Studying Discrimination

Javier Polavieja

The D-Lab
Discrimination & Inequality Lab

uc3m | Universidad Carlos III de Madrid

Discrimination

- Discrimination (D) is both unjust and inefficient. It hurts people, it heightens inequality and it hampers economic growth
- Discriminatory behaviours take many forms, but they all involve unequal treatment + (often) some form of exclusion or rejection
- Many basis for D→ e.g. age, gender, sexual orientation, religion, ethnicity, phenotype, looks, etc
- Many realms and agents of D
- …but researchers concerned with socio-economic inequality typically focus on…
  1. D in access to crucial assets/resources (e.g. D against minority children in schools, housing market, etc)
  2. D in the labor market (DLM) (e.g. access to employment, promotion opportunities and pay)

!

D in the labour market (LMD) is the main focus of today's talk

!

D is only but one possible explanation of LM gaps

!

As we will see, estimating **D** is a complex task

Outline

1. Intro
2. Types of D (theory)
    1. Taste based D
    2. Statistical D
    3. Consumer-driven D
    4. Implicit D
3. Field experiments: why we need them
    1. Limitations of standard observational data
       An example: The Oaxaca-Blinder decomposition method
    2. Field-experiments
        1. What are field-experiments?
        2. Types
        3. Correspondence studies
            1. Full factorial and fractional factorial designs
            2. Pair and unpaired designs
4. Summary of findings on gender and ethno-racial D in the LM
    - The GEMM study
5. Q&A

2. Types of labour market **D**

Theory

!

Theories of LMD focus on discriminatory practices by firms (employers, managers & directors) in hiring, promoting and paying workers from specific social groups (e.g. women, ethnic/racial minorities)

Types of LMD

1. Discrimination by taste (Becker 1993[1964])

- Firms discriminate against particular groups (e.g. women/minorities) due to
    1. the firms' (i.e. employers) dislike for them
    2. the firms' employees' dislike for them
    3. The firm's customers dislike for them
- In the literature DbT is often used as referring only to 1) but Becker spoke of the three forms (I follow the conventional view and treat consumer-driven D as a different case below)
- Because D by taste is based on prejudice, it is **irrational** from an economic point of view

  → In equilibrium competitive markets should penalize firms that discriminate by taste → This is actually not the case in consumer-driven D

Types of LMD

2. Statistical discrimination (Arrow 1971; Phelps 1972; Aigner and Cain 1977)

- Under incomplete/asymmetric information, rational employers might still discriminate against individuals from some groups if such groups are believed to be…

  1) *on average* less productive (e.g. some international migrants in communicational-intensive tasks) –or having a higher average probability of interrupting their careers (e.g. women) OR…
  2) If the *variance* in the distribution of unobserved skills is expected to differ by group (e.g. more dispersion in motivation) OR…
  3) If the signal employers receive for judging expected productivity is noisier for some groups (e.g. test scores from foreign schools)

Types of LMD

2. Statistical discrimination (Phelps 1972; Arrow 1973; Aigner and Cain 1977)

- Statistical D is based on information deficits, not taste
- Yet employers' assessments of the distribution of unobserved qualities are often based on biased beliefs (stereotypes)

  → Processes of status categorization typically involved in stereotypes (i.e. beliefs about performance, behaviours, capabilities→ e.g. "gypsies are lazy"; "women can't handle pressure"…)

  Stereotypes are widely shared by members of the in-group/dominant culture/majority population

  D reinforces stereotyping because people interpret differences in outcomes as proof of their prior stereotypical beliefs

- IMPLICATION: Yet for stat-D theory, reducing information deficits should always reduce D

Types of LMD

3. Customer-driven discrimination

- Rational firms might still discriminate against particular individuals to comply with customers/clients own prejudiced preferences
  - Example: British Oil companies working in the Persian Gulf did not hire women not to upset their main clients
- This is a situation where firms rationally adapt to their costumers' irrational tastes (this is why it is considered distinct from DbT)
- Recent experiments with employers suggest customer-driven discrimination plays a significant role in shaping employers' decisions (Baert & De Pauw 2014)

