Understanding and Overcoming Biases Against Marginalized Groups: Behavioral and Experimental Evidence From The Netherlands and Burkina Faso

Luis D. Artavia-Mora

ISBN 978-90-6490-123-2

Understanding and Overcoming Biases Against Marginalized Groups: Behavioral and Experimental Evidence From the Netherlands and Burkina Faso

Begrijpen en bestrijden van vooroordelen tegen gemarginaliseerde groepen: resultaten van gedrags- en experimenteel onderzoek in Nederland en Burkina Faso

Thesis

to obtain the degree of Doctor from the

Erasmus University Rotterdam

by command of the

Rector Magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board

The public defence shall be held on

30 November 2020 at 16.00 hrs

by

Luis Daniel Artavia-Mora

born in Minnesota, USA

International Institute of Social Studies

Erasmus University Rotterdam

Doctoral Committee

Doctoral dissertation supervisor

Prof. A.S. Bedi

Other members

Prof. H. Kazianga, Oklahoma State University

Prof. E. Nillesen, Maastricht University

Dr P.R. Nikiema, Université Norbert Zongo de Koudougou

Prof. E.K.A. van Doorslaer

Co-supervisor

Dr M. Rieger

Table of Contents

“The growth and development of people is the highest calling of leadership.” — H. Ford

To my loved ones, to the people that helped me along the way, and,

to all inspiring researchers and mentors around the World

Acknowledgements

The completion of my PhD is the merit of many people around me. I am thankful to my Professors, research assistants, doctoral colleagues, the DEC group, and everyone that helped me during this period. I am grateful to EUR for giving me the opportunity to pursue my Doctoral Studies. This experience has allowed me to become a better researcher and to enhance my skills.

Three professors are at centre of my PhD progress: Professor Arjun Bedi, Dr. Matthias Rieger, and Dr. Natascha Wagner. I am extremely thankful to you Arjun for your guidance, support, and candid mentorship and coaching to help me become a better scholar and to grow during this process. I am truly grateful for your close engagement in my process, and for your encouraging, patient, and uplifting approach to promote students. Thanks Matthias, for following my areas of work and interest, and for your remarks since the master. Thank you Natascha for your motivation, positive attitude, and sincere interest in the development and success of your young students. I am very fortunate to have worked together in the wonderful and inspiring health project in Burkina Faso. I will always treasure this life-changing learning experience. In sum, I am grateful to the three of you for allowing me to enhance my skills; and for your genuine desire and capacity to collaborate, learn and work together in a wide range of topics. In sum, everything I have learned from each of you will help and inspire my career.

The completion of my PhD was only possible thanks to the grit, loving support and true friendship of Dr. A. Floridi, Dr. R. Toppo, the members of the Fellowship, Sush, Windy, Lize, Jime, and Dr. S. Kasahara. Andrea, I will always remember our heated debates, and the long cooking workshops we shared together. Our walks and discussions during Corona times were invaluable through that period. To Richard, Sush, Windy, Lize, Jime, and the Fellowship, there are no words to express my love, gratitude, and appreciation. I have learned a lot from each of you. And Cape, I will always remember our mid-night talks and your delicious Japanese meals.

I also want to show my highest appreciation to many important people at ISS, in especial to Professor Mansoob Murshed, my PhD colleagues, the DEC group, the IT department, Sandy Kamerling, and Institute's Management. The combination of different research topics, cultures and characters constantly challenged me to adapt, learn and grow. For a behavioural and development economist it was attractive the potential to generate positive academic

synergies in such an atmosphere; and undoubtedly, the skills I developed during this period will help my career and personal life. Thank you Professor M. Murshed for supporting additional work and for the good conversations and advice the past years. Lastly, I am very thankful to Sheraz Malik for rescuing my work several times, and to Sandy for allowing me to refine my entrepreneurial skills at Butterfly. The collaboration from all of you was invaluable throughout my experience, we shared wonderful moments that I will cherish for my lifetime.

Furthermore, I want to recognize my co-authors and field assistants, the editors and referees in my publications, and the members of my Doctoral Committee. I am especially thankful to the research group in Burkina Faso, in particular to Dr. B. Thiombiano and Professor D. Ouedraogo. I was very lucky to have the opportunity of working together thanks to the support of Natascha and Arjun. Equally importantly, I want to acknowledge my dedicated research assistants Zsófi Kelemen, Elisa Vreugdenhil, and especially Maria Dafnomili for your patience and caring help during our fieldworks. In addition, I am grateful for the constructive comments from Dr. Robert Sparrow, and the excellent collaboration from Dr. Binyam Demena in our work. Lastly, I am extremely thankful to the members of my Doctoral Committee Prof. E.K.A van Doorslaer, Prof. E. Nillesen, Prof. H. Kazianga, Dr. P. Nikiema, and also Prof. M. Dekker. I cherish your dedication and time to read my thesis, and the extremely valuable advice you provided. All of you allowed me to accomplish my Doctoral degree, to enhance my academic skills, and to become a better person.

