituting direct discrimination as long as it can be shown that "ethnic origin determined the decision to impose the

[8] Art. 2(1) RED.

[9] Art. 2(2)(a) RED.

[10] Art. 2(2)(b) RED.

[11] Art. 3(1)(e) RED.

[12] Art. 2(2)(b) RED.

[13] C-54/07, *Centrum voor gelijkheid van kansen en voor racismebestrijding v. Firma Feryn NV* [2008] ECR I-5187.

[14] Case C-83/14 *CHEZ Razpreldelenie Bulgaria AD v Komisia za Zastita ot Diskriminatsia* [2015] EU:C:2015:480.

treatment."[15] The ECJ listed some factors that the national court should take into account when deciding whether the practice was "introduced for reasons relating to racial or ethnic origin."[16] These include the "compulsory, widespread and lasting nature of the practice", the fact that it was introduced only in Romani-dominated districts,[17] as well as the fact that the electricity provider failed to provide evidence supporting the allegation of meter tempering in Romani-dominated districts, claiming that it was "common knowledge."[18]

4.2. Fairness interventions as direct discrimination?

A crucial consequence of the fact that EU law allows for a finding of direct discrimination in the absence of discriminatory intent is that algorithmic fairness techniques can, in certain circumstances, fall under the scope of direct discrimination. This can happen when fairness measures involve manipulating the output of the algorithm based on a protected characteristic. As evidenced by the preceding sections, in the case of clinical algorithms, some interventions, even when adopted with fairness in mind, can in fact exacerbate the inequalities and violate the principle of equal treatment. Thus, this and the proceeding sections draw attention to the largely understudied topic of the legality of the fairness measures themselves, exploring how techniques that make explicit use of race are accommodated under the RED.

As a preliminary caveat, it should be noted that the majority of fairness interventions are not likely to raise legal concerns. For instance, in the case of ML, data redistribution techniques aimed to remedy imbalances in clinical datasets, or data purification techniques aimed to detect and remove human biases encoded into data should not be deemed to constitute discrimination [16]. Although these interventions require race consciousness, they do not involve making decisions about an individual based on race. Conversely, fairness methods that take account of the patient's race or ethnicity at the prediction stage are more prone to legal challenges. However, as argued by Kim, it is too simplistic to draw a liability line "between race-awareness at training time versus prediction" [29]. Following *CHEZ*, whether a fairness intervention that considers race at the time of prediction constitutes a *prima facie* direct discrimination is likely to depend on whether race is a directly determinative factor in the decision. For instance, a complex ML algorithm might use interactions between race and other features to predict the outcome taking into account factors relevant to different groups [29]. This case is unlikely to amount to discrimination because the factor of "race" alone does not alter the decision. On the other hand, an algorithm that shifts the outcome of prediction based solely on race is likely to violate the principle of equal treatment.

Based on these remarks, let us now consider the legality of race correction algorithms under the RED. Race correction adjusts the outcome of prediction based solely on race or ethnic origin, leading to different recommendations for otherwise similarly situated patients. Thus, as stated in *CHEZ*, since racial or ethnic origin determines the decision to impose a given treatment, direct discrimination takes place. In other words, the practice of race adjustment in clinical algorithms can lead to one person being treated less favorably than another in a comparable situation on grounds of racial or ethnic origin.[19] Representatives of a racial or ethnic group that are likely to suffer disadvantage because of the race correction, including the increased probability of misdiagnosis or delayed treatment, can bring a direct discrimination claim under the RED. Pursuant to Art. 8(1) of the Directive, if the claimant is able to establish facts from which it can be presumed that discrimination took place, it is for the respondent to prove that the principle of equal treatment has not been breached. As reiterated in *Feryn*, the Court cannot take into account the intention of algorithm developers. It is thus immaterial if they are racist or truly believe that race correction is in the best interest of the racial and ethnic minority patients. On the other hand, what the Court could take into account, following *CHEZ*, is the problematic history of the race correction practice and the presence or absence of scientific evidence supporting it. It should also

[15] *CHEZ*, para 76.
[16] *CHEZ*, para 95.
[17] CHEZ, para 84.
[18] CHEZ, para 83, 117.
[19] Art. 2(2)(a) RED.

be underlined that this evidence must go beyond common knowledge or established practice which can be, in fact, a result of historic bias.

