# Correspondence Studies

Pieter-Paul Verhaeghe

## Contents

**Abstract**

Correspondence tests are the golden standard to examine discriminatory behavior in the real world. This chapter starts with distinguishing correspondence tests from related methods, such as situation tests, mystery calls, and mystery visits on the basis of three criteria. Afterwards, it briefly discusses the history of correspondence tests and its main applications during the past decades on the labor, housing, and consumer markets. Next, it considers a few methodological issues that should be taken into account while conducting correspondence tests:

P.-P. Verhaeghe (✉)
Department of Sociology, Vrije Universiteit Brussel, Brussels, Belgium
e-mail: Pieter-Paul.Verhaeghe@vub.be

1

matched versus randomly assigned testing; signaling the discrimination ground; getting ethical approval for testing; and the representativeness of the tested subjects. Finally, this chapter ends with recommendations. Future research should combine correspondence tests with other methods of data collection, expand the scope of contexts and groups, trade-off the advantages of situation tests against the advantages of correspondence tests, and apply correspondence tests for policy applications and evaluations.

## Introduction

For scholars it is often hard to examine discrimination profoundly by means of traditional survey questions, official complaints or qualitative methods. Many employers or landlords are reluctant to openly talk about discrimination, because of social desirability and antidiscrimination legislation. Moreover, some among them are even not aware of their discriminatory behavior because it is driven by unconscious prejudices. In addition, it is also difficult for job or rental candidates to figure out whether they are discriminated against or not. On the one hand, they could not be invited for a job interview or a visit of a rental dwelling, which would lead them to think they were not sufficiently capable to do the job or rent the house, while in fact they were discriminated against. On the other hand, they might also be turned down and think they were discriminated against, although it was for professional reasons they were not invited. In other words, discrimination can be both under- and overestimated by people. Finally, also for NGO's and policy makers it is often challenging to provide clear and convincing evidence of discrimination when they intervene in the public debate, design their antidiscrimination policies, or start legal procedures in court.

However, since the 1960s scholars, activists, and politicians have a strong instrument at their disposal to objectively measure discrimination: correspondence tests. These tests refer to a field experimental technique in which pairs of candidates apply by mail or e-mail for a job, house, or another service. The candidates are similar on all relevant characteristics, except for the discrimination ground under scrutiny. One candidate originates from the test group (e.g., ethnic minorities, women, or older people), the other from the control group (e.g., ethnic majority, men, or younger people). Afterwards, scholars examine whether both candidates were treated equally or not. Unequal treatment is assumed to be due to discrimination on the basis of the discrimination ground.

Discrimination can be defined as the unequal, adverse, and unjustifiable treatment of people on the basis of a protected discrimination ground. As a method, correspondence tests have several assets to examine discrimination according this definition. Firstly, correspondence tests directly measure discriminatory behavior and not related but different concepts as attitudes or prejudices. Secondly, the controlled experimental design allows causal interpretations between the discrimination ground and the (un)equal treatment. This results in a high internal validity of correspondence

tests. Thirdly, because they measure behavior into the "field," they have a high external validity too. Finally, since the tested subjects (e.g., employers or landlords) are not aware that they are tested, they do not suffer from social desirability problems like surveys or interviews. For all these reasons, they are called the "golden standard" to measure discrimination (Heath and Di Stasio 2019) and are increasingly used by academic researchers and advocacy groups to examine discrimination in housing and labor markets and beyond (Riach and Rich 2002; Rich 2014; Zschirnt and Ruedin 2016; Flage 2018; Gaddis 2018; Quillian and Midtbøen 2021).

This chapter starts with distinguishing correspondence tests from related field methods, such as situation tests, mystery calls, and mystery visits. Afterwards, it briefly discusses the history of correspondence tests and its main applications during the past decades. In the following section, it gives a concise, but profound overview of a few potential difficulties and pitfalls that come along with performing correspondence tests. Finally, this chapter ends with making a few general recommendations for further research and application. This chapter has a general public of scholars, activists, and policy makers in mind. Therefore, the aim is not to provide an in-depth analysis of all methodological and substantial issues of correspondence testing, but to give a sound and updated introduction to the method. Nevertheless, even scholars and practitioners experienced with the method might find this chapter insightful.

## A Typology of Discrimination Tests

Correspondence tests are a specific method that falls under the broader umbrella of discrimination tests and field experiments. It should also be distinguished from related techniques to measure discrimination in the field, such as situation tests or mystery shopping. To further add the complexity, different words are used for the same method in different parts of the world. This section aims to shed light on the different methods which measure discriminatory behavior into the field on the basis of three criteria. This typology of discrimination tests has been previously published as part of the overview article of Verhaeghe and Van der Bracht (2017).

A first important difference in the typology is between correspondence or situation tests on the one hand and mystery shopping on the other. This distinction is based on two criteria: (1) Is there a discriminatory request or not? and (2) Is there a test and control group or not? In the case of mystery shopping a discriminatory request is asked and there is no control group. In the case of correspondence or situation testing there is no discriminatory question, but there are both test and control groups (and consequently an experimental design). An additional distinction is based on a third criterion: (3) Is there personal contact or not between the tester and tested subject? There can be personal contact (by phone or in real life) or no personal contact (by mail or e-mail). Mystery calls or mystery visits refer to mystery shopping with personal contact, while mystery mails have reference to mystery shopping without personal contact. Situation tests is the appropriate term when there is an experimental design with test and control groups, no discriminatory request, and

**Fig. 1** A typology of discrimination tests

Discriminatory request?

| | YES | NO |
|---|---|---|
| YES | Mystery calls & Visits | Situation tests |
| NO | Mystery mails | Correspondence tests |
| | NO | YES |

Personal contact?

Test- and control group?

personal contact between tester and tested subjects during the experiment. Finally, correspondence tests refer to the same experimental design, but without personal contact. These three criteria result in a typology of discrimination tests depicted in Fig. 1 below.

