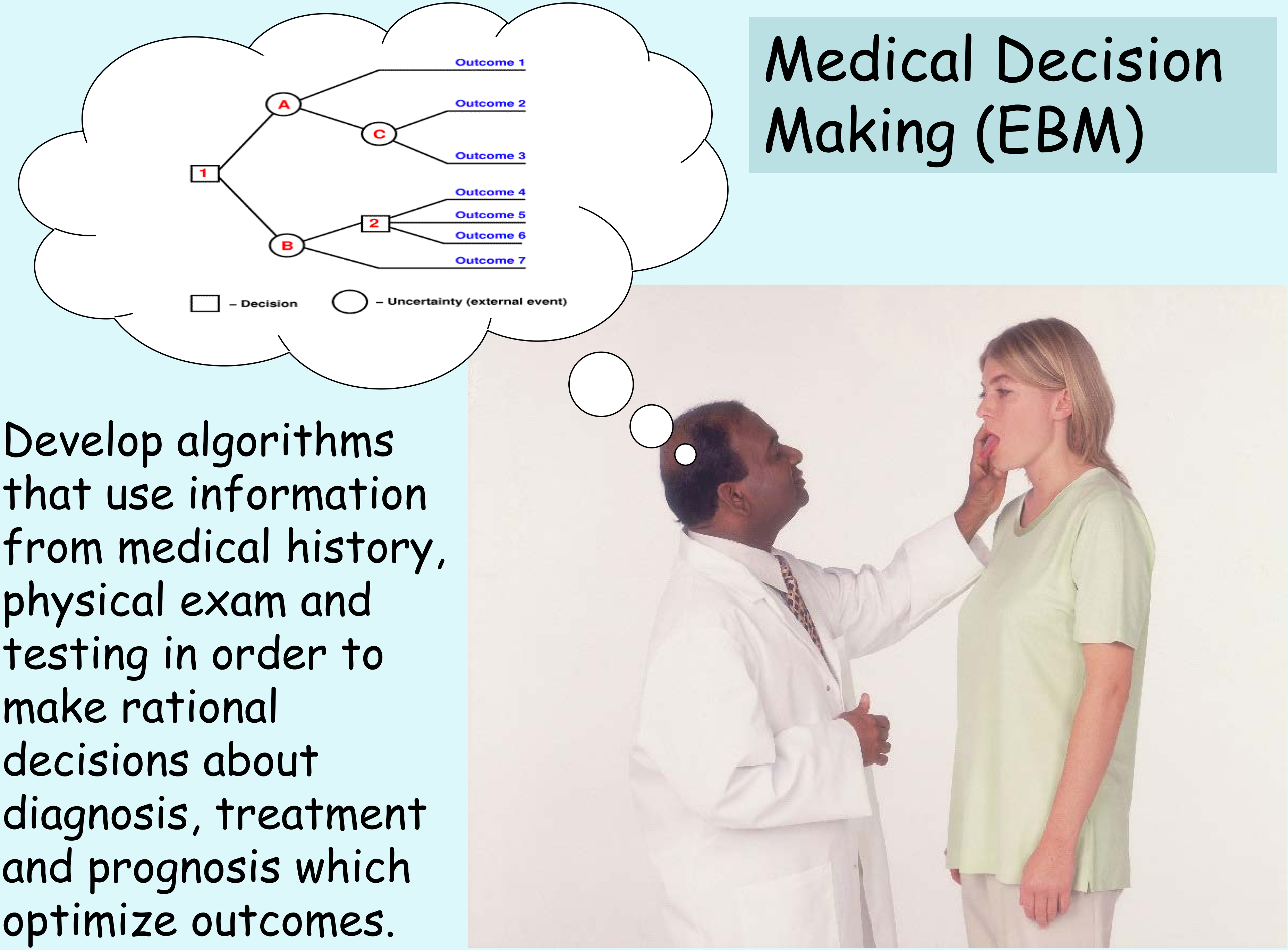
Discrimination in Algorithmic Evaluations and Treatments

Jay S. Kaufman, PhD
Department of Epidemiology, Biostatistics, and
Occupational Health
McGill University
1020 Pine Ave West
Montreal, Quebec H3A 1A2

Friday 7 June 2019, 9:00-10:20 AM
Wrong at the Root
Simons Institute and Sloan Foundation
University of California at Berkeley

Medical Decision Making (EBM)



Develop algorithms that use information from medical history, physical exam and testing in order to make rational decisions about diagnosis, treatment and prognosis which optimize outcomes.

Fairness (absence of discrimination)

Weak (direct):

Algorithm does not rely explicitly on protected characteristics or classes

Strong (indirect):

Algorithm produces decisions that yield equally advantageous results for all strata of protected characteristics or classes

Council NR. 2004. Measuring Racial Discrimination. Washington, DC: NAP
Barocas S, Selbst AD. 2016. Big data's disparate impact. Calif Law Rev. 104.

Often these two definitions are in conflict

In which case, considered ethical to use protected characteristics or classes in diagnostic or treatment algorithms in pursuit of more equal outcomes.

For example, 1) Screening Algorithms:

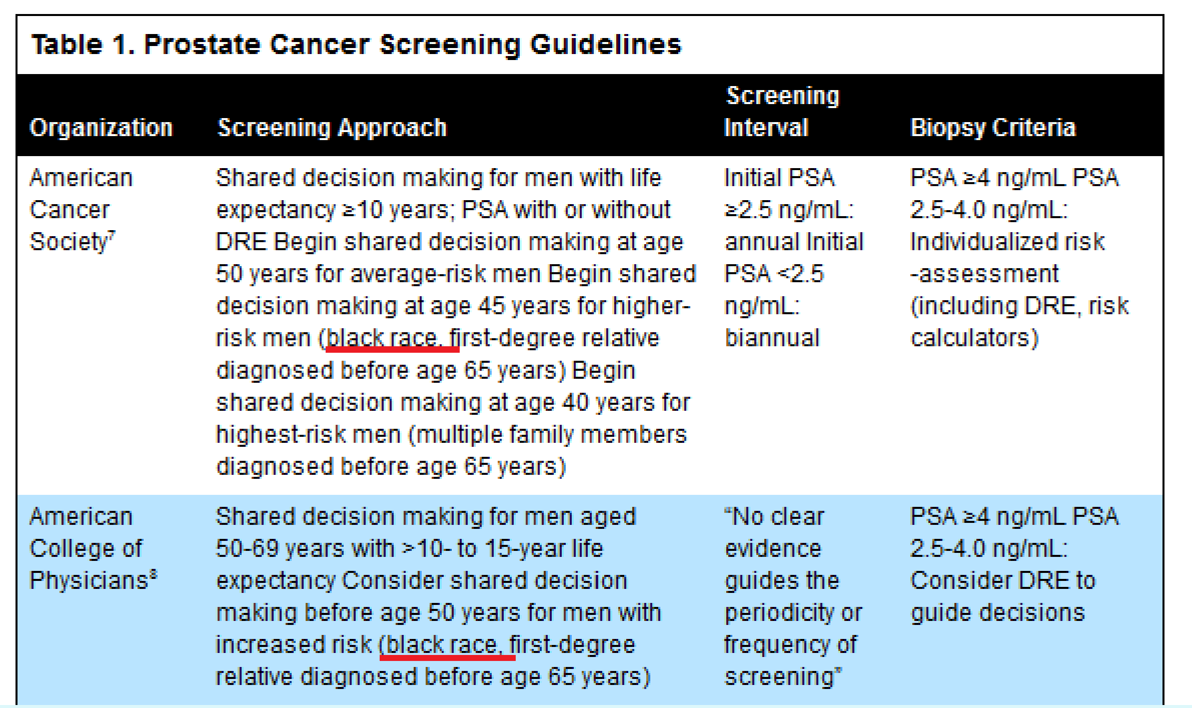
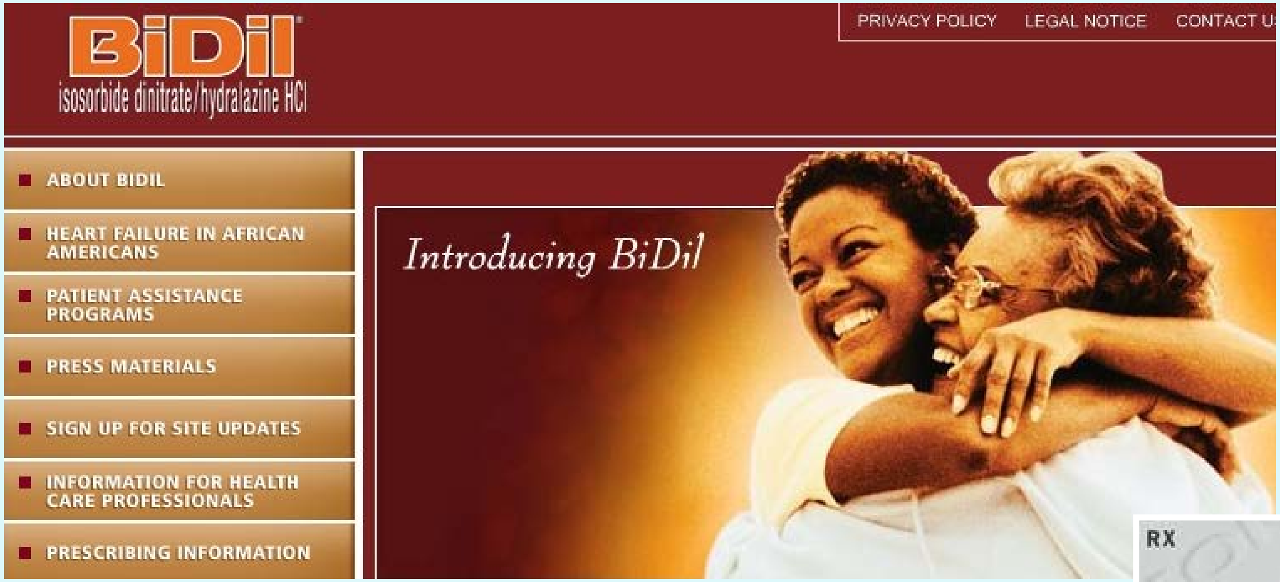