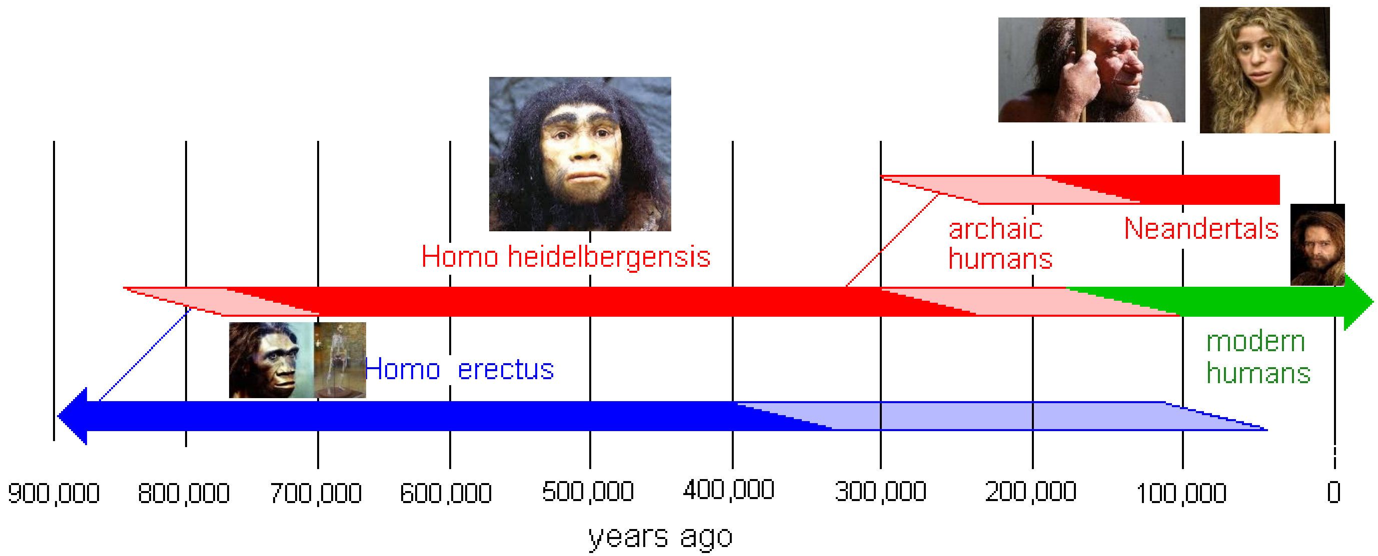
Types of LMD

4. Implicit D by employers

- Note both DbT and Stat D imply conscious assessments of the applicants' qualities by employers
- But research in cognitive and social psychology shows people often categorize, stereotype and D others on the basis of implicit mental associations of which they are largely (if not fully) unaware (see e.g. Richeson and Sommers 2015; Phelps and Thomas 2003; Reskin 2000)
- Evolutionary psychologists argue that the "computational machinery" that triggers race, sex and age categorization & stereotyping of others is a universal feature of human cognition, which can be explained by its adaptive function (see e.g. discussion in Kurzban et al. 2001; Neuberg and Schaller 2016)
  - Age, gender, and race would be "primitive" dimensions which the mind activates in an automatic and mandatory fashion when encountering others

IMPLICATION: D might be harder to eradicate

A very long human evolution



IMPLICATIONS

- Human males and females transmitted their genes under different reproductive circumstances. Genetic adaptation in prehistory led to sexual dimorphism → **sex-specific traits** (i.e. nurturance vs aggressiveness and competition)
- The capacity to recognize outgroup members and to predict their behaviour was crucial for survival under extremely harsh and competitive conditions→ **outgroup recognition**
- The ability to recognize healthy and potentially fertile mates→ **age recognition**

→ Age, sexual and phenotypic categorization are "primitive" dimensions of cognition operating in the 'automatic' area of our brains (where thinking fast takes place)→ Little conscious control over age, sex, and phenotypic categorization→ Strong forces leading to implicit bias (but see e.g. Reskin 2000; Vaisey 2009; for a "cultural" version of implicit bias)

3. Field experiments: Why we need them

?

Can we measure D quantitatively using observational data (i.e. surveys)?

?

NO,we cannot *because*…

!

…surveys do not contain all the characteristics that employers observe when hiring, promoting, or setting wages…

!

Hence we can never be
sure that the minority and nonminority
workers being compared are truly
similar

!

…Le's now see this Q in greater detail by looking at the standard wage gap decomposition method…

Wage decomposition (Oaxaca-Blinder)

Men's factual earnings→ $S_i^{\text{♂}} = \beta_i^{\text{♂}} X_i^{\text{♂}} + e_i^{\text{♂}}$ (i.e. the wages men really get in the LM)

Women's factual earnings→ $S_i^{\text{♀}} = \beta_i^{\text{♀}} X_i^{\text{♀}} + e_i^{\text{♀}}$ (i.e. the wages women really get in the LM)

Women's counterfactual earnings→ $S^{*}{}_i^{\text{♀}} = \beta_i^{\text{♂}} X_i^{\text{♀}} + e_i^{\text{♀}}$ (i.e. what women would get in a neutral LM)

$$\underbrace{S_i^{\text{♂}} - S_i^{\text{♀}}}_{} = \underbrace{\beta_i^{\text{♂}}(X_i^{\text{♂}} - X_i^{\text{♀}})}_{} + \underbrace{(\beta_i^{\text{♂}} - \beta_i^{\text{♀}})X_i^{\text{♀}}}_{}$$

Gross wage gap= Differences in assests + Differences in returns

Explained component + Unexplained component

(This has often been interpreted as capturing discrimination)

The problem is…

- The effect of any unobserved characteristic affecting wage differences that is not captured by the explained component will necessarily appear in the residual component!!

An example from Farkas & Vicknair (1996)

Table 1. Coefficients from Regression of (ln) Hourly Wage on Selected Independent Variables, and Percent of Wage Gap Explained: Full-Time Black Male Workers, Ages 26 to 33 in 1991

Independent Variable	Mean Difference (Black minus White)	Coefficients		Percent of Wage Gap Explained	
		Model 1	Model 2	Model 1	Model 2
Cognitive skill (1980)	–1.000	—	.109***	—	40.43
Years of school	–.627	.069***	.045***	16.11	10.41
Work experience (weeks)	–75.861	.001***	.001***	22.48	22.48
Mother's education (years)	–1.255	.013*	.010	6.18	4.56
Age in 1979	–.110	–.002	–.007	–.07	–.29
Lives in rural area	–.122	–.019	–.016	–.86	–.74
Lives in the South	.320	–.168**	–.166**	19.91	19.68
Health limitation	–.005	–.038	–.050	–.07	–.10
Married	–.219	.179***	.164***	14.52	13.30
Grew up in South	.341	–.014	.014	1.80	–1.73
Number of children under age 18	.304	–.012	–.007	1.36	.81
Has preschool child	.045	.022	.006	–.36	–.09
R^2 (adjusted)	—	.274	.295	—	
Total	—	—	—	80.99	108.72

100-(Total)= % unexplained

Note: The mean of the dependent variable for this sample is 6.75. The dependent variable is ln(100 × dollars per hour). The sample size is 602.

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$ (two-tailed tests)

The problem is…

- The effect of any unobserved characteristic affecting wage differences that is not captured by the explained component will necessarily appear in the residual component!!

D cannot be properly identified with observational data

Measuring discrimination in the labour market

The advantages of field experiments

What are Field Experiments?