With respect to the used terminology, there might be some variation between scholars, disciplines, and geographical regions. Some scholars tend to use "audit studies" to denote only situation tests (e.g., Riach and Rich 2002; Fibbi et al. 2021), while other – especially in the United States – use "audit studies" to refer to the broader category of both situation and correspondence tests (e.g., Gaddis 2018). In addition, some scholars prefer to use "field experiments" as an umbrella term to refer to both situation and correspondence tests (e.g., Pager 2007; Rich 2014; Zschirnt and Ruedin 2016; Heath and Di Stasio 2019; Quillian and Midtbøen 2021).

## Discriminatory Request or Not?

The answer to this question is the main criterion to distinguish mystery shopping from other discrimination tests. Mystery shopping originates from the private world in which commercial or production processes where tested (e.g., mystery guests are testing the customer services in shops by asking questions to salesclerks). The technique can, however, also be applied to test discrimination. On the rental housing market, for example, a (fictious) landlord can ask a real estate agent to let a rental dwelling, but "preferably not to foreigners" (see Van den Broeck and Heylen 2015;

Verstraete and Verhaeghe 2020 for applications). In contrast to field experiments, mystery shopping has been much less used to examine discrimination, because one can only test organizations where one can "play" a client and ask a discriminatory request (e.g., real estate agencies, taxi companies, temporary employment agencies, or selection offices). Another disadvantage is that a positive answer to a discriminatory request in a mystery call only reveals a willingness to discriminate, but not real discriminatory behavior. A taxi company might, for example, answer that they will send a native driver in order to get the client, but send a driver of foreign origin afterwards to the client. Although this might seem very unlikely, it is not impossible.

## Test and Control Group or Not?

This criterion is central to the experimental design of correspondence and situation tests. The test group usually consists of the candidates against which discrimination might occur (e.g., ethnic minorities, females, older candidates, or people who are disabled), while the control group involves the candidates who are usually not discriminated against (e.g., natives, males, younger candidates, or people without disabilities). By using a control group ("null treatment") and test group ("positive treatment") whose difference only lies in the discrimination ground ("treatment condition"), one can draw causal conclusions from the discrimination ground to the potential (un)equal treatment (Fix and Struyk 1993; Pager 2007; Baldassarri and Abascal 2017). This results in a high internal validity of field experimental data. It is, however, key that units are randomly assigned to the test and control groups at every stage of a study in order to eliminate possible systematic errors (Jackson and Cox 2013). For correspondence tests on hiring, for example, this implies that the résumés should be made as similar as possible and be assigned randomly to test and control groups. In addition, also vacancies should be allocated at random to the test and control groups, in the case each vacancy is only tested once.

## Personal Contact or Not?

Key in the design of field experiments is that test and control groups only differ with respect to their discrimination ground under scrutiny. It is precisely this feature that is difficult to control when there is personal contact between the tester and tested subjects in the case of situations tests. Heckman and Siegelman (1993) are criticizing the method for – among other things – being unable to control all characteristics of candidates that might affect the decision-making process of employers or landlords. First of all, scholars do not have a list of all characteristics that might influence the perceived productivity for doing a job or reliability for renting a dwelling. In addition, even when they train extensively the testers – often volunteers or actors – to behave as similar as possible during a house visit, a phone call, or a job interview, it is still possible that they differ on verbal or nonverbal elements that matter for the outcome. Finally, situation tests are not double blind because the testers know the

purpose of the study. This raises the possibility that testers alter their behavior in order to find discrimination. In the same vein, a similar critique can be made for mystery shopping, where mystery callers could also insist in their discriminatory request to confirm the study purpose of finding discrimination. Therefore, most scholars increasingly prefer to use correspondence tests without personal contact instead of situation tests (Zschirnt 2016; Gaddis 2018). Additional reasons are that correspondence tests are less time-consuming to perform, and the growing popularity of online applications on both housing and labor markets makes it more valid to perform correspondence tests.

## Its History and Applications

The technique of discrimination tests dates back to two studies of British sociologists at the end of the 1960s (Daniel 1968; Jowell and Prescott-Clarke 1970). They used matched pairs of white and non-white applicants to uncover discrimination on the housing and labor markets. From the seventies, the method has been increasingly used by university researchers, independent research agencies, and advocacy groups (Riach and Rich 2002; Quillian and Midtbøen 2021). In his overview, Gaddis (2018) differentiates four waves of situation and correspondence tests.

The first wave encompasses the period from the early 1970s to the mid-1980s (Gaddis 2018). It consists of a limited number of tests performed mainly by civil rights organizations in partnership with academics and often commissioned by the government, such as the Department of Housing and Urban Development (HUD) and the Urban Institute in the USA. In the USA, the emphasis was mainly on housing discrimination, while in Europe discrimination in hiring was more the research focus. Zschirnt (2016) explains this divergence from the different antidiscrimination laws in these regions with the Civil Rights Act (also known as the Fair Housing Act) of 1968 in the USA and the influential Race Relations Act of 1968 (which specifically addressed the issue of employment) in the UK. In addition, most discrimination tests during the first wave were performed in-person and examined ethnic or racial discrimination, though some studies also examined gender, disability, and marital status as discrimination grounds (Gaddis 2018).

The second wave of discrimination tests ranges from the late 1980s through the late 1990s (Gaddis 2018). During this wave, there was the implementation of a few large-scale projects, such as the HUD housing and the Urban Institute's employment studies in the USA and the research conducted through the International Labour Office (ILO) in Europe. Next to the larger scale, the US housing studies also started to examine discrimination against Hispanics and racial steering. While most tests were still in-person with trained assistants who physically participated in the selection processes, more efficient and cheaper telephone-based and correspondence tests were increasingly used during this period.