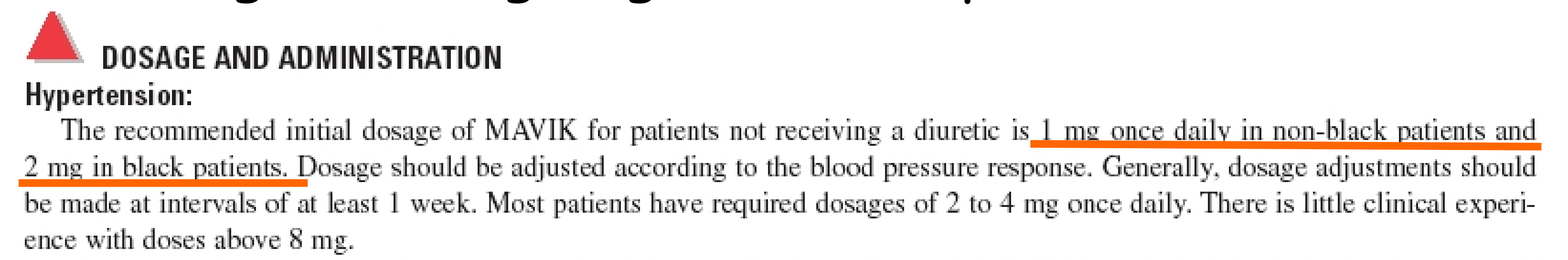
Table 1. Prostate Cancer Screening Guidelines

Organization	Screening Approach	Screening Interval	Biopsy Criteria
American Cancer Society[7]	Shared decision making for men with life expectancy ≥10 years; PSA with or without DRE Begin shared decision making at age 50 years for average-risk men Begin shared decision making at age 45 years for higher-risk men (black race, first-degree relative diagnosed before age 65 years) Begin shared decision making at age 40 years for highest-risk men (multiple family members diagnosed before age 65 years)	Initial PSA ≥2.5 ng/mL: annual Initial PSA <2.5 ng/mL: biannual	PSA ≥4 ng/mL PSA 2.5-4.0 ng/mL: Individualized risk -assessment (including DRE, risk calculators)
American College of Physicians[8]	Shared decision making for men aged 50-69 years with >10- to 15-year life expectancy Consider shared decision making before age 50 years for men with increased risk (black race, first-degree relative diagnosed before age 65 years)	"No clear evidence guides the periodicity or frequency of screening"	PSA ≥4 ng/mL PSA 2.5-4.0 ng/mL: Consider DRE to guide decisions

2) Choice of Drug (e.g., BiDil, ACE-I)


BiDil
isosorbide dinitrate/hydralazine HCl
PRIVACY POLICY
LEGAL NOTICE
CONTACT U
ABOUT BIDIL
HEART FAILURE IN AFRICAN AMERICANS
PATIENT ASSISTANCE PROGRAMS
PRESS MATERIALS
SIGN UP FOR SITE UPDATES
INFORMATION FOR HEALTH CARE PROFESSIONALS
PRESCRIBING INFORMATION
Introducing BiDil

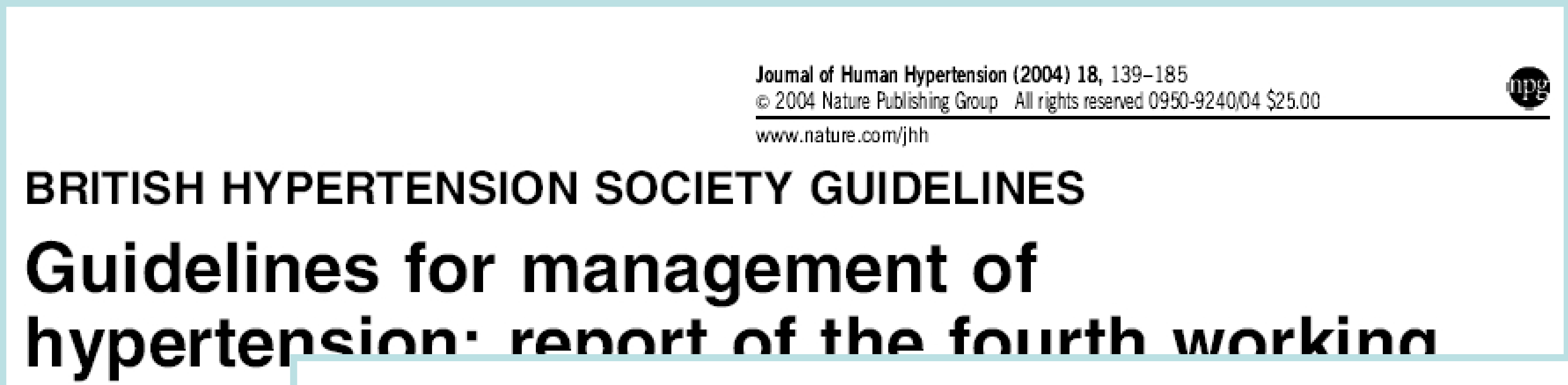
3) Dosage of Drug (e.g., Trandolapril)



DOSAGE AND ADMINISTRATION

Hypertension:

The recommended initial dosage of MAVIK for patients not receiving a diuretic is 1 mg once daily in non-black patients and 2 mg in black patients. Dosage should be adjusted according to the blood pressure response. Generally, dosage adjustments should be made at intervals of at least 1 week. Most patients have required dosages of 2 to 4 mg once daily. There is little clinical experience with doses above 8 mg.



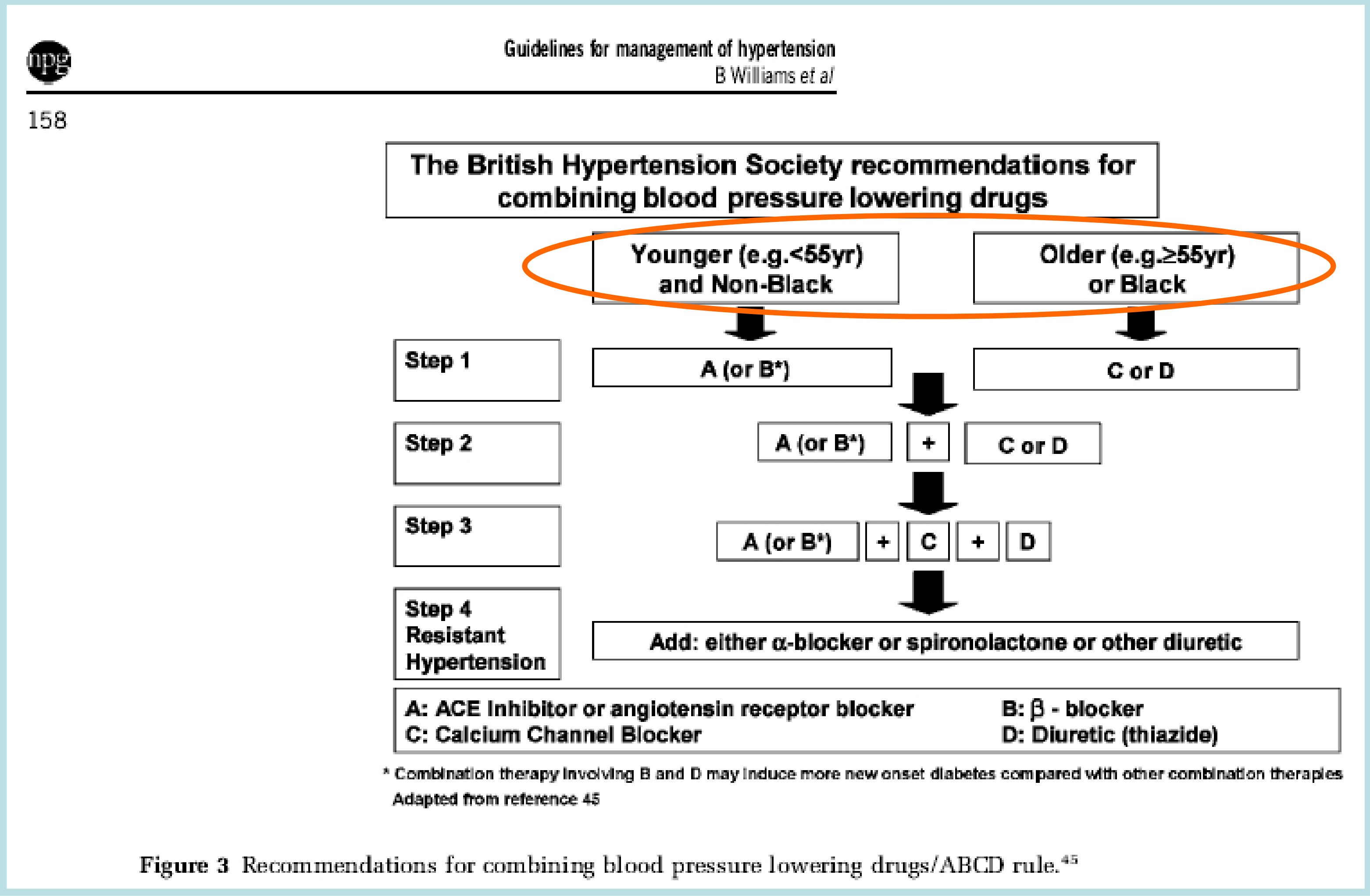
Journal of Human Hypertension (2004) 18, 139–185
© 2004 Nature Publishing Group All rights reserved 0950-9240/04 $25.00
www.nature.com/jhh

