

IZA DP No. 8584

**What Do Field Experiments of Discrimination in
Markets Tell Us?
A Meta Analysis of Studies Conducted since 2000**

Judith Rich

October 2014

What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000

Judith Rich

*University of Portsmouth
and IZA*

Discussion Paper No. 8584
October 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000

Sixty-seven field experiments of discrimination in markets conducted since 2000 across seventeen countries were surveyed. Significant and persistent discrimination was found on all bases in all markets. High levels of discrimination were recorded against ethnic groups, older workers, men applying to female-dominated jobs and homosexuals in labour markets. Minority applicants for housing needed to make many more enquiries to view properties. Geographical steering of African-Americans in US housing remained significant. Higher prices were quoted to minority applicants buying products. More information made no significant improvement to minority applicant outcomes. Clear evidence of statistical discrimination was found only in product markets.

JEL Classification: J7, C93

Keywords: field experiments, discrimination, survey, meta analysis

Corresponding author:

Judith Rich
Economics and Finance
University of Portsmouth Business School
Richmond Building
Portsmouth PO1 3DE
United Kingdom
E-mail: judy.rich@port.ac.uk

This paper provides a survey and assesses the more general significance of sixty-seven field experimental studies of discrimination in markets conducted over the last decade or more, enabling a comparison of their results by presenting them in the form of either net discrimination or callback rates. The survey documents the modifications and innovations which have been made to these types of field experiments over this period, so as to consider a broader range of bases of discrimination, such as gender, age and sexual orientation. A number of these recent studies have also attempted to gather more information on decision-making behaviour so as to confront the criticism that the experimental technique cannot capture all aspects of the enquirer (unobservable characteristics) which may influence the decision-maker. A mathematical test to obtain an unbiased estimate of discrimination suggested by Neumark (2012) is discussed briefly as it enables a more robust testing of experimental results. The survey also considers how the experimental technique has been adapted so as to distinguish between the diverse theories of discrimination. Overall, the pattern of discrimination documented in the survey reflects that found in an earlier survey of experiments conducted from 1966 to 2000 (Riach and Rich 2002). This is a disturbing finding, given that the later experiments were conducted in a period when the coverage of anti-discrimination legislation had been broadened.

The first section of the paper provides a general background to explain the experimental technique used in the various studies. The next three sections discuss the findings of the studies in labour, housing and product markets. Section five considers the Neumark test for the impact of unobservable characteristics of an applicant on the results from the studies, while section six examines whether the studies provide evidence of statistical or animus-based discrimination. The paper concludes by considering future directions of the experimental approach and the implications of this work for anti-discrimination legislation.

1. General Background to Field Experiments

Field experiments enable researchers to investigate whether discrimination exists within a market. Harrison and List (2004), Levitt and List (2009) and List and Rasul (2011) have discussed the current contribution and potential of the full range of field experiments, of which the studies discussed here are a subset. The type of field

experiments surveyed in this paper have used three approaches to undertake research on discrimination which can be conducted in labour, housing and product markets.

The first approach uses audits or in-person tests, where carefully matched testers (actors, but more usually undergraduate students) apply in person for jobs in the labour market, for rental accommodation or to purchase a house or flat in the housing market, and in the product market, to purchase goods or services. In applying for jobs, for example, the matching of applicants controls for all aspects of an individual that would affect their work productivity such as schooling, work experience, training, socio-economic background etc., as well as their personal appearance such as attractiveness, height, weight, etc. Testers are trained in job specific requirements and interview technique for labour market tests, in presentation and enquiry technique for housing market tests, and in presentation style and bargaining techniques for product market tests. Testers are then subjected to independent scrutiny for verification that they are of equivalent presentation which covers their personal attributes such as attractiveness and composure, as well as job specific attributes. The end result is that the matched pair of testers are to all intents and purposes presenting for a market transaction as identical except for the basis of discrimination which the researcher is investigating such as race, ethnicity, gender, age, sexual orientation or disability. In the second approach, in-person tests are conducted over the telephone where, once again, the paired applicants are matched. Aspects of the transaction process are carefully recorded by the testers and researchers in both these in-person approaches.

The third approach is to use correspondence or written tests where, in the labour market, curriculum vitae (résumés) are constructed which carefully match for socio-economic background, educational qualifications, work experience, marital status, age, hobbies and interests so that the applicants are identical except for the basis of discrimination which the researcher is investigating. Real CVs on publicly available websites are often used to create the experimental CVs. A computer programme can then be used to randomly generate the CVs by using a stored bank of necessary matched information (Banerjee, Bertrand, Datta and Mullainathan 2009; Bertrand and Mullainathan 2014; Carlsson and Eriksson 2014; Lahey 2008; Oreopoulos 2011). Lahey and Beasley (2009) provide a guide to this computerised randomisation process and creation of CVs. The CVs are often vetted by professionals in the occupations to

be tested, as well as by others involved in hiring, such as employment agencies. Usually entry-level jobs in manufacturing, construction and the service sectors, which are typically unskilled or semi-skilled, are applied for. This is because correspondence testing is difficult to apply to highly skilled jobs which require proof of identity or qualifications. Tests in the housing market involve sending a written enquiry, usually by email, in response to a flat or house advertised for rent or sale. The applications provide relevant information, for decision-making on the part of the landlord/owner/real estate manager, which is controlled so that the researcher can select the basis for investigation of any possible differential treatment of applicants. Responses from employers and real estate agents are carefully recorded.

Riach and Rich (2002) provided details of fifty studies of field experiments of discrimination conducted between 1966 and 2000, while Pager (2007) surveyed a small number of studies investigating race discrimination. Over the last twelve years, however, there have been many field experiments of discrimination, sixty-seven of which are surveyed in this paper: forty in labour, nineteen in housing and eight in product markets. In the labour market, the recent studies have conducted tests for discrimination in hiring on the basis of race, ethnicity, gender, age, disability, sexual orientation, obesity, caste and religion. In the housing market, work has been undertaken on offers to rent or view accommodation, testing for discrimination on the basis of race, ethnicity, sexual orientation and disability. In product markets, research has focused on the impact of race, ethnicity, gender, disability and age in pricing outcomes for transactions. Many of the recent studies have sought to modify the technique so as to gain a deeper understanding of differential treatment and the causes of discrimination, seeking, in particular, to distinguish between taste-based (animus) and statistical discrimination.

2. Field Experiments of the Labour Market, 2000 to 2012

By far the most researched area using the experimental technique is hiring in the labour market. Forty studies are surveyed in this section: thirty-five used the written approach and five used the in-person approach. These studies aimed to test for discrimination in hiring across the bases of race/ethnicity (twenty), gender (two), age (five), gender and age (two), disability (one), sexual orientation (seven), caste, religion (two) and obesity (one). Aspects of the technique are discussed in section 2.1, and the results are

presented by bases of race/ethnicity, gender, age, sexual orientation, caste and religion, obesity and disability in sections 2.2 to 2.8. Other aspects of the experiments and tests such as regression models, other dimensions of differential treatment and dishonest concealment are discussed in sections 2.9 to 2.11.

2.1 General Aspects of the Technique Used

Written matched pair tests were used in twenty-four studies, multiple CVs were used in nine studies and a further seven created multiple CVs but sent only one enquiry randomly selected from the CVs. Those sending multiple CVs aimed to test for the existence of discrimination against applicants from multiple ethnic backgrounds, multiple bases of discrimination such as the interaction of gender and age or the impact of particular information such as qualifications. The studies that tested for discrimination against multiple ethnic backgrounds in the one field experiment and sent two or more applications ensured that one applicant was always from the majority group when responding to a job vacancy. In the case of the immigrant/ethnic studies, all applicants were said to be fluent in the native language and educated in the country being tested. Once the set of CVs were constructed, the names to be used to signify race, ethnicity, gender, and caste were randomly assigned to a CV. The majority of studies created applicants who were young (early to mid-twenties), with the exception of the sexual orientation tests (late-twenties to early thirties) and the age tests. Over half of the written tests applied for jobs via email, the others used a combination of email and mail applications, or faxed applications. The matched pair tests alternated the CVs between the applicants, with the exceptions of the age and sexual orientation tests, and an equal proportion from each group made the first application. The applications for the matched set were sent at intervals usually of one hour to half a day, although a small number were sent with an interval of one to two days (such as Baert 2013; Baert, Cockx, Gheyle and Vandamme 2013; Kass and Manger 2011; Oreopoulos).

Personal approaches were used in five studies of race/ethnicity. Three of these were conducted by the ILO in France, Italy and Sweden and two in the USA by Pager (2003) and Pager, Western and Bonikowski (2009). In the ILO tests, matched pairs of testers were used where the minority applicant always made the first approach. In the USA tests, Pager (2003) used a matched pair of testers while Pager, Western and Bonikowski

(2009) used matched sets of three testers to investigate multiple aspects of hiring on the basis of race, ethnicity and criminal record.

The statistical significance of either responses or difference in responses were tested on the basis of race, ethnicity, age, gender, sexual orientation or disability, using chi-square/Fisher exact tests, binomial tests, bootstrap technique or test of difference for the null hypothesis that there is no difference in the callback rate between the groups. Statistical significance of responses was also tested for any impact from letter-type, tester pair, city, or other aspects controlled in the CVs.

Many studies provide full details of responses and the net discrimination level which is defined as the responses where only the majority applicant was invited to interview minus the responses where only the minority applicant was invited to interview. However, studies which sent multiple applications to a job advertisement report the callback rate, which is defined as the number of positive responses to an applicant (such as being asked to submit more information or invited to interview) as a proportion of total applications made by the applicant. Tables 1 to 9 here have been constructed to show either net discrimination levels or callback rates. Location and dates for the tests, the cities in which they took place, the occupations applied to (if data are published), the minority group tested and the researchers, are identified in the tables. If a study has not reported either of these then the study's findings are discussed but not presented in any table.

2.2 Race and Ethnicity Tests

Those studies reporting net discrimination levels conducted in Belgium, France, Germany, Greece, Ireland, Italy, Poland, Sweden and the USA are discussed in 2.2.1 and those studies reporting callback rates for Australia, Canada, China, France, Great Britain, Sweden and the USA are discussed in 2.2.2. Consistent features of all studies for race/ethnicity discrimination are identified at the end of the sub-sections.

2.2.1 Net discrimination levels

Table 1 reports net discrimination rates found in the ILO studies of race/ethnicity discrimination conducted in 2003 in Italy (Allasino, Reyneri, Venturini and Zincone 2004) and in 2005-2006 in France (Cediey and Foroni 2008) and Sweden (Attström

2007). These studies followed those the ILO conducted in the early to mid-1990s for all the major cities in Belgium, the Netherlands and Spain, and tested at all stages of the hiring process, using in-person tests as designed by Bovenkerk (1992). The initial approach was done by telephone, with additional approaches, in the case of the French tests, done by mailing applications as well as leaving CVs with an employer. The matched pairs of testers applied to low-skilled or semi-skilled jobs in the occupational categories of healthcare services and sales, hotel and restaurant, office and clerical, retail and trade, transport and warehousing. In the Swedish and French studies either all female or all male pairs applied to jobs. In Italy, all testers were young males.

Significant discrimination against racial and ethnic minorities was found in all stages of hiring except for the job offer stage in Sweden. The level of net discrimination recorded was very similar across the ILO studies: against Africans in France, between 41% and 54%; against Moroccans in Italy, 53%; against Middle-Eastern groups in Sweden, 45%. The levels of net discrimination against Moroccans were very similar to those found in the ILO studies conducted a decade earlier for Belgium (51%), the Netherlands (44%) and Spain (47%) (Riach and Rich 2002, p. F495). No significant differences were found by tester pair or letter type, city, or gender (Allasino *et al.* 2004, pp. 41, 51; Attström 2007, pp. 41, 45-6; Cediey and Foroni 2008, p. 100). These studies confirmed that the vast majority of discrimination detected in hiring occurred at the initial stage of vetting the curriculum vitae of applicants with 85.6% in the French study and 93.6% in the Italian. These levels are comparable to the earlier ILO studies in Belgium, the Netherlands and Spain which recorded an average of 90% (Riach and Rich 2002, p. F494). The findings provide strong support that written approaches are testing hiring decisions in the labour market where the majority of discriminatory practices occur.

Table 1: Results for Net Discrimination in the Labour Market Studies of Race/Ethnicity conducted by the ILO, 2003 to 2006

Country and Study	Year and Location of test	Minority	Neither invited	Usable tests	Equal treatment	Discrimination against minority	Discrimination against majority	Net Discrimination ^{1,2}	
								(1)	(2)
						No.	No.	No.	%
<i>France</i>									
Cediey and Forini (2008)	2005-6	North/							
Personal/in-person-interview offer	Lille	Sub-	60	50	7	33	10	23	46.0***
Telephone/In-person- interview offer	Lyon	Saharan	543	873	216	507	150	357	40.9***
Written- interview offer	Marseille	African	467	227	16	167	44	123	54.2***
In-person – job offer	Nantes		183	149	38	96	15	81	54.3***
	Paris								
	Strasbourg								
<i>Italy</i>									
Allasino, Reyneri, Venturini and Zincone (2004)	2003	Moroccan							
Telephone – interview offer	Naples		250	383	135	226	22	204	53.3***
In-person – job offer	Rome		59	123	46	16	2	14	11.4***
	Turin								
<i>Sweden</i>									
Attström (2007)	2005-6	Middle-							
Telephone – interview offer	Gothenburg	Eastern	804	417	55	274	88	186	44.6***
In-person – job offer	Malmö		11	33	17	7	9	-2	-6.1
	Stockholm								

1. Tests of statistical significance are indicated * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level.

2. A negative value indicates discrimination against the majority applicant

Table 2 reports net discrimination rates found in studies using matched pair written approaches conducted in Belgium (Baert *et al.*), Germany (Kass and Manger), Greece (Drydakis and Vlassis 2010), Ireland (McGinnity and Lunn 2011), Poland (Wysienska-Di Carlo and Karpinski 2014), Sweden (Carlsson and Rooth 2007) and the USA (Bertrand and Mullainathan, which is discussed in more detail in sub section 2.2.2). The experiments were conducted over the period 2005 to 2012, except for Bertrand and Mullainathan. Entry level jobs in occupations were applied for such as accountant, administrative assistant, cleaner, clerical officer, computer programmer, customer service, financial analyst, health care, information technology, marketing/sales, nurse, office assistant, receptionist, restaurant staff, sales, student internship, teacher, teaching assistant and vehicle driver. While varying the racial or ethnic name of the applicants in the set of CVs, a mix of either all male or all female applicants were sent to vacancies and, if an occupation was identified as dominated by one gender, the applications were made only by that gender. For example, Bertrand and Mullainathan used female names only for jobs in clerical areas and male and female names for administrative, customer service and sales jobs; Baert *et al.*, Carlsson and Rooth, Drydakis and Vlassis, and Kass and Manger used only male names; McGinnity *et al.* used only female names for jobs in administration and retail sales and only male names for jobs in accountancy.

The results in Table 2 show significant levels of net discrimination against race and ethnic groups: Africans in Ireland, 48.2%; African-Americans in the USA, 41.1%; Albanians in Greece, 24.2% to 65.7%; Asians in Ireland, 35%; those of Middle-Eastern background in Sweden, 28.9%; Turkish applicants in Belgium and Germany, 28% and 10.1% respectively; and Ukrainians and Vietnamese in Poland, 23% and 11% respectively.