This tendency consolidated during the third wave, which covered the period from the early 2000s through the late 2000 (Gaddis 2018). Because rental and employment applications were progressively taking place over the internet also the situation

tests with personal contact made way for the more standardized and time-efficient correspondence tests without personal contact. The consequence was a strong increase in the number of correspondence studies in both the USA and Europe (Gaddis 2018; Rich 2014; Zschirnt and Ruedin 2016; Quillian and Midtbøen 2021). A good overview of all studies can be found in the meta-analysis of Judith Rich in 2014. In addition, scholars aimed more and more to not only document the existence of discrimination, but also to explore the underlying mechanisms and conditions of discrimination (Gaddis 2019a).

The current wave of correspondence tests took off around the early 2010s (Gaddis 2018). During this wave, the number of studies grew exponentially with more countries, more domains, and more discrimination grounds being investigated through correspondence tests. Moreover, scholars paid more attention to methodological issues of testing (e.g., how to validly signal ethnic origin in written applications); combining correspondence tests with other methods in order to further examine the mechanisms and drivers of discrimination; the geographical and market contexts of discrimination; and the types of discrimination (e.g., tasted-based or statistical discrimination).

During the past decades, most situation and correspondence tests have been applied to examine discrimination on the labor and housing markets (Riach and Rich 2002; Rich 2014). Recent meta-analyses have shown that there is substantial ethnic discrimination in hiring. Zschirnt and Ruedin (2016) found that across OECD countries minority groups have 49% lower odds to be invited for a job interview than their majority groups. In the United States, there appears to be no change in the level of hiring discrimination against African Americans and a modest decline in discrimination against Latinos (Quillian et al. 2017). And also in the United Kingdom, there is no significant diminution in discrimination over time for candidates with typically sounding Caribbean or South-Asian names (Heath and Di Stasio 2019). Nevertheless, Lippens and his colleagues (2021) found a recent decrease in hiring ethnic discrimination in Europe.

Next to ethnic and racial minorities, also other groups are discriminated against on the labor market. In his extensive meta-analysis of correspondence studies, Louis Lippens and his colleagues (2021) examined and compared multiple grounds of discrimination between 2005 and 2020 (during the last two waves in Gaddis's typology of 2018). They found that there is hiring discrimination against older candidates, candidates with disabilities and less psychically attractive candidates as much as there is against ethnic minorities, in addition to more modest discrimination on the basis of religion, wealth, and marital status. The finding about religion corroborates with the meta-analysis conducted by Bartkoski and his colleagues (2018), which found discrimination against Muslim and Arab people on the labor market. Moreover, Lippens et al. (2021) found no overall evidence of hiring on the basis of gender, motherhood, or military service/affiliation. Furthermore, Flage (2020) found in his meta-analysis substantial discrimination against gay job applicants and to a lesser degree also against lesbian candidates in the OECD countries. Lippens et al. (2021), however, claim that this is because of their activism in LGB+ rights organization rather than their sexual orientation in itself.

For the housing market, Flage (2018) showed in his meta-analysis that ethnic majority candidates are almost twice as likely to be invited to view a vacant dwelling than ethnic minority candidates. In addition, two meta-analyses have found decreasing levels of ethnic discrimination during the past decades (Auspurg et al. 2019; Quillian et al. 2020a), especially the more exclusionary forms of discrimination (e.g., not receiving a response and so-called unavailability of advertised unit) appear to decline. With respect to gender, females tend to be more invited than males (Flage 2019) and these gender differences are more pronounced among ethnic minorities (Flage 2018). With respect to sexual orientation, Flage (2021) found that gay males are significantly discriminated against while lesbians are not. In general, private landlords tend to discriminate less than real estate agents (Flage 2018).

More recently, correspondence and situation tests have also been applied in other domains than work and housing. Without having the intention to be exhaustive, important studies have been carried out to examine ethnic, social class, and LGBTI discrimination in the access to kindergartens and primary schools (e.g., Bourabain et al. 2020; de Lafuente 2021; Diaz-Serrano and Meix-Llop 2016; Olsen et al. 2020), ethnic and social class discrimination in primary health care (e.g. Olah et al. 2013), racial discrimination on online market places (e.g. Ayres et al. 2015; Zussman 2013), ethnic and gender discrimination during shopping (e.g. Bourabain and Verhaeghe 2019), ethnic discrimination in the way elected representatives or local civil servants respond to their constituents (e.g., Gell-Redman et al. 2018; Simonovits et al. 2021), ethno-religious discrimination in the wedding venue business (e.g., Carol et al. 2021), ethnic discrimination in the access to sport teams (e.g., Nesseler et al. 2019; Gomez-Gonzalez et al. 2021), and racial discrimination in the sharing economy (e.g., Edelman et al. 2017; Simonovits et al. 2018; Cui et al. 2020). Although there is little tradition in conducting correspondence tests in these domains, these studies are promising and contribute to the understanding of discrimination.

## Methodological Issues to Consider

Although correspondence tests are the "golden standard" to measure discriminatory behavior, there are a few methodological issues that should be taken into account while conducting them. Below four important aspects are discussed.

## Random Assignment or Pairwise Matched?

Correspondence tests differ in the extent to which they are "pairwise matched" versus "randomly assigned" to job or accommodation vacancies. In the case of pairwise matched tests, the two candidates apply for the same vacancy, whereas with randomly assigned, unmatched tests the candidates apply for different vacancies. There are benefits associated with both types of designs.

From the perspective of statistical efficiency, scholars aim to examine discrimination with an as small as possible sample size. Following this statistical point of

view, one might think that matched designs are the most efficient because every tested vacancy results in two observations (one for the test group and one for the control group), while unmatched designs only deliver one observation per tested vacancy. Vuolo and his colleagues (2016, 2018) have shown, however, that this is not always true. It depends on the level of concordance in your matched design, which refers to the proportion of vacancies where the test and control groups are treated equally (both not invited or both invited). When the concordance is above 0.5 in the population, pairwise matched tests are statistically more efficient. When the concordance is below 0.5, the unmatched, randomly assignment design is more efficient. This concordance level can be anticipated on the basis of previous studies or by means of a small pilot test.