BRITISH HYPERTENSION SOCIETY GUIDELINES

Guidelines for management of hypertension: report of the fourth working party of

2004—B

B Williams[1], NR
S McG Thom[2]



Guidelines for management of hypertension
B Williams *et al*

158

The British Hypertension Society recommendations for combining blood pressure lowering drugs

Younger (e.g.<55yr) and Non-Black

Older (e.g.≥55yr) or Black

Step 1: A (or B*) | C or D

Step 2: A (or B*) + C or D

Step 3: A (or B*) + C + D

Step 4 Resistant Hypertension: Add: either α-blocker or spironolactone or other diuretic

A: ACE inhibitor or angiotensin receptor blocker
B: β - blocker
C: Calcium Channel Blocker
D: Diuretic (thiazide)

* Combination therapy involving B and D may induce more new onset diabetes compared with other combination therapies

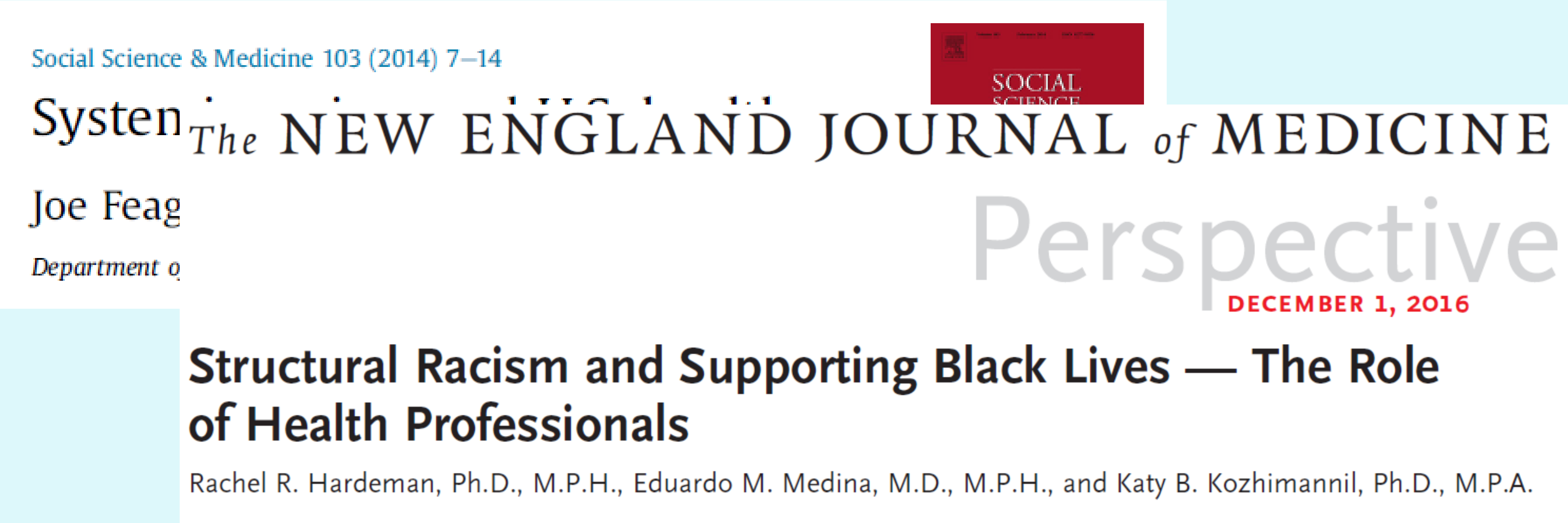
Adapted from reference 45

Figure 3 Recommendations for combining blood pressure lowering drugs/ABCD rule.[45]

In medicine (unlike employment, law enforcement, etc), use of race in algorithms is PROMOTED as long as the goal is equality of outcomes (e.g. NIMHD)

Argument often made (e.g. Sally Satel) that it would be unethical to IGNORE race in decision-making.

But at the same time, there are copious data on racism in medical practice, such that groups are treated unequally in physically and psychologically harmful ways.



Social Science & Medicine 103 (2014) 7–14

SOCIAL SCIENCE

Systen

Joe Feag

Department o

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

DECEMBER 1, 2016

Structural Racism and Supporting Black Lives — The Role of Health Professionals

Rachel R. Hardeman, Ph.D., M.P.H., Eduardo M. Medina, M.D., M.P.H., and Katy B. Kozhimannil, Ph.D., M.P.A.

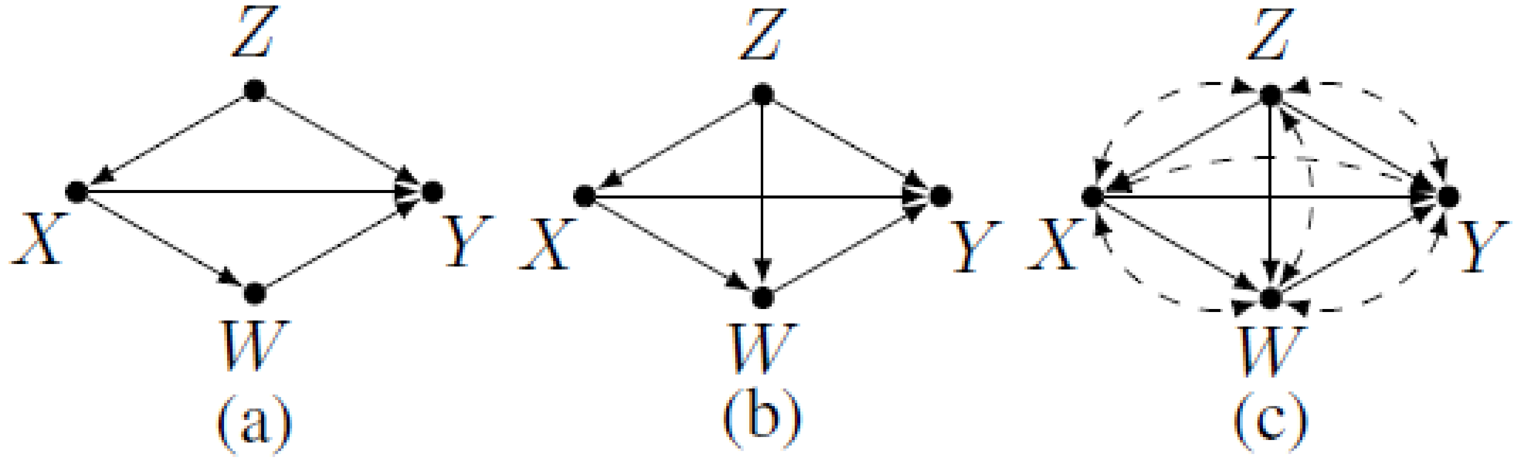
So for medical treatment, what is the logic used in:

Identifying a practice difference that is "unfair"?

Excluding alternative explanations?
- by measured factors?
- by unmeasured factors?

Accounting for knowledge of a previous difference in justifying a future difference?

Planning interventions to diminish the difference?



Association of protected class X (e.g. race) and outcome Y.

Can be direct (X → Y)
Can be mediated by other factors (X → W → Y)
Can be confounded by an observed covariate (X ← Z → Y)
Can be confounded by unobserved covariates (DAG c)

Zhang J, Bareinboim E. Fairness in decision-making—the causal explanation formula. 32nd AAAI Conference on Artificial Intelligence 2018 Apr 25. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewPaper/16949

Z & B:

$$TV_{x0,x1}(y) = P(y|x1) - P(y|x0)$$

$$ETT_{x0,x1}(y) = P(y_{x1}|x0) - P(y|x0)$$

Kunser et al: $ETT_{x0,x1}(y)^* = P(y_{x1}|x0,z,w) - P(y|x0,z,w)$

Datta et al: $CDE_{x0,x1}(y_{z,w}) = P(y_{x1,z,w}) - P(y_{x0,z,w})$

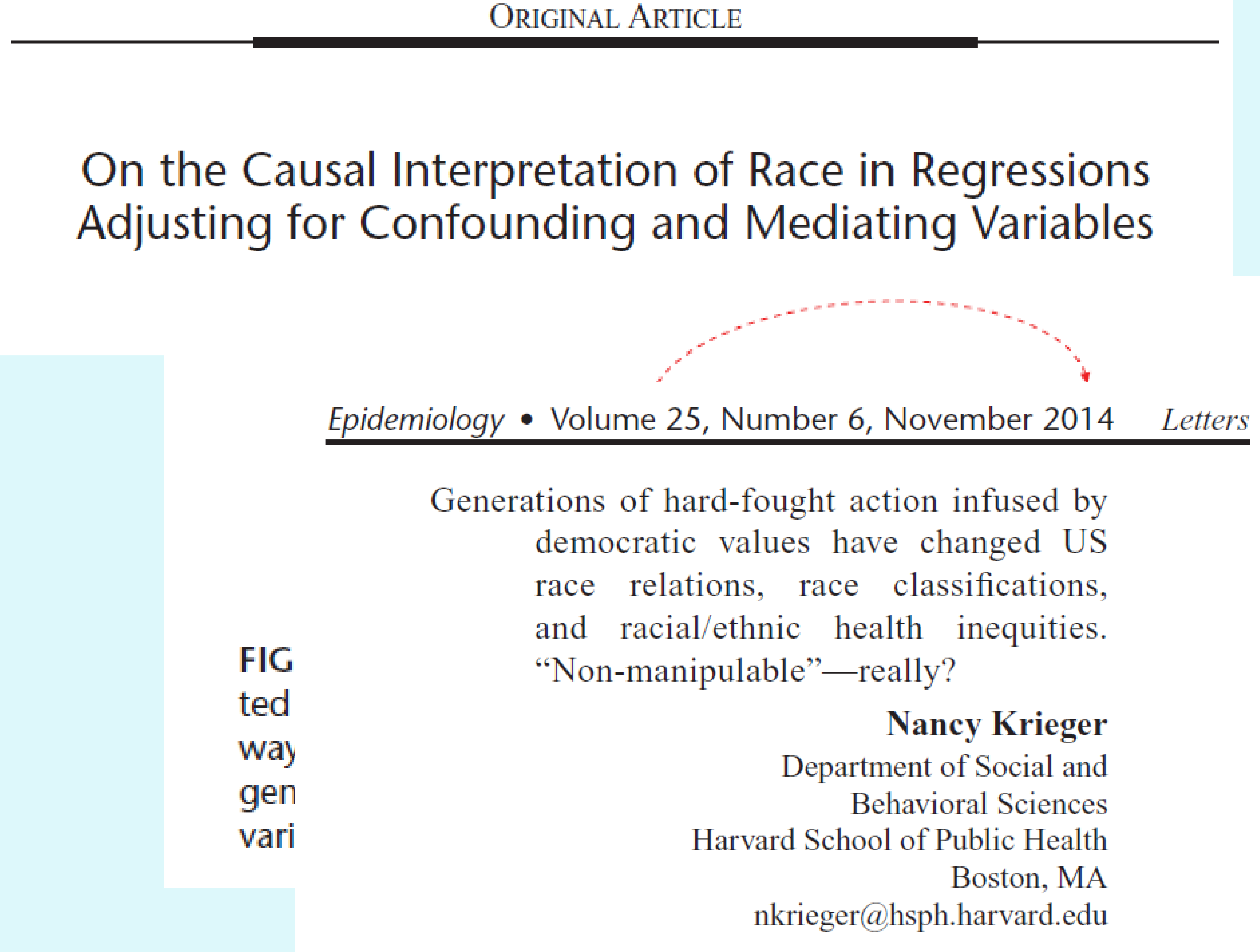
Pearl: $NDE_{x0,x1}(y) = P(y_{x1},w_{x0}) - P(y_{x0})$

$NIE_{x0,x1}(y) = P(y_{x0},w_{x1}) - P(y_{x0})$

Kusner MJ, et al. 2017. Counterfactual fairness. arXiv preprint 1703.06856.

Datta A, Sen S, Zick Y. 2016. Algorithmic transparency via quantitative input influence. In Security and Privacy (SP), 2016 IEEE Symp., 598–617.

Pearl J. 2009. Causality. New York: Cambridge University Press. 2nd ed.



ORIGINAL ARTICLE

On the Causal Interpretation of Race in Regressions Adjusting for Confounding and Mediating Variables

Epidemiology • Volume 25, Number 6, November 2014 *Letters*

Generations of hard-fought action infused by democratic values have changed US race relations, race classifications, and racial/ethnic health inequities. "Non-manipulable"—really?

Nancy Krieger
Department of Social and
Behavioral Sciences
Harvard School of Public Health
Boston, MA
nkrieger@hsph.harvard.edu

FIG
ted
way
gen
vari

Mediation example for gender

Kolev et al (2019) studied blinded evaluation of grant proposals sent to Gates Foundation 2008-2017.

Female applicants scored lower. Difference not explained by reviewer characteristics, proposal topics, or measures of applicant quality.

Differences explained by text-based measures of titles and descriptions, specifically: usage of broad and narrow words.

Text-based measures that predict higher reviewer scores do not also predict higher ex-post performance.

Kolev J, Fuentes-Medel Y, Murray F. IS BLINDED REVIEW ENOUGH? HOW GENDERED OUTCOMES ARISE EVEN UNDER ANONYMOUS EVALUATION NBER Working Paper 25759, May 2019.

Use of Experiment:

Adams et al. show that art made by women sells for lower prices at auction, and demonstrate that this is not a function of talent or thematic choices. It is solely because the artists are female.

Adams RB, Kräussl R, Navone MA, Verwijmeren P. Is gender in the eye of the beholder? Identifying cultural attitudes with art auction prices. 2017 Dec 6. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3083500

To test the proposed explanation that women are intrinsically less talented than men, the authors conducted experiments.

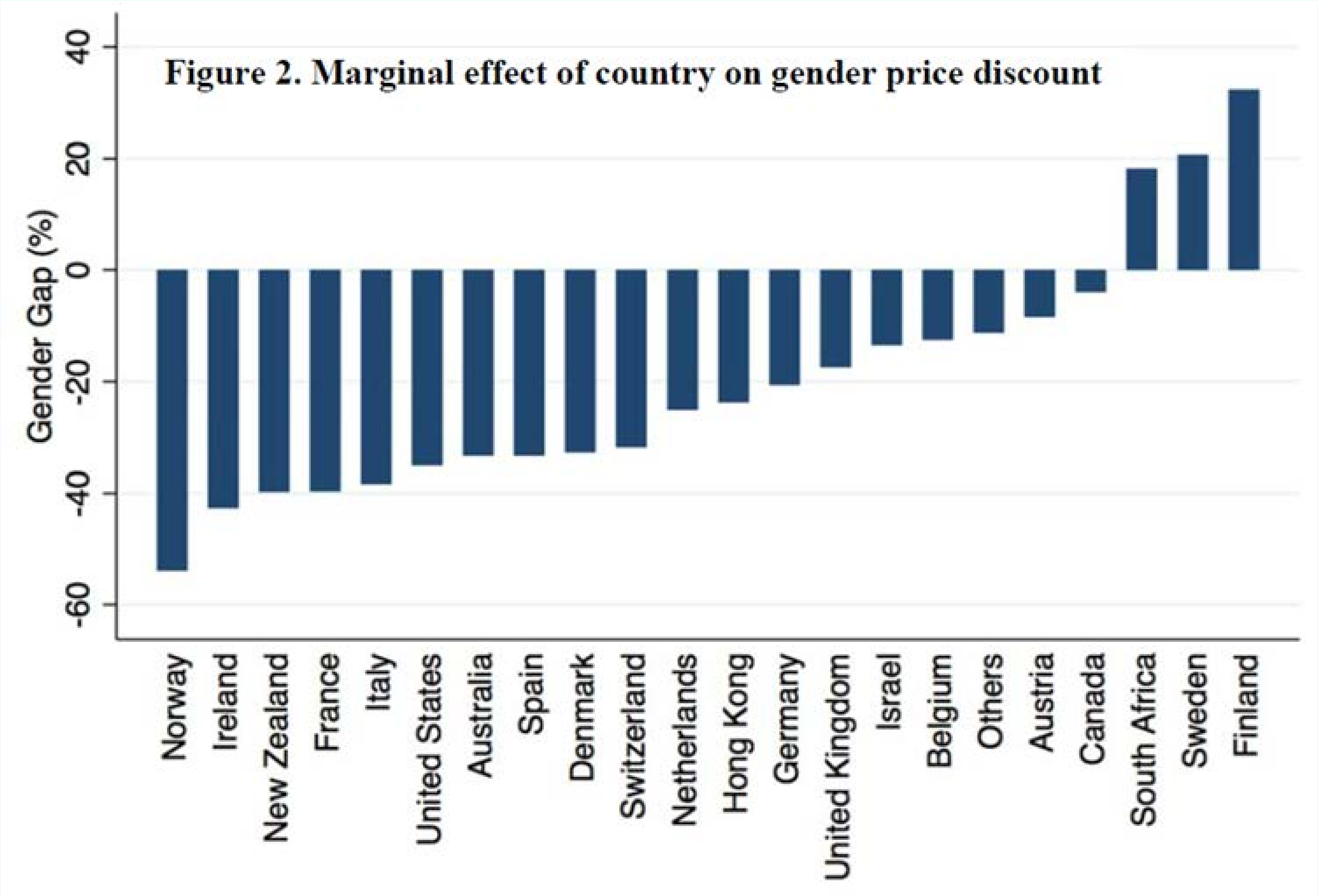
1) They showed sets of lesser-known paintings to large n of participants asking them to guess the gender of the artists. Respondents did no better than chance.

2) They used a computer program to generate paintings and randomly designate the "artists" with male or female names. Asked large n of participants to rate the paintings and assign a value. Female artists systematically earned a lower valuation.

Perhaps participants knew that female works are valued less and then they made their appraisals accordingly. This could be deemed "rational", even if not fair ("statistical discrimination").

Adams RB, Kräussl R, Navone MA, Verwijmeren P. Is gender in the eye of the beholder? Identifying cultural attitudes with art auction prices. 2017 Dec 6. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3083500

The gap was also variable across countries and changed over time.