Most importantly, I want to thank my parents Luis Artavia and Elsi Mora for your leading example, wisdom and precious advice; and to my brother Diego Artavia for your love throughout this process. I hope we can share more time together from now on.

Summary

The marginalization and exclusion of certain social groups remains a challenge in modern societies. Vulnerable groups often experience prominent and systematic obstacles while trying to access opportunities, rights, and resources. These obstacles may emerge as a result of ethnicity, class, skin color, religion, or health inequalities. These barriers to secure wellbeing may become insidious as they overlap and reinforce each other across generations. In an attempt to overcome such barriers, in the last few decades there has been growing interest in launching policies that mitigate the vicious circles of exclusion and deprivation in today's societies. For instance, the Sustainable Development Goals 2030 has dedicated Goal #10 to "reduce inequality within and among countries" with the purpose of establishing cost-efficient strategies to tackle historical social inequalities in low- and high-income countries.

To help achieve such outcomes the last two decades has seen pioneering work in behavioral economics/science in international development. This area of research has enhanced our understanding of the barriers to development and led to the emergence of a wider and richer theoretical and empirical framework to inform human decision making. This framework builds on fields such as sociology, anthropology, psychology, economics, and political science. Two of the last three Nobel Prizes in Economics (2017 and 2019) have been awarded to Behavioral and Experimental economists working on development-related issues. Results from this body of work have been used by academics, governments, and international organizations to design evidence-based policies in tax collection, human cooperation, healthcare, education, energy consumption and finance.

Set against this background, the present thesis provides results based on a series of small and large-scale interventions designed to understand the challenges experienced by various marginalized and excluded groups as well as to propose ways to improve their wellbeing. The various essays comprising this thesis deal with a range of issues. They make use of a variety of methods and data, and they are set in the context of a developed (The Netherlands) and a developing (Burkina Faso) country. The unifying theme is the attention to and concern for typically excluded groups.

The first half of the dissertation (Chapters 2 and 3) comprises small-scale studies on human cooperation and housing discrimination in The Netherlands. The second half of this thesis (Chapters 4 to 7) relies on four-rounds of nationally representative panel data collected over

two years. These data, which are from an experimental health intervention in Burkina Faso are used to examine, among other issues, whether a system of message reminders sent to People Living with HIV (PLHIV) improves their bio-physical traits and psychosocial measures of wellbeing. The succeeding paragraphs describe the various chapters in more detail.

Promoting prosocial behavior in The Netherlands

The multicultural and tolerant nature of Dutch society offers an opportunity to test theoretical principles about human prosociality and to identify instruments that may enhance access to the housing market for ethnic minorities. Based on a field experiment, Chapter 2 of the thesis shows that despite incurring costs, people engage in prosocial behavior and help strangers. The analysis shows that a shorter time span to decide whether to help (or not) encourages helping behavior while a longer time span reduces the probability of helping strangers. The evidence is consistent with that obtained from laboratory studies. These findings challenge the idea that humans are driven mainly by self-interest and they suggest that human beings are intuitively helpful.

The third chapter implements two experiments designed to test ways of mitigating housing discrimination against ethnic minorities in Amsterdam. These experiments evaluate whether seeking housing through a secondary, less prominent, housing platform (Craigslist) is associated with less discrimination and whether a (positive) reference letter improves access to housing for Turkish and Moroccan minorities. The results show no discrimination against Turkish and/or Moroccan candidates on houses offered through Craigslist. This is in marked contrast to the literature which shows high and systematic prejudice in the Dutch labor market and in leading European housing markets. The study also finds that attaching a reference letter from a previous landlord does not influence access to housing. Hence, this study proposes that the use of similar auxiliary websites may lessen ethnic discrimination typically present in more prominent housing platforms across Europe.

Promoting wellbeing in PLHIV in Burkina Faso

The second half of the thesis consists of four chapters which draw on information from a randomized controlled trial to improve health outcomes in PLHIV in Burkina Faso. Using information from the baseline survey, Chapter 4 examines the process of adaptation and

biomedical transition in measures of subjective and objective health as patients undergo antiretroviral treatment. The findings indicate that subjective and objective measures of health capture different aspects of wellbeing. The broader subjective health measure provides an overly optimistic picture while the narrower objective measure underestimates the beneficial effects of access to ART.