- **Field experiments** combine experimental methods (to improve causal ID) with real-life contexts (to enhance external validity) (Gerber & Green, 2011)
  - "*A data collection strategy that employs manipulation and random assignment to investigate preferences and behaviors in naturally occurring contexts*" (Baldassarri & Abascal 2017:43)
- **Random assignment** of participants into treatment conditions **excludes** the possibility of **unobserved confounders** affecting the outcome, except by calculable chance
- **Randomization allows** for **causal identification** of the effect of the treatment
- **Participants** are typically **unaware** of the experiment and **this excludes** the possibility of **desirability bias** (no observer effects!)…

…but raises **ethical concerns**!

Field Experiments

Compared to observational and lab experiments, **field experiments have**:

- Greater internal validity (i.e. greater potential for causal identification) than observational data
- Greater external validity (i.e. greater generalizability) than lab experiments
  - But researchers have lower control over implementation than in lab settings
- But lower external validity than observational data
  - Note no single concrete experiment is generalizable!
  - Generalizability is achieved by replication across settings

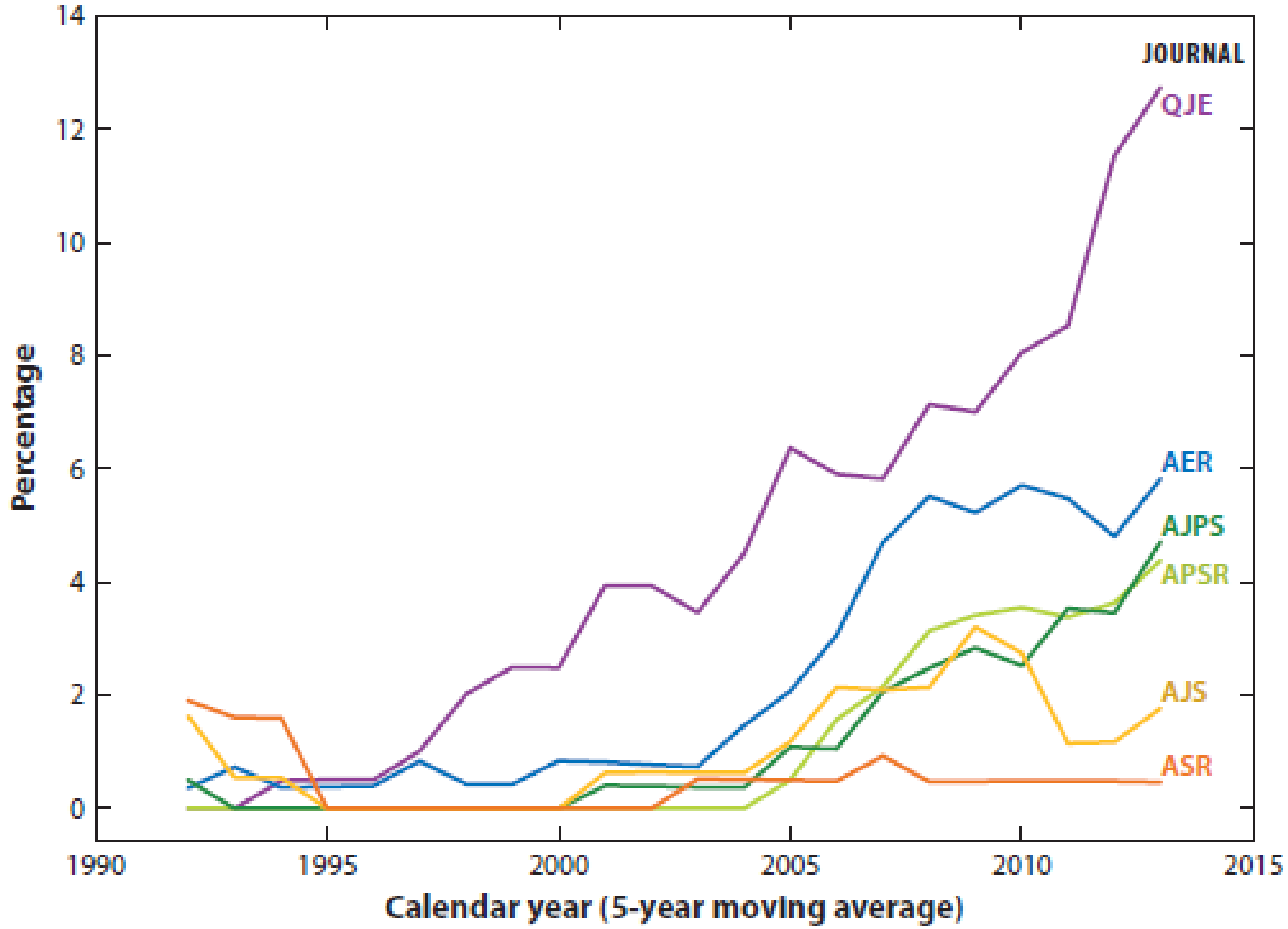


Figure 1

The percentage of research articles reporting field experiments. Abbreviations: AER, *American Economic Review*; AJPS, *American Journal of Political Science*; APSR, *American Political Science Review*; AJS, *American Journal of Sociology*; ASR, *American Sociological Review*; QJE, *Quarterly Journal of Economics*.