Adams RB, Kräussl R, Navone MA, Verwijmeren P. Is gender in the eye of the beholder? Identifying cultural attitudes with art auction prices. 2017 Dec 6. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3083500

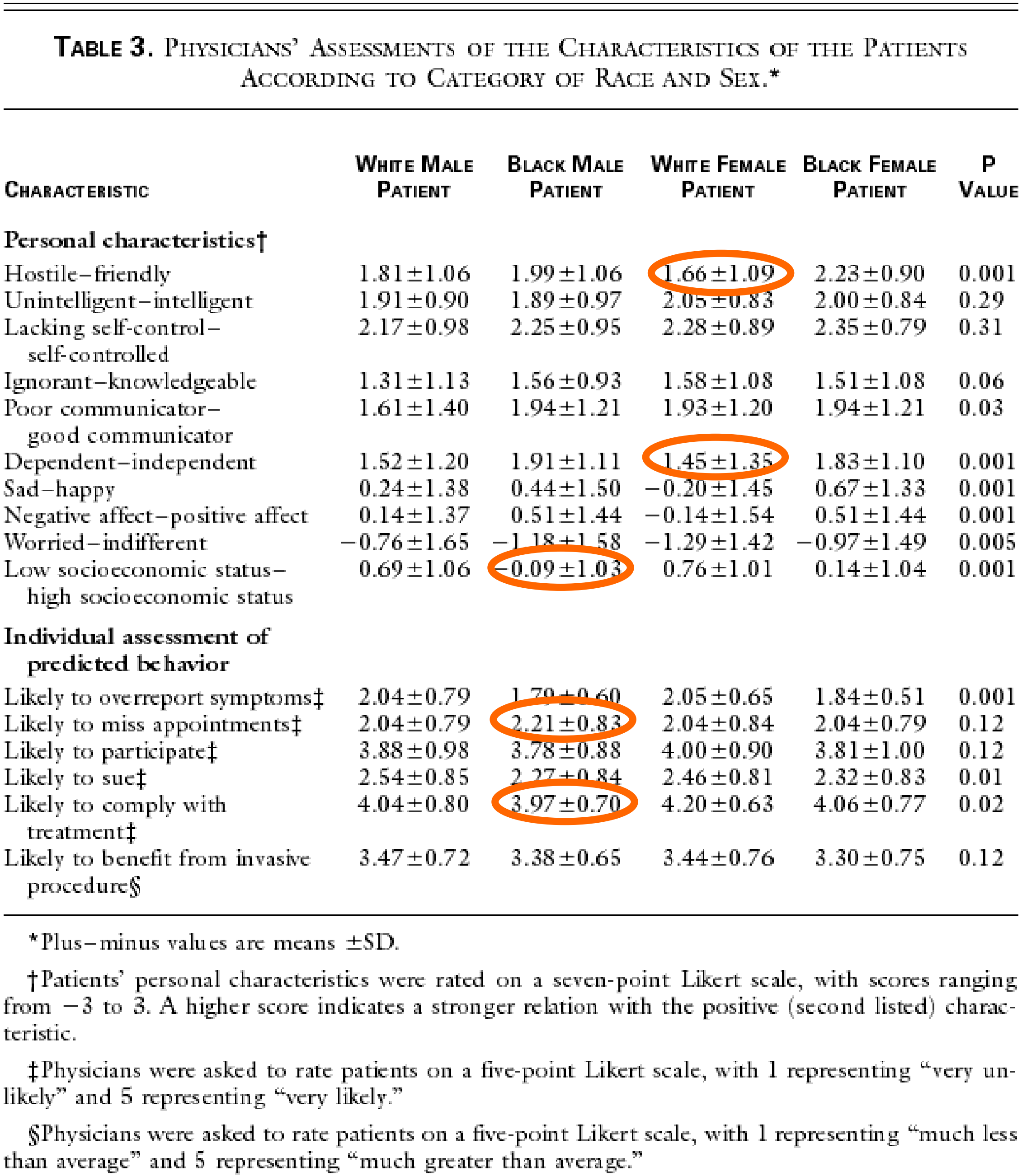


TABLE 3. PHYSICIANS' ASSESSMENTS OF THE CHARACTERISTICS OF THE PATIENTS ACCORDING TO CATEGORY OF RACE AND SEX.*

CHARACTERISTIC	WHITE MALE PATIENT	BLACK MALE PATIENT	WHITE FEMALE PATIENT	BLACK FEMALE PATIENT	P VALUE
Personal characteristics†					
Hostile–friendly	1.81±1.06	1.99±1.06	1.66±1.09	2.23±0.90	0.001
Unintelligent–intelligent	1.91±0.90	1.89±0.97	2.05±0.83	2.00±0.84	0.29
Lacking self-control–self-controlled	2.17±0.98	2.25±0.95	2.28±0.89	2.35±0.79	0.31
Ignorant–knowledgeable	1.31±1.13	1.56±0.93	1.58±1.08	1.51±1.08	0.06
Poor communicator–good communicator	1.61±1.40	1.94±1.21	1.93±1.20	1.94±1.21	0.03
Dependent–independent	1.52±1.20	1.91±1.11	1.45±1.35	1.83±1.10	0.001
Sad–happy	0.24±1.38	0.44±1.50	−0.20±1.45	0.67±1.33	0.001
Negative affect–positive affect	0.14±1.37	0.51±1.44	−0.14±1.54	0.51±1.44	0.001
Worried–indifferent	−0.76±1.65	−1.18±1.58	−1.29±1.42	−0.97±1.49	0.005
Low socioeconomic status–high socioeconomic status	0.69±1.06	−0.09±1.03	0.76±1.01	0.14±1.04	0.001
Individual assessment of predicted behavior					
Likely to overreport symptoms‡	2.04±0.79	1.79±0.60	2.05±0.65	1.84±0.51	0.001
Likely to miss appointments‡	2.04±0.79	2.21±0.83	2.04±0.84	2.04±0.79	0.12
Likely to participate‡	3.88±0.98	3.78±0.88	4.00±0.90	3.81±1.00	0.12
Likely to sue‡	2.54±0.85	2.27±0.84	2.46±0.81	2.32±0.83	0.01
Likely to comply with treatment‡	4.04±0.80	3.97±0.70	4.20±0.63	4.06±0.77	0.02
Likely to benefit from invasive procedure§	3.47±0.72	3.38±0.65	3.44±0.76	3.30±0.75	0.12

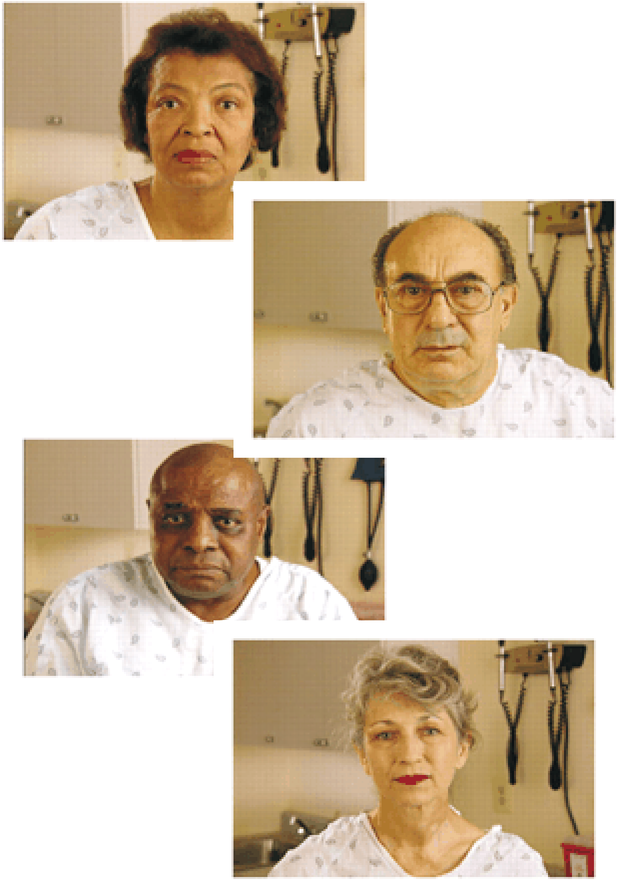
*Plus–minus values are means ±SD.

†Patients' personal characteristics were rated on a seven-point Likert scale, with scores ranging from −3 to 3. A higher score indicates a stronger relation with the positive (second listed) characteristic.

‡Physicians were asked to rate patients on a five-point Likert scale, with 1 representing "very unlikely" and 5 representing "very likely."

§Physicians were asked to rate patients on a five-point Likert scale, with 1 representing "much less than average" and 5 representing "much greater than average."

Schulman KA, et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. N Engl J Med. 1999 Feb 25; 340(8):618-26.



Medical Example: GFR estimation [mentioned by Dorothy Roberts on Wednesday]

Glomerular filtration rate = overall index of kidney function.

GFR cannot be measured directly in clinical practice, so it is estimated from serum levels of endogenous filtration markers.

Several equations have been developed:

Cockcroft-Gault equation
MDRD equation
CKDEPI equation (most recommended)
Cystatin C equation

Cockcroft-Gault equation (1976)

Estimates GFR from age, sex, body weight, and serum creatinine.

Original study population included 249 US white men.

Adjustment factor for women based on the assumption of 15% lower creatinine generation due to lower muscle mass.

This equation does not contain a variable for race, and on average underestimates GFR in African Americans.

$$\text{CrCl (mL/min)} = \frac{(140 - \text{age}) \times \text{lean body weight [kg]}}{\text{Cr [mg/dL]} \times 72}$$

Inker, L.A., Fan, L. & Levey, A.S. Comprehensive Clinical Nephrology. 5th edn (Elsevier Saunders, Philadelphia, PA, 2015).