Chapter 6 presents the results of the mHealth intervention. The intervention examines the effect of four message reminders that vary in content (text or image) and frequency (once or twice a week) to promote bio-physical, treatment-related, and psychosocial outcomes in PLHIV undergoing antiretroviral therapy over two years. The pooled sample results show no global impact on primary outcomes (retention, adherence, and physical health) nor within follow-up surveys. In contrast, there is evidence of a large and positive impact of the intervention on a wide range of psychosocial measures. These results extend the discussion about the cost-effectiveness of mHealth to a relatively unexplored dimension of health. The bulk of the literature focuses on a narrow set of standard bio-physical and treatment-related indicators without recognizing the deeper and important psychosocial benefits.

Chapter 7 explores the association between HIV-related stigma and subjective health of PLHIV. Based on patient-level fixed effects models, the study shows that stigma has a negative and statistically significant association with subjective health (2.3%-points, p-value=0.090). Results indicate that income, household size and sexual activity are also positive predictors of subjective health. The analysis shows that retention in care reduces experienced stigma while regular participation in PLHIV-support groups led to increased awareness of stigma. Thus, participation in HIV self-help groups presents a double-edged sword. While self-help groups support patients during their recovery and have been shown to improve their physical health in Burkina Faso (Artavia-Mora et al., 2020), participation in these groups also has negative repercussions in prompting negative social perceptions and in exacerbating their experiences of stigma.

Samenvatting

De marginalisering en uitsluiting van bepaalde sociale groepen blijft een uitdaging in moderne samenlevingen. Kwetsbare groepen krijgen vaak te maken met aanzienlijke en systematische hindernissen wanneer ze proberen toegang te krijgen tot mogelijkheden, rechten en voorzieningen. Deze hindernissen kunnen het gevolg zijn van etniciteit, sociale klasse, huidskleur, religie of ongelijkheid op het gebied van gezondheid. Het zijn verraderlijke belemmeringen voor het veiligstellen van het welzijn omdat ze er door de generaties heen insluipen en elkaar versterken. Om deze belemmeringen weg te nemen is er de afgelopen decennia steeds meer belangstelling voor beleid dat de vicieuze cirkel van uitsluiting en achterstelling in moderne samenlevingen doorbreekt. Het tiende doel van de Sustainable Development Goals 2030 (Werelddoelen voor duurzame ontwikkeling) is bijvoorbeeld het 'Verminderen van ongelijkheid binnen en tussen landen'. Dit doel moet worden bereikt door kostenefficiënte strategieën te ontwikkelen om de historische sociale ongelijkheden in landen met een laag of hoog inkomen aan te pakken.

In dit kader is er de afgelopen twee decennia pionierswerk verricht in de gedragseconomie/gedragswetenschap op het gebied van internationale ontwikkeling. Dit onderzoek heeft geleid tot meer inzicht in de belemmeringen voor ontwikkeling en tot een breder en rijker theoretisch en empirisch kader om de besluitvorming te onderbouwen. Dit kader omvat wetenschapsgebieden als de sociologie, antropologie, psychologie, economie en politieke wetenschappen. Twee van de laatste drie Nobelprijzen voor economie (in 2017 en 2019) zijn toegekend aan economen die zich bezighouden met gedrags- en experimenteel onderzoek naar ontwikkelingsvraagstukken. Wetenschappers, overheden en internationale organisaties gebruiken de resultaten van dit onderzoek om empirisch onderbouwd beleid te ontwikkelen op het gebied van belastinginning, samenwerking, gezondheidszorg, onderwijs, energieverbruik en financiën.

Tegen deze achtergrond beschrijft dit proefschrift onderzoeksresultaten op basis van een reeks klein- en grootschalige interventies die bedoeld zijn om de uitdagingen van verschillende gemarginaliseerde en uitgesloten groepen te begrijpen en methoden te ontwikkelen om hun welzijn te verbeteren. De verschillende essays waaruit dit proefschrift bestaat, gaan over uiteenlopende onderwerpen. Er zijn verschillende methoden en gegevens gebruikt, en het onderzoek is gedaan in een ontwikkeld land (Nederland) en een

ontwikkelingsland (Burkina Faso). Het verbindende thema is de aandacht en zorg voor buitengesloten groepen.