4.3. Race correction as a positive action?

It can be argued that in certain cases race adjustment in clinical algorithms can serve equity purposes, for instance by prioritizing racial or ethnic minority patients for certain screening programs in which they have been underrepresented. Yet, such practice would still entail a breach of the principle of formal equality and thus could attract direct discrimination claims by White patients. However, as an instrument aimed at eliminating inequalities, the RED aims to achieve not only formal but also substantive equality, including the equality of outcomes. Thus, under the Directive, differential treatment based on race can be lawful, if the measure falls under the scope of "positive action." Art. 5 of the RED provides: "With a view to ensuring full equality in practice, the principle of equal treatment shall not prevent any Member State from maintaining or adopting specific measures to prevent or compensate for disadvantages linked to racial or ethnic origin."

The topic of positive action under the RED has not been yet explored in caselaw of the ECJ, which focuses predominantly on gender discrimination in the area of employment. This leaves room for uncertainty, as it is not evident if the same legal standards would be applied to positive action based on race or ethnicity in the area of healthcare [57]. However, assuming that textual similarities will lead to similar interpretations of the positive action rules between the equality directives, developers purporting to justify race correction in clinical algorithms under Art. 5 RED will likely face a high evidentiary burden.

Firstly, the race adjustment must be designed to remove a genuine disadvantage stemming from inadequate access to healthcare or race-based stereotypes. Notably, in order to satisfy the requirements of positive action, a race correction has to target "a disadvantage linked to racial or ethnic origin," as opposed to the ground of racial or ethnic origin itself. Thus, developers cannot claim lawfulness of the race correction based on a vague assertion that it benefits the racial minority patient. Rather, they are required to identify a concrete disadvantage linked to the minority status that they are planning to remedy. For instance, if developers of an algorithm that suggest a dose of painkiller want to automatically adjust the score for race by raising it in the case of Black patients, they need to prove that under-prescription is linked to the race of patients.

This position is confirmed by the gender *acquis* on positive action. For instance, in *Commission v France*,[20] the Court favored a strict link between the positive action in question and a concrete disadvantage. It held that France had not sufficiently shown that generic provision of special rights for women, such as obtaining leave for a sick child, would, in fact, reduce concrete inequalities in social life. Conversely, in *Loomers*,[21] a sufficient link between the positive action and the concrete disadvantage was found in the case of a scheme that prioritized access of women employees to subsidized nursery places in light of the proven insufficiency of affordable daycare facilities.

The correct operationalization of race is the prerequisite for establishing the link between race correction in clinical algorithms and a specific disadvantage. Thus, the developers must start with choosing the dimension of race that is the most appropriate for studying the disparity in question and ensure that the racial or ethnic data present in their dataset is correctly and consistently measured for the dimension chosen. Moreover, the developers must prove that race correction has an effect of "preventing or compensating" for the disadvantage suffered. For instance, if the correlation between a disadvantage and race is questionable, they shall consider if the race correction will not in fact exacerbate the disparities.

Secondly, the positive action in the form of race correction must satisfy the test of proportionality. In *Kalanke*,[22] a case concerning positive action based on gender under the old Equal Treatment Directive,[23] the ECJ held that positive action has to be construed strictly, as a derogation from the right

[20] Case 312/86 *Commission v France* [1988] ECR 6315.

[21] Case C-476/99, *H. Lommers v. Minister van Landbouw, Natuurbeheer en Visserij* [2002] ECR I-2891.

[22] Case C-450/93 *Kalanke v Freie Hansestadt Bremen* [1995] ECR I-3051.