Next to purely statistical arguments, there are also other methodological aspects that should be considered. First of all, the burden of testing for the tested people (e.g., employers, landlords, or schools) is twice as large in the case of matched tests compared with unmatched tests. If the application process requires a substantial amount of time, energy, or financial resources for the tested subjects, this is an ethical argument in favor of randomly assigned, unmatched designs (for the ethical reasoning, see the third issue of this section). Secondly, it could be that the population of subjects or organizations that could be tested is limited (e.g., limited number of advertisements or schools in a particular area). In that case, the matched design in which each subject or organization is tested twice could be more useful (Vuolo et al. 2018). Some studies even test the same subject three or four times in order to increase the sample size (e.g., Bertrand and Mullainathan 2004). Thirdly, during correspondence testing fictious candidates apply for a job, rental dwelling or another service, without the tested subjects knowing that they are tested. If the tested subjects would know that they are tested, they could alter their behavior with biased results as the consequence. The challenge for scholars is, therefore, not to be discovered. Vuolo and his colleagues (2018) argue that utilizing unmatched designs would minimize the likelihood of being discovered because in this approach no single subject would be confronted with two applications that look so similar as to raise suspicion.

## Testing the Signal

Correspondence tests heavily rely on the idea that two applicants are similar, except for the discrimination ground under scrutiny. This discrimination ground is manipulated by a signal to an employer or landlord. In a résumé or rental application, for example, background characteristics could be explicitly mentioned, such as the age, sex, family composition, disability, or ethnic origin. However, since the seminal studies of Bertrand and Mullainathan (2004) and Carpusor and Loges (2006) on the housing market, researchers often rely on names to implicitly signal the ethnic origin of applicants (Gaddis 2018). The assumptions here are that the tested people reliably and validly perceive the ethnicity of applicants through their names and these names only signal ethnicity and/or sex and no other relevant characteristics (such as

socioeconomic status or religiosity) that could impact the outcome. Moreover, these assumptions are seldom tested and recent research showed that they are often fallacious.

In the United States, it appears that the correct ethnic interpretation of black and Hispanic names depends on the gender, the popularity, the type of first and last named used, and the social class associated with the name (Gaddis 2017a, b, 2019c). Names more commonly given by highly educated black mothers are, for example, less likely to be perceived as black than names given by less educated black mothers (Gaddis 2017a) and the combination of Hispanic first and last names are better interpreted as Hispanic than mixed first and last names (Gaddis 2017b). Furthermore, mixed names also signal another generational status than ethnically homogeneous names (Gaddis 2019b). Moreover, the correct ethnic interpretation of names is dependent of the racial composition of the surrounding context (Crabtree and Chykina 2018).

The perception of names has been less examined outside the USA yet, but Martiniello and Verhaeghe (2021) have found in Belgium that people only succeed in properly distinguishing Belgian from non-Belgian names, but that there is substantial error in attributing the correct specific ethnic origin to non-Belgian names. Moreover, there appears to be also much heterogeneity in the perception of names within each ethnic group in terms of social class, educational level, gender, and religiosity.

This variation might affect the outcomes of correspondence studies. For example, minority candidates with working-class names might be less likely to be invited by an employer for a job interview than majority candidates with middle class names (or vice versa), not (only) because of their ethnic origin, but rather due to the social class connotations of the names used in the correspondence tests. Therefore, the general recommendation is to test beforehand the perception of names used in correspondence studies among the research population (e.g., employers or realtors) instead of just assuming that they are perceived properly (Gaddis 2018; Martiniello and Verhaeghe 2021).

## Getting Ethical Approval

Correspondence tests rely on covert research in which tested subjects are not aware that they are tested and are to some extent misled by fictious candidates. Therefore, conducting correspondence tests violate key principles for ethical research, such as informed consent, voluntary participation, and the absence of deception (Banton 1997; Riach and Rich 2004). In an in-depth discussion of these ethical aspects, Eva Zschirnt (2019) shows that most scholars succeed in getting ethical approval from ethical commissions in most countries to conduct correspondence tests under strict circumstances. The core arguments are that covered correspondence tests are the only method to measure discriminatory behavior in a reliable and valid way, that the negative consequences for the tested subjects are minimal, that informing subjects beforehand would bias the results, and that conducting research about discrimination

is a legitimate goal. Following this argumentation, most ethical committees tend to approve covered correspondence testing without informed consent. However, the outcome of requests for ethical approval always remains uncertain and might depend on the national context and the specific composition of ethical committees.

Zschirnt (2019) really emphasizes the importance of keeping the negative consequences for tested subjects as minimal as possible. It is against this background that scholars could prefer unmatched above matched test designs in order to minimize the time investments of, for example, landlords or employers in fake applications from correspondence tests (Vuolo et al. 2018). In the same vein, this chapter would also make a plea to never schedule concrete appointments for job interviews with employers or house visits with landlords, or – if necessary – cancel these appointments politely well in advance. In addition, data from correspondence tests should be anonymized or pseudomized during the research process and results should be published anonymously afterwards – preferable on the aggregate and not the individual level – in order to prevent reputational damage or legal consequences for tested scholars (Zschirnt 2019). However, this last recommendation only applies for scholars who conduct correspondence tests for research aims and not for activist purposes, such as legal enforcement.