Statistically significant levels of discrimination against Albanians were found in all four occupations tested in Greece. In Sweden, statistically significant levels of discrimination against Middle Eastern applicants were found in ten of the twelve occupations tested. Kass and Manger's finding of statistically significant discrimination against Turkish undergraduates in Germany relates to student internships which are positions that can increase the chances of an individual obtaining their first graduate job. Baert *et al.* grouped occupations into 'bottleneck and non-bottleneck', as defined by the employment website from which they obtained vacancies, to investigate

whether discrimination varied by labour market tightness. While they found a high level of net discrimination against Turkish applicants in the non-bottleneck group (50%), they found no discrimination in the occupations deemed bottleneck, ‘i.e. occupations which take a long time to fill’ (Baert *et al.* 2013, p. 3). Taking a long time to fill a vacancy, though, may reflect difficulty recruiting because the inherent nature of the job is unpleasant for a variety of factors and may not necessarily solely reflect a shortage of labour for jobs on an aggregate level. Some caution is thus needed in interpreting these results as indicating discrimination varied with aggregate labour market tightness.

Table 2: Results for Net Discrimination in Labour Market Studies of Race/Ethnicity Conducted 2001 to 2012

Country and Study	Year and Location of test	Minority	Occupation	Neither invited	Usable tests	Equal treatment	Discrimination against minority	Discrimination against majority	Net Discrimination ¹	
							(3)	(4)	(3)-(4)	(3)-(4)/(1)
							No.	No.	No.	%
<i>Belgium</i>										
Baert, Cockx, Gheyle and Vandamme (2013) Written	2011-2 Flanders	Turkish	Bottleneck	144	37	24	7	6	1	3.0
			Non-bottleneck	153	42	19	22	1	21	50.0***
			All	297	79	43	29	7	22	28.0***
<i>Germany</i>										
Kass and Manger (2011) Written	2007-8 2008-9 Germany	Turkish	Student internships	270	258	134	75	48	26	10.1*
<i>Greece</i>										
Drydakis and Vlassis (2010) Written	2006-7 Athens	Albanian	Industries	129	88	47	38	3	35	39.7**
			Office jobs	74	99	30	67	2	65	65.6**
			Restaurant workers	117	124	84	35	5	30	24.1**
			Shop sales	81	77	32	42	3	39	50.6**
			All occupations	401	388	193	182	13	169	43.5**
<i>Ireland</i>										
McGinnity and Lunn (2011) Written	2008 Dublin area	African		54	27	4	18	5	13	48.2**
			Asian	46	34	8	19	7	12	35.3*
			German	47	32	11	18	3	15	46.9***
			All	147	93	23	55	15	40	43.0***
<i>Poland</i>										
Wysienska-Di Carlo and Karpinski (2014) Written	2011-2 Poland	Ukrainian	All	571	148	62	60	26	34	23.0***
			Vietnamese	838	210	77	78	55	23	11.0**

1. Tests of statistical significance are indicated * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level

Table 2 continued: Results for Net Discrimination in Labour Market Studies of Race/Ethnicity Conducted 2001 to 2012

Country and Study	Year and Location of test	Minority	Occupation	Neither invited	Usable tests	Equal treatment	Discrimination against minority	Discrimination against majority	Net Discrimination ¹	
									(3)-(4)	(3)-(4)/(1)
					(1)	(2)	(3)	(4)	No.	%
							No.	No.	No.	
<i>Sweden</i>										
Carlsson and Rooth (2007)	2005-6	Middle	Accountant	155	31	10	14	7	7	22.6
Written	Gothenburg Stockholm	Eastern	Business sales assistant	164	114	57	39	18	21	18.4**
			Computer professionals	71	35	9	14	12	2	5.7
			Construction workers	44	20	7	12	1	11	55.0**
			Motor vehicle drivers	59	19	6	13	0	13	68.4**
			Nurses	95	55	30	20	5	15	27.3**
			Restaurant workers	128	12	3	8	1	7	58.3*
			Shop sales assistant	167	33	5	24	4	20	60.6**
			Teachers language	26	34	9	19	6	13	38.2**
			Teachers maths/science	16	26	17	7	2	5	19.2
			Teachers preschool	64	120	76	36	8	28	23.3**
			Teachers upper second	41	23	10	11	2	9	39.1*
			All occupations	1030	522	239	217	66	151	28.9**
			<i>USA</i>							
Bertrand and Mullainathan (2004)	2001-2	African	All	1103	220	63	111	46	65	41.1 ^{2***}
Written	Boston Chicago	American								

1. Tests of statistical significance are indicated * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level

2. Results constructed from Bertrand and Mullainathan (2004, Table 2, p. 999)

2.2.2 *Callback rates*

The research that reports callback rates highlights the average number of applications a job candidate needs to make to achieve a positive response from an employer. This research then investigates differences in callback rates between various applicants, rather than net discrimination. Table 3 reports callback rates for studies testing for discrimination on the basis of race/ethnicity, where statistical significance indicates the testing of the difference in callback rate between the minority and the majority applicant for equivalent CVs.

Four studies reporting callback rates have been conducted in the USA with tests spanning the period 2001 to 2010. Bertrand and Mullainathan's study, conducted from 2001 to 2002 in the US cities of Boston and Chicago, sent four applications to each job vacancy where one of the matched pair gave greater information on skills. They created two matched pairs of applicants, one pair with low-quality background and another pair with high-quality background: African-American/White low quality background; African-American/White high quality background. Table 3 indicates that Bertrand and Mullainathan found statistically significant lower callback rates to African-Americans. A white applicant needed to apply to ten (1/9.7) jobs to obtain a positive response from an employer, whereas an African-American applicant needed to apply to fifteen (1/6.5) or, fifty percent more jobs. Further, having a higher qualification made a significant improvement in callbacks for whites (a further 2.5 percent) but virtually no difference for African-Americans (a further 0.5 percent). The impact of qualifications on an applicant's ability to obtain a positive response from an employer can be investigated by designing the CVs as in the Bertrand and Mullainathan study. The first researchers to use written tests, Jowell and Prescott-Clarke (1970), had considered this issue. Of the total matched pair applications they sent in 1969, half gave the immigrant applicant a higher qualification than the British white applicant and the other half gave the immigrant and British applicants equivalent qualifications. Jowell and Prescott-Clarke found that although the higher qualified immigrant did receive more positive responses than the less qualified immigrant, there was little improvement when compared to the number of responses the white applicant received. A statistically significant difference in response rates remained, with the higher qualified immigrant receiving less positive responses as compared to the British applicant.

The subsequent tests conducted by Jacquemet and Yannelis (2012) in Chicago, sending three CVs to one hundred and ten job adverts, found the same callback ratio for African-American/White with the difference in callback rates statistically significant. They also found lower callback rates for foreign-sounding names which were similar to those for African-Americans. Jacquemet and Yannelis demonstrated that employers in the suburbs of Chicago accounted for the majority of the discrimination.

Pager (2003) and Pager *et al.* have developed the in-person approach in the USA by introducing further information such as an applicant having a criminal record. Pager (2003) used male testers and created two teams, one black, one white, where one tester in the pair was given a criminal record, and the criminal record was alternated within the pair each week. The testers applied to entry-level jobs in ten occupational categories in Milwaukee in 2001. The callback rate to the black without a criminal record was nearly eighteen percent lower than the callback rate to the white with a criminal record (14% and 17% respectively) and nearly sixty percent lower than to the white with no criminal record (14% and 34% respectively, with this difference highly statistically significant). Pager *et al.* also used two all male tester teams to apply for low-wage jobs in New York in 2004, one team consisting of a White, a Latino and a Black, the other a White with a criminal record, a Latino and a Black. Their findings, on the difference in callback rates reported in Table 3, reflected those of the earlier study by Pager in that blacks with no criminal record received fewer callbacks for jobs than whites with a criminal record.

Over the period 2006 to 2011, studies have tested for discrimination against multiple ethnic/race groups over the same time period in Australia (Booth, Leigh and Varganova 2012), Canada (Oreopolous), China (Maurer-Fazio 2012), France (Duguet, Leandri, L'Horty and Petit 2010), Great Britain (Wood, Hales, Purdon, Sejerson and Hayllar 2009) and Sweden (Bursell 2007). Table 3 provides full details of the race/ethnic groups investigated and the type of CVs constructed. In the Chinese and Swedish studies a matched-pair of applications was sent. Multiple CVs (numbers in brackets after the study) were sent to each vacancy in the Australian (four), Canadian (four), French (four) and Great British (three) studies. They applied to a range of entry-level jobs in accounting, IT, sales and service with either all male or all female

sets of applicants. The Australian and Swedish tests ensured all applicants had been educated in schools, and were proficient in the language, of the country of residence.

In 2006-7 in Sweden, Middle-Eastern/African applicants received statistically significant fewer callbacks than Swedish applicants (20.1% and 36.6% respectively, findings similar to Carlsson and Rooth who tested these groups in Sweden a year earlier). Of the fifteen occupations tested by Bursell only two, senior school teacher and receptionist, recorded differences in callback rates that were not statistically significant (Bursell 2007, Table 2, p. 15). In 2007 in Australia, lower callback rates were recorded for all the racial/ethnic groups as compared to White Australians with this difference statistically significant in the case of Chinese, Indigenous and Middle-Eastern groups. The callback rates of 22.0% for Middle Eastern and 21.0% for Chinese applicants are similar to levels found for these groups in Sweden. The tests in Great Britain over the period 2008-9 found statistically significant differences in callback rates against Chinese, Black African, Black Caribbean, Indian and Pakistani/Bangladeshi. No statistically significant differences in treatment were found by occupation or in public sector employment. In France, the tests conducted by Duguet *et al.* (2010) in Paris in 2006 found very low callback rates for Moroccan applicants, with the French applicant receiving, on average, a four times greater callback rate than those applicants with a Moroccan surname. The recent tests in China in 2011 for ethnic discrimination found in aggregate, significant lower callback rates for Mongolian, Tibetan and Uighur minorities as Table 3 indicates. Ethnic discrimination varied across the six cities with no discrimination (Hohhot and Urumqi), discrimination against only Tibetans (Nanjing), discrimination against Tibetan and Uighur (Kunming), discrimination against all groups (Chengdu and Shenzhen). Differences in callback rates between Han and ethnic minorities across the three occupations indicated that, where there was a shortage of labour (administrative assistant specialist positions), the Mongolian and Uighur applicants were treated similarly to Han applicants (Maurer-Fazio 2012, p. 10-11). Tibetan applicants were thus the least preferred of all ethnic groups.

Table 3: Results for Callback Rates in Labour Market Studies of Race/Ethnicity Conducted 2001 to 2011

Country and Study	Year and location of test	Basis of test	Type of CV	Ethnic or Minority group	Callback Rate ¹ %				
<i>Australia</i>									
Booth, Leigh and Varganova (2010) Written	2007 Sydney Melbourne Brisbane	Ethnicity		Chinese	21.0*				
				Indigenous	26.0*				
				Italian	32.0				
				Middle Eastern	22.0*				
				White Australian	35.0				
<i>Canada</i>									
Oreopoulos (2011) Written	2008 Toronto	Immigrant status	Canadian education and experience	Chinese	11.3*				
				Indian	11.5*				
				Pakistani	11.0*				
				White Canadian	16.0				
				CN/Cedu/Cexp ²	14.2				
				FN/Cedu/Cexp	10.0***				
				FN/Fedu/Cexp	9.3***				
FN/Fedu/C+Fexp	8.1***								
FN/Fedu/Fexp	6.0***								
<i>China</i>									
Maurer-Fazio (2012) Written	2011 Chengdu Hohhot Kunming Nanjing Shenzhen Urumqi	Ethnicity		Mongolian	2.6**				
				Tibetan	3.7**				
				Uighur	4.5**				
				Han	8.2				
				<i>France</i>					
				Duguet, Leandri, L'Horty and Petit (2010) Written	2006 Paris	Ethnicity		MMM ³	0.4**
FMM	1.8								
FMF	2.9								
FFF	7.3								
<i>Great Britain</i>									
Wood, Hales, Purdon, Sejerson and Hayllar (2009) Written	2008-9 Birmingham Bradford Bristol Glasgow Leeds London Manchester	Ethnicity/ Race		All Ethnic Groups	6.2 ⁴ ***				
				All White British	10.7				
				Black African	7.8***				
				White British	12.8				
				Black Caribbean	5.3***				
				White British	9.8				
				Chinese	5.6***				
				White British	10.4				
				Indian	5.9***				
				White British	11.2				
				Pakistani/Bangladeshi	6.4***				
				White British	9.5				

1. Statistical significance refers to the difference in callback rate to the majority applicant for equivalent CV: * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level

2. The full details are: Canadian name/ Canadian education/ Canadian experience; Foreign name/ Canadian education/ Canadian experience; Foreign name/ Foreign education/ Canadian experience; Foreign name/ Foreign education/ Canadian and Foreign experience; Foreign name/ Foreign education/ Foreign experience

3. Moroccan nationality/Moroccan surname/Moroccan first name; French nationality/Moroccan surname/Moroccan first name; French nationality/Moroccan surname/French first name; French nationality/French surname/French first name

4. The callback rates are reported matched first for ethnic groups in aggregate (all) as compared to white British in aggregate (all) and then each ethnic group as compared to the white British group.

Table 3 continued: Results for Callback Rates in Labour Market Studies of Race/Ethnicity Conducted 2001 to 2011

Country and Study	Year and location of test	Basis of test	Type of CV or Occupation	Ethnic or Minority group	Callback Rate ¹ %
<i>Sweden</i>					
Bursell (2007) Written	2006-7 Stockholm	Ethnicity	All	Middle- Eastern/African Swedish	20.1* 36.6
<i>USA</i>					
Bertrand and Mullainathan (2004) Written	2001-2 Boston Chicago	Race	High quality High quality Low quality Low quality All All	African-American White African-American White African-American White	6.7 11.0 6.2 8.5 6.5 9.7
Jacquemet and Yannelis (2012) Written	2009-10 Chicago	Ethnicity/ Race		African-American Foreign White	15.8 16.4 23.0
Pager (2003) In-person	2001 Milwaukee	Race	No criminal record No criminal record Criminal record Criminal record	Black White Black White	14.0* 34.0 5.0 17.0
Pager, Western and Bonikowski (2009) In-person	2004 New York	Race	No criminal record No criminal record No criminal record Criminal record Criminal record Criminal record	Black Latino White Black Latino White	15.2** 25.2 31.0 13.0 15.4 17.2

1. Statistical significance refers to the difference in callback rate to the majority applicant for equivalent CV: * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level

Oreopoulos investigated differential treatment of immigrants in Toronto, Canada in 2008 by constructing five CVs (detailed in Table 3) for applicants, controlling for experience and education as well as ethnicity. The foreign applicant was randomly assigned one of the three ethnic groups of Chinese, Indian or Pakistani and the set of applications, with variations in education and experience, were then sent to each job vacancy. Table 3 indicates lower callback rates for Chinese (11.3%), Indian (11.5%) and Pakistani (11.0%) immigrants as compared to Canadians (16.0%) were found, with these differences being statistically significant. The greater the foreign education and job experience of immigrants, the fewer the callbacks: immigrants with both Canadian education and experience received a callback rate over twice that of immigrants with both foreign education and experience.