De eerste helft van het proefschrift (hoofdstuk 2 en 3) is gewijd aan kleinschalige studies over samenwerking en huisvestingsdiscriminatie in Nederland. De tweede helft van het proefschrift (hoofdstuk 4 tot en met 7) omvat een beschrijving van de resultaten van panelonderzoek waarbij landelijk representatieve data in vier ronden binnen twee jaar tijd zijn verzameld. Deze data zijn verzameld in een experimentele gezondheidsinterventie in Burkina Faso en worden onder andere gebruikt om na te gaan of de biofysische eigenschappen en het psychosociale welzijn van mensen met hiv verbeteren als er herinneringen aan hen worden gestuurd. Hieronder worden de verschillende hoofdstukken nader beschreven.

Bevorderen van prosociaal gedrag in Nederland

Het multiculturele en tolerante karakter van de Nederlandse samenleving biedt de mogelijkheid om theoretische inzichten over prosocialiteit van mensen te toetsen en instrumenten te vinden die de woningmarkt toegankelijker kunnen maken voor etnische minderheden. Hoofdstuk 2 van het proefschrift beschrijft een veldexperiment waaruit blijkt dat mensen zich prosociaal gedragen en vreemden helpen, ook al zijn er kosten aan verbonden. Uit het onderzoek blijkt dat mensen meer geneigd zijn te helpen wanneer ze minder tijd hebben om te beslissen om wel of niet te helpen, terwijl een langere beslistijd de kans op het helpen van vreemden verkleint. Dit is in overeenstemming met de resultaten van laboratoriumonderzoek. Deze bevindingen zijn in tegenspraak met het idee dat mensen vooral handelen uit eigenbelang en wijzen erop dat mensen intuïtief hulpvaardig zijn.

Het derde hoofdstuk behandelt twee experimenten waarin werd onderzocht op welke manieren huisvestingsdiscriminatie van etnische minderheden in Amsterdam tegengegaan kan worden. In deze experimenten werd onderzocht of mensen bij het zoeken naar huisvesting via een minder bekende website (Craigslist) minder last hebben van discriminatie. Ook werd nagegaan of een (positieve) aanbevelingsbrief de toegang tot huisvesting voor Turkse en Marokkaanse minderheden verbetert. Uit de resultaten blijkt dat Turkse en/of Marokkaanse belangstellenden niet worden gediscrimineerd bij woningen die via Craigslist worden aangeboden. Dit staat in schril contrast met de literatuur waarin sterke en systematische vooroordelen op de Nederlandse arbeidsmarkt en op de Europese woningmarkt worden gerapporteerd. Verder blijkt uit het onderzoek dat het bijvoegen van

een aanbevelingsbrief van een vorige huisbaas geen invloed heeft op de toegang tot huisvesting. Op grond van deze resultaten kan worden geconcludeerd dat etnische discriminatie, die veel voorkomt op de prominentere huisvestingwebsites in Europa, mogelijk minder optreedt bij gebruik van soortgelijke aanvullende websites.

Bevorderen van welzijn bij mensen met hiv in Burkina Faso

De tweede helft van het proefschrift bestaat uit vier hoofdstukken waarin resultaten worden beschreven van een gerandomiseerd en gecontroleerd experiment om de gezondheidstoestand van mensen met hiv in Burkina Faso te verbeteren. Op basis van informatie uit het referentieonderzoek beschrijft hoofdstuk 4 het proces van aanpassing en biomedische transitie in metingen van de subjectieve en objectieve gezondheid van patiënten die een behandeling met antiretrovirale middelen ondergaan. Uit de resultaten blijkt dat subjectieve en objectieve maatstaven voor gezondheid verschillende aspecten van welzijn omvatten. De bredere subjectieve meting van gezondheid geeft een te optimistisch beeld, terwijl de beperktere objectieve meting een onderschatting van de gunstige effecten van de toegang tot antiretrovirale therapie (ART) oplevert.

Hoofdstuk 5 schetst de resultaten van de interventie mHealth. Deze interventie bestaat uit vier herinneringsberichten die variëren in inhoud (tekst of beeld) en frequentie (één of twee keer per week) met als doel om verbetering op biofysische, behandelingsgerelateerde en psychosociale indicatoren te realiseren bij mensen met hiv die twee jaar lang antiretrovirale therapie ondergaan. Uit de samengevoegde steekproefresultaten blijkt geen algemeen effect op de primaire variabelen (zorgbehoud, therapietrouw en fysieke gezondheid), ook niet in vervolgonderzoeken. Er zijn daarentegen wel aanwijzingen dat de interventie een groot en positief effect heeft op een breed scala aan psychosociale metingen. Deze resultaten voegen een nieuwe, relatief onbekende dimensie toe aan de discussie over de kosteneffectiviteit van mHealth. De meeste literatuur is gericht op een beperkte reeks standaard biofysische en behandelingsgerelateerde indicatoren, waarbij de fundamentele en belangrijke psychosociale voordelen buiten beschouwing blijven.