MDRD Equation (1999)

Estimates GFR indexed for BSA from age, sex, race (African American vs. white and other) and serum creatinine.

Original study population included 1,628 US men and women

Studies in have showed that the MDRD equation is substantially more accurate compared to the Cockcroft-Gault equation.

$$\text{GFR, in mL/min per } 1.73\ \text{m}^2 = 175 \times \text{SCr} (\exp[-1.154]) \times$$

$$\text{Age} (\exp[-0.203]) \times (0.742 \text{ if female}) \times (1.21 \text{ if black})$$

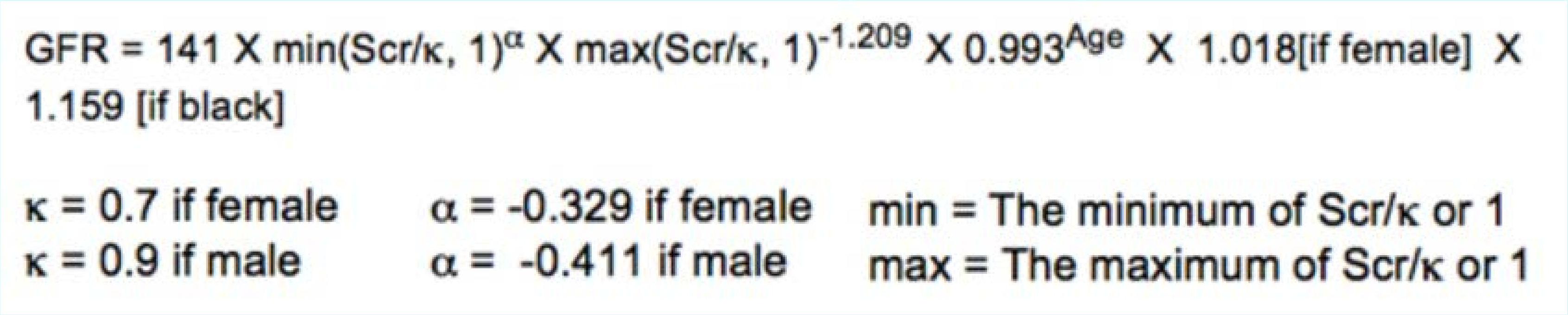
Stevens, L.A. et al. Impact of creatinine calibration on performance of GFR estimating equations in a pooled individual patient database.
Am. J. Kidney Dis. 50, 21–35 (2007).

CKD-EPI Equation (2003)

NIDDK assembled a pooled dataset of n = 12,150 from diverse studies in North America and Europe, including individuals with and without kidney disease and with diabetes.

Same variables as in MDRD equation, but functions and coefficients differ. Again, race variable is African Americans vs. whites and others.

Evaluation of the CKD-EPI vs. the MDRD equation in the validation population showed improved accuracy, **but performance of both equations was worse outside North America**.



$$\text{GFR} = 141 \text{ X } \min(\text{Scr}/\kappa, 1)^{\alpha} \text{ X } \max(\text{Scr}/\kappa, 1)^{-1.209} \text{ X } 0.993^{\text{Age}} \text{ X } 1.018[\text{if female}] \text{ X } 1.159 \text{ [if black]}$$

$\kappa = 0.7$ if female	$\alpha = -0.329$ if female	min = The minimum of Scr/κ or 1
$\kappa = 0.9$ if male	$\alpha = -0.411$ if male	max = The maximum of Scr/κ or 1

Earley, A., Miskulin, D., Lamb, E.J., Levey, A.S. & Uhlig, K. Estimating equations for glomerular filtration rate in the era of creatinine standardization: a systematic review. Ann. Intern. Med. 156, 785–95 (2012).

GFR ESTIMATION USING CYSTATIN C

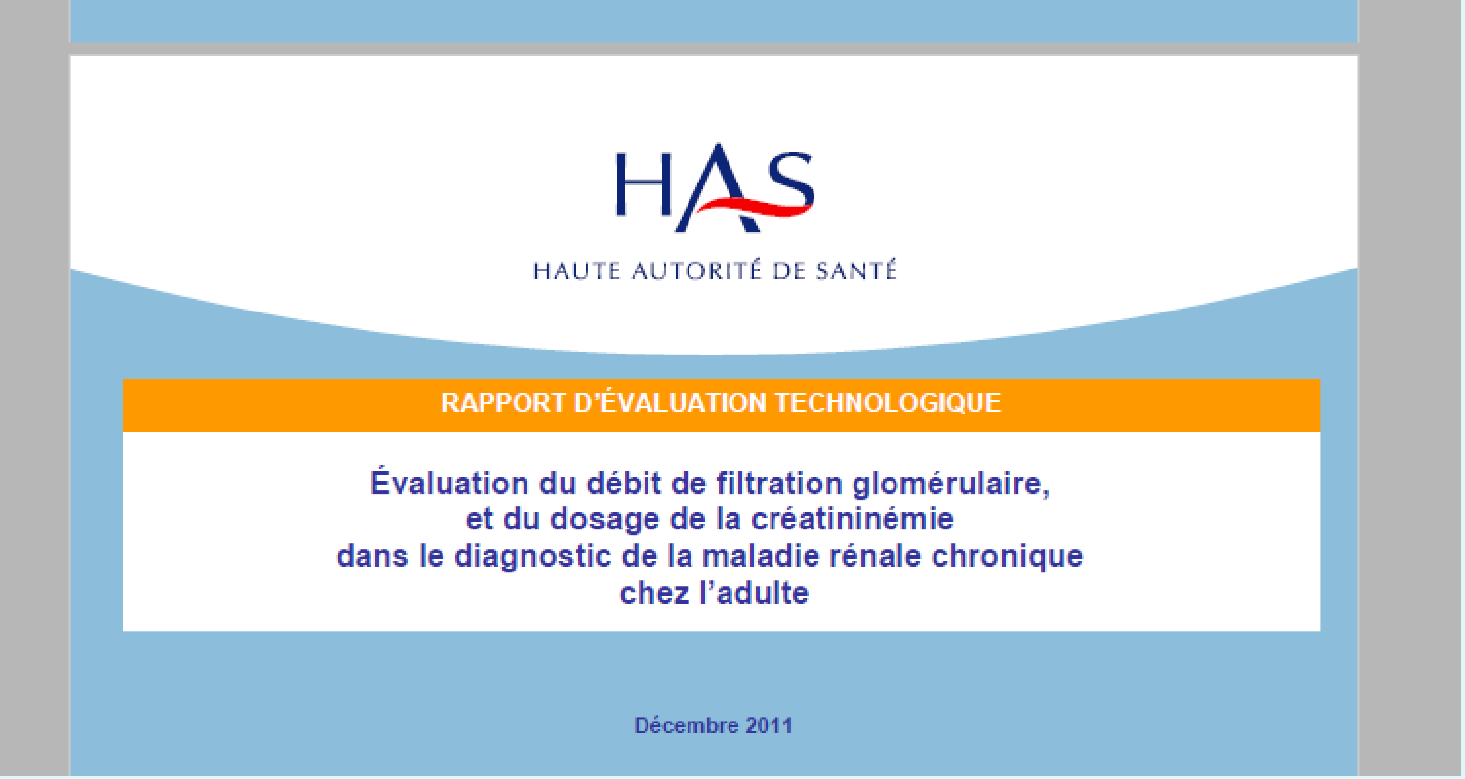
Cystatin C identified in 1979 and proposed as a filtration marker in 1985. Still, not common in practice.

Cystatin C not affected by muscle mass or diet, and, thus, is more strongly correlated with measured GFR than creatinine, and less strongly associated with age, sex, and race.

But strongly affected by smoking, inflammation, adiposity, thyroid diseases, etc.

Studies confirmed the findings that estimated GFR with cystatin C and creatinine is more precise than using creatinine alone **and no longer requires a local coefficient for racial or ethnic groups**.

Levey, A.S., Inker, L.A. & Coresh, J. GFR estimation: from physiology to public health. Am. J. Kidney Dis. 63, 820–834 (2014).