These callback rate differences for the race/ethnicity studies can be summarised as follows. Dominant majority applicants received a greater level of positive responses to job enquiries as compared to the minority applicants in all countries. This meant that lower callback rates characterised outcomes for Chinese, Indigenous and Middle-Eastern applicants in Australia; Chinese, Indian and Pakistani applicants in Canada; Tibetan applicants in China; Moroccan applicants in France; black minority ethnic groups in Great Britain; African and Middle Eastern applicants in Sweden; and African-American applicants in the USA. Higher qualifications did not lead to a significant improvement in the callback rate to African-Americans in the USA. Immigrant groups were discriminated against despite being educated in schools, and proficient in the language, of the country of residence. In Canada, the level of discrimination against the immigrant increased, the greater the foreign education and job experience. The callback rates between the race/ethnic groups were not significantly different with the exception of the Chinese study.

Three consistent features can be identified from all the race/ethnicity studies surveyed in this section. First, Middle Eastern and Moroccan groups were discriminated against across Europe. Second, race/ethnic minority groups needed to apply for nearly twice as many jobs as the majority group to get a positive response. Third, race/ethnic groups were discriminated against in all types of occupations, although there was no pattern of higher rates of discrimination in jobs requiring customer contact.

2.3 Gender Tests

This section discusses those studies reporting net discrimination levels conducted in China, England and France and then the study reporting callback rates for Spain. Consistent features of all studies for gender discrimination are identified at the end of the sub-sections.

2.3.1 Net discrimination levels

Table 4 reports Chinese, English and French results on net discrimination in hiring on the basis of gender. The recent tests for discrimination on the basis of gender in China conducted by Zhou, Zhang and Song (2013) from December 2010 to May 2011, applied to approximately nineteen thousand vacancies online using the three largest Chinese job websites. Matched pairs of applicants, always living locally to the firm, applied to vacancies in four occupations (accounting, IT, marketing and secretary) across six major cities. In the English tests, Riach and Rich (2006a) used a matched pair of applicants and applied to vacancies for engineers, computer analyst programmers and secretaries as well as sending unsolicited enquiries for accountant positions. In the French tests, Petit (2007) sent three matched pair applicants (that is six applications) to each vacancy found on the largest national employment website (ANPE) for seven high skill and five low skill occupations in administrative and commercial areas often in the financial sector. The CVs were designed to capture effects from motherhood and age as well as gender as Table 4 indicates (the findings for age discrimination are discussed later).

In China, Zhou *et al.* distinguished responses to low-ranked and high-ranked jobs within the occupations. A range of net discrimination of -48.1% to +38.1% was recorded, all statistically significant except for low-ranked IT. Men were discriminated against in all jobs, low or high rank positions, with the exception of low-rank positions in accountancy. There was, however, no consistent pattern to this discrimination against men on the basis of rank: for IT jobs, discrimination was greater in high-rank positions; for marketing, discrimination was greater in low-rank positions. The Chinese tests recorded a level of net discrimination against males in secretarial positions of 40% and against females in accountant positions of 9%. In England net discrimination against men applying to secretarial jobs (-43%) was at a level nearly twice that experienced by women applying to engineer jobs (23%). The

discrimination against men recorded in the English tests for chartered accountant (-20%) and computer analyst programmer (-35%) can be compared to the Chinese tests for high-rank accountant (-19%) and high-rank IT (-12%). This interesting similarity could be indicative of occupational segregation arising from gender stereotyping. Consistent with previous studies of gender discrimination, the highest levels of net discrimination in both these studies were found for men applying to female-dominated jobs. These results for China and England reflect those of other studies on gender discrimination (Riach and Rich 2002, p. F504-F505). In France, a statistically significant level of net discrimination of 20% was found against young, single females in high skilled administrative jobs, in particular those jobs with long-term contracts.

2.3.2 *Callback rates*

Table 6 reports the Spanish results on callback rates in hiring on the basis of gender. Albert, Escot and Fernandez-Cornejo (2011) sent five male/female pairs of applicants to jobs in Madrid for accountant, administrative assistant, marketing and secretarial positions. While there was little difference in callback rates for females and males for the accountant, marketing and sales representative jobs, females were significantly favoured in the lower status, female-dominated positions. Callback ratios for female/males of 15.8/5.0 were recorded for secretary and 10.5/3.4 for administrative assistant. These results reflect the findings of the other studies on gender discrimination, that is, statistically significant discrimination against men in the female-dominated jobs which is of a much higher order than any found for the integrated occupations or against females applying to male-dominated jobs.

Two consistent features can be identified from all the gender studies surveyed in this section. First, men applying for strongly female-stereotyped jobs need to make between twice to three times as many applications as do women to receive a positive response for these jobs. Second, women applying to male-dominated jobs face lower levels of discrimination in comparison to men applying to female-dominated jobs.

Table 4: Results for Net Discrimination in the Labour Market Studies of Gender Conducted 2000 to 2011

Country and Study	Year and Location of test	Minority	Occupation	Neither invited	Usable tests	Equal treatment	Discrimination against female	Discrimination against male	Net Discrimination ^{1,2}	
									(1)	(2)
							No.	No.	No.	%
<i>China</i>										
Zhou, Zhang and Song (2013)	2010-11	Female	Accountant low-rank	3012	97	22	56	19	37	38.1***
Written	Beijing		Accountant high-rank	3102	101	32	25	44	-19	-18.8***
	Chengdu		Accountant – All	6114	198	54	81	63	18	9.1*
	Guangzhou		IT – low-rank	1238	119	50	31	38	-7	-5.9***
	Shanghai		IT high-rank	1134	215	110	40	65	-25	-11.6***
	Shenzhen		IT – All	2372	334	160	71	103	-32	-9.6***
	Wuhan		Marketing low-rank	2666	215	36	58	121	-63	-29.3***
			Marketing high-rank	1364	130	124	0	6	-6	-4.6**
			Marketing – All	4030	345	160	58	127	-69	-20.0***
			Secretary low-rank	2246	104	22	16	66	-50	-48.1***
			Secretary high-rank	2554	130	12	37	81	-44	-33.8***
		Secretary - All	4800	234	34	53	147	-94	-40.2***	
<i>England</i>										
Riach and Rich (2006a)	2000-1	Female	Chartered accountant	284	55	22	11	22	-11	-20.0*
Written	England		Computer analyst prog	96	34	14	4	16	-12	-35.3**
			Engineer	134	39	12	18	9	9	23.1*
			Secretary	180	51	13	8	30	-22	-43.1***
<i>France</i>										
Petit (2007)	2004-5	Female	Low skill	37	27	16	3	8	-5	-18.5
Written	Paris	Single 25	High skill	38	55	28	19	8	8	20.0**
		No child	All	75	82	44	22	16	6	7.3
		Female	Low skill	47	17	5	8	4	4	23.5
		Single 37	High skill	71	21	7	6	8	-2	-9.5
		No child	All	118	38	12	14	12	2	5.3
		Female	Low skill	55	5	2	2	1	1	20.0
		Married 37	High skill	69	24	8	9	7	2	8.3
		3 children	All	124	29	10	11	8	3	10.3

1. Tests of statistical significance are indicated * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level.

2. A negative value indicates net discrimination against the male applicant

2.4 Age Tests

The application of the experimental approach to testing for age discrimination has to confront the issue that experience/human capital will necessarily differ between older and younger workers. The original Jowell and Prescott Clarke technique was to control for characteristics demonstrating productivity/human capital, ensuring that the impact on the hiring decision of the basis being investigated, such as race/ethnicity, was measured as the only remaining difference between the applicants. Ahmed, Andersson and Hammarstedt (2012), Petit and Tinsley (2012) created younger and older applicants who had the same experience relevant to the job applied for, but gave the older applicant other experience in an area unrelated to the job applied to, such as employment in other occupations (Petit 2007, p. 376, footnote 3) or military service (Ahmed *et al.* 2012, p. 404). Riach and Rich created more realistic older applicants who had greater experience than the younger applicants, whilst ensuring equivalence on all other components of human capital. They argued that controlling for experience creates an unrealistic older applicant and introduces a problem for interpreting any findings of differential treatment (Riach and Rich, 2002, p. F505-507). That is, controlling experience to be the same for the younger and older applicants risks sending a signal that the older worker has not been successful in employment and therefore may be less productive, which would see an economically rational preference for the younger worker and differential treatment interpreted as evidence of statistical discrimination. Using realistic human capital differences, however, allows differential treatment favouring younger applicants with significantly less experience to be interpreted as evidence of prejudicial rather than cost-based behaviour, which would see an economically rational preference for older workers. Riach and Rich confronted ageism in their age tests by stating flexibility, agility, robustness and health in the interests section of the CVs (Riach and Rich 2010, p. 170-171). Lahey tested for age discrimination from age thirty-five which enabled her to use a realistic employment history of ten years for each candidate. She also confronted age stereotypes in the design of the CVs.

Studies testing for age discrimination in England, France, Spain, Sweden and the USA are discussed in the first sub-section below, followed by a discussion of separate studies in England, France and Spain that reported callback rates.

Consistent features identified from all studies for age discrimination are made at the end of the sub-sections.

2.4.1 *Net discrimination levels*

Studies from England, France, Spain and Sweden which examined age discrimination in the labour market and published net discrimination levels are reported in Table 5. A further study in the USA could not be reported in the table format. In England over the period 2002 to 2004, Riach and Rich (2010) sent a matched pair of females applicants, one twenty-one the other thirty-nine, to graduate positions and retail manager jobs and a matched pair of male applicants, one twenty-seven the other forty-seven, to apply for waiter positions. Riach and Rich (2006b, 2007) also conducted tests in Spain over 2005 and in France over 2006 enquiring for waiter positions using a matched pair of male applicants. Petit applied to a range of high-skill and low-skill jobs in 2004-5 as discussed earlier. In Sweden, Ahmed *et al.* (2012) used a matched pair of male applicants, both married with no children, one aged thirty-one and the other aged forty-six, to apply for jobs in restaurants and sales in Sweden. Lahey conducted tests of hiring discrimination on the basis of age in the USA over the period 2004-2005 in two cities, Boston and St. Petersburg. She sent a matched pair of female applicants to apply for entry level positions, where age was assigned by a range of thirty-five, forty-five, fifty, fifty-five and sixty-two (indicated by year of high school graduation on the CV).

The results from these studies on the basis of age were: in England a range of 28.8% (waiters) to 59.6% (graduates), with discrimination in favour of the older applicant for retail manager jobs; in France a level of 29.0% (waiter) and 20.0% to 23.5% (against females in low-skill jobs but not statistically significant); in Spain a level of 64.5% (waiter); in Sweden a range of 59.4% (restaurant worker) to 66.7% (sales assistant). These studies in England and Europe indicated very high levels of net discrimination against older applicants who were thirty-seven to forty-seven years old. In the USA, an average 44% difference in invitation to interview in favour of the younger (less than fifty) applicant was recorded.

2.4.2 *Callback rates*

Further studies from England, France and Spain which examined age discrimination in the labour market and published callback rates are reported in Table 6. Tinsley (2012) tested in England in 2011 by sending matched pairs of female younger/older applicants to over one thousand job vacancies for either personal assistant or bar work. Statistically significant differences in callback rates were found in favour of younger applicants in their mid-twenties compared to older applicants who were fifty or fifty-one, although the difference was greater for the speculative enquiries for bar work compared to applications for vacancies for personal assistant (respectively 2.25 times and 1.5 times greater positive response rate to the younger applicant). Petit found in French tests controlling for age that, compared to the older applicant, the younger applicant received a callback rate that was at least 1.6 times greater in low skilled jobs and at least 2.25 times greater in high skilled jobs. This indicated that a younger applicant needed to apply to approximately two jobs, whereas an older applicant needed to apply to approximately four, to receive a positive response. In Spain, Albert *et al.* found statistically significant differences in callback between the twenty-eight and thirty-eight year-olds. In aggregate, the callback ratio of younger/older workers of 10.9/6.1 indicated that older applicants had to apply to nearly twice as many jobs as younger applicants to get a positive response. These tests suggest that, in France and Spain, older applicants for secretarial and administrative positions need to make twice as many applications as younger applicants to obtain a positive response from an employer, whereas in England it was slightly lower at approximately one and a half. For waiter positions in France, Spain, and Sweden, an older applicant needed to apply to approximately three times more vacancies than a younger applicant to obtain a positive response from an employer.

One consistent finding is identified from these studies of age discrimination which is that an older worker needs to make between two to three times as many job applications as a young worker to get a positive response, where 'old' can range between late thirties to early fifties.

Table 5: Results for Net Discrimination in the Labour Market Studies of Age Conducted 2002 to 2011

Country and Study	Year and Location of test	Minority	Occupation	Neither invited	Usable tests	Equal treatment	Discrimination against older	Discrimination against younger	Net Discrimination ^{1, 2}	
					(1)	(2)	(3)	(4)	(3)-(4)	(3)-(4)/(1)
					No.	No.	No.	No.	No.	%
<i>England</i>										
Riach and Rich (2010)	2002-4	Older ³	Graduate	373	47	15	30	2	28	59.6 ^{***}
Written	England		Retail manager	273	27	3	8	16	-8	-29.6 [*]
	London		Waiter	390	80	11	46	23	23	28.8 ^{***}
			Waiter, London		22	3	17	2	15	68.2 ^{***}
<i>France</i>										
Riach and Rich (2006b)	2006	Older ⁴	Waiter	314	31	1	24	11	9	29.0 [*]
Written	Cities of France									
<i>Spain</i>										
Riach and Rich (2007)	2005	Older ⁴	Waiter	309	31	5	23	3	20	64.5 ^{***}
Written	Cities of Spain									
<i>Sweden</i>										
Ahmed, Andersson and Hammarstedt (2012)	2011	Older ⁵	Restaurant worker	231	32	7	22	3	19	59.4 ^{***}
Written	Sweden		Sales assistant	188	15	1	12	2	10	66.7 [*]
			All	419	47	8	34	5	29	61.7 ^{***}

1. Tests of statistical significance are indicated * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level

2. A negative value indicates net discrimination against the younger applicant

3. Older 39 for graduate and retail and 47 for waiter; younger 21 for graduate and retail manager and 27 for waiter

4. Older 47 and younger 27

5. Older 46 and younger 31

Table 6: Results for Callback Rates in the Labour Market Studies on Gender and Age Conducted 2005 to 2011