[23] Council Directive 76/207/EEC of 9 February 1976 on the implementation of the principle of equal treatment for men and women as regards access to employment, vocational training and promotion, and working conditions.

to equal treatment. As stated by the Court in *Loomers*, such a derogation "must remain within the limits of what is appropriate and necessary in order to achieve the aim in view."[24] Thus, the developer must show that race correction is a suitable way to address the disadvantage in question and that no less restrictive alternative was available to fulfill the same purpose. In light of the variety of algorithmic fairness techniques which might involve race in a more subtle manner, the latter requirement might turn out particularly difficult to fulfill. Of course, what is proportionate will depend on the scale and nature of the disadvantage in question. The bigger the disadvantage, the more proportional race correction might be as a response.

Interestingly, the ECJ seems to apply the strictest scrutiny in cases where the positive action is likely to disadvantage the more privileged group. For instance, in *Abrahamsson*, the Court held that a Swedish scheme that prioritized sufficiently qualified candidates of the under-represented gender to be appointed for a public post in preference to a better-qualified candidate of the opposite gender was "disproportionate to the aim pursued."[25] It is probable that the Court could adopt an even stricter threshold when the potential adverse effects consider health, as opposed to employment opportunities.

Moreover, Waddington and Bell argue that, although it is not directly specified under EU law, the nature of positive action should generally be temporary [57]. The aim of positive action is to achieve substantive equality of outcomes, which presupposes that when such equality is reached, the measure should cease. Thus, the developers of clinical algorithms should consider that even if race correction can be successfully justified as a positive action at a given moment, its legality might change in the future.

Summing up, the developers of clinical algorithms which use race adjustment in their outputs are likely to face a very high evidentiary burden to successfully defend their design choice as a positive action under Art. 5 of the RED. Moreover, proving the presence of a significant disadvantage and the proportionality of race adjustment requires robust statistical data on racial and ethnic inequalities, which, as discussed above, is scarcely available in the European context.

5. Conclusions and recommendations

The use of race and ethnicity in medical research and practice remains controversial because the very concept of race was created to underline biological differences between peoples and justify the subjugation of those considered biologically "inferior" to the biologically "superior" White Europeans. These pseudo-scientific arguments influenced the development of medicine, leading to violent practices, the racialization of diseases, and harmful assumptions about non-White bodies. The legacy of racism in healthcare has led to the development of stereotypes that are still deeply embedded in modern medical practice and often scientifically legitimized, for instance through genetic essentialism. Thus, modern race-aware medicine, although deployed with equity in mind, can in fact perpetuate racial stereotypes and deepen inequalities.

In light of these considerations, the use of race and ethnicity as variables in clinical algorithms deserves heightened attention. While the rise of ML algorithms in clinical practice could offer significant benefits to patients, these emerging technologies can share the pitfalls of their knowledge-based predecessors, including the entrenchment of racial bias. Thus, this contribution has outlined two crucial foundational issues regarding the inclusion of race and ethnicity in clinical algorithms - the operationalization of race and the use of race in the design of the algorithm.

Based on the consensus regarding the fact that race is not a natural attribute, but a complex social construct, this contribution has argued that its correct operationalization is a crucial, yet often overlooked, methodological step in the design of clinical algorithms. Different dimensions of race require different measurement methods and are appropriate for capturing different kinds of discriminatory treatment. Thus, wrong or inconsistent operationalization of race for a given clinical problem will be at best ineffective at pursuing equity concerns and at worst harmful to racial and ethnic minorities.

The contribution has submitted that in spite of the controversies surrounding the concept of race, racial and ethnic classifications can still be clinically useful and should be included in clinical

[24] *Loomers*, para 39.

[25] Case C-407/98 *Abrahamsson v Fogelqvist* [2000] ECR I-5539, para 55.

algorithms. Erasing the category of race is not a viable option, as the "fairness through unawareness" technique does not prevent algorithms from discriminating based on racial proxies. Moreover, replacing race with a neutral category, such as socio-economic background, does not always capture the nature of racial disadvantages, for instance in the case of generational trauma or intersectional disadvantage.