## Representativeness of the Tested Subjects

The aim of correspondence tests is to examine discriminatory behavior by a particular population in a particular market or life domain. For example, to study discrimination by employers on the labor market in a country. It is key to select the tested subjects in such way that they are representative for the research population. This is in most studies challenging, because they have to rely for parsimonious reasons on sampling frames which do not precisely represent the research population. Correspondence studies about hiring and rental discrimination, for example, typically rely on online platforms with job or rental advertisements from which they draw at random advertisements that could be tested (Lahey and Beasley 2018). These online job and rental sites as sampling frames are, however, not representative for all selection processes of employers or landlords in a country. Firstly, they only cover the officially advertised jobs and rental dwellings and leave the informal vacancies out of the picture. To the extent that there is more discrimination occurring in these informal selection processes, these correspondence studies might underestimate the level of discrimination. Moreover, the penetration of these job and rental websites is not equal across geographic markets or economic sectors. For example, real estate agents may make more use of rental platforms than smaller private landlords, or rental advertisements from urban areas may be more posted on the national rental sites, while advertisements from the countryside may use more local sites or offline channels. Boeing (2019), for example, showed that most cities' online housing markets in the USA are digitally segregated by race and class. The same goes for job advertisements; some economic sectors may rely on national job sites to post their vacancies, while others may utilize more sector-specific platforms. A recent

study in the USA has, however, found that racial discrimination does not vary by the source of online job postings (Pedulla et al. 2021). Nevertheless, these caveats have important consequence for the representatives of correspondence studies and scholars should always be aware about the extent to which their tests are (not) representative for their research population.

## Directions for Further Research and Applications

Notwithstanding the long and rich research tradition of correspondence studies, there are still many questions and topics that need to be investigated. Below five possible avenues for future research are spelled out.

## Combination with Other Methods of Data Collection

Although correspondence tests are still the "golden standard" to investigate the occurrence and patterns of discrimination (Heath and Di Stasio 2019), they are less suited to examine the underlying mechanisms and preceding causes of discrimination: the so-called Why? and How? (Gaddis 2018, 2019a). A promising avenue is, therefore, to combine correspondence data about discriminatory behavior with additional data about the mechanisms and drivers of discrimination. There are at least two ways to do this.

On the one hand, scholars could combine correspondence test data with data about the context in which discrimination occurs, such as the market, political, or geographical context. Recent studies have, for example, examined the association between discrimination and market tightness or slackness (e.g., Baert et al. 2015; Vuolo et al. 2017), local anti-immigrant attitudes (e.g., Carlsson and Rooth 2012; Carlsson and Eriksson 2017; Verhaeghe and De Coninck 2021), the socioeconomic or ethnic neighborhood composition (e.g., Carlsson and Eriksson 2014; Hanson and Santas 2014; Ghekiere and Verhaeghe 2022), and the Covid-19 pandemic (e.g., Verhaeghe and Ghekiere 2021).

On the other hand, researchers could add individual level factors to the correspondence test data, such as personal attitudes, legitimations, personal history, or voting behavior. In this strategy, the same sample of subjects that has been tested through correspondence tests should be further examined through surveys, in-depth interviews, or mystery shopping. It is, however, very challenging to carry out uninformed correspondence tests and informed surveys or interviews among the same sample of landlords or employers, without biases due to selective drop-out or cross-mode test effects. Only a handful studies have managed to make this individual-level combination of data. Noteworthy examples are the studies of Agerström and Rooth (2011) and Rooth (2010) which combined correspondence tests about hiring discrimination with implicit association tests (IATs) and explicit attitudes questions; the study of Zussman (2013) investigated discrimination on the automobile market by means of both correspondence tests and a telephone survey;

and a few studies interviewed landlords and employers who have been tested through correspondence tests (e.g. Midtbøen 2015; Bonnet et al. 2016).

## Expand the Scope of Studied Contexts and Groups

Although correspondence tests have been widely used during the past decades, there are still many contexts and groups which have not been examined yet. The vast majority of published correspondence studies have been conducted in Northern America, Australia, or Europe. Large parts of the African, South-East Asian, and South-American continents have not been covered yet by correspondence studies (exceptions are a handful studies about Brazil, Mexico, Peru, Argentina, Philippines, Singapore, Malaysia, and Algeria, see Lippens et al. 2021 and Quillian and Midtbøen 2021 for overviews with respect to hiring discrimination). Some influential meta-analyses also limit themselves to include only the OECD countries (e.g., Zschirnt and Ruedin 2016; Flage 2018) or Europe and/or Northern America (e.g., Auspurg et al. 2019; Quillian et al. 2017; Quillian et al. 2020a). And even within Europe, there is a geographical imbalance with only few correspondence studies in East and Central Europe. This causes a fundamental Western bias in the knowledge about discrimination. The levels and patterns of discrimination might be totally different depending on the national context, and the larger the variation in investigated contexts, the better scholars could examine the role of contextual factors.

In addition, also the discrimination against several minority groups has been little investigated yet through correspondence tests. With respect to discrimination on the grounds of ethnic origin and citizenship status, for example, there is a remarkable lack of correspondence studies that examines discrimination against refugees, asylum seekers, irregular migrants, and the Roma communities in Europe (Bartoš et al. 2016 is a noteworthy exception for discrimination against Roma in Czechia). In the same vein, there are almost no correspondence studies about discrimination on the basis of political orientation or union membership (Lippens et al. 2021). This scarcity also constrains the understanding of discrimination. Therefore, this chapter recommends to expand both the geographical scope and groups that are studied with correspondence tests.

## Using Again More Situation Tests

Following the Heckman and Siegelman (1993) critique, scholars increasingly prefer to use correspondence tests without personal contact instead of situation tests with personal contact (Zschirnt 2016; Gaddis 2018). A major drawback of correspondence tests is that it could only capture part of human behavior prone for discrimination. On the labor and housing markets, it usually only measures discrimination in the initial contact phase of the selection process: getting an appointment for a job interview or a visit of the rental dwelling. To the extent that discrimination also occurs during and after a job interview and a house visit, correspondence tests

systematically underestimate the prevalence of discrimination (Cherry and Bendick 2018). A recent study, for example, found that there is considerable additional ethnic discrimination in hiring after the invitation for a job interview (Quillian et al. 2020b). Moreover, it could be that recent trends in decreasing ethnic discrimination on the labor and housing markets (Auspurg et al. 2019; Quillian et al. 2020a; Lippens et al. 2021) only reveal a decrease in discrimination in the initial contact phase, but not in the later stages of the selection process. In other words, it could be that ethnic discrimination has not been diminished, but only shifted. In addition, many aspects of workplace discrimination are left out of the picture with correspondence tests, such as bullying and promotion. In the same vein, other domains of life are very hard to examine through correspondence tests, such as discrimination during shopping or restaurant visits, in college admissions, and in the access to bars and dance clubs. These are all reasons to make more use again of situation tests in future research. Scholars should trade-off the disadvantages of less standardization and larger time investment of situation tests with personal contact against the advantages of having a more complete assessment of selection procedures and more possibilities to examine other life domains.