Country and Study	Year and location of test	Basis of test	Occupation	Applicant group	Number of CVs sent	Callback Rate ^{1,2} %
<i>Spain</i>						
Albert, Escot and Fernandez-Cornejo (2011) Written	2005-6 Madrid	Gender	Accountant	Female	830	7.0
				Male	830	6.4
			Accountant assistant	Female	990	11.2***
				Male	990	7.8
			Administrative assist	Female	880	10.5***
				Male	880	3.4
			Executive secretary	Female	400	15.8***
				Male	400	5.0
			Marketing technician	Female	1080	2.3
				Male	1080	2.3
			Sales rep	Female	1130	16.4
				Male	1130	17.0
			<i>England</i>			
Tinsley (2012) Written	2011 England London	Age	Personal assistant	Older, 50	250	7.3
				Younger, 24	250	16.4***
			Bar work	Older, 51	977	4.4
				Younger, 25	977	6.4***
<i>Spain</i>						
Albert, Escot and Fernandez-Cornejo (2011) Written	2005-6 Madrid	Age	Accountant	Older, 38	396	4.8
				Younger, 28	396	6.9
			Accountant assistant	Older, 38	792	6.1
				Younger, 28	792	12.1
			Administrative assist	Older, 38	704	4.0
				Younger, 28	704	8.7
			Executive secretary	Older, 38	320	6.9
				Younger, 28	320	12.2
			Marketing technician	Older, 38	864	1.0
				Younger, 28	864	3.1
			Sales rep	Older, 38	904	13.4
				Younger, 28	904	21.2
				All	Older, 38	4248
		Younger, 28	4248	10.9		

1. Difference in callback rate to the majority applicant for like CV: significant at the 0.05*; 0.01 **; 0.001***

2. It was not possible to calculate statistical significance for results for Spain in the format reported in this table

2.5 Sexual Orientation Tests

Researchers in this area acknowledge that construction of the CV to signal homosexuality may also signal other qualities which could be perceived as having a negative impact at work. This perception, in itself, could constitute prejudice against openly homosexual individuals (Drydakis 2011a, p. 95-96; Drydakis 2012, p. 8; Tilcsik 2011, p. 596-599). Tilcsik argues cogently that stereotypes of gays need to be confronted on the CV, mirroring the argument by Riach and Rich on the need to confront ageism in age tests (Tilcsik 2011, p. 618). In the interests section of the CV, sexual orientation was signalled by stating either membership or volunteer activity with a homosexual association. However, to control that this did not send a negative signal, Tilcsik used university career service advice to undergraduates who engage in volunteer activity and emphasised skills the applicant used in the volunteering position such as financial skills, rather than merely stating the activity (Tilcsik 2011, p. 597, footnote 3). He then ensured similar signalling was controlled for on the heterosexual applicant's CV. Weichselbaumer (2013) signalled being lesbian by either stating the applicant was single and a volunteer in a recognised, large gay and lesbian organisation, or partnered in a registered partnership. Baert avoided the uncertainty of either a negative/positive signal being attached to the signal of homosexuality/heterosexuality, indicating sexual orientation by explicitly stating the name of the spouse/registered partner and alternated this signal between the lesbian and heterosexual applicant.

Drydakis has been at the forefront of testing for discrimination on the basis of sexual orientation and his studies of Cyprus and Greece are discussed in the first sub-section together with that for Belgium by Baert. The second sub-section discusses studies conducted in Germany, Sweden and the US which reported callback rates. Reflection is again made of all the studies at the end of the sub-sections.

2.5.1 Net discrimination levels

Studies on discrimination in hiring on the basis of sexual orientation using matched pairs, conducted in Belgium, Cyprus and Greece are reported in Table 7. These studies all sent a matched pair of applicants (to advertised jobs) who were thirty years old in the case of Drydakis (2011a; 2012) and new graduates in their early twenties in the case of Tilcsik (2011). Baert was interested in exploring the impact of motherhood

and her female applicants were all married but assigned an age of twenty-five or thirty-seven and assigned no child or one child.

In Greece Drydakis (2009; 2011a) tested for discrimination based on sexual orientation by applying to around two thousand eight hundred jobs in Athens over the period 2006 to 2008. He found very high levels of discrimination against lesbians (54.2%) and gay men (64.3%). The tests conducted by Drydakis (2012) in Cyprus over the period 2010-11 constructed two matched pairs, one providing basic information on the applicant, the other more information in the way of work commitment and personality traits. Net discrimination on the basis of sexual orientation was found for both types of CVs. In the case of basic information, net discrimination levels of 76.7% against lesbians and 71.9% against gay men were recorded. In the case of the more informative CV emphasising positive work attributes, slightly higher net discrimination was recorded against lesbians and gay men of 77.3% and 72.1% respectively. These were higher levels of net discrimination than found in the Greek studies. In both countries, the levels varied across the four occupations tested but remained very high (occupational results were not reported in the case of Cyprus). In particular, the lowest levels of net discrimination were recorded for the jobs with customer contact - restaurant/café services (lesbian and gay men) and shop sales (gay men).

The only statistically significant net discrimination Baert found in the tests conducted over 2012-13 was against the heterosexual female manual worker (43%), a male-dominated occupation, which conformed with expected perceptions of lesbians, although she also found heterosexuals were preferred for the male-dominated occupation of engineer (Baert 2013, p. 13). Net discrimination of 8% against the heterosexual in aggregate, and over a range of 12% to 17% for three other occupations were not significant. While a preference was found for the younger lesbian (with or without a child), this was not statistically significant at conventional levels (Baert 2013, Table 1, p. 30).

2.5.2 *Callback rates*

Three studies published callback rates but only that for the USA by Tilcsik is reported in Table 8 (the others for Germany and Sweden sent a single enquiry and so are not

reported in table format). Tilcsik applied to nearly one thousand eight hundred jobs across seven states in 2005. In aggregate he found a statistically significant difference in callback rates favouring the heterosexual applicant (11.5% compared to 7.2%). This indicated that, to obtain a positive response, a gay applicant needed to make fifty percent more applications than a heterosexual applicant. The difference in callback rates was statistically significant in the states of Florida, Ohio and Texas but not in California, Nevada, Pennsylvania and New York.

Discrimination against gay men and lesbians was investigated in Sweden over a six month period from August 2010 (Ahmed, Andersson and Hammarstedt 2013) and against single and married lesbians in Germany over a fifteen month period from May 2011 (Weichselbaumer 2013). Both these studies sent only a single application, chosen at random from a carefully designed set of applications, to each identified vacancy. Weichselbaumer initially sent three applications (one single heterosexual, one married heterosexual and one lesbian) to each vacancy but stopped this design after three months when detection became an issue for reasons specific to the German environment. Ahmed *et al.* (2013) did not find any statistically significant differences in callback rates (from heterosexual applicants) in aggregate or across the ten occupational groups with the exception of lesbians in cleaner positions, although a less favourable treatment of lesbians in female-dominated positions and gay men in male-dominated jobs was indicated. Weichselbaumer found statistically significant lower call back rates to lesbians, single or partnered, in Munich but not in Berlin (Weichselbaumer 2013, pp. 16-19).

All these studies investigating discrimination in hiring on the basis of sexual orientation indicate levels of net discrimination on the basis of sexual orientation which are as high or higher than those found in the age and some of the race/ethnicity tests, making them amongst the highest recorded for labour market studies of discrimination. No consistent pattern of discrimination by occupation emerged, such as in favour of lesbians for male-dominated jobs or against lesbian and gay men in customer contact jobs.

2.6 Caste and Religion Tests

In the first study of discrimination on the basis of either caste or religion, Banerjee *et al.* applied via email to software and call centre jobs identified in newspapers or online in Delhi, India in 2004. They tested for discrimination on the basis of social class by using surnames to distinguish three caste groups: upper, scheduled and 'other backward'. The surnames used to identify caste were carefully checked to ensure they were recognised as the correct caste. They created CVs for male and female applicants with both high and low quality backgrounds. They tested for discrimination on the basis of religion by using both first names and surnames to identify Muslim applicants. Siddique (2011) also studied caste-based discrimination in India over 2006 using a matched pair to apply, via email, to over five hundred vacancies advertised online for customer service and office/administration jobs.

Table 8 reports the callback rates for these caste and religion tests. Banerjee *et al.* found no statistically significant differences (at the conventional five percent level) in callback rates in either of the occupation groups on the basis of caste or religion. Siddique's study did identify statistically significant callback rate differences between upper caste and lower caste applicants and the responses by matched pair found a statistically significant net discrimination rate of 18.8% against low caste applicants.

2.7 Obesity Tests

Swedish results on discrimination on the basis of obesity are reported in Table 9. Rooth (2009) conducted tests over 2006 in Gothenburg and Stockholm, applying to jobs in seven occupations using male and female matched pair applicants. He signalled obesity by attaching to one CV a digitally enhanced photo of an individual who was made to look obese, while using the normal photo of the same individual, who had been judged to be attractive, for the matched CV. Statistically significant discrimination against obese female (16.8%) and male (15.2%) applicants was recorded which varied across the seven occupations included in the tests. The statistically significant discrimination against men occurred in three job areas requiring customer contact (business and sales assistants and restaurant workers) while for females it was in the three occupations of accountant, pre-school teacher and restaurant worker.

2.8 Disability Tests

Hiring discrimination on the basis of disability has been investigated in Scotland by MacRae and Lavery (2006). This consisted of matched pairs tests in which CVs were sent to one hundred and twenty advertised job vacancies in 2006. While full details of the responses have not been published, they report that the able-bodied applicant received sixty-nine percent of the invitations to interview, that is, over twice as many as the disabled applicant (MacRae and Lavery 2006, p. 7).

Summarising these studies (that is, other than for race/ethnicity), discrimination was found on the basis of age in England, France, Spain and Sweden; on the basis of gender in England and France; on the basis of sexual orientation in Cyprus, Greece and the USA; and in a less substantive group of studies, on the basis of caste in India, disability in Scotland and obesity in Sweden.

2.9 Regression Analysis of Responses

A detailed discussion of the regression models developed in a number of the studies covered is not possible within this survey although some basic observations can be made. Many of the researchers obtained information related to hiring such as firm size, location, gender or race/ethnicity of recruiters, the proportion of the minority group living in the area of the firm location, to test the impact of these together with the applicants' characteristics on the probability of a positive response. A range of hypotheses were investigated such as: firms with a stated equal employment opportunity policy would be more likely to have goals of ensuring equality of opportunity and therefore less likely to discriminate; large firms would be less likely to discriminate because they would be more likely to have personnel trained in equality and diversity awareness; recruiters will favour applicants from their own group, so, for example, male recruiters would be more likely to discriminate against females; firms located in areas with a high proportion of minority group residents would be less likely to discriminate against this group. Apart from confirming the statistical significance of the bases of discrimination they were testing, no patterns of impact from these other variables on the probability of a positive response emerged, with few finding any of these variables statistically significant at 5% level or less (see for example, Bertand and Mullainathan 2004, Table 7 p.1004; Carlsson and Rooth 2007, Table 3, p. 724; Kass and Manger 2011, Table 1, p. 8).

Table 7: Results for Net Discrimination in the Labour Market Studies of Sexual Orientation Conducted 2006 to 2013

Country and Study	Year and Location of test	Minority	Occupation	Neither invited	Usable tests	Equal treatment	Discrimination against minority	Discrimination against majority	Net Discrimination ^{1,2}		
									(1)	(2)	(3)
							No.	No.	No.	%	
<i>Belgium</i>											
Baert (2013) Written	2012-13 Flanders	Lesbian	Engineer	72	24	10	9	5	4	17.0	
			Ergotherapist	62	34	23	3	8	-5	-15.0	
			Management assistant	83	13	7	2	4	-2	-15.0	
			Manual worker	75	21	6	3	12	-9	-43.0*	
			Nanny	70	26	13	8	5	3	12.0	
			Secretary	80	16	8	3	5	-2	-13.0	
			All occupations	442	134	67	28	39	-11	-8.0	
<i>Cyprus</i>											
Drydakis (2012) Written	2010-11 Larnaca, Limassol Nicosia Paphos	Gay	All occupations								
			Less informative	564	652	163	479	19	469	71.9**	
			More informative	499	701	181	513	7	506	72.1**	
		Lesbian	All occupations								
			Basic informative	519	521	109	406	6	400	76.7**	
			More informative	467	601	120	473	8	465	77.3**	
<i>Greece</i>											
Drydakis (2011) Written	2007-8 Athens	Lesbian	Industrial jobs	176	135	56	77	2	75	55.6**	
			Office jobs	153	123	49	69	5	64	52.0**	
			Restaurant/café services	106	150	74	73	3	70	46.7**	
			Shop sales	89	125	41	82	2	80	64.0**	
			All occupations	524	533	220	301	12	289	54.2**	
Drydakis (2009) Written	2006-7 Athens	Gay	Industrial jobs	215	131	40	89	2	87	66.4**	
			Office jobs	268	187	46	140	1	139	74.3**	
			Restaurant/café services	342	169	57	110	2	108	63.9**	
			Shop sales	193	209	87	118	4	114	54.5**	
			All occupations	1018	696	230	457	9	448	64.3**	

1. Tests of statistical significance are indicated * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level

2. A negative value indicates discrimination against the heterosexual applicant

Table 8: Results for Callback Rates in the Labour Market Studies of Sexual Orientation, Caste and Religion Conducted 2004 to 2006

Country and Study	Year and location of test	Basis of test	Occupation	Applicant group	Number of CVs sent	Callback Rate ¹ %
<i>USA</i>						
Tilcsik (2011) Written	2005 California Florida New York Nevada, Ohio Pennsylvania	Sexual Orientation		Gay Heterosexual male	1776	7.2 ^{***} 11.5
<i>India</i>						
Banerjee, Bertrand, Dutta and Mullainathan (2009)	2004 Delhi	Caste	Software IT	Upper Caste Scheduled Caste Upper Caste Other Backward Caste	459 465 323 327	4.8 ² 6.0 4.6 4.3
			Call Centre	Upper Caste Scheduled Caste Upper Caste Other Backward Caste	195 205 149 169	20.0 14.6 20.8 13.0
		Religion	Software IT	Upper Caste Muslim	418 415	5.3 5.8
			Call Centre	Upper Caste Muslim	96 96	21.9 20.8
Siddique (2011)	2006 India	Caste	All	Upper Caste Lower Caste	523 523	16.1 13.6

1. Difference in callback rate to the majority applicant for like CV: significant at the 0.05*; 0.01 **; 0.001***

2. All results for Banerjee *et al.* (2009) are from Tables 2a, 2b and 2c, p. 21

2.10 *Other Dimensions of Differential Treatment*

Evidence of other dimensions of differential treatment were found in many of the labour market studies. Using an innovative approach, Drydakis (2009, 2011a, 2012) and Drydakis and Vlassis (2010) gathered further information on wage offers and insurance cover by having testers available to receive telephone calls from employers wanting to arrange an interview. Albanians in Greece were offered on average eleven percent lower wages and had a twenty-five percent lower probability of getting insurance cover from an employer. Drydakis found statistically significant lower wage offers to gay men and lesbians in all his studies of sexual orientation. Pager *et al.* also document similar types of differential treatment and developed a useful typology of the steps of hiring to categorise the treatment they recorded in the US. Step one is the 'initial point of contact' where categorical exclusion can occur; step two is 'assessment of qualifications' where shifting standards can arise; step three is 'job placement' where channelling can occur such as suggesting other (lower status, lower pay) jobs are available (Pager *et al.*, 2009, p. 792). Earlier studies in the labour market also found differential treatment occurred such as minority applicants being treated less courteously at interviews or being offered lower wages (Riach and Rich 2000, p. F509).