Source: Baldassarri & Abascal (2017)

Types of FExs

1. **Randomized Control Trials** (RCTs)→ The gold standard to evaluate policy interventions

   E.g. Perry Preschool Project and the MTO (desegregation) experiments in the US; PROGRESA in Mexico; Anti-poverty experiments in Africa by the MIT Poverty Action Lab, etc

2. **Social Norms Experiments**

   e.g. Broken-Window Experiments on social norms; Lost-Letter Experiments on trust, etc;

3. **Political Mobilization experiments**

   e.g. Get-Out-the Vote Experiments in the US (Green & Gerber 2008)

4. **Behavioral Games in the field**

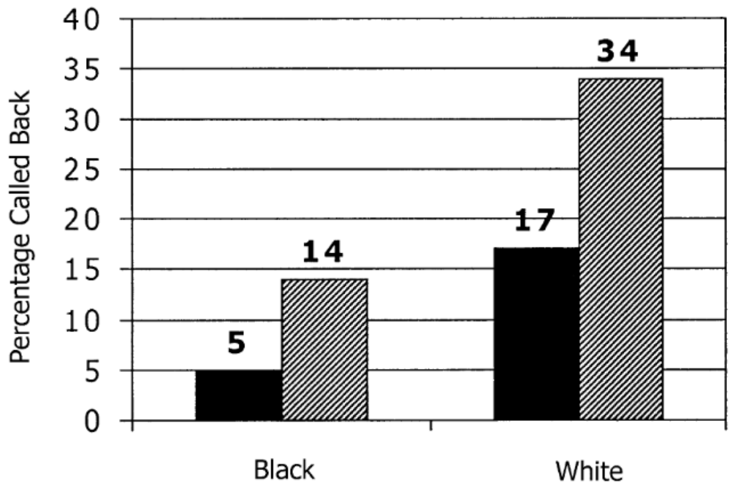
   e.g. cultural differences in trust, cooperation, competitiveness, reciprocity, sanctioning, monitoring, etc

5. **Discrimination experiments**→ The best tool to study market discrimination

   1. Audit (simulation) studies
   2. Correspondence studies

***1. Audit Studies** for hiring D*

- Two or more trained employees of the researcher -auditors or testers- apply for real entry-level jobs
- Auditors are matched for all relevant personal characteristics other than those tested for discrimination (e.g. gender, race)
  - E.g. Pager (2003) investigates the effect of applicants' race and criminal records on employers' hiring decisions in the US using an audit study



Black bars represent criminal record; striped bars represent no criminal record.
The main effects of race and criminal record are statically significant (P>.01). The interaction between the two is not significant in the full sample.

Weaknesses of audit studies

- **Testers** from different groups may **not** appear **identical** to employers (Heckman & Siegelman, 1993; Heckman, 1998)

- **Tester bias**→ Audits are not double-blind, this could generate (un)conscious motives to generate data consistent with their beliefs about labour market discrimination

- Due to **high running costs**, audit studies can only deal with **small n** of treatments

- **Ethically questionable**

2. Correspondence studies for hiring D

- Involve sending **written applications** of fictitious job applicants to real potential employers, varying only the treatment(s) under study
  - Ethnic origin & gender of the applicant is typically manipulated with the **applicant's name**
- There is now **strict comparability** across groups for all information seen by employers
- But **only** accounts for **discrimination at the initial stage** of the job seeking process
- YET audit tests show about 90 % of D takes place at this stage (Riach & Rich 2002:494)

Research designs for correspondence studies

Depending on matching method

- **Paired design**→ 2 identical CVs but for the treatment are sent to each vacancy. PROBLEM: High detection risks & "treatment-salience" bias
- **Unpaired design**→ only 1 CV to each vacancy. Captures overall D, no value in court as evidence of D. Lower detection risk & no "treatment-salience bias" BUT requires larger N

Depending on treatment randomization

- A full **factorial design**→ uses 2 or more randomized factors/treatments (e.g. race, gender, past incarceration, religion) and all experimental units take on all possible combinations across all such factors

  → **each factor is orthogonal to the others** (e.g. fictitious CVs include all possible combinations of race gender criminal record and religiosity and each combination is represented in the experiment with equal probability due to **randomization**)

- **Fractional factorial design**→ some of the possible combinations are omitted for realism or efficiency

  → Not all combinations of treatments are possible—e.g. gender and ethnicity (migrant origin) are orthogonal but e.g. religion cannot be realistically orthogonal to ethnicity