However, it has been acknowledged that algorithms that take race into account at prediction time should be subject to strict scrutiny, as they risk treating people differently based on their race, and thus can constitute a *prima facie* case of direct discrimination. In particular, focusing on race correction in clinical algorithms, the paper has argued that the decision to adjust an outcome of the prediction based on race often stems from the problematic assumption that race is a biological or genetic factor. Instead, justifications for race correction should be built around race as a social category, capturing the essence of racial disadvantages.

Analyzing the problem of race and ethnicity in clinical algorithms from the European perspective unearths new challenges. The European paradox of fighting racism without acknowledging the existence of race fails to capture the social dimension of race, leading to definitional conundrums in anti-discrimination law. At the same time, the EMA's interest in ethical differences in medicines effect can potentially lead to the popularization of race-aware medicine. This is highly problematic given the lack of European regulatory guidelines on using the categories of race and ethnicity in clinical contexts and the lack of systemic collection of data on racial and ethnic inequalities in healthcare. Both of these issues make the correct operationalization of race in clinical algorithms more difficult.

Therefore, the EU and Member States should cooperate in order to establish coherent standards for collecting disaggregated data on racial and ethnic minorities, with the purpose of studying discrimination patterns in Europe, in particular in the area of healthcare. This is especially important in the advent of the European Health Data Space Regulation (EHDS) proposal,[26] which aims at establishing standards for sharing of data for secondary purposes, including research and development of AI technologies. While the EHDS presents an opportunity to increase the availability of data concerning underrepresented racial minorities, it needs to be accompanied by strategies aimed at the trustworthy collection of these data.

Under the RED, algorithmic fairness interventions that use race at the prediction time could constitute discrimination if race plays a leading role in the decision-making process. Thus, it has been argued that in the case of race correction algorithms that change the outcome of the decision based solely on race, a *prima facie* direct discrimination will likely be established. The developers of clinical algorithms that use race correction could attempt to justify the measure under the positive action doctrine. This would require them to identify a concrete disadvantage linked to racial and ethnic origin and show that race correction could plausibly remedy it.

However, the ECJ's caselaw suggests that positive action is interpreted very strictly, as a derogation from the principle of equal treatment. Thus, in light of the development of more sophisticated methods of incorporating race in fairness interventions through ML algorithms, race correction techniques are likely to fail the proportionality test, even if they pursue a scientifically justified substantive equality aim. This outcome is reminiscent of the wider criticism directed at the RED, namely, that is it designed to protect mainly formal equality [20]. As noted by Howard, the voluntary character and unclear scope of the positive action mechanism deter public and private entities from adopting it in order to address and rectify historic injustices [25]. This entrenches the view that: (1) racist behaviors are predominantly isolated incidents, as opposed to institutionalized practices, and (2) that racism has mainly biological, as opposed to social and cultural dimensions. Therefore, in its future case law, the ECJ should consider adopting a more flexible approach towards positive action, giving effect to the substantive equality aspirations of the RED. This would allow for accommodating a broader spectrum of fairness measures aimed at achieving racial equity in health.

6. References

[26] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space COM(2022) 197 final.

[1] Adamson, Adewole S., and Avery Smith. ‘Machine Learning and Health Care Disparities in Dermatology’. JAMA Dermatology 154, no. 11 (1 November 2018): 1247. https://doi.org/10.1001/jamadermatol.2018.2348.

[2] Adams-Prassl, Jeremias, Reuben Binns, and Aislinn Kelly-Lyth. ‘Directly Discriminatory Algorithms’. *The Modern Law Review* 86, no. 1 (2023): 144–75. https://doi.org/10.1111/1468-2230.12759.

[3] Bell, Mark. ‘“Race”, Ethnicity, and Racism in Europe’. In Racism and Equality in the European Union, Mark Bell (Ed.). Oxford University Press, 2009. https://doi.org/10.1093/acprof:oso/9780199297849.003.0002.