2.11 *Dishonest Concealment of Rejection*

Dishonest concealment of rejection of applicants has been reported by Riach and Rich (2006, p. 14-15) and the ILO studies (Allasino *et al.* 2004, pp. 43-47; Cediey and Foroni 2008, pp. 78-86). These type of responses tell the minority applicant either that they have been rejected because the job vacancy has been filled (only to contact the majority applicant the following day to ask them to make a time for interview), or that they are over-qualified for the job (at the same time inviting the equivalently qualified majority applicant to an interview). Whilst it is illegal in many countries for an employer to specify the desired race, gender, or age of a job applicant in a job advertisement, this dishonest concealment of rejection, together with the significant differences in callback rates recorded on all these bases, alludes to its covert existence.

Table 9: Results of Net Discrimination in the Labour Market Study of Obesity Discrimination Conducted 2006

Country and Study	Year and Location of test	Minority	Occupation	Neither invited	Usable tests	Equal treatment	Discrimination against minority	Discrimination against majority	Net Discrimination ^{1,2}	
					(1)	(2)	(3)	(4)	(3)-(4)	(3)-(4)/(1)
					No.	No.	No.	No.	No.	%
<i>Sweden</i>										
Rooth (2009)	2006	Obese	Accountant	47	6	1	3	2	1	16.7
Written	Gothenburg, Stockholm	Male	Business sales assist	37	43	28	13	2	11	25.6**
			Computer professional	23	13	6	5	2	3	23.1
			Nurse	24	55	42	6	7	-1	-1.8
			Restaurant worker	66	27	9	13	5	8	29.6*
			Shop sales assistant	93	18	9	7	2	5	27.8
			Teacher pre-school	27	48	35	9	4	5	10.4
			All occupations	317	210	130	56	24	32	15.2**
		Female	Accountant	34	16	7	8	1	7	43.8*
			Business sales assist	57	56	27	18	11	7	12.5
			Computer professional	24	13	6	2	5	-3	-23.1
			Nurse	20	25	19	2	4	-2	-8.0
			Restaurant worker	42	29	11	17	1	16	55.2**
			Shop sales assistant	30	3	2	1	0	1	33.3
			Teacher pre-school	42	67	46	15	6	9	13.4*
All occupations	249	209	118	63	28	35	16.8**			

1. Tests of statistical significance are indicated * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level

2. A negative value indicates discrimination against the normal weight applicant

3. Field experiments of the housing market 2000 to 2012

There have been fewer studies of discrimination in housing markets as compared to the number conducted in labour markets since 2000. Nineteen studies are surveyed with twelve using the written approach and seven using the in-person approach. The studies aimed to test for discrimination in access to housing across the bases of race/ethnicity (fourteen), disability (two), and sexual orientation (three). Aspects of the technique are discussed in section 3.1 and the results of the studies are considered by bases in sections 3.2 to 3.4. Other dimensions of differential treatment are discussed in section 3.5 and the dishonest concealment of rejection in section 3.6.

3.1 General Aspects of the Technique Used

The ability to use written approaches for housing tests has been facilitated by the development of the internet and the posting of rental adverts on websites which can be responded to via email. Prior to this, the in-person approach to real estate agents and landlords was the more successful mode of acquiring responses to (fictitious) enquiries for housing for rent or purchase. Studies using the written approach have tested for discrimination on the basis of race or ethnicity in Italy, Norway, Spain, Sweden and the USA and on the basis of sexual orientation in Sweden. Seven studies using the written approach sent a single enquiry chosen at random from a set of profiles while the other five studies sent matched-pairs. Many studies also tested the impact of providing further information on the applicant. The in-person approach to test for discrimination on the basis of race/ethnicity was used in Greece and the USA, of sexual orientation in the USA, and of disability, again, in the USA.

The statistical significance of responses and difference in responses was tested on the basis of race, ethnicity, sexual orientation or disability, using chi-square/Fisher exact tests, binomial tests, or test of difference for the null hypothesis that there is no difference in the callback rate between the groups. Statistical significance of responses was also tested for impact from tester pair, city, or other aspects controlled in the CVs. Full details of responses and the net discrimination level are provided in many studies and these are reported in Tables 10 and 11. Location and dates for the tests, the cities in which they took place, the minority group tested, the type of profile/information sent and the researchers, are also identified in the tables. However, not all the studies provided responses that could be presented in this format. This was the case for

experiments which sent single enquiries to a rental or sales advertisement. This was also the case for the audits of race/ethnicity and sexual orientation conducted by fair housing centres in Boston, Michigan, New Orleans, Seattle and Vermont which are discussed in section 3.5.

3.2 Race/Ethnicity Tests

Section 3.2.1 reports the responses from the studies conducted in Greece, Spain, Sweden and the USA, which sent matched-pairs to make enquiries either in-person or written, while section 3.2.2 reports the responses from those studies conducted in Italy, Norway, Sweden and the USA, which sent a single written enquiry.

3.2.1 Matched pair tests

In the USA, the Urban Institute conducted three separate studies of race/ethnicity using in-person matched pairs of testers to enquire about renting or purchasing flats or houses. The first phase in 2000 comprised over one thousand audits in twenty cities for African-American/White pairs and over seven hundred audits in ten cities for Hispanic/White pairs (Turner, Ross, Galster and Yinger 2002). Phase two conducted the first tests of discrimination against Asian and Pacific Islanders in housing, undertaking nearly nine hundred audits across eleven cities over 2000 to 2001 (Turner and Ross 2003a). Phase three, in 2002, involved four hundred audits in eight cities of three states using a matched pair of Native American and White testers (Turner and Ross 2003b). In contrast, testing by Hanson and Hawley (2011) sent email enquiries from a matched African-American and White pair with information on social class (randomly assigned from the six profiles created), to over three thousand online adverts for rental properties across ten major cities of the USA in 2009. Social class was signalled by the expression used in the enquiry.

Drydakis (2011b) used the in-person approach in Greece, where matched pairs of Albanian and Greek testers made telephone enquiries to adverts in newspapers, the major source of housing availability. He conducted close to five thousand tests over the period 2006 and 2007, with female testers enquiring to rent accommodation in working class, middle class and upper class areas of Athens. Pilot tests and surveys were conducted to ensure the ethnicity of testers could be correctly identified by

accent. In Spain, Bosch, Carnero and Farré (2010) conducted around two thousand matched enquiries in 2008 and 2009, testing for differential treatment between native Spaniards and Moroccan immigrants. Applicants were assigned common Spanish or Moroccan names to signal ethnicity and the matched batch of enquiries were either all females or all males to avoid conflating observations with gender discrimination. The applications were constructed to investigate whether providing a greater amount of information to landlords, indicating reliability and income of the applicant, had an impact on the callback rate. An applicant was given either standard/basic information or high quality/more information such as professional background (details of the profiles are provided in Table 10). In Sweden, Ahmed and Hammarstedt (2008) investigated ethnic discrimination in housing over the year 2007 by sending three applications to each advert for a rental apartment: Swedish male/ Swedish female/ Middle-Eastern male.

The results found for net discrimination in these studies are reported in Table 10. In reporting the findings for the (US) Urban Institute studies, the hierarchical measure of net discrimination is used. This measure is a summary statistic that captures differential treatment across a broad range of indicators (see Turner and Ross 2003a, p. 2-18-2-22). The tests in 2000 for housing sales, found that African-Americans were discriminated against at a statistically significant level but Hispanics were not, whereas both groups were discriminated against in rental enquiries. Further, Turner *et al.* (2002) found the incidence of geographic steering of African-Americans was statistically significant and had increased from 1989 to 2000. They concluded that “black homeowners may be more likely to receive favourable treatment on housing availability and housing inspections than they were a decade ago, but they are apparently also more likely to be steered to neighbourhoods that are more predominately black than those recommended to comparable white homebuyers” (Turner *et al.* 2002, p. 3-12). This may go some way towards explaining why residential racial segregation remains a continuing feature of cities in the USA (Bayer, McMillan and Rueben 2001; Seitles 1998). Statistically significant discrimination was found against Asian and Pacific Islanders in housing sales but not for rental properties, in the 2000-2001 Phase two tests. In the 2002 Phase three tests, statistically significant discrimination was found against Native Americans in rental

properties, although no significant difference was found in the limited housing sales tests.

Hansen and Hawley's findings are broadly consistent with those of the Urban Institute. Table 10 shows they found statistically significant net discrimination in aggregate against African-Americans of 4.54% which was explained by low class enquiries, recording no significant difference for high class enquiries. Hansen and Hawley also shed some light on geographic steering by grouping responses by the percentage of white residents in the neighbourhood. They did this to investigate whether "neighborhoods exhibit tipping behavior when the minority share is between 5% and 20% - at that point white residents move almost entirely out of neighborhoods. If this is true, and landlords want to prevent neighborhood tipping, they may be more likely to discriminate in neighborhoods where the share of non-whites is between 5% and 20%, than in other neighborhoods" (Hansen and Hawley 2011, p. 107). Table 10 indicates Hansen and Hawley found the highest levels of discrimination in these 'tipping neighbourhoods'. This provides evidence of geographic steering of African-Americans occurring across the US in 2009, consistent with the Urban Institute Phase 1 study of 2000. Moreover, discrimination was sensitive to the type of rental property as statistically significant discrimination was recorded for flats and condominiums (Hansen and Hawley 2011, Table 9 p. 108).

The tests in all the other countries found statistically significant discrimination in rental housing against the ethnic minority. In Greece, where greater differential treatment against Albanians was recorded as rent increased, the levels of net discrimination were: working class, 25.6%, middle class, 36.6%, upper class, 46.4%. In Sweden net discrimination levels against Middle-Eastern men of between 24.8% to 35.2% were recorded (as well as discrimination in favour of Swedish females). Probit estimates also found that differential treatment became greater as rent increased (Ahmed and Hammarstedt 2008, p. 371). In Spain, providing information on ability to pay rent (high-quality applicant) reduced but did not eliminate differential treatment between natives and Moroccans (H.Q. native/standard immigrant net discrimination of 20.0% compared to H.Q. native/H.Q. immigrant net discrimination of 12.0%).

3.2.2 *Single Enquiry Tests*

None of the results for the seven studies discussed here are reported in table format. There have been two further studies of race discrimination in housing in the USA using written approaches which sent single enquiries to advertised rental properties. Carpusor and Loges (2006) created an enquiry message and then randomly assigned a name from either an African-American, Arab or White when responding to just over one thousand rental advertisements in Los Angeles over ten weeks from March 2003. They found significantly fewer responses were received by the Arab and African-American enquiry as compared to the White, with the African-American receiving the least. There was evidence of greater differential treatment of African-Americans as rent increased to the middle range. A much larger study conducted by Ewens, Tomlin and Wang (2014) in 2009 responded to over fourteen thousand rental adverts over seven States. They constructed three types of enquiries for each racial group which contained information on the enquirer that was either basic, positive or negative and then sent one of these to enquire for a rental property. The intent of this study was to test various hypotheses which were consistent with either statistical or taste-based discrimination. The control for 'no signal' meant that the response to applicants was based only on their name. When name only was used, black applicants received 16% fewer positive responses than white applicants. Positive information had a beneficial impact on responses to both groups but the response gap remained the same. Negative information did, however, reduce the difference in treatment between white and black enquirers. The implications of these findings for theories of discrimination are discussed in section 6.

Three studies of ethnic discrimination in housing in Sweden which sent a single enquiry to rental adverts were conducted from 2008 to 2011. Ahmed, Andersson and Hammarstedt (2010) conducted a study for three months in early 2008 to investigate the impact on callback rates of providing more information on applicants such as employment, education, marital status and income when applying to rent a property. They designed four applications: a matched pair of Middle-Eastern male/ Swedish male with basic information and a matched pair of Middle-Eastern male/ Swedish male with enhanced information. One application was randomly selected and sent to an advertised rental property, with over one thousand enquires made. This additional information reduced, but did not eliminate, differential treatment. The difference in

callback rates between Middle-Eastern and Swedish males for the basic application was 16% as compared to 12% for the enhanced application, with the differences being statistically significant (Ahmed *et al.* 2010, Table 1, p. 84). Bengtsson, Iverman and Hinnerich (2012) tested two years later, for three months from early to mid-2010, sending one of four profiles, female/male and Arab/Swedish name to adverts for apartments to rent. Females with Arabic names were the only group discriminated against and the discrimination was concentrated in the suburbs of Stockholm. Carlsson and Eriksson conducted tests of ethnicity (age, gender and employment status also) from late 2010 to early 2011, sending a randomly generated enquiry created by a computer programme to nearly six thousand advertised rental properties. Arab enquiries had statistically significant lower rates of invitation to view properties than Swedish enquiries. The addition of positive information did not have an impact on the size of ethnic discrimination. If the landlord was Swedish, then the male Arab faced substantial discrimination. Consistent with Bengtsson *et al.* ethnic discrimination was concentrated in areas outside the metropolitan/city centre.

In Norway, Andersson, Jakobsson and Kotsadam (2012) tested for ethnic discrimination between late 2009 and early 2010 by creating eight profiles in terms of a Norwegian or Arab name, male or female, and occupation of economist or warehouse worker. They responded to nearly one thousand adverts for rental apartments, randomly selecting one of the eight profiles each time. Statistically significant discrimination against Arab names was found and confirmed ethnic discrimination rather than gender discrimination explained this. There was no statistical significant difference in the positive response rate between Norwegian males (58.5%) and Norwegian females (65.8%) nor between Arab males (46.4%) and Arab females (52.7%).

Baldini and Federici (2011) conducted tests of ethnic discrimination in Italy for four months from March 2010 responding to over three thousand six hundred adverts for rental properties. One profile, randomly selected from twelve fictitious profiles (constructed using characteristics of Arab/Eastern European/Italian ethnicity, gender and socio-economic information) was sent to adverts for rental property across Italy. Positive responses were recorded as follows: Italian, 62%, Eastern European 50% and Arab 44%. The lower response rate for Arab as compared to Italian names was

statistically significant for the Northern region of Italy with no significant difference in central and southern Italy. The addition of socio-economic information increased the response rate to all.

Four consistent features are identified from all these studies of discrimination on the basis of race/ethnicity in housing. First, persistent discrimination against, and geographic steering of, African-Americans in the USA. Second, prevalent discrimination against Middle Eastern and Moroccan ethnic groups across Europe. Third, positive information on socio-economic circumstances improved but did not eliminate differential treatment of ethnic minorities in access to rental properties. Fourth, discrimination in housing in Sweden tended to be concentrated in the non-metropolitan areas of cities.

Table 10: Results for Net Discrimination Rates in the Housing Market Studies on Race/Ethnicity Conducted 2000 to 2009

Country and Study	Year and location of test	Basis of test	Type of profile/information	Number of tests	Net Discrimination Rate ^{1,2} %								
<i>Greece</i>													
Drydakís (2011b)	2006-7	Ethnicity	Working-class (average rent €394.68)	1639	25.6 ^{**}								
In-person - telephone	Athens	Albanian	Middle-class (average rent €472.29)	1826	36.6 ^{**}								
			Upper-class (average rent €589.33)	1419	46.4 ^{**}								
<i>Spain</i>													
Bosch, Carnero and Farré ³ (2010) Written	2009 20 large cities in Spain	Moroccan Immigrants	Standard native/ standard immigrant	532	15.0 ⁵								
			H.Q native/ H.Q. immigrant ⁴	986	12.0 ⁵								
			H.Q. immigrant/ standard immigrant	1382	8.0 ⁵								
			H.Q. native/ standard native	105	7.0 ⁵								
			H.Q. native/ standard immigrant	986	20.0 ⁵								
			H.Q. immigrant/ standard native	105	-3.0 ⁵								
<i>Sweden</i>													
Ahmed and Hammarstedt (2008) Written	2007 Sweden	Ethnicity Middle-Eastern	Swedish male/ ME male	104	24.8 ^{6***}								
			Swedish female/ Swedish male	226	-10.4 ^{6***}								
			Swedish female/ ME male	92	35.2 ^{6***}								
<i>USA</i>													
Turner, Ross, Galster, and Yinger (2002)	2000	Race	African-American	20 cities 16 cities	1152 1112								
						Housing rental	7.9 ^{7*}						
						Housing sales	8.3 ^{7*}						
						In-person	2000	Race	Hispanic	10 cities 10 cities	731 759		
												Housing rental	15.1 ^{7*}
												Housing sales	4.9 ⁷

1. * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level.