4. The state of the art on LMD

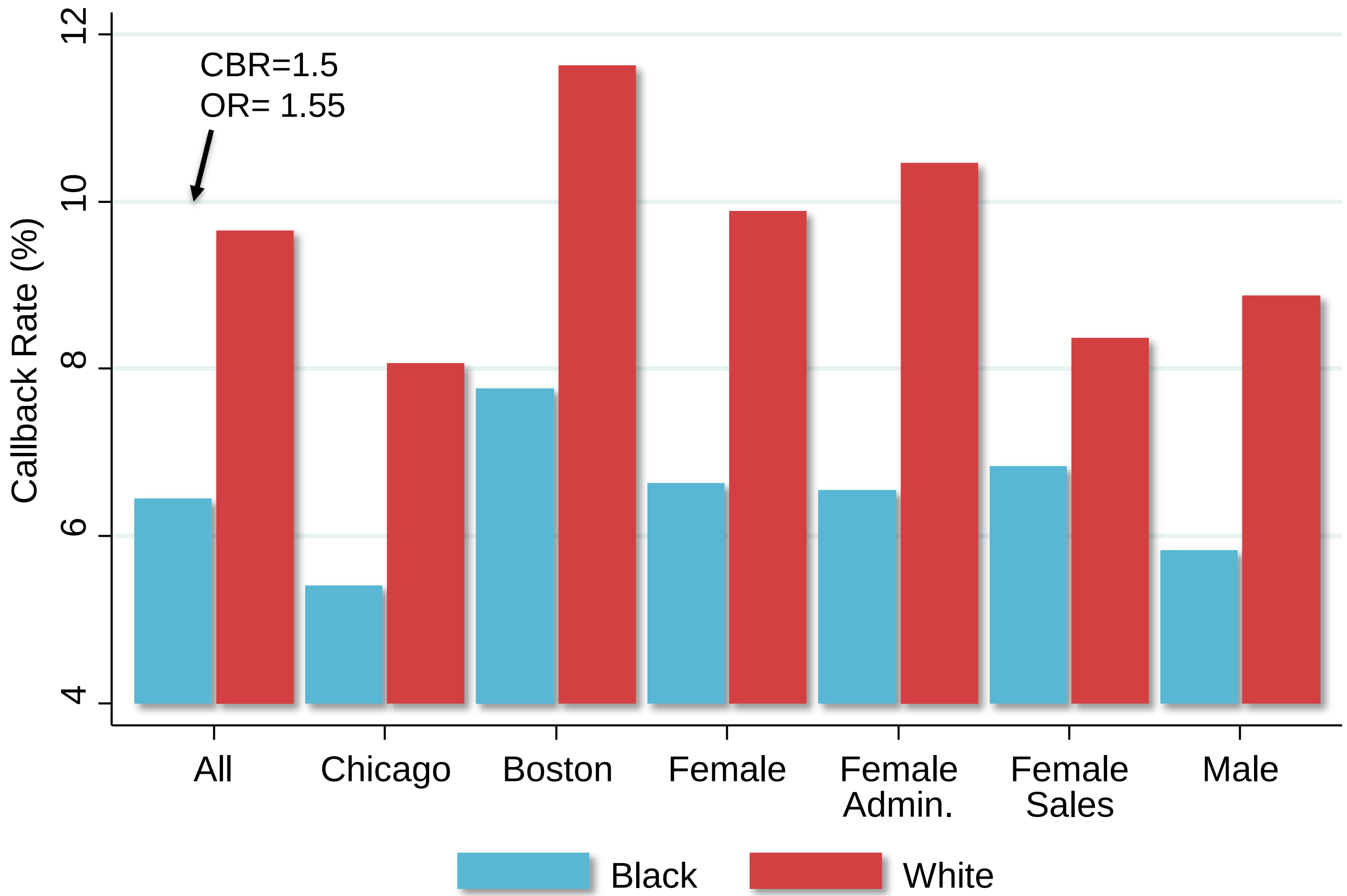
Recent contributions and findings

Some recent examples…

Bertrand and Mullainathan's (2004) send fictitious resumes to newspaper ads in Boston and Chicago

- Applicants' race is signaled with African-American (Lakisha/Jamal) or White (Emily/Greg) sounding names
- White names receive 50 percent more callbacks for interviews
- For White names, a higher quality resume elicits 30 percent more callbacks whereas for African Americans, it elicits a far smaller increase

Job Callback Rates by Race for Resumes with Otherwise Identical Credentials, US



Source: Bertrand and Mullainathan (AER 2004)

2 measures of discrimination

- D estimates measure differences in employers' callback across treatment conditions, i.e. for majority and minority applicants
- 2 measures:

1. **Callback Ratio** (CBR)→ Intuitive, most widely used

CBR= $\frac{N\ callbacks_{maj}/N\ applicants_{maj}}{N\ callbacks_{min}/N\ applicants_{min}}$

, where maj= Majority applicants; min=minority applicants

2. **Odds Ratio** (OR)→ Less intuitive but preferable for comparing D estimates across contexts with large differences in overall callback rates

OR= $\frac{P\ callback_{maj} \big/ 1-(P\ callbacks_{maj}}{P\ callback_{min} \big/ 1-(P\ callbacks_{min}}$

, where maj= Majority applicants, and min=minority applicants

Some recent examples…

Bertrand and Mullainathan's (2004) send fictitious resumes to newspaper ads in Boston and Chicago

- Applicants' race is signaled with African-American (Lakisha/Jamal) or White (Emily/Greg) sounding names
- White names receive 50 percent more callbacks for interviews
- For White names, a higher quality resume elicits 30 percent more callbacks whereas for African Americans, it elicits a far smaller increase

Riach and Rich (2006) send fictitious resumes to advertised positions in the English labour market to test for gender D

- They find D in sex-stereotyped occupations: against men in the 'female occupation' secretary, and against women in the 'male occupation' - engineer
- D against men also found in two 'mixed occupations' - trainee chartered accountant and computer analyst programmer

Baert et al. (2015) test the relationship between hiring D and labour market tightness at the level of the occupation

- No D in against candidates with foreign-sounding names in occupations for which vacancies are difficult to fill but sig D for occupations for which labour market tightness is low

Summary of findings on ethnic and gender D

(see meta-analyses by Zschirnt & Ruedin (2016) and Riach and Rich (2002))

- Widespread discrimination in hiring for ethnic and racial minority groups
  - Equivalent minority candidates need to send around 50 per cent more applications to be invited for an interview than majority candidates
- Strong gender discrimination in sex-stereotyped occupations
- Taste-based (or perhaps implicit) discrimination remains dominant for both ethnic and gender discrimination, although in some instances there is evidence that statistical discrimination also plays a role
- More extensive and standardised procedures of job application seem to reduce statistical discrimination (e.g. labour market in Germany vs other countries' labour markets) but not tasted-based/implicit discrimination
- Suggests importance of

  1) Employers' considerations about consumers' tastes (Baer & De Pauw 2014)

  2) LM tightness at the level of occupations (Baer et al 2015)
- More research is needed!!