## Policy Applications and Evaluations

Correspondence and situation tests were initially developed against the background of activist scholarship that aimed to tackle racism, sexism, and discrimination (Fix and Struyk 1993; Cherry and Bendick 2018). They were used to provide evidence for discrimination in order to advocate for antidiscrimination legislation and subsequently to legally sue perpetrators. For these more activist purposes, scholars often collaborated with community groups. Cherry and Bendick (2018) observe, however, that since the 2000s these historical activist purposes of discrimination tests have been de-emphasized under pressure of methodological rigorism and the academic rat-race. They make a plea to go beyond purely academic goals again and use discrimination tests to tackle discrimination by using them in public policy debates, legal enforcement and community organizing.

In the same vein, this chapter would like to add another purpose for using correspondence and situation tests, which is policy evaluation. There are multiple national and international antidiscrimination laws, conventions, and declarations. At the same time, many (local) governments, firms, schools, and other organizations have applied specific measures to mitigate discrimination. The effectivity of these legislations and measures are, however, seldom adequately investigated. There is a high need for sound evaluation studies. Correspondence and situation tests are very suited for this goal, because they measure discriminatory behavior (and not just intentions or attitudes) in the field. A possible application is that meta-analyses about time trends in discrimination investigate to which extent discrimination has been changed since the implementation of a particular legislation or measure. In their meta-analysis, Zschirnt and Ruedin (2016) found, for example, no evidence that ethnic discrimination decreased on the labor market after the adaptation of two EU

directives against direct and indirect discrimination (Directive 2000/43/EC and 2000/78/EC). Another application is that scholars conduct correspondence tests before and after the implementation of antidiscrimination policies and measures in order to have a proper pre- and post-measurement. The challenge here is, however, that academics often arrive "too late" (after the introduction of a policy or measure) to conduct a proper pre-measurement, while non-academics often do not evaluate their policies because of a lack of time, budget, and/or methodological expertise. For this latter application, scholars need to cooperate again closer with policy makers, lawyers, NGOs, employers, or landlord organizations.

## Summary

The aim of this chapter was to give a sound and updated introduction to the method of correspondence tests. These tests refer to a field experimental method in which pairs of candidates apply by mail for a job, house, or another service. The candidates are similar on all relevant characteristics, except for the discrimination ground under scrutiny. If the candidates are treated unequally by employers, landlords, or other service providers, this inequality is assumed to be due to discrimination. Because correspondence tests measure discriminatory behavior in a very standardized, time-efficient, and controlled way in the field, they are superior to related field techniques, such as situation tests and mystery shopping. Therefore, they are increasingly used to examine discrimination in housing and labor markets (Riach and Rich 2002; Rich 2014; Zschirnt and Ruedin 2016; Flage 2018; Gaddis 2018; Quillian and Midtbøen 2021). Moreover, they are also progressively used to investigate discrimination in other life domains, such as education, sport, health care, and shopping. Along with their growing popularity, also their methodological aspects were more and more the subject of research, such as the choice between matched and unmatched tests (e.g. Vuolo et al. 2018), signaling the discrimination ground (e.g., Gaddis 2017a; Crabtree and Chykina 2018; Martiniello and Verhaeghe 2021), and the representativeness of the tested subjects (e.g., Lahey and Beasley 2018; Auspurg et al. 2020; Pedulla et al. 2021).

Correspondence studies have been predominantly conducted to examine the prevalence and patterns of discrimination, but are less suited to investigate the underlying mechanisms and preceding drivers of discrimination (Gaddis 2019a). In the combination with additional methods of data collection (such as surveys or interviews), scholars could have strong research designs to shed new light on the why and how of discrimination. This is a very promising avenue for future research.

Notwithstanding the exponential growth of correspondence studies, they are still limited to mainly applications in the West. Moreover, some groups and discrimination grounds are still very little or even not investigated through correspondence tests, such as discrimination against Roma communities, refugees, asylum seekers, irregular migrants, and political minorities. These biases constrain the understanding of discrimination. Another limitation is that they are seldom used by scholars for policy evaluations or to tackle discrimination. This is in contrast to the more activist

scholarship tradition of situation and correspondence testing before the turn of the century (Cherry and Bendick 2018). Finally, the growing popularity of correspondence tests because of their high levels of standardization and time-efficiency is at the expense of the study of discrimination in latter stages of selection processes and in life domains where written correspondence is less common. In these situations, scholars should trade-off the disadvantages of less standardization and larger time investment of situation tests with personal contact against the advantages of having a more complete assessment of selection procedures and more possibilities to examine other life domains.

## Cross-References

▶ Design and Analysis of Experiments
▶ Discrimination Against Transgender Employees and Jobseekers
▶ Discrimination Due to Sexual Orientation
▶ Ethnicity, Race, and Minorities
▶ Gender Discrimination
▶ Group Identity, Ingroup Favoritism, and Discrimination
▶ Skin Color Discrimination