2. A negative sign means discrimination in favour of the minority

3. Bosch, Carnero and Farre (2010, p. 14) made enquiries using matched standard native/ standard immigrant to 427 flats; matched high quality immigrant/ standard immigrant to 396 flats; matched high quality native /high quality immigrant /standard native to 881 flats; high quality native /high quality immigrant /standard native/ standard immigrant to 105 flats

4. H.Q. refers to high quality

5. Percentage of emails in net terms favouring the standard native (Bosch, Carnero and Farre 2010: footnote b to table 2, 15)

6. Results reported for 'Contact' (Ahmed and Hammarstedt 2008, Table 2 p. 368)

7. Results reported for hierarchical measure are respectively, exhibit 3-5, p. 3-4; exhibit 3-16, p. 3-14; exhibit 3-10, p. 3-9; exhibit 3-22, p. 3-19 (Turner *et al.* 2002)

Table 10 continued: Results for Net Discrimination Rates in the Housing Market Studies on Race/Ethnicity Conducted 2000 to 2009

Country and Study	Year and location of test	Basis of test	Type of profile/information	Number of tests	Net Discrimination Rate ^{1,2} %
<i>USA</i>					
Turner and Ross (2003a)	2000-1	Race			
In-person	11 cities	Asian and Pacific Islanders		481	4.3 ^{2*}
Housing rental				408	19.6 ^{2*}
Housing sales					
Turner and Ross (2003b)	2002	Race			
In-person	8 cities	Native Americans		297	18.4 ^{3*}
Housing rental				100	5.2 ³
Housing sales					
Hanson and Hawley (2011)	2009	Race			
Written	10 cities	African-American	All audits	3153	4.54 ^{4***}
			High class enquiry	790	1.77 ⁵
			Low class enquiry	788	5.96 ^{5***}
			Neighbourhood < 50% white residents		5.81 ^{6**}
			Neighbourhood 50% < 80% white residents		4.43 ^{6**}
			Neighbourhood 80% < 95% white residents		7.02 ^{6***}
			Neighbourhood 95% to 100% white residents		2.58 ^{6*}

1. * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level.

2. Results reported for hierarchical measure are for rental, exhibit 3-5, p. 3-3; sales exhibit 3-12, p. 3-8 (Turner and Ross 2003a)

3. Results reported for hierarchical measure are for rental, exhibit 4-1, p. 4-2; sales exhibit 3-21, p. 3-31 (Turner and Ross 2003b)

4. Hanson and Hawley (2011), Table 5, p. 105.

5. Hanson and Hawley (2011), Table 6, p. 105.

6. Hanson and Hawley (2011), Table 8, p. 107.

3.3 Sexual Orientation Tests

A matched-pair of applications were used in all these studies. The results found for net discrimination in the three studies, two in Sweden and one in the USA, are reported in Table 11. The studies in Sweden each sent two matched-pair applications via email consisting of a homosexual male couple/heterosexual couple in the case of Ahmed and Hammarstedt (2009) and lesbian couple/heterosexual couple in the case of Ahmed *et al.* (2008). Homosexual men were discriminated against when they applied to rent a property, a level of net discrimination of 19.9% recorded on the measure 'contact following an enquiry'. In contrast, virtually no difference in treatment between lesbian and heterosexual couples was found. Ahmed, Andersson and Hammarstedt (2008) concluded that this was consistent with previous experiments they had conducted that recorded preferences for females to rent property.

3.4 Disability Tests

The Urban Institute in the US conducted in-person, matched pair tests on access to housing for the disabled in 2003, where the disability was either being deaf or in a wheelchair (Turner, Herbig, Kaye, Fenderson and Levy 2005). Approximately one hundred audits in the Chicago area were conducted for each type of disability and the results for the summary measure of consistent adverse treatment are reported in Table 11. In comparison to the African-American race tests conducted in 2000, levels of net discrimination on the basis of disability were at least three times higher [26.7% (deaf) and 30.3% (wheelchair)].

3.5 Other Dimensions of Differential Treatment

The Boston Fair Housing tests of race/ethnicity (2004-5) recorded differential treatment against African-Americans and Latinos in 47% of tests, such as differences in access to properties and agents, mortgage required and encouragement and scrutiny (Fair Housing Center of Greater Boston 2005). In an audit of African-Americans access to housing in New Orleans conducted in 2006-2007, differential treatment of the African-American was recorded in 57.5% of the forty tests (Greater New Orleans Fair Housing Action Center 2005). The tests conducted in Seattle in 2011 recorded inconsistencies favouring whites in 69% of tests such as quoting higher rents to African-Americans, withholding information and more intense scrutiny of African-

Americans (Seattle Office for Civil Rights 2011). The Vermont Fair Housing matched-pair audits (2003-4) of discrimination in housing against immigrants from Islamic and non-Islamic backgrounds, recorded differential treatment against both groups of the type outlined above (CVOEO Fair Housing Project 2004, pp. 5-6). In the audits for discrimination on the basis of sexual orientation conducted by the Michigan Fair Housing Centers in 2006, differential treatment of the same-sex couple, who were better qualified than the heterosexual couple, was reported in 27% of tests. (Fair Housing Center of Southeastern Michigan 2007). The Seattle Office for Civil Rights (2011) investigating differential treatment on the basis of disability (in a wheelchair or having a service animal) found differential treatment in 36% of audits, consisting of refusal to accommodate an animal, withholding information on properties and failure to provide parking dedicated to the disabled.

3.6 Dishonest Concealment of Rejection

Dishonest concealment of rejection, similar to that found in the labor markets tests, has been recorded in the housing market tests. Tests by the Urban Institute in 2000 and 2005 report instances of the minority tester being told that applications for renting were not being taken, only for the majority tester to be told, when enquiring hours later, that applications for renting were being accepted (Turner *et al.*, 2005, pp. 35-38). Other examples were: the minority tester's enquiry not receiving a response/not asked for further information, hanging up the phone in the case of the disability tests, while the later contact from the majority tester was responded to and a viewing arranged; minority applicants being told the property had been rented, but when the majority applicant enquired subsequently, they were told the property was available.

Table 11: Results for Net Discrimination Rates in the Housing Market Studies on Sexual Orientation and Disability Conducted 2004 to 2007

Country and Study	Year and location of test	Basis of test	Type of profile/information	Number of tests	Net Discrimination Rate ¹ %
<i>Sweden</i>					
Ahmed and Hammarstedt (2009) Written	2007 Sweden	Sexual orientation	Homosexual male/ Heterosexual male	408	19.9 ^{2**}
Ahmed, Andersson and Hammarstedt (2008) Written	2007 Sweden	Sexual orientation	Lesbian/Heterosexual female	423	0.0 ³
<i>USA</i>					
Turner, Herbig, Kaye, Fenderson, and Levy (2005) In-person Housing rental	2004 Chicago	Disability	Deaf Wheelchair	101 99	26.7 ^{4*} 30.3 ^{4*}

1. * significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level.

2. Results reported for 'Callback' (Ahmed and Hammarstedt 2009, Table 2, p. 593)

3. Results reported for 'Contact' (Ahmed, Andersson and Hammarstedt 2008, Table 2 p. 237)

4. Results reported are for the summary measure of consistent adverse treatment (Turner, Herbig, Kaye, Fenderson, and Levy, 2005, Table 4-5, p. 41 and Table 5-6, p. 51)

4. Field experiments of the product market

There have been far fewer studies of discriminatory treatment in product markets since 2000 and the vast majority of these have been conducted in the USA. Seven studies in the USA focused on investigating differential treatment on the basis of race (five, one of which tested multiple bases for discrimination), disability (one) and sexual orientation (one). One study was conducted in Israel which investigated differential treatment based on ethnicity. The in-person approach was used and in most studies the testers were real buyers who made enquiries to purchase products, but recent studies have created fictitious sellers or buyers to transact on online websites. In the main, the focus of these tests was the price outcomes for the different groups in the experiment. Prices are recorded at various stages of the transaction process and any differences in price outcomes are tested for statistical significance. Each experiment is carefully designed to test hypotheses as to why different price outcomes may arise between different groups, particularly whether the behaviour of transactors is driven by cost/profit considerations or by distaste when interacting with specific groups. Gneezy and List have been at the forefront in developing such tests that can reflect on the causes of discriminatory behaviour, thereby distinguishing taste-based from statistical discrimination.

List (2004) conducted tests on trading in the sportscard market in the USA from June 2002 to July 2003. Dealers and buyers at sportscard meets were recruited for the tests and voluntarily participated in recording the offers, initial to final, made in transactions. Specific instructions were given to participants on how to conduct enquiries and bargain in transactions, using training to control for bargaining method, mannerisms and technique. Price offers and other aspects of the transactions were recorded immediately after the bargaining concluded. List found that women, older people and minorities, paid more for sports cards than did young white men. Black buyers, for example, recorded final price offers that were 4.2 percent higher than those achieved by young white buyers, with these differences being statistically significant.

Gneezy and List (2004) conducted tests on the basis of disability in 2003. They sent matched pairs of male testers, one disabled (in a wheelchair), one able-bodied, to auto repair shops in Chicago to obtain quotes for car repair. Car make and model were

equivalent and repair was controlled to body work. No bargaining was allowed with testers simply asking for quotes to repair a car. The difference in average price quoted to the disabled tester of \$1425 compared to \$1212 to the able-bodied tester was statistically significant. Gneezy, List and Price (2012) report two further studies of discrimination by sending matched testers to purchase new cars, one testing for race, the other for sexual orientation. In the race tests, statistically significant higher initial and final prices were made to minority testers as compared to white testers for high-end cars but no difference was found for low-end cars. In the sexual orientation tests, the outcome was found to be dependent on the race of the salesperson: statistically significant higher initial and final offers to testers characterised as homosexual were made when the car dealer was from a minority group.

In an audit of nightclubs and bars in New Orleans in 2005, African-American and White testers, matched in appearance, age, height and trained in demeanour, were sent to ask for the same drinks (Greater New Orleans Fair Housing Action Center 2005). The white team was followed ten minutes later by the African-American team. In forty percent of the forty tests, statistically significant differential treatment was recorded: African-American testers were charged higher prices for drinks in a range of 17.7% to 24% more, while dress and entry conditions were enforced although they were lax for white testers.

In an experiment conducted for one year from March 2009 selling iPods on an online advertising site, Doleac and Stein (2013) signalled the race of the seller by the skin-tone of the seller's hand holding the iPod. They found that iPods offered for sale by black sellers received 13% fewer responses and 17% fewer offers than those made to white sellers.

Nunley, Owens and Howard (2011) investigated racial discrimination for most of 2009 by setting-up auctions (in total two hundred and eighty-eight) on eBay with matched pairs of sellers selling the identical product, using names to signal whether the seller was Black or White. They found products targeted to white buyers received higher prices when sold by a white seller than a black seller and for products targeted to black buyers received higher prices when sold by a black seller than a white seller. This bias in favour of same-race seller was mainly due to the lack on the part of the

seller of a credible reputation. Once buyer feedback reached a certain score and the credibility of the seller was established no statistically significant price differences remained.

In Israel, Zussman (2013) tested for ethnic discrimination in an online market for used cars by sending enquiries from fictitious buyers signalling ethnicity by a distinct first name. The experiment was conducted from August 2009 to June 2010 and responded to eight thousand advertisements. The Jewish buyer received a 22% higher response rate than the Arab buyer. Arab buyers offering to buy at the stated price received the same response rate as Jewish buyers offering to buy at a 5-10% discount (Zussman 2013, p. F447).

5. Neumark test for the impact of unobservable characteristics of an applicant

One of the more notable observations from our survey of the sixty-seven studies of discrimination in different markets and seventeen countries is the very small number of tests where no net discrimination was found. The question has been raised, however, of whether the existence of unobservable characteristics (to the researchers) has possibly contributed to an overestimation of discrimination in the results. It has been argued that factors not controlled for but influencing a worker's productivity or an individual's ability to pay (for rent or products) may explain why the minority applicant would be a less desirable candidate. Assume productivity is normally distributed around the mean for the group and each applicant has the same average productivity arising from both the observed characteristic and the unobserved characteristic, but the variation from the average for the unobserved characteristic is much greater for the minority applicant. In this case, if a high standard/quality is set on the experimenter's constructed CVs to increase the chance of a positive response from an employer, a greater risk will be associated with choosing the minority applicant as an individual with a much lower productivity could be chosen. A bias is introduced into the testing favouring the majority applicant and thus an overestimation of discrimination against the minority may arise. If, on the other hand, a low standard/quality is set on the experimenter's constructed CVs to avoid any over qualification for the jobs being applied to, a lower risk is associated with choosing the minority applicant as an individual with a higher productivity could be chosen. A bias is then introduced into the testing, favouring the minority applicant and an

underestimation of discrimination against the minority may occur. Total discrimination which has been measured by the experiment could be due to statistical discrimination and/or distaste. The statistical discrimination could be due in part to differences in the average productivity of the unobservable characteristics between majority and minority applicants and in part to an artifice of the experiment through the difference in variance of the unobservable characteristics between majority and minority applicants.

Neumark (2012) provides a technique for statistical analysis to account for the impact of unobservable characteristics on the responses gathered in field experiments, although details of that mathematically complex procedure are not reproduced here. It is possible to use this technique if the experimental study explores the impact of different productivity characteristics by, for example, creating applicants who have different levels of qualifications. This information allows the effect of the difference in variance between the groups' unobserved characteristics on the responses to be isolated. The method developed by Neumark calculates the impact of being a member of a group on the probability of an interview for a job, keeping the variance of productivity constant and the impact of being a member of a group and allowing the variance of productivity from the unobservable characteristic. Isolating the impact of the variance of productivity arising from unobservable characteristics can assess whether the total discrimination recorded in the experiment has been compromised.