HAS

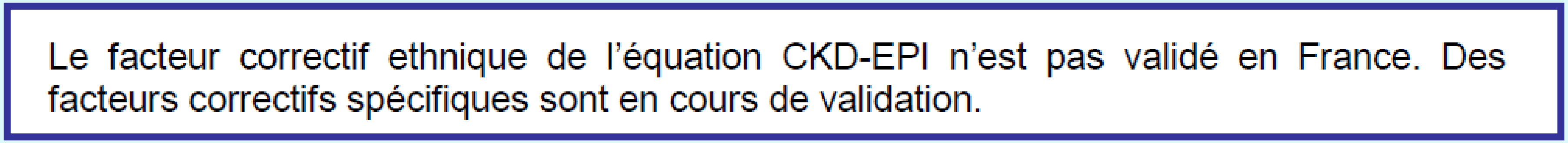
HAUTE AUTORITÉ DE SANTÉ

RAPPORT D'ÉVALUATION TECHNOLOGIQUE

Évaluation du débit de filtration glomérulaire,
et du dosage de la créatininémie
dans le diagnostic de la maladie rénale chronique
chez l'adulte

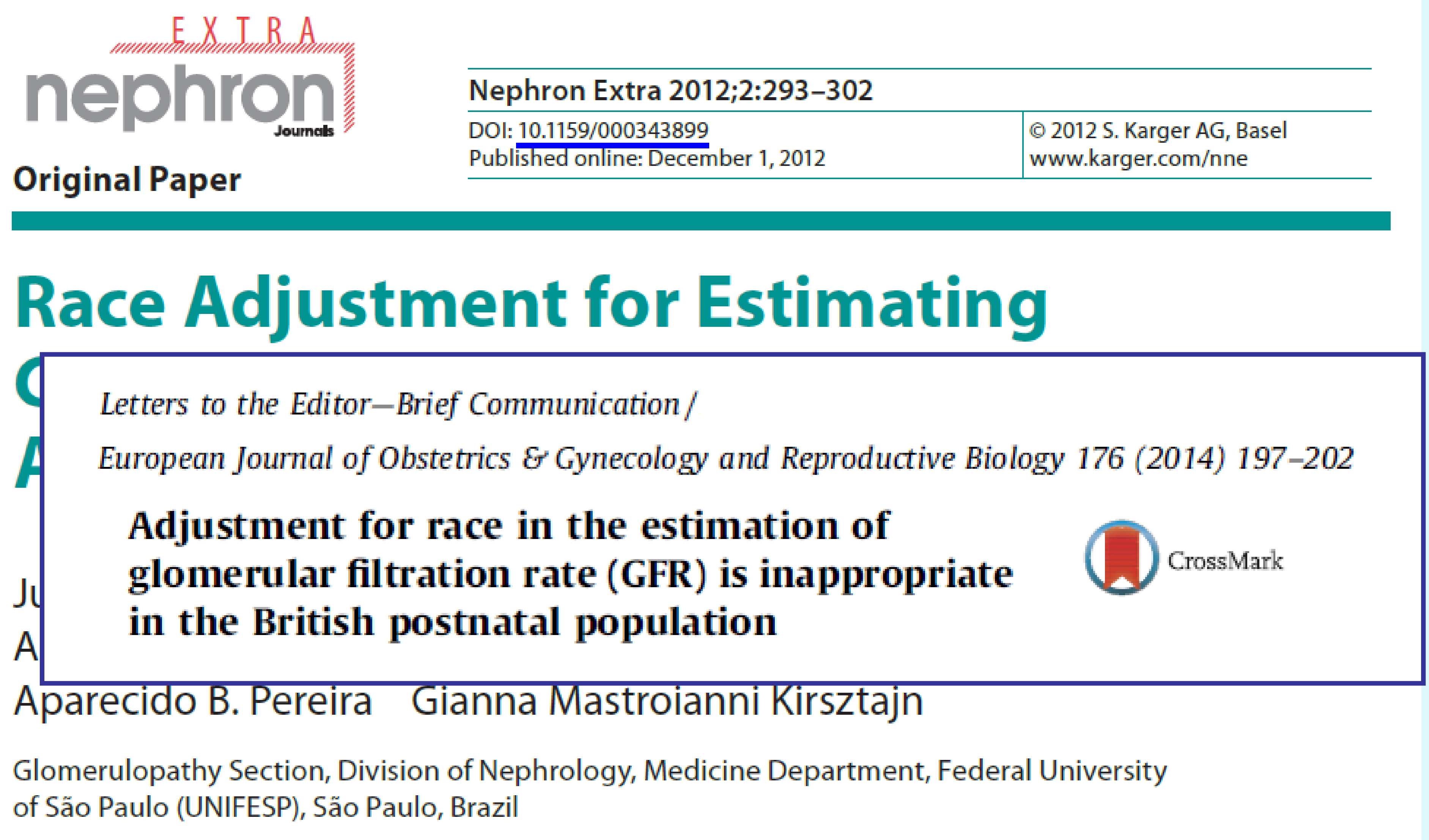
Décembre 2011

According to the French *Haute Autorité de Santé,* the US correction factor for race in the CKD-EPI equation should NOT be applied in the French population



Le facteur correctif ethnique de l'équation CKD-EPI n'est pas validé en France. Des facteurs correctifs spécifiques sont en cours de validation.

Likewise, studies in Brazil and UK have shown that no race term is needed in the model in these settings:



EXTRA nephron Journals

Original Paper

Nephron Extra 2012;2:293–302

DOI: 10.1159/000343899
Published online: December 1, 2012

© 2012 S. Karger AG, Basel
www.karger.com/nne

Race Adjustment for Estimating

C

A

Letters to the Editor—Brief Communication /

European Journal of Obstetrics & Gynecology and Reproductive Biology 176 (2014) 197–202

Adjustment for race in the estimation of glomerular filtration rate (GFR) is inappropriate in the British postnatal population

CrossMark

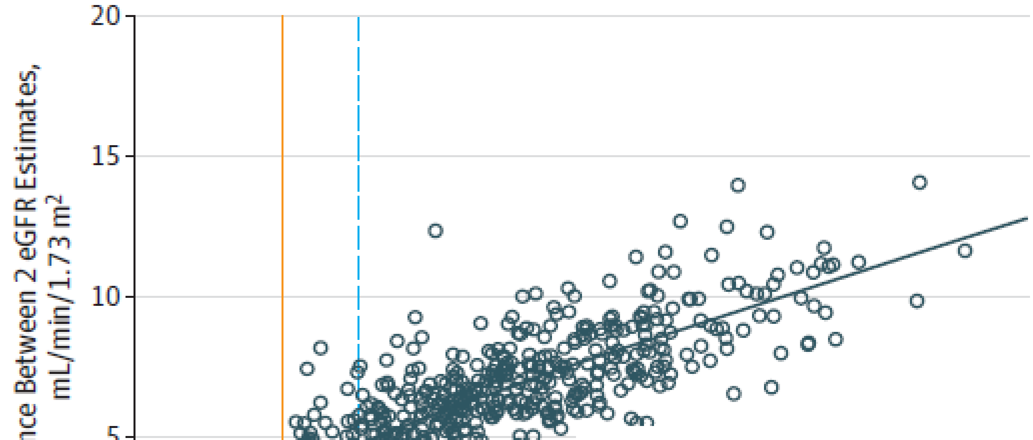
Ju

A

Aparecido B. Pereira Gianna Mastroianni Kirsztajn

Glomerulopathy Section, Division of Nephrology, Medicine Department, Federal University of São Paulo (UNIFESP), São Paulo, Brazil

Figure. Relationship Between Racial Categories and Estimation of Kidney Function Across the Spectrum of Chronic Kidney Disease



Circles indicate how much higher estimated
for patients when assigned black race instea
calculated twice for self-identified black adu
black race and then assigning them nonblac
kidney function lower than (ie, to the left of
for the kidney transplant waiting list. The Ki
Outcomes (KDIGO) guidelines recommend
with kidney function lower than (left of) the
adult participants in the Chronic Renal Insuf
urinary ^{125}I-iothalamate clearance testing, a
of kidney filtration function. Estimated GFR
Chronic Kidney Disease Epidemiology Colla

Nwamaka Denise Eneanya, MD, MPH[1,2]; Wei Yang, PhD[3]; Peter Philip Reese, MD, MSCE[1,3]

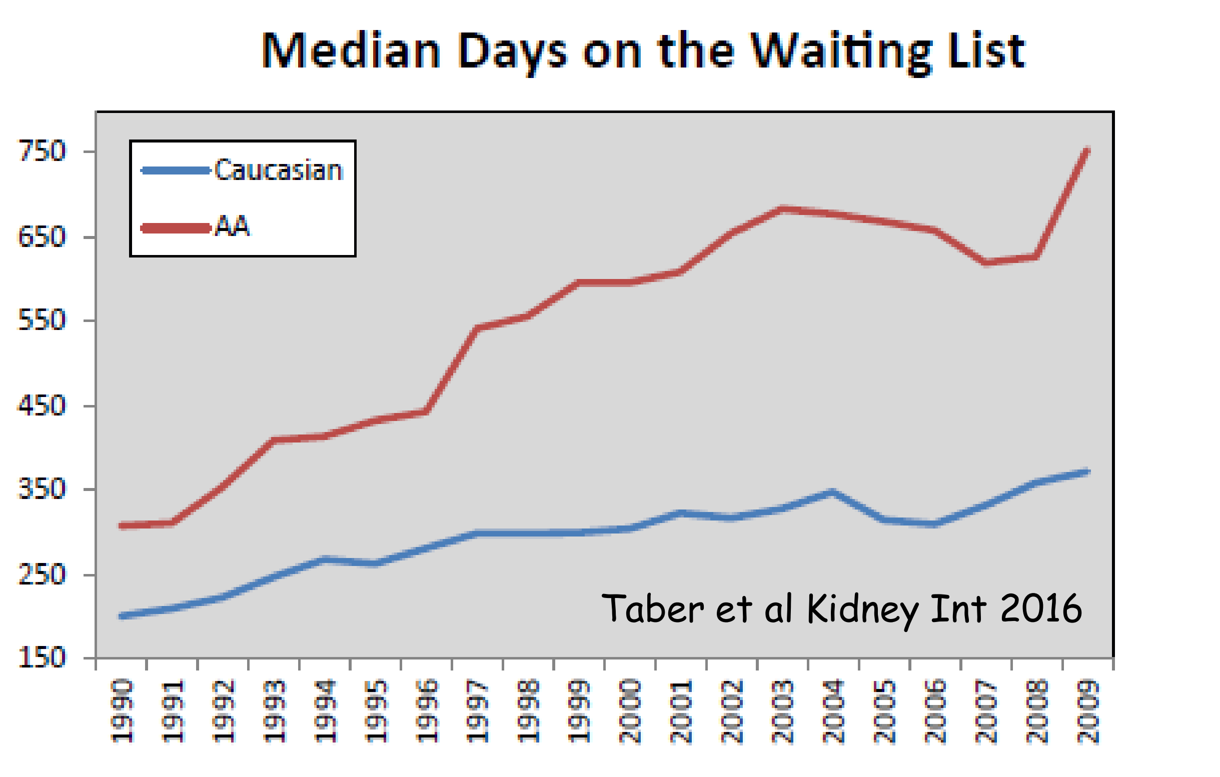
Reconsidering the Consequences of Using Race to Estimate Kidney Function

JAMA.