## References

Agerström J, Rooth DO (2011) The role of automatic obesity stereotypes in real hiring discrimination. J Appl Psychol 96:790–805
Auspurg K, Schneck A, Hinz T (2019) Closed doors everywhere? A meta-analysis of field experiments on ethnic discrimination in rental housing markets. J Ethn Migr Stud 1:95–114
Auspurg K, Schneck A, Thiel F (2020) Different samples, different results ? How sampling techniques affect the results of field experiments on ethnic discrimination. Res Soc Stratification Mobil 65:100444
Ayres I, Banaji M, Jolls C (2015) Race effects on eBay. Rand J Econ 46:891–917
Baert S, Cockx B, Gheyle N, Vandamme C (2015) Is there less discrimination in occupations where recruitment is difficult? ILR Rev 68:467–500
Baldassarri D, Abascal M (2017) Field experiments across the social sciences. Annu Rev Sociol 43:41–73
Banton M (1997) The ethics of practice-testing. New Commun 23:413–420
Bartkoski T, Lynch E, Witt C, Cort R (2018) A meta-analysis of hiring discrimination against Muslims and Arabs. Pers Assess Decis 4:1–16
Bartoš V, Bauer M, Chytilová J, Matějka F (2016) Attention discrimination: theory and field experiments with monitoring information acquisition. Am Econ Rev 106:1437–1475
Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Am Econ Rev 94:991–1013

Boeing G (2019) Online rental housing market representation and the digital reproduction of urban inequality. Environ Plann A: Econ Space 52:449–468

Bonnet F, Lalé E, Safi M, Wasmer E (2016) Better residential than ethnic discrimination! Reconciling audit and interview findings in the Parisian housing market. Urban Stud 53:2815–2833

Bourabain D, Verhaeghe PP (2019) Could you help me, please? Intersectional field experiments on everyday discrimination in clothing stores. J Ethn Migr Stud 45:2026–2044

Bourabain D, Verhaeghe PP, Stevens P (2020) School of choice or schools' choice? Intersectional correspondence testing on ethnic and class discrimination in the enrolment procedure to Flemish kindergarten. Race, Ethn Educ. https://doi.org/10.1080/13613324.2020.1798379

Carlsson M, Eriksson S (2014) Discrimination in the rental market for apartments. J Hous Econ 23: 41–54

Carlsson M, Eriksson S (2017) Do attitudes expressed in surveys predict ethnic discrimination? Ethn Racial Stud 40:1739–1757

Carlsson M, Rooth DO (2012) Revealing taste-based discrimination in hiring: a correspondence testing experiment with geographic variation. Appl Econ Lett 19:1861–1864

Carol S, Kuipers C, Koesling P, Kaspers M (2021) Ethnic and religious discrimination in the wedding venue business: evidence from two field experiments in Germany and Austria. Soc Probl. https://doi.org/10.1093/socpro/spab032

Carpusor AG, Loges WE (2006) Rental discrimination and ethnicity in names. J Appl Soc Psychol 36:934–952

Cherry F, Bendick M (2018) Making it count: discriminating auditing and the activist scholar tradition. In: Michael Gaddis S (ed) Audit studies: behind the scenes with theory, method, and nuance. Springer, Cham, pp 45–62

Crabtree C, Chykina V (2018) Last name selection in audit studies. Sociol Sci 5:21–28

Cui R, Li J, Zhang D (2020) Reducing discrimination with reviews in the sharing economy: evidence from field experiments on Airbnb. Manag Sci 66:1071–1094

Daniel W (1968) Racial discrimination in England. Penguin Books, Middlesex

De Lafuente DM (2021) Cultural assimilation and ethnic discrimination: an audit study with schools. Labour Econ 72:102058

Diaz-Serrano L, Meix-Llop E (2016) Do schools discriminate against homosexual parents? Evidence from a randomized correspondence experiment. Econ Educ Rev 53:133–142

Edelman B, Luca M, Svirsky D (2017) Racial discrimination in the sharing economy: evidence from a field experiment. Am Econ J Appl Econ 9:1–22

Fibbi R, Midtbøen AH, Simon P (2021) Migration and discrimination. IMISCOE research series. Springer, Cham

Fix M, Struyk RJ (1993) Clear and convincing evidence: measurement of discrimination in America. The Urban Institute, Washington, DC

Flage A (2018) Ethnic and gender discrimination in the rental housing market: evidence from a meta-analysis of correspondence tests, 2006–2017. J Hous Econ 41:251–273

Flage A (2019) Why is Vince less likely to get housing than Emily? An analysis of the causes of discrimination against male names. Revue d'Économie Polique 130:633–657

Flage A (2020) Discrimination against gays and lesbians in hiring decisions: a meta-analysis. Int J Manpow 41:671–691

Flage A (2021) Discrimination against same-sex couples in the rental housing market: a meta-analysis. Econ Bull 41:643–653

Gaddis M (2017a) How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies. Sociol Sci 4:469–489

Gaddis M (2017b) Racial/ethnic perceptions of Hispanic names: selecting names to test for discrimination. Socius 3:1–11

Gaddis M (2018) An introduction to audit studies in the social sciences. In: Michael Gaddis S (ed) Audit studies: behind the scenes with theory, method, and nuance. Springer, Cham, pp 3–44

Gaddis M (2019a) Understanding the "how" and "why" aspects of racial-ethnic discrimination: a multimethod approach to audit studies. Sociol Race Ethn 5:443–455

Gaddis M (2019b) Assessing immigrant generational status from names: evidence for experiments racial/ethnic and immigrant discrimination. Soc Sci Res Netw:1–33. https://doi.org/10.2139/ssrn.3022217

Gaddis M (2019c) Signalling class: an experiment examining social class perceptions from names used in correspondence audit studies. Soc Sci Res Netw:1–29. https://doi.org/10.2139/ssrn.3350739

Gell-Redman M, Visalvanich N, Crabtree C (2018) It's all about race: how state legislators respond to immigrant constituents. Polit Res Q 71:517–531

Ghekiere A, Verhaeghe PP (2022) How does ethnic discrimination on the housing market differ across neighborhoods and real estate agencies? J Hous Econ 55:101820

Gomez-Gonzalez C, Nesseler C, Dietl HM (2021) Mapping discrimination in Europe through a field experiment in amateur sport. Humanit Soc Sci Commun 8:95

Hanson A, Santas M (2014) Field experiment tests for discrimination against Hispanics in the U.S. rental housing market. South Econ J 81:135–167