5. The GEMM project

Our research in context

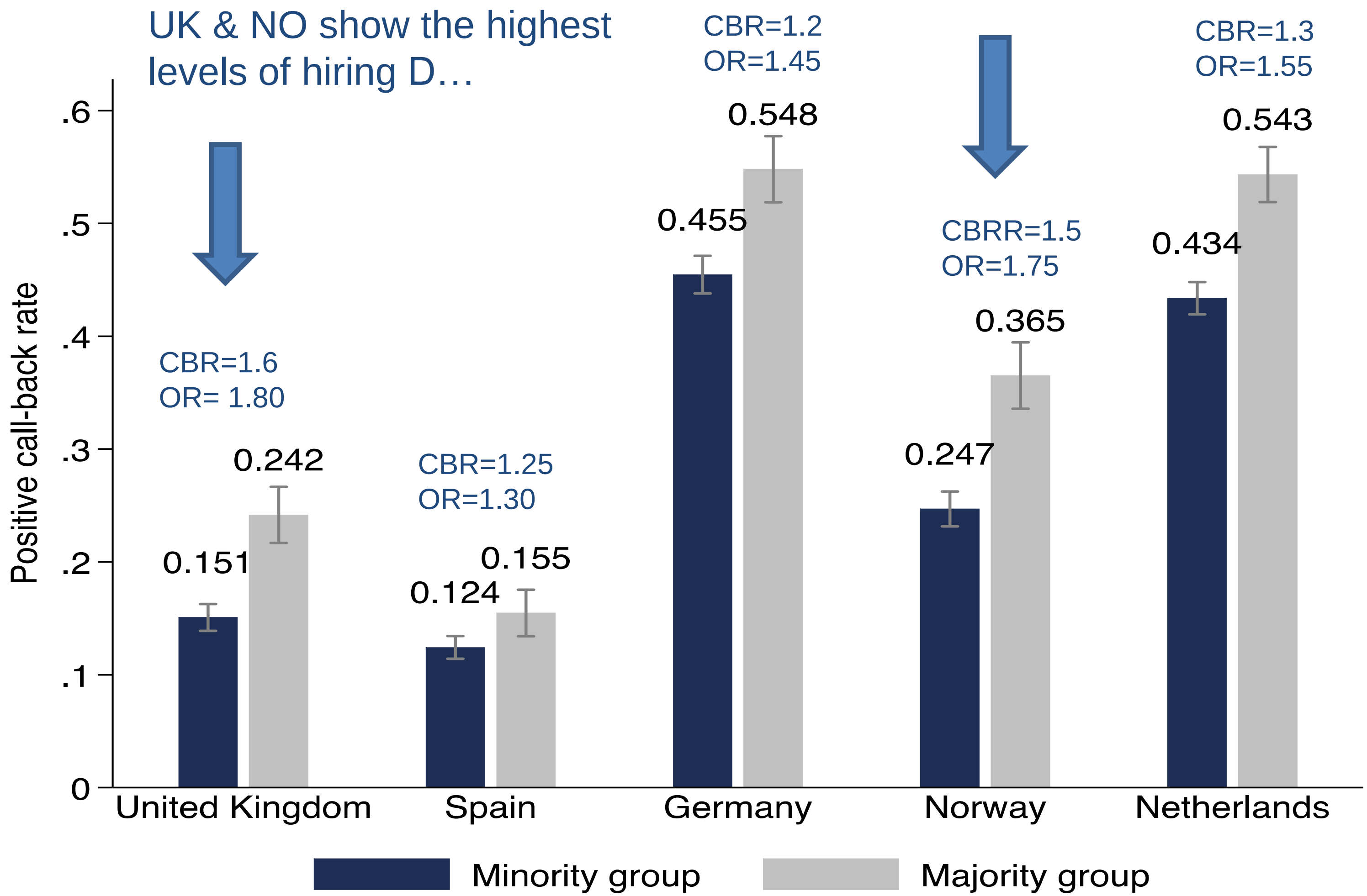
The GEMM study

www.gemm2020.eu

GEMM PROJECT
GROWTH, EQUAL OPPORTUNITIES, MIGRATION & MARKETS

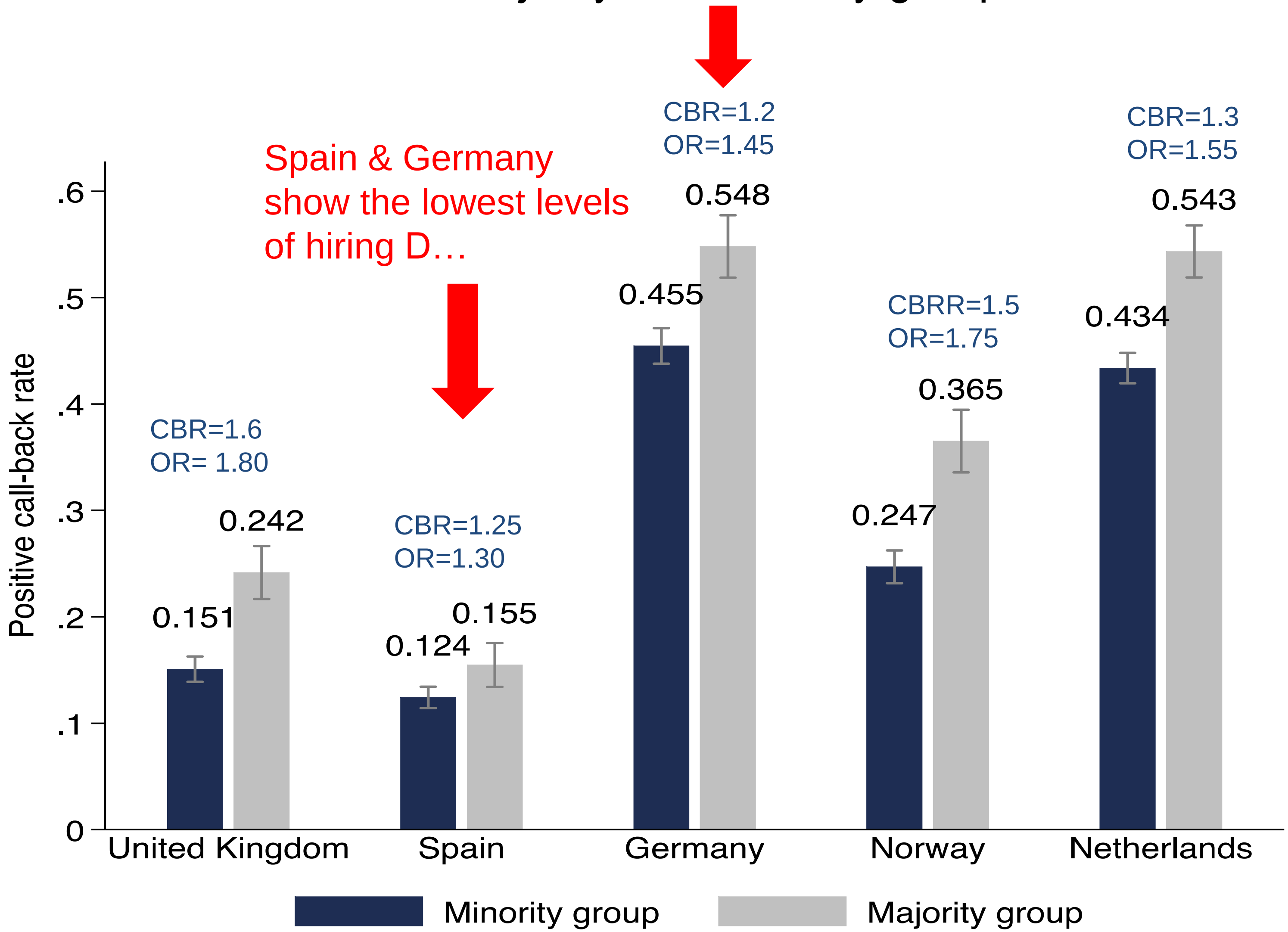
- The largest comparative field experiment on hiring D for children of migrants ever carried out in Europe (over 19,000 European firms targeted)
- Conducted **simultaneously** and with a **fully harmonised** design in **5 European countries**: Germany, the Netherlands, Norway, Spain and the UK over a period of approximately 18 months (ES→ Nov2016 until May 2018)
- Unique in scope, complexity and theoretical ambition
  - Involving 6 institutions: **Oxford** University, **Uc3m, WZB**, University of **Olso**, University of **Utrech,** University of **Amsterdam**
  - Large investments in human capital, infrastructure & IT development (e.g. D-Lab in uc3m)
  - Strict ethical clearance procedures (many bodies involved) https://www.d-labsite.com/ethics
  - A host of ancillary validity tests required (photograph ratings on attractiveness and friendliness, phenotype plausibility tests, name recognition surveys, extension of fieldwork, etc..)
- Unpaired fractional design; multiple treatments (origin, phenotype, religion, gender), 6 occupations, 53 different national origin groups