[4] Bell, Mark. ‘Social Inclusion: Education, Health, and Housing’. In Racism and Equality in the European Union, Mark Bell (Ed.). Oxford University Press, 2009. https://doi.org/10.1093/acprof:oso/9780199297849.003.0007.

[5] Benthall, Sebastian, and Bruce D. Haynes. ‘Racial Categories in Machine Learning’. In Proceedings of the Conference on Fairness, Accountability, and Transparency, 289–98. FAT* ’19. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3287560.3287575.

[6] Boulamwini, Joy and Timnit Gebru. ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’. Proceedings of Machine Learning Research 81 (2018): 1–15.

[7] Cerdeña, Jessica P., Emmanuella Ngozi Asabor, Marie V. Plaisime, and Rachel R. Hardeman. ‘Race-Based Medicine in the Point-of-Care Clinical Resource UpToDate: A Systematic Content Analysis’. EClinicalMedicine 52 (1 October 2022). https://doi.org/10.1016/j.eclinm.2022.101581.

[8] Choy, Kay Weng. ‘From Race-Based to Race-Conscious Medicine: Proceed with Care’. Internal Medicine Journal 51, no. 2 (2021): 309–309. https://doi.org/10.1111/imj.15176.

[9] Commissioner’s Office. ‘Collection of Race and Ethnicity Data in Clinical Trials’. U.S. Food and Drug Administration. FDA, 11 August 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/collection-race-and-ethnicity-data-clinical-trials.

[10] Crenshaw, Kimberlé. ‘Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics’. *University of Chicago Legal Forum* 1989 (1 January 1989): 139.

[11] ECRI. ‘Deep Dive: Racial and Ethnic Disparities in Health and Healthcare Executive Brief’. Accessed 29 January 2023. https://d84vr99712pyz.cloudfront.net/p/images1/ecri-trusted-voice-healthcare.jpg.

[12] Eneanya, Nwamaka Denise, Wei Yang, and Peter Philip Reese. ‘Reconsidering the Consequences of Using Race to Estimate Kidney Function’. JAMA 322, no. 2 (9 July 2019): 113–14. https://doi.org/10.1001/jama.2019.5774.

[13] European Commission. Directorate General for Justice and Consumers. and European network of legal experts in gender equality and non discrimination. The meaning of racial or ethnic origin in EU law: between stereotypes and identities. LU: Publications Office, 2016. https://data.europa.eu/doi/10.2838/83148.

[14] European Union Agency for Fundamental Rights. ‘Handbook on European Non-Discrimination Law – 2018 Edition’, 2 March 2018. http://fra.europa.eu/en/publication/2018/handbook-european-non-discrimination-law-2018-edition.

[15] European Union Agency for Fundamental Rights. Inequalities and Multiple Discrimination in Access to and Quality of Healthcare. LU: Publications Office, 2013. https://data.europa.eu/doi/10.2811/17523.

[16] Feng, Qizhang, Mengnan Du, Na Zou, and Xia Hu. ‘Fair Machine Learning in Healthcare: A Review’. arXiv, 16 August 2022. https://doi.org/10.48550/arXiv.2206.14397.

[17] Flanagin, Annette, Tracy Frey, Stacy L. Christiansen, and AMA Manual of Style Committee. ‘Updated Guidance on the Reporting of Race and Ethnicity in Medical and Science Journals’. JAMA 326, no. 7 (17 August 2021): 621–27. https://doi.org/10.1001/jama.2021.13304.

[18] Fosch Villaronga, Eduard, Hadassah Drukarch, Pranav Khanna, Tessa Verhoef, and Bart Custers. ‘Accounting for Diversity in AI for Medicine’. *Computer Law & Security Review* 47 (1 November 2022): 105735. https://doi.org/10.1016/j.clsr.2022.105735.

[19] Gerards, Janneke, and Raphaële Xenidis. Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law : A Special Report, 2021.

[20] Goodwin, Morag. Romani Marginalisation after the Race Equality Directive In EU Anti-Discrimination Law beyond Gender, Uladzislau Belavusau and Kristin Henrard (Eds.). Hart, 2018.