Published online June 6, 2019.

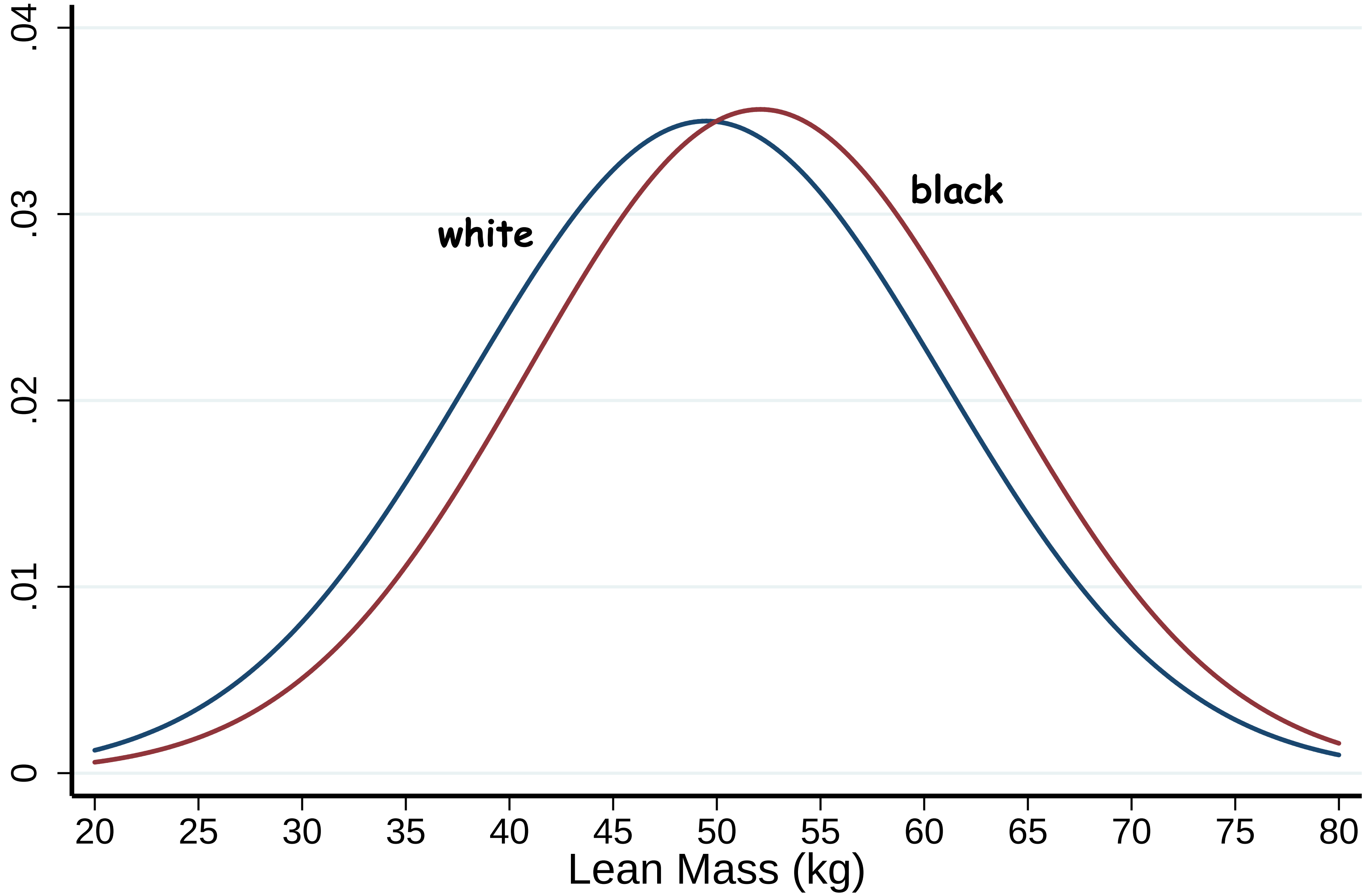
doi:10.1001/jama.2019.5774

Median Days on the Waiting List



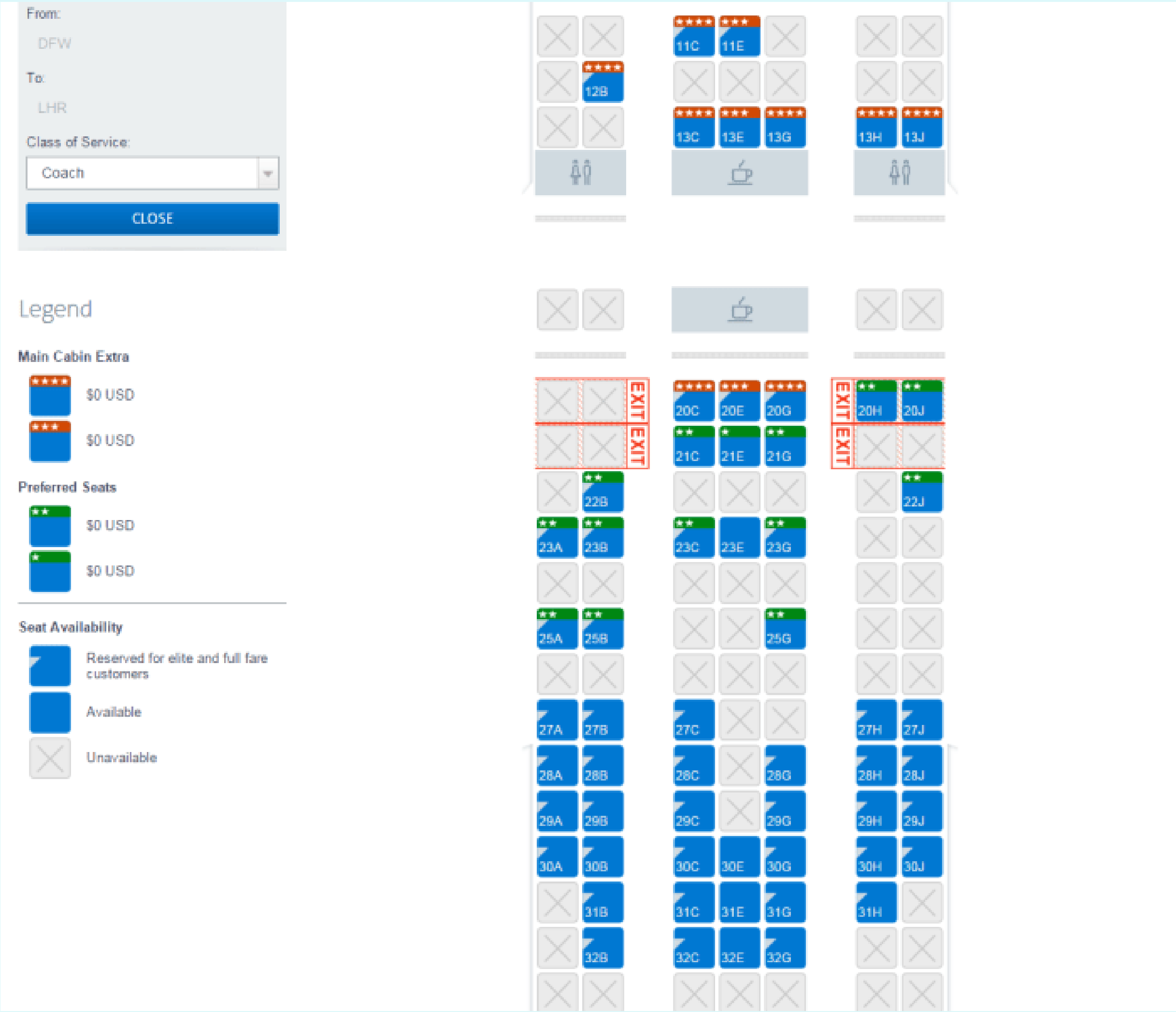
Taber et al Kidney Int 2016

Non-Hispanic White versus Black Lean Mass from DXA



NHANES, men and women

Few people in the population sit right at the population average.



From:
DFW
To:
LHR
Class of Service:
Coach

CLOSE

Legend

Main Cabin Extra

★★★★ $0 USD

★★★ $0 USD

Preferred Seats

★★ $0 USD

★ $0 USD

Seat Availability

Reserved for elite and full fare customers

Available

☒ Unavailable

What is "fair" for the mean is not necessarily "fair" for everyone else.

Observations from the Example

It should not be considered ethical to use a weak proxy that systematically disadvantages a large proportion of the population based on a readily refutable mischaracterization.

Race is used in this algorithm not because it is the optimal quantity in any rational sense, but rather because of its historical and ideological saliency.

Overall Summary

Causal framework of direct and indirect effects has a concrete experimental foundation, but does not encompass forms of "statistical discrimination" that are based on algorithms designed to optimize one (arbitrary) function at the expense of many others.