Neumark applied this technique to reexamine the results from Bertrand and Mullainathan (2004), a study which did control for variation in information on qualifications of applicants, thus enabling the impact of the unobservable characteristic to be measured. Neumark found that Bertrand and Mullainathan did not underestimate or overestimate the level of net discrimination in their tests on racial discrimination in the USA. Carlsson, Fumarco and Rooth (2013) then applied Neumark's technique to their own written tests on ethnicity/race and gender which had been conducted in 2007 and 2005 respectively. They found no bias in the estimation of ethnic discrimination. They suggest, however, that variance of unobservables may differ between high skill and low skill jobs and may affect the results found for discrimination. Their reexamination of the gender discrimination tests found it overestimated discrimination against males (Carlsson, Fumarco and

Rooth 2013, p. 22-23). They surmised that this overestimation arose because low-quality applications had been used in the correspondence test and so biased the responses in favour of the group with higher variance of unobservables, that is, females. The application of Neumark's technique by Baert *et al.* (2013, p. 19) to their responses in Belgium found that discrimination was, if anything, underestimated and certainly not overestimated. Baert also confirmed her own results were not compromised (Baert 2013, footnote 13, p. 18). Future studies, designed to provide further information on applicants (such as Ahmed *et al.*; Bertrand and Mullainathan; Bosch *et al.*) enabling the application of Neumark's technique, should provide additional insight on the impact of characteristics other than the basis of discrimination being tested.

At the same time, the diversity of the studies surveyed here raises the question of what possible unobservable characteristics could explain the persistence of lower variance in productivity for the majority applicant over minority applicant. Such characteristics would have to apply over different bases (for example, female, African-American, Albanian, Chinese, Moroccan, Turkish, obese, older, disabled, homosexual, etc.), across different markets (labour, housing, and product) and over time. The identification of perceived attitudes may allude to what unobservable characteristics could be affecting decision-making. For example, is it that minority applicants present as more tentative or reticent and therefore are considered less suitable for senior leadership positions; could they be less sociable or adaptable and therefore thought less suitable for working in teams such as call centre positions; could they be thought more likely to have family/financial/other socio-economic problems which influence their concentration at work? These attitudes influencing unobservable characteristics raise the possibility that some of them may be explained by discriminatory attitudes (see for example Wozniak's 2012 analysis of hiring in firms testing for use of illegal drugs). If there is no or little evidence to support actual differences in the variance of unobservable characteristics from the same mean for applicant groups, then use of these unobservable characteristics as a difference in productivity between the groups is in itself further evidence of discrimination in treatment.

6. Theory: Taste-based or statistical discrimination?

Guryan and Charles (2013) provide a comprehensive discussion of how correspondence and audit studies can be at the forefront of providing insight into the underlying behaviour explaining discrimination. In discussing the recent product market tests, they make cogent suggestions for developing the field experiments to gather such information. Many of the recent experimental studies discussed in this survey analysed their results to explore whether a convincing explanation was provided by taste-based (animus) or statistical discrimination. The findings from labour market studies will be considered first, then those of the housing market and the product market.

Bertrand and Mullainathan felt both statistical and taste-based theories struggled to provide a plausible explanation for their findings of discrimination against African-Americans. Neither theory could adequately explain their finding of no significant difference in discrimination against African-Americans across occupations and industries. Other aspects of statistical discrimination were also hard to accept given that the same résumé characteristics were used for both groups tested (Bertrand and Mullainathan 2004, pp. 1010-1011). Carlsson and Rooth (2011) reexamined their findings on ethnic discrimination in Sweden using a large publicly available attitude survey, specifically using the question on immigrants and whether they are thought valuable or not. Investigating regional variation in responses to Swedish and Middle-Eastern names from their field experiment, with regional variation in attitudes to immigrants suggested that taste-based discrimination could well be a factor that explained employer responses. Maurer-Fazio suggested that her finding of similar treatment for ethnic minorities in China when labour markets were tight provided support for taste-based discrimination, as employers faced with a shortage of labour had less ability to indulge their preferences (Maurer-Fazio 2012, p. 11). Petit concluded from her study in France that statistical discrimination explained the findings of discrimination against young, single females with no children, in high-skill jobs in the financial sector, as the likely explanation was the potential costs involved in hiring young women. The banking sector in France faced legal obligations regarding paid maternity leave and the fertility rates in France at the time of the tests meant there was a general presumption that young females were likely to have children (Petit 2007, footnote 1, p. 374). Albert *et al.* (2011, p. 369-370) also

interpreted their findings on age discrimination in Spain as providing support for statistical discrimination, although the reasons they give for why an employer would avoid hiring older workers are based on assertions regarding those workers' capabilities. This does not provide strong support for evidence of productivity differences as such attitudes could also be considered prejudicial or taste-based. Lahey did not find compelling evidence to support statistical discrimination in her age tests as the vast majority of variables used to confront age stereotypes, which one could argue would have a negative effect on productivity and costs, were not statistically significant. On the other hand she did not present compelling evidence against taste-based discrimination as the variables used to capture employer, employee and customer distaste were too broad and, as such, not entirely convincing. Riach and Rich concluded from their studies of age and gender discrimination that there were strong implications that decisions were based on distaste. Their age tests had ensured the creation of realistic older workers with greater experience than the younger worker and they hypothesised that this would bias employer preferences to the older workers, which occurred in only one of the three occupations they tested. In the case of their gender tests this was because the discriminatory outcomes for men who applied to female-dominated jobs was suggestive of strong gender stereotyping. Drydakis and Tilcsik in their studies of sexual orientation discrimination and Rooth in his study of obesity discrimination surmised that prejudicial attitudes were a more plausible explanation for their results, particularly demonstrated in the pattern of discrimination against men in customer service jobs from Rooth's obesity tests.

The housing market tests conducted in Italy, Norway, Spain and Sweden (by Ahmed and Hammarstedt) found that the provision of enhanced information reduced, but did not remove, discrimination against the ethnic minority. The reduction in the level of discrimination suggests that some proportion of differential treatment may be explained by statistical discrimination. For example, Ahmed and Hammarstedt surmised that landlords used an applicant's characteristics as a signal of ability to pay rent and that Middle-Eastern applicants with lower employment rates may be perceived to have lower ability to pay (Ahmed and Hammarstedt 2008, p. 371). However in their study, the reduction in differential treatment from enhanced information on applicants was less than fifty percent, leaving the majority of differential treatment unexplained. This suggested the presence of taste-based

discrimination, that is, an animosity towards Moroccans. This was the case for the ethnic minority in the other studies in Italy, Norway and Spain. The US study by Ewens *et al.* concluded that statistical discrimination explained the pattern of responses to the different information they provided. Testing their responses for various hypotheses conformed more with predictions from statistical than taste-based discrimination. This was particularly the case for the finding that positive information did not close the difference in response rates although it improved them for both groups, while negative information closed the difference by reducing responses to whites. This would support the notion that landlords find available information a better predictor of reliability for white than black renters and do not use the average reliability of the group for whites to any great extent. For black applicants, landlords use the average reliability of the group to a greater extent and discount the information provided on the individual. The Urban Institute tests of both race and disability, however, suggested the presence of prejudicial attitudes. In the case of the disability tests even when the disabled enquirer offered to meet costs that may be involved in required modifications to the property, they were rejected. Hansen and Hawley explained their findings on discrimination by type of rental property as perhaps revealing landlord behaviour based on perceived discrimination of tenants (Hansen and Hawley 2011, p. 107). This would also suggest the presence of taste-based discrimination. Carlsson and Eriksson concluded that their findings that ethnic discrimination in housing was concentrated largely in non-metropolitan areas and exercised by Swedish landlords, were more indicative of taste-based discrimination, as attitude surveys confirmed higher levels of prejudicial attitudes in these areas and by that group.

The product market tests of Gneezy *et al.*, List, Nunley *et al.* and Zussman provide compelling evidence that statistical discrimination rather than prejudice explains the differential treatment of buyers by sellers. Sellers appear to take advantage of differences in search costs that exist between groups of buyers as was the case in Gneezy and List's tests for disability. Buyers may mitigate negative attitudes by not revealing information on their ethnicity as Zussman reported. Price differences due to lack of information on sellers disappeared once improved information on reputation was provided, as was the case with Nunley *et al.*. Doleac and Stein's study is consistent with this finding from Nunley *et al.* as they found evidence of (lack of)

trust which affected the interactions of buyers with (black) sellers. Zussman conducted a survey of Jewish sellers who had responded to the fictitious Arab buyers in his experiment, to elicit their attitudes to ethnic minority buyers. He designed the survey so that attitudes could be parsed into those expressing distaste/prejudice as distinct from those expressing perceptions of behaviour (which if accurate would mean a seller transacting with such a buyer would incur monetary costs). He then examined the relationship between these seller's responses and their attitudes to Arabs. The only attitude that was significantly related to discriminatory behaviour was that of the perception of cheating, or lack of trustworthiness, evidence once again, consistent with statistical discrimination (Zussman 2013, p. F451-452). This evidence from product market tests is supported by Bayer, Casey, Ferreira and McMillan (2012) who analysed data on house sale prices in four major cities of the US for twenty years (1990-2008). The average three percent higher price charged to black and Hispanic buyers was primarily explained by statistical discrimination.

7. Concluding comments

The sixty-seven studies conducted between 2000 and 2012 over seventeen countries which have been surveyed in this paper found that access to jobs was restricted in particular for racial minorities, women, homosexuals, and older workers. Persistent and enduring high levels of discrimination against Moroccans, African-Americans, older workers (bearing in mind these were between thirty-seven to forty-seven years old) were recorded, as well as strong gender-stereotyping of jobs. Higher qualifications or better employment backgrounds did not significantly improve the outcomes for minority groups. Access to housing was restricted for racial and ethnic minorities, homosexuals and disabled individuals. Geographic steering of African-Americans in housing rental and sales remains persistent across the USA. A caveat needs to be added as the housing market tests conducted in 2000 in the USA found that net discrimination had declined for both African-Americans and Hispanics since the tests conducted in 1989. African-Americans and disabled individuals, in particular, in the USA and Arabs in Israel, paid more for products. This survey mirrors the pattern of discrimination evidenced in Riach and Rich (2002). The impact of repeated failure to obtain jobs or housing on an individual are only too well documented. The effects include low self-esteem, low morale, lower productivity, discouraged job

seeking, poor health and social exclusion. The findings of this survey document the persistence of not only individual but also group differential treatment and exclusion.

In designing future studies, it would be useful for comparative purposes if those testing multiple ethnic groups at the same time used matched-pair applications and reported both callback rates and matched pair net discrimination levels as in the recent Chinese study by Maurer-Fazio and the Swedish study by Bursell. Single enquiry tests should aim to make substantial enquiries (such as Ewens *et al.*) given that the response of the same employer or real estate agent to a different, matched applicant is not known. One could argue, however, that if an employer or landlord did want to exercise differential treatment, they could do so when faced with an individual from a group they disliked or discounted their productivity without there being any other enquirer, particularly one from a favoured group. By undertaking widespread testing Ewens *et al.* were able to demonstrate patterns of differential treatment across a market.

In view of ageing populations and policies to increase the retirement age, the findings of high levels of net discrimination against older applicants for jobs are discouraging when considering the likely effects of such policies on the welfare of older workers. More testing in this area could provide a valuable contribution to policy considerations. The ILO three-stage tests in the labour market are another important contribution, establishing that the majority of discrimination occurred at the initial stage of hiring. The Urban Institute tests in the housing market are also an important guide to repeat experiments as they enabled tracking of discrimination over time, establishing that African-Americans continued to be steered to what are considered to be suitable residential neighbourhoods. Drydakis and Drydakis and Vlassis gathered further information on aspects of differential treatment in hiring by having testers available to receive telephone calls from employers wanting to arrange an interview. More of this type of information would be extremely useful in assessing the range of differential treatment. Zussman's follow-up survey of (majority population) sellers who had been approached in the experiment to elicit their attitudes to an ethnic minority, gathered information that enabled an analysis of his results on seller responses to buyer offers that could distinguish animus-based from statistical discrimination. This additional aspect of the testing is valuable for providing insight

into possible theoretical implications of the findings of field experiments. Neumark's statistical technique for analysing the impact of unobservable characteristics on the estimation of discrimination provides a basis for future work improving the reliability of the measure of discrimination obtained from the experiment. This requires designing the CVs so that applicants can signal for example, different levels of qualifications or skills (such design is evidenced in a number of recent studies as well as Bertrand and Mullainathan).

Future research design should consider how testing could provide insight on aspects of decision-making. Rooth (2010) provides a discussion regarding the usefulness of the implicit attitude test (IAT) in distinguishing reasons for hiring preferences. He combined the IAT with his own field experiments of ethnic discrimination by contacting employers who had been tested in the original experiment, requesting their participation to undertake the IAT. He found a clear association between a negative response to ethnic male and the lower probability of invitation to a job interview. Additional research needs to be done in this area to provide more conclusive evidence of the robustness of the IAT and its contribution to the theory of discrimination. This is particularly the case as there is by no means a consensus of opinion that the IAT does measure racial prejudice as it may measure belonging to, or exclusion from, a group rather than ethnicity or nationality (Azar 2008; Kaufman 2011). Rooth and Guryan and Charles nevertheless draw attention to the benefit economists could gain from collaborating with social psychologists who have a long history of research on prejudice, stereo-typing and discrimination (Guryan and Charles 2013, p. F428; Rooth 2010, p.530). A valuable area would be investigating perceived attitudes affecting decision-making that possibly underpin unobservable characteristics.

The latest findings are from an era where many countries have passed anti-discrimination legislation, making it illegal to discriminate against individuals on various bases and where there has been much discussion about discrimination in the public arena. It is thus rather alarming that the recent studies replicate the findings of tests conducted over the period 1966 to 2000 and document responses that indicated dishonest concealment of the reason(s) for rejection (see pages 33 and 45). It should also be recognised that one of the impacts of anti-discrimination legislation is that no sensible employer would behave in a clear and overt fashion if engaging in

discriminatory practices. The research of social psychologists (Al Ramiah, Hexstone, Davidio and Penner 2010; Ellemers and Barreto 2009) indicates that there has been a shift in such practices so that they have become more subtle and hence much more difficult for individuals to detect. So too does the research by Pager *et al.* who found minority applicants qualifications and experience were devalued as compared to white applicants (Pager *et al.* 2009, p. 793). The dishonest concealment of rejection and the subtle forms of discrimination mean it would be extremely difficult for an individual to gain *prima facie* evidence to instigate legal action under current legislation where it is complaint-based. This raises the issue of the legislative framework of complaints-based anti-discrimination laws which legislators, equal employment bodies and others could pursue more vigorously.

References

- Ahmed, A., Andersson, L. and Hammarstedt, M. (2013). 'Are gay men and lesbians discriminated against in the hiring situation?', *Southern Economic Journal*, vol. 79 (3), pp.565-585.
- Ahmed, A., Andersson, L. and Hammarstedt, M. (2012). 'Does age matter for employability? A field experiment on ageism in the Swedish labor market', *Applied Economics Letters*, vol. 19, pp. 403-406.
- Ahmed, A., Andersson, L. and Hammarstedt, M. (2010). 'Can discrimination in the housing market be reduced by increasing the information about the applicants', *Land Economics*, vol. 86(1), pp. 79-90.
- Ahmed, A., Andersson, L. and Hammarstedt, M. (2008). 'Are lesbians discriminated against in the rental housing market? Evidence from a correspondence testing experiment', *Journal of Housing Economics*, vol. 17, pp. 234-238.
- Ahmed, A. and Hammarstedt, M. (2009). 'Detecting discrimination against homosexuals: Evidence from a field experiment on the internet', *Economica*, vol. 76(303), pp. 588-597.
- Ahmed, A. and Hammarstedt, M. (2008). 'Discrimination in the rental housing market: A field experiment on the internet', *Journal of Urban Economics*, vol. 64(2), pp. 362-372.
- Al Ramiah, A., Hexstone, M., Dovidio, J. F., and Penner, L. A. (2010). 'The social psychology of discrimination: Theory, measurement and consequences', in (L. Bond, F. McGinnity and H. Russell eds.) *Making Equality Count: Irish and International Research Measuring Equality and Discrimination*. Dublin, Ireland, The Equality Authority.