Callback rates for majority and minority groups, GEMM



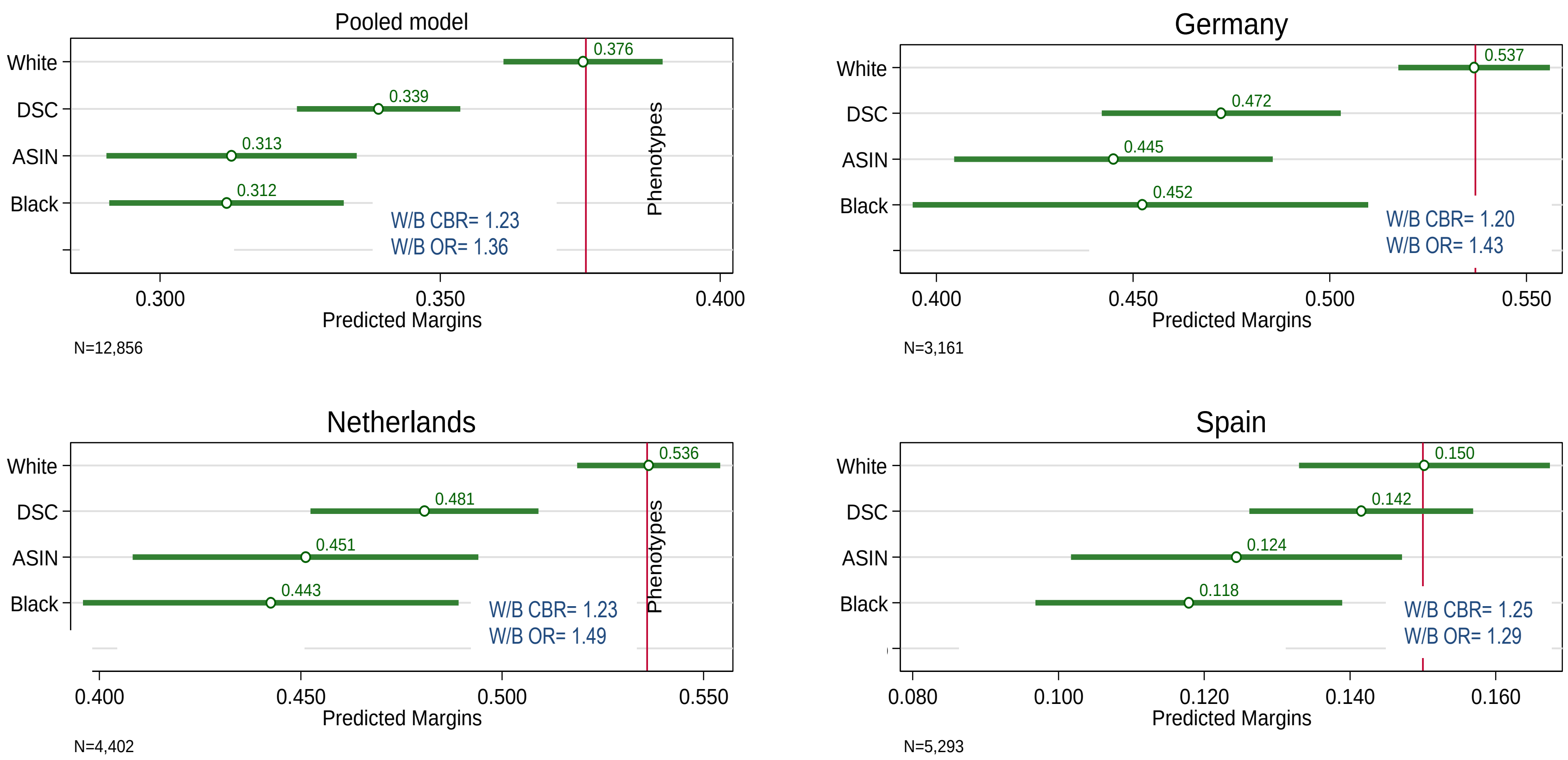
Result as in Feb 2018 N=19,181

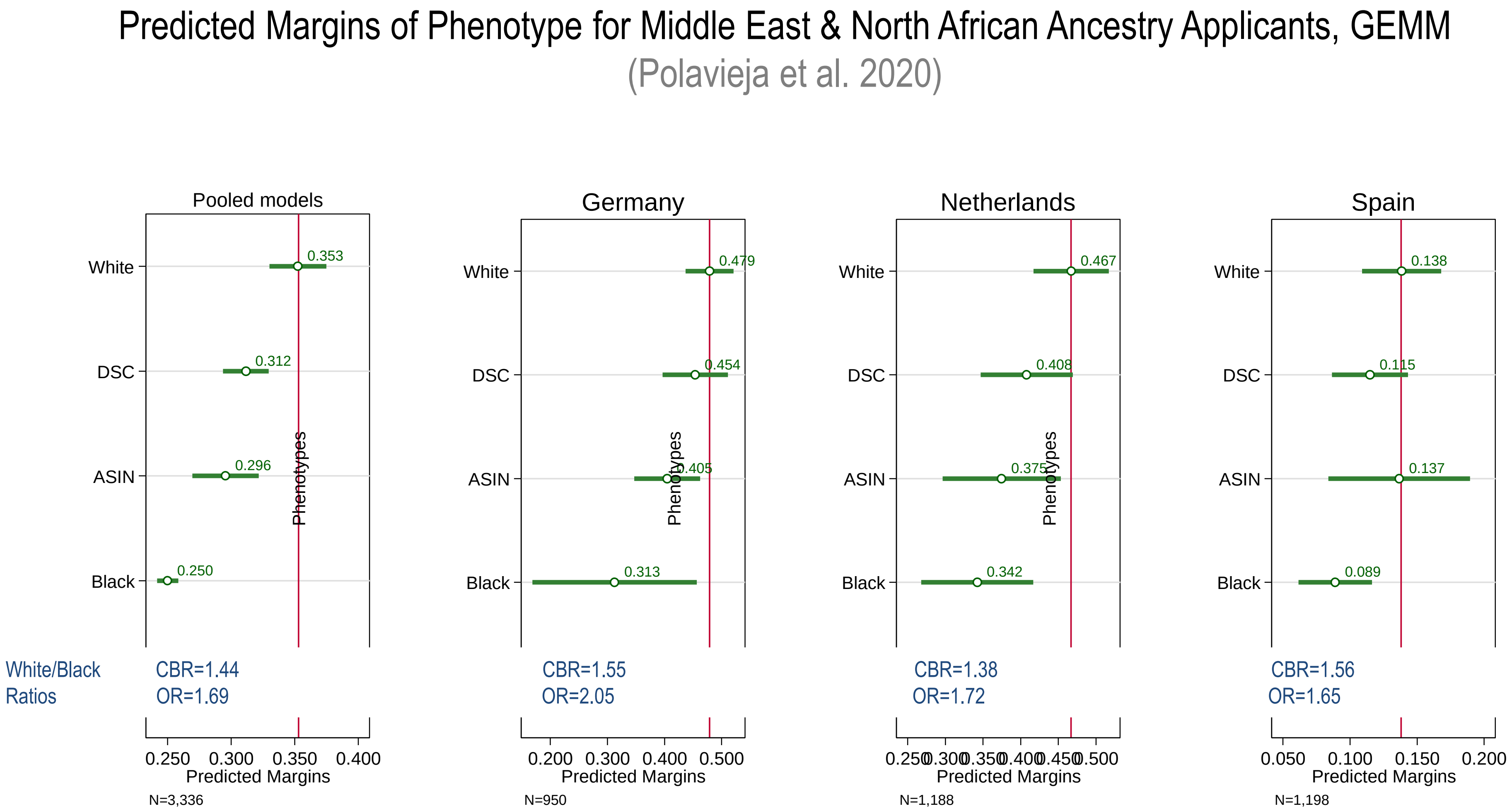
Callback rates for majority and minority groups, GEMM



Result as in Feb 2018 N=19,181
(Excluding data for Catalunya)

Cross-ancestry racial discrimination estimates, GEMM study
(Polavieja et al. 2020)




Predicted Margins of Phenotype for Middle East & North African Ancestry Applicants, GEMM
(Polavieja et al. 2020)
Pooled models
Germany
Netherlands
Spain
White
DSC
ASIN
Black
0.353
0.312
0.296
0.250
0.479
0.454
0.405
0.313
0.467
0.408
0.375
0.342
0.138
0.115
0.137
0.089
Phenotypes
White/Black
Ratios
CBR=1.44
OR=1.69
CBR=1.55
OR=2.05
CBR=1.38
OR=1.72
CBR=1.56
OR=1.65
Predicted Margins
N=3,336
N=950
N=1,188
N=1,198

Summary of findings, GEMM study

- Ethnic discrimination in all countries; substantial differences across countries
- Big difference across ethnicities → ethnic hierarchies (not shown)
- Differences across occupations (not shown)
- Evidence of phenotypic discrimination
  - Less phenotypic D in Spain than in Germany or the Netherlands
  - Suggestive of ethnicity*phenotype intersections (not shown)
- Discriminations against Muslim applicants (not shown)
- Evidence suggests discrimination against men! (not shown)
- Evidence predominantly in line with non-rational explanations of discrimination (taste-based or implicit D)

To see further research currently carried out at the D-Lab check:

https://www.d-labsite.com

That's all

Many thanks for your attention!

The D-Lab

Discrimination & Inequality Lab

uc3m | Universidad Carlos III de Madrid