[21] Hacker, Philipp. ‘Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law’. *Common Market Law Review* 55, no. 4 (1 August 2018). https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2018095.pdf.

[22] Hoffmann, Anna Lauren. ‘Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse’. Information, Communication & Society 22, no. 7 (7 June 2019): 900–915. https://doi.org/10.1080/1369118X.2019.1573912.

[23] Hoffman Sharona, and Andy Podgurski. ‘Artificial Intelligence and Discrimination in Health Care’. Yale Journal of Health Policy, Law, and Ethics 19, no. 3 (2020): 1–49.

[24] Hong, Chuan, Michael J. Pencina, Daniel M. Wojdyla, Jennifer L. Hall, Suzanne E. Judd, Michael Cary, Matthew M. Engelhard, et al. ‘Predictive Accuracy of Stroke Risk Prediction Models Across Black and White Race, Sex, and Age Groups’. *JAMA* 329, no. 4 (24 January 2023): 306–17. https://doi.org/10.1001/jama.2022.24683.

[25] Howard, Erica. ‘The EU Race Directive: Time for Change?’ *International Journal of Discrimination and the Law* 8, no. 4 (1 March 2007): 237–61. https://doi.org/10.1177/135822910700800403.

[26] Johnson, Kirk A. Medical Stigmata: Race, Medicine, and the Pursuit of Theological Liberation. Singapore: Palgrave Macmillan, 2019.

[27] Julian, B. A., R. S. Gaston, W. M. Brown, A. M. Reeves-Daniel, A. K. Israni, D. P. Schladt, S. O. Pastan, S. Mohan, B. I. Freedman, and J. Divers. ‘Effect of Replacing Race With Apolipoprotein L1 Genotype in Calculation of Kidney Donor Risk Index’. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 17, no. 6 (June 2017): 1540–48. https://doi.org/10.1111/ajt.14113.

[28] Kahn, Jonathan. Race in a Bottle: The Story of BiDil and Racialized Medicine in a Post-Genomic Age. Columbia University Press, 2012.

[29] Kim, Pauline T.; ‘Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action’, 2022. https://doi.org/10.15779/Z387P8TF1W.

[30] Kirkbride, James B., Yasir Hameed, Konstantinos Ioannidis, Gayatri Ankireddypalli, Carolyn M. Crane, Mukhtar Nasir, Nikolett Kabacs, et al. ‘Ethnic Minority Status, Age-at-Immigration and Psychosis Risk in Rural Environments: Evidence From the SEPEA Study’. Schizophrenia Bulletin 43, no. 6 (21 October 2017): 1251–61. https://doi.org/10.1093/schbul/sbx010.

[31] Kolossváry, Márton, Vineet K. Raghu, John T. Nagurney, Udo Hoffmann, and Michael T. Lu. ‘Deep Learning Analysis of Chest Radiographs to Triage Patients with Acute Chest Pain Syndrome’. Radiology 306, no. 2 (February 2023): e221926. https://doi.org/10.1148/radiol.221926.

[32] Kuo, Wen-Hua. ‘Understanding Race at the Frontier of Pharmaceutical Regulation: An Analysis of the Racial Difference Debate at the ICH’. The Journal of Law, Medicine & Ethics: A Journal of the American Society of Law, Medicine & Ethics 36, no. 3 (2008): 498–505. https://doi.org/10.1111/j.1748-720X.2008.297.x.

[33] Ladbury, Colton, Arya Amini, Ameish Govindarajan, Isa Mambetsariev, Dan J. Raz, Erminia Massarelli, Terence Williams, Andrei Rodin, and Ravi Salgia. ‘Integration of Artificial Intelligence in Lung Cancer: Rise of the Machine’. Cell Reports Medicine 0, no. 0 (3 February 2023). https://doi.org/10.1016/j.xcrm.2023.100933.