- Albert, R., Escot L. and Fernandez-Cornejo, J. (2011). 'A field experiment to study sex and age discrimination in the Madrid labour market', *The International Journal of Human Resource Management*, vol. 22(2), pp. 351-375.
- Allasino, E., Reyneri, E., Venturini, A. and Zincone, G. (2004). 'Labour market discrimination against migrant workers in Italy', *International Migration Papers 67*, Geneva: International Labour Office.
- Andersson, L., Jakobsson, N. and Kotsadam, A. (2012). 'A field experiment of discrimination in the Norwegian housing market: Gender, class and ethnicity', *Land Economics*, vol. 88(2), pp. 233-240.
- Attström, K. (2007). 'Discrimination against native Swedes of immigrant origin in access to employment', *International Migration Papers 86E*, Geneva: International Labour Office.
- Azar, B. (2008). 'IAT: Fad or fabulous?', *American Psychological Association*, vol. 39(7), p. 44.
- Baert, S. (2013). 'Career lesbians: Getting hired for not having kids?', *IZA Discussion Paper No. 7767*.
- Baert, S., Cockx, B., Gheyle, N. and Vandamme, C. (2013). 'Do employers discriminate less if vacancies are difficult to fill?: Evidence from a field experiment', *IRES Discussion Paper 2013-1*, Université catholique de Louvain.
- Baldini, M. and Federici, M. (2011). 'Ethnic discrimination in the Italian rental housing market', *Journal of Housing Economics*, vol. 20, pp. 1-14.
- Banerjee, A., Bertrand, M., Datta, S. and Mullainathan, S. (2009). 'Labor market discrimination in Delhi: Evidence from a field experiment', *Journal of Comparative Economics*, vol. 37, pp. 14-27.

- Bayer, P., Casey, M., Ferreira, F. And McMillan, R. (2012). ‘ Price discrimination in the housing market’,
www.real.wharton.upenn.edu/fferreir/documents/bcfm_4may12.pdf
- Bayer, P., McMillan, R. and Rueben, K. (2001). *The Causes and Consequences of Residential Segregation: An Equilibrium Analysis of Neighborhood Sorting*, Public Policy Institute of California.
- Bengtsson, R., Iverman, E. and Hinnerich, B. T. (2012). ‘Gender and ethnic discrimination in the rental housing market’, *Applied Economic Letters*, vol. 19(1), pp. 1-5.
- Bertrand, M. and Mullainathan, S. (2004). ‘Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination’, *American Economic Review*, vol. 94(4), pp. 991-1013.
- Booth, A., Leigh, A. and Varganova, E. (2012). ‘Does ethnic discrimination vary across minority groups? Evidence from a field experiment’, *Oxford Bulletin of Economics and Statistics*, vol. 74(4), pp. 547-573.
- Bosch, M., Carnero, M. and Farré, L. (2010). ‘Information and discrimination in the rental housing market: Evidence from a field experiment’, *Regional Science and Urban Economics*, vol. 40, pp. 11-19.
- Bovenkerk, F. (1992). *Testing Discrimination in Natural Experiments: A manual for international comparative research on discrimination on the grounds of ‘race’ and ethnic origin*. Geneva: International Labour Office.
- Bursell, M. (2007). *What’s in a Name? A field experiment test for the existence of ethnic discrimination in the hiring process*. The Stockholm University Linnaeus Center for Integration Studies, University of Stockholm, Working Paper 2007:7.

- Carlsson, M. and Eriksson, S. (2014). 'Discrimination in the rental market for apartments', *Journal of Housing Economics*, vol. 23, pp. 41-54.
- Carlsson, M., Fumarco, L. and Rooth, D.-O. (2013). 'Artifactual evidence of discrimination in correspondence test: A replication of the Neumark method', *IZA Discussion Paper* No. 7619.
- Carlsson, M. and Rooth, D.-O. (2011). 'Revealing taste-based discrimination in hiring: A correspondence testing experiment with geographic variation', *IZA Discussion Paper* no. 6153.
- Carlsson, M. and Rooth, D.-O. (2007). 'Evidence of ethnic discrimination in the Swedish labor market using experimental data', *Labour Economics*, vol. 14(4), pp. 716-729.
- Carpusor, A. and Loges, W. (2006). 'Rental discrimination and ethnicity in names', *Journal of Social Psychology*, vol. 36(4), pp. 934-952.
- Cediey, E. and Foroni, F. (2008). 'Discrimination in access to employment on grounds of foreign origin in France', *International Migration Papers 85E*, Geneva: International Labour Office.
- CVOEO Fair Housing Project. (2004). *National origin discrimination in rental housing in Vermont*. www.cvoeo.org/vti/fair.htm
- Doleac, J. and Stein, L. (2013). 'The visible hand: Race and online market outcomes', *Economic Journal*, vol.123, F469-F492.
- Drydakis, N. (2012). 'Sexual orientation discrimination in the Cypriot labour market: Distastes or uncertainty?', *IZA Discussion Paper* No. 6684. (forthcoming *International Journal of Manpower*).
- Drydakis, N. (2011a). 'Women's sexual orientation and labor market outcomes in Greece', *Feminist Economics*, vol. 17(1), pp. 89-117.

- Drydakis, N. (2011b). 'Ethnic discrimination in the Greek housing market', *Journal of Population Economics*, vol. 24, pp. 1235-1255.
- Drydakis, N. (2009). 'Sexual orientation discrimination in the labour market', *Labour Economics*, vol. 16, pp. 364-372.
- Drydakis, N. and Vlassis, M. (2010). 'Ethnic discrimination in the Greek labor market: Occupational access, insurance coverage, and wage offers', *Manchester School*, vol. 78(3), pp. 201-218.
- Duguet, E., Leandri, N., L'Horty, Y. and Petit, P. (2010). 'Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area', *Annales d'Economie et Statistique*, vol. 99/100, pp. 187-215.
- Ellemers, N. and Baretto, M. (2009). 'Collective action in modern times: How modern expressions of prejudice prevent collective action', *Journal of Social Issues*, vol. 65(4), pp. 749-768.
- Ewens, M., Tomlin, B. and Wang, L-C. (2014). 'Statistical discrimination or prejudice? A large sample field experiment', *Review of Economics and Statistics*, vol. 96 (1), pp. 119-134.
- Fair Housing Center of Greater Boston. (2005). *You don't know what you're missing ... A report on discrimination in the Greater Boston home sales market.*
www.bostonfairhousing.org
- Fair Housing Center of Southeastern Michigan. (2007). *Sexual orientation and housing discrimination in Michigan.* www.fhcmichigan.org
- Gneezy, U. and List, J. (2004). 'Are the disabled discriminated against in product markets? Evidence from field experiments', www.aeaweb.org/annual_mtg.

- Gneezy, U., List, J. and Price, M. (2012). 'Toward an understanding of why people discriminate: Evidence from a series of natural field experiments', NBER Working Paper No. 17855.
- Greater New Orleans Fair Housing Action Center. (2005). *An Audit of Discriminatory Treatment of African-American Males by Bourbon Street Nightclubs and Bars*. New Orleans: GNOFHAC.
- Guryan, J. and Charles, K. (2013). 'Taste-based or statistical discrimination: The economics of discrimination returns to its roots', *Economic Journal*, vol. 123, pp. F433-F468.
- Hanson, A. and Hawley, Z. (2011). 'Do landlords discriminate in the rental housing market? Evidence from an Internet field experiment in U.S. cities', *Journal of Urban Economics*, vol. 70(2-3), pp. 99-114.
- Harrison and List, J. (2004). 'Field experiments', *Journal of Economic Literature*, vol. 42(4), pp.1009-1055.
- Heckman, J. (1998). 'Detecting discrimination', *Journal of Economic Perspectives*, vol. 12, pp. 101-116.
- Kaufman, S. (2011). 'Does the implicit association test (IAT) really measure racial prejudice? Probably not', *Psychology Today*, January 28.
- Jacquemet, N. and Yannelis, C. (2012). 'Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market', *Labour Economics*, vol.19, pp. 824-832.
- Jowell, R., and Prescott-Clarke, P. (1970). 'Racial discrimination and white-collar workers in Britain', *Race*, vol. 11(4), pp. 397-417.
- Kaas, L. and Manger, C. (2011). 'Ethnic discrimination in Germany's labour market: A field experiment', *German Economic Review*, vol. 13(1), pp. 1-20.

- Lahey, J. (2008). 'Age, women and hiring: An experimental study', *Journal of Human Resources*, vol. 43(1), pp. 30-56.
- Lahey, J. and Beasley, R. (2009). 'Computerizing audit studies', *Journal of Economic Behavior and Organisation*, vol. 70, pp. 508-514.
- Levitt, S. and List, J. (2009). 'Field experiments in economics: The past, the present, and the future', *European Economic Review*, vol. 53, pp. 1-18.
- List, J. (2004). 'The nature and extent of discrimination in the marketplace: Evidence from the field', *Quarterly Journal of Economics*, vol. 119(1), pp. 49-89.
- List, J and Rasul, I. (2011). 'Field experiments in labor economics', *Handbook of Labor Economics*, vol. 4a, chapter 2.
- McGinnity, F. and Lunn, P. (2011). 'Measuring discrimination facing ethnic minority job applicants: an Irish experiment', *Work, Employment and Society*, vol. 25(4), pp. 693-708.
- MacRae, G. and Lavery, L. (2006). *Discrimination Doesn't Work: Disabled people's experience of applying for jobs in Scotland*. Scotland: Leonard Cheshire.
- Maurer-Fazio, M. (2012). 'Ethnic discrimination in China's internet job board labor market', *IZA Journal of Migration*, 1:12
- Neumark, D. (2012). 'Detecting discrimination in audit and correspondence studies', *Journal of Human Resources*, vol. 47, pp. 1128-1157.
- Nunley, J., Owens, M. and Howard, R.S. (2011). 'The effects of information and competition on racial discrimination: Evidence from a field experiment', *Journal of Economic Behavior and Organisation*, vol. 80, pp. 670-679.

- Oreopoulos, P. (2011). 'Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes', *American Economic Journal: Economic Policy*, vol. 3, pp. 148-171.
- Pager, D. (2007). 'The use of field experiments for studies of employment discrimination: Contributions, critiques and directions for the future', *The Annals of the American Academy of Political and Social Sciences*, vol. 609(1), pp. 104-133.
- Pager, D. (2003). 'The mark of a criminal record', *American Journal of Sociology*, vol. 108, pp. 937-975.
- Pager, D., Western, B. and Bronowski, B. (2009). 'Discrimination in a low-wage labor market: A field experiment', *American Sociological Review*, vol. 74 (October), pp. 777-799.
- Petit, P. (2007). 'The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector', *Labour Economics*, vol. 14, pp. 371-391.
- Riach, P. and Rich, J. (2010). 'An experimental investigation of age discrimination in hiring in the English labour market', *Annales d'Economie et Statistique*. vol. 99/100, pp. 169-185.
- Riach, P. and Rich, J. (2007). 'An experimental investigation of age discrimination in hiring in the Spanish labor market', *IZA Discussion Paper* No. 2654.
- Riach, P. and Rich, J. (2006a). 'An experimental investigation of sexual discrimination in hiring in the English labor market', *B.E. Journal of Economic Analysis and Policy*, 6(2).
www.bepress.com/bejeap/advances/6,issue2
- Riach, P. and Rich, J. (2006b). 'An experimental investigation of age discrimination in hiring in the French labor market', *IZA Discussion Paper* No. 2522.

- Riach, P. A. and Rich, J. (2002). 'Field experiments of discrimination in the market place', *Economic Journal*, vol. 112(483), pp. F480-F518.
- Rooth, D-O. (2010). 'Automatic associations and discrimination in hiring: Real world evidence', *Labour Economics*, vol. 17, pp. 523-534.
- Rooth, D-O. (2009). 'Obesity, attractiveness, and differential treatment in hiring: A field experiment', *Journal of Human Resources*, vol. 44(3), pp. 710-735.
- Seattle Office for Civil Rights (2011). Fair housing testing conducted by the Seattle Office for Civil Rights (SCOR).
www.fairhousingcoach.com/article/seattle-tests-show-frequent-rental-housing-discrimination
- Seitles, M. (1998). 'The perpetuation of residential racial segregation in America: Historical discrimination, modern forms of exclusion, and inclusionary remedies', *Journal of Land Use and Environmental Law*, vol. 14(1), pp. 89-123.
- Siddique, Z. (2011). Caste-based discrimination: Evidence and policy', *Labour Economics*, vol. 18(S1), pp. S146-S159.
- Tilcsik, A. (2011). 'Pride and prejudice: Employment discrimination against openly gay men in the United States', *American Journal of Sociology*, vol. 117 (2), pp. 586-626.
- Tinsley, M. (2012). 'Too much to lose: Understanding and supporting Britain's older workers'. Policy Exchange,
<http://www.policyexchange.org.uk/publications/category/item/too-much-to-lose-understanding-and-supporting-britain-s-older-workers>
- Turner, M. and Ross, S. (2003a). *Discrimination in Metropolitan Housing Markets: Phase 2 – Asians and Pacific Islanders Final Report* March. Washington DC: The Urban Institute Press.

- Turner, M. and Ross, S. (2003b). *Discrimination in Metropolitan Housing Markets: Phase 3 – Native Americans* September. Washington DC: The Urban Institute Press.
- Turner, M., Herbig, C., Kaye, D., Fenderson, J. and Levy, D. (2005). *Discrimination against persons with disabilities: Barriers at every step*. Washington DC: The Urban Institute Press.
- Turner, M., Ross, S., Galster, G. and Yinger, J. (2002). *Discrimination in Metropolitan Housing Markets: National results from phase 1 HDS 2000. Final Report* November. Washington DC: The Urban Institute Press.
- Weichselbaumer, D. (2013). 'Testing for discrimination against lesbians of different marital status: A field experiment', *IZA Discussion Paper* No. 7425.
- Wood, M., Hales, J., Purdon, S., Sejersen, T. and Hayllar, O. (2009). *A Test for Racial Discrimination in Recruitment Practice in British Cities*. Department of Works and Pensions Research Report No. 607.
- Wozniak, A. (2012). 'Discrimination and the effects of drug testing on black employment', *IZA Discussion Paper* No. 6605.
- Wysienska-Di Carlo, K. and Karpinski, Z. (2014). 'Discrimination facing immigrant job applicants in Poland - results of a field experiment', The XVIII ISA World Congress of Sociology (July 13-19).
- Zhou, X., Zhang, J. and Song, X. (2013). 'Gender discrimination in hiring: Evidence from 19,130 resumes in China', <http://mpra.ub.uni-muenchen.de/43543/>
- Zussman, A. (2013). 'Ethnic discrimination: Lessons from the Israeli online market for used cars', *Economic Journal*, vol. 123, pp. F433-F468.