Heath A, Di Stasio V (2019) Racial discrimination in Britain, 1969–2017: a meta-analysis of field experiments on racial discrimination in the British labour market. Br J Sociol 70:1774–1798

Heckman JJ, Siegelman P (1993) The Urban Institute audit studies: their methods and findings. In: Fix M, Struyck RJ (eds) Clear and convincing evidence: measurement of discrimination in America. The Urban Institute Press, Washington, DC, pp 187–258

Jackson M, Cox DR (2013) The principles of experimental design and their application in sociology. Annu Rev Sociol 39:27–49

Jowell R, Prescott-Clarke P (1970) Racial discrimination and white-collar workers in Britain. Race & Class 11:397–417

Lahey J, Beasley R (2018) Technical aspects of correspondence studies. In: Michael Gaddis S (ed) Audit studies: behind the scenes with theory, method, and nuance. Springer, Cham, pp 81–102

Lippens L, Vermeiren S, Baert S (2021) The state of hiring discrimination: a meta-analysis of (almost) all recent correspondence experiments. GLO discussion paper, no. 972. Global Labor Organization (GLO), Essen

Martiniello B, Verhaeghe PP (2021) Signaling ethnicity and race through names? The perception of names from an intersectional perspective. Manuscript submitted to PLOS One

Midtbøen A (2015) The context of employment discrimination: interpreting the findings of a field experiment. Br J Sociol 66:193–214

Nesseler C, Gomez-Gonzalez C, Dietl H (2019) What's in a name? Measuring access to social activities with a field experiment. Palgrave Commun 5:160

Olah ME, Gaisano G, Hwang SW (2013) The effect of socioeconomic status on access to primary care: an audit study. Can Med Assoc J 185:E263–E269

Olsen AL, Kyhse-Andersen JH, Moynihan D (2020) The unequal distribution of opportunity: a national audit study of bureaucratic discrimination in primary school access. Am J Polit Sci. https://doi.org/10.1111/ajps.12584

Pager D (2007) The use of field experiments for studies of employment discrimination: contributions, critiques and directions for the future. Ann Am Acad Polit Soc Sci 609:104–133

Pedulla D, Munoz J, Wullert K, Dias F (2021) Field experiments and job posting sources: the consequences of job database selection for estimates of racial discrimination. Sociol Race Ethn. https://doi.org/10.1177/23326492211029336

Quillian L, Midtbøen A (2021) Comparative perspectives on racial discrimination in hiring: the rise of field experiments. Annu Rev Sociol 47:391–415

Quillian L, Pager D, Hexel O, Midtbøen AH (2017) Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. Proc Natl Acad Sci U S A 114:10870–10875

Quillian L, Lee JJ, Honoré B (2020a) Racial discrimination in the U.S. housing and mortgage lending markets: a quantitative review of trends, 1976–2016. Race Soc Probl 12:13–28

Quillian L, Lee JJ, Oliver M (2020b) Evidence from field experiments in hiring shows substantial additional racial discrimination after the callback. Soc Forces 99:732–759

Riach PA, Rich J (2002) Field experiments of discrimination in the market place. Econ J 112: 480–518

Riach PA, Rich J (2004) Deceptive field experiments of discrimination: are they ethical? Kyklos 57: 457–470

Rich J (2014) What do field experiments of discrimination in markets tell us? A meta-analysis of studies conducted since 2000. IZA Discussion paper no. 8584

Rooth DO (2010) Automatic associations and discrimination in hiring: real world evidence. Labour Econ 17:523–534

Simonovits B, Shvets I, Taylor H (2018) Discrimination in the sharing economy: evidence from a Hungarian field experiment on the topic of ridesharing. Corvinus J Sociol Soc Policy 9:55–79

Simonovits G, Simonovits B, Vig A, Hobot P, Nemeth R, Csomor G (2021) Back to "normal": the short-lived impact of an online NGO campaign of government discrimination in Hungary. Polit Sci Res Meth. https://doi.org/10.1017/psrm.2021.55

Van den Broeck K, Heylen K (2015) Differential treatment of rental home seekers according to their sociodemographic and economic status by real estate agencies in Belgium. Eur J Homelessness 9:39–62

Verhaeghe PP, De Coninck D (2021) Rental discrimination, perceived threat and public attitudes towards immigration and refugees. Ethnic Racial Stud. https://doi.org/10.1080/01419870.2021. 1939092

Verhaeghe PP, Ghekiere A (2021) The impact of the Covid-19 pandemic on ethnic discrimination on the housing market. Eur Soc 23(suppl 1):S384–S399

Verhaeghe PP, Van der Bracht K (2017) Praktijktesten: van onderzoeksmethode naar beleidsinstrument tegen discriminatie? Sociologos 38:182–200

Verstraete J, Verhaeghe PP (2020) Ethnic discrimination upon request? Real estate agents' strategies for discriminatory questions of clients. J Housing Built Environ 35:703–721

Vuolo M, Uggen C, Lageson S (2016) Statistical power in experimental audit studies: cautions and calculations for matched tests with nominal outcomes. Sociol Methods Res 45:260–303

Vuolo M, Uggen C, Lageson S (2017) Race, recession, and social closure in the low-wage labor market: experimental and observational evidence. Res Sociol Work 30:141–183

Vuolo M, Uggen C, Lageson S (2018) To match or not to match? Statistical and substantive considerations in audit design and analysis. In: Michael Gaddis S (ed) Audit studies: behind the scenes with theory, method, and nuance. Springer, Cham, pp 119–140

Zschirnt E (2016) Measuring hiring discrimination – a history of field experiments in discrimination research. NCCR working paper #7

Zschirnt E (2019) Research ethics in correspondence testing: an update. Res Ethics 15:1–21

Zschirnt E, Ruedin D (2016) Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. J Ethn Migr Stud 42:1115–1134

Zussman A (2013) Ethnic discrimination: lessons from the Israeli online market for used cars. Econ J 123:433–468