[34] Lebret, Audrey. ‘Allocating Organs through Algorithms and Equitable Access to Transplantation—a European Human Rights Law Approach’. *Journal of Law and the Biosciences* 10, no. 1 (1 January 2023): lsad004. https://doi.org/10.1093/jlb/lsad004.

[35] Lee, Michelle S., Lisa N. Guo, and Vinod E. Nambudiri. ‘Towards Gender Equity in Artificial Intelligence and Machine Learning Applications in Dermatology’. *Journal of the American Medical Informatics Association* 29, no. 2 (12 January 2022): 400–403. https://doi.org/10.1093/jamia/ocab113.

[36] Lee, Seung Mi, Garam Lee, Tae Kyong Kim, Trang Le, Jie Hao, Young Mi Jung, Chan-Wook Park, et al. ‘Development and Validation of a Prediction Model for Need for Massive Transfusion

During Surgery Using Intraoperative Hemodynamic Monitoring Data'. JAMA Network Open 5, no. 12 (14 December 2022): e2246637. https://doi.org/10.1001/jamanetworkopen.2022.46637.

[37] Malinowska, Joanna K., and Tomasz Żuradzki. 'Reductionist Methodology and the Ambiguity of the Categories of Race and Ethnicity in Biomedical Research: An Exploratory Study of Recent Evidence'. Medicine, Health Care and Philosophy, 9 November 2022. https://doi.org/10.1007/s11019-022-10122-y.

[38] Malinowska, Joanna K., and Tomasz Żuradzki. 'The Practical Implications of the New Metaphysics of Race for a Postracial Medicine: Biomedical Research Methodology, Institutional Requirements, Patient-Physician Relations'. American Journal of Bioethics 17, no. 9 (September 2017): 61–63. https://doi.org/10.1080/15265161.2017.1353181.

[39] Malinowska, Joanna K., and Tomasz Żuradzki. 'Towards the Multileveled and Processual Conceptualisation of Racialised Individuals in Biomedical Research'. Synthese 201, no. 1 (28 December 2022): 11. https://doi.org/10.1007/s11229-022-04004-2.

[40] Meghani, Salimah H., Eeeseung Byun, and Rollin M. Gallagher. 'Time to Take Stock: A Meta-Analysis and Systematic Review of Analgesic Treatment Disparities for Pain in the United States'. Pain Medicine 13, no. 2 (1 February 2012): 150–74. https://doi.org/10.1111/j.1526-4637.2011.01310.x.

[41] Möschel, Mathias. Law, Lawyers and Race: Critical Race Theory from the United States to Europe. Routledge, 2014.

[42] Mulinari, Shai, Andreas Vilhelmsson, Piotr Ozieranski, and Anna Bredström. 'Is There Evidence for the Racialization of Pharmaceutical Regulation? Systematic Comparison of New Drugs Approved over Five Years in the USA and the EU'. Social Science & Medicine 280 (1 July 2021): 114049. https://doi.org/10.1016/j.socscimed.2021.114049.

[43] Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations'. Science (New York, N.Y.) 366, no. 6464 (25 October 2019). https://doi.org/10.1126/science.aax2342.

[44] OHCHR. 'OHCHR | A/77/197: Report by the Special Rapporteur on the Right of Everyone to the Enjoyment of the Highest Attainable Standard of Physical and Mental Health - Racism and the Right to Health'. Accessed 18 November 2022. https://www.ohchr.org/en/documents/thematic-reports/a77197-report-special-rapporteur-right-everyone-enjoyment-highest.

[45] Omi, Michael, and Howard Winant. Racial formation in the United States. Routledge, 2014.

[46] Perez-Rodriguez, Javier, and Alejandro de la Fuente. 'Now Is the Time for a Postracial Medicine: Biomedical Research, the National Institutes of Health, and the Perpetuation of Scientific Racism'. The American Journal of Bioethics: AJOB 17, no. 9 (September 2017): 36–47. https://doi.org/10.1080/15265161.2017.1353165.

[47] Raghu, Vineet K., Anika S. Walia, Aniket N. Zinzuwadia, Reece J. Goiffon, Jo-Anne O. Shepard, Hugo J. W. L. Aerts, Inga T. Lennes, and Michael T. Lu. 'Validation of a Deep Learning–Based Model to Predict Lung Cancer Risk Using Chest Radiographs and Electronic Medical Record Data'. JAMA Network Open 5, no. 12 (28 December 2022): e2248793. https://doi.org/10.1001/jamanetworkopen.2022.48793.

[48] Ramirez-Goicoechea, Eugenia. 'Life-in-the-Making: Epigenesis, Biocultural Environments and Human Becomings'. In Biosocial Becomings: Integrating Social and Biological Anthropology, edited by Gisli Palsson and Tim Ingold, 59–83. Cambridge: Cambridge University Press, 2013. https://doi.org/10.1017/CBO9781139198394.005.

[49] Roth, Wendy. 'Methodological Pitfalls of Measuring Race: International Comparisons and Repurposing of Statistical Categories'. Ethnic and Racial Studies 40, no. 13 (21 October 2017): 2347–53. https://doi.org/10.1080/01419870.2017.1344276.

[50] Roth, Wendy. 'The Multiple Dimensions of Race'. Ethnic and Racial Studies 39, no. 8 (20 June 2016): 1310–38. https://doi.org/10.1080/01419870.2016.1140793.

[51] Samorani, Michele, and Linda Goler Blount. 'Machine Learning and Medical Appointment Scheduling: Creating and Perpetuating Inequalities in Access to Health Care'. American Journal of Public Health 110, no. 4 (April 2020): 440–41. https://doi.org/10.2105/AJPH.2020.305570.

[52] Schmidt, Insa M., and Sushrut S. Waikar. 'Separate and Unequal: Race-Based Algorithms and Implications for Nephrology'. *Journal of the American Society of Nephrology* 32, no. 3 (March 2021): 529. https://doi.org/10.1681/ASN.2020081175.

[53] Schönberger, Daniel. 'Artificial intelligence in healthcare: A critical analysis of the legal and ethical implications,' International Journal of Law and Information Technology 27/2 (2019).

[54] Shachar, Carmel, and Sara Gerke. 'Prevention of Bias and Discrimination in Clinical Practice Algorithms'. JAMA 329, no. 4 (24 January 2023): 283–84. https://doi.org/10.1001/jama.2022.23867.

[55] Vyas, Darshali A., Leo G. Eisenstein, and David S. Jones. 'Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms'. New England Journal of Medicine 383, no. 9 (27 August 2020): 874–82. https://doi.org/10.1056/NEJMms2004740.

[56] Wachter, Sandra, Brent Mittelstadt, and Chris Russell. 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI'. SSRN Electronic Journal, 2020. https://doi.org/10.2139/ssrn.3547922.

[57] Waddington, Lisa, and Mark Bell. 'Exploring the Boundaries of Positive Action under EU Law: A Search for Conceptual Clarity'. Common Market Law Review 48, no. 5 (1 October 2011). https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2011059.pdf.

[58] White, Kellee, Jourdyn A. Lawrence, Nedelina Tchangalova, Shuo J. Huang, and Jason L. Cummings. 'Socially-Assigned Race and Health: A Scoping Review with Global Implications for Population Health Equity'. *International Journal for Equity in Health* 19, no. 1 (10 February 2020): 25. https://doi.org/10.1186/s12939-020-1137-5.

[59] Williams, David R., and Michelle Sternthal. 'Understanding Racial-Ethnic Disparities in Health: Sociological Contributions'. Journal of Health and Social Behavior 51 Suppl, no. Suppl (2010): S15-27. https://doi.org/10.1177/0022146510383838.

[60] Xenidis, Raphaële. Multiple Discrimination in EU Anti-Discrimination Law: Towards redressing complex Inequality? In EU Anti-Discrimination Law beyond Gender, Uladzislau Belavusau and Kristin Henrard (Eds.). Hart, 2018.