Hoofdstuk 7 gaat over het verband tussen hiv-gerelateerde stigmatisering en de subjectieve gezondheid van mensen met hiv. In het onderzoek is een fixed-effects-model op patiëntniveau gebruikt, en hieruit blijkt dat stigmatisering negatief en statistisch significant samenhangt met subjectieve gezondheid (2.3%-punten, p-waarde=0.090). De resultaten

wijzen er verder op dat inkomen, grootte van het huishouden en seksuele activiteit positieve voorspellers zijn van subjectieve gezondheid. Uit het onderzoek blijkt dat behoud van de zorg ervaren stigmatisering vermindert, terwijl regelmatige deelname aan lotgenotengroepen voor mensen met hiv leidt tot een verhoogd bewustzijn van stigmatisering. Deelname aan hiv-zelfhulpgroepen is dus een tweesnijdend zwaard. Hoewel zelfhulpgroepen patiënten in Burkina Faso ondersteunen tijdens hun herstel en een positief effect hebben op de fysieke gezondheid (Artavia-Mora et al., 2020), heeft deelname aan deze groepen ook negatieve gevolgen. Deelnemers hebben last van negatieve sociale percepties en hun ervaringen met stigmatisering verergeren.

Chapter 1: Introduction

The marginalization and exclusion of certain social groups remains a challenge in modern societies. Vulnerable groups often experience prominent and systematic obstacles while trying to access opportunities, rights, and resources (in employment, housing, healthcare, and education among others). These obstacles may emerge as a result of ethnicity, class, skin color, religion, or health inequalities (Akerlof and Kranton, 2000; Tajfel and Billic, 1974). These barriers to secure wellbeing may become insidious as they overlap and reinforce each other across generations. In an attempt to overcome such barriers, in the last few decades there has been growing interest in launching policies that mitigate the vicious circles of exclusion and depravation in today's societies. For instance, the Sustainable Development Goals 2030 has dedicated Goal #10 to "*reduce inequality within and among countries*" with the purpose of establishing cost-efficient strategies to tackle historical social inequalities in low- and high-income countries (United Nations, 2020).

To help achieve such outcomes the last two decades has seen pioneering work in behavioral economics/science in international development. This area of research has enhanced our understanding of the barriers to development and led to the emergence of a wider and richer theoretical and empirical framework to inform human decision making. This framework builds on fields such as sociology, anthropology, psychology, economics, and political science (Ruggeri, 2018). Two of the last three Nobel Prizes in Economics (2017 and 2019) have been awarded to Behavioral and Experimental economists working on development-related issues.[1] Results from this body of work have been used by academics, governments, and international organizations to design evidence-based policies in tax collection, human cooperation and conflict, healthcare, education, energy consumption and finance (Kazianga and Wahhaj, 2020; Barr et al., 2020; Haile et al., 2020; Nikiema, 2019; Espinosa et el., 2019; Barr et al., 2019; Wagstaff et al., 2019; Voor et al., 2012; Nillesen et al., 2019; Ruggeri, 2018; Bicchieri, 2016; Kazianga et al., 2012).

Set against this background, the present thesis provides results based on a series of small and large-scale interventions designed to understand the challenges experienced by various marginalized and excluded groups as well as to propose ways to improve their wellbeing. The various essays comprising this thesis deal with a range of issues, they make use of a variety

[1] Professor Richard Thaler won the prize in 2018, and Professors Abhijit Banerjee, Esther Duflo and Michael Kremer won the prize in 2019 (for more information review: Nobel, 2020).

of methods and data and are set in the context of a developed (The Netherlands) and a developing (Burkina Faso) country. The unifying theme is the attention to and concern for typically excluded groups.

The first half of the dissertation (Chapters 2 and 3) comprises small-scale studies on human cooperation and housing discrimination in The Netherlands. The first study (Chapter 2) examines reactions to strangers in Dutch society by investigating whether strangers are helped or punished and whether specific cognitive mechanisms promote or discourage such behaviors in daily life. The third chapter explores discrimination against ethnic minorities in the Dutch housing market. The second half of this thesis (Chapters 4 to 7) relies on panel data from a large-scale and long-term experimental health intervention in Burkina Faso to test whether a system of message reminders sent to People Living with HIV (PLHIV) improves their bio-physical traits and psychosocial measures of wellbeing. The following section describes the various chapters in more detail.

1.1 Promoting prosocial behavior in The Netherlands

The Netherlands (NL) is a high-income European country that is recognized for the harmonious coexistence of multicultural groups. Despite the historical and international perception that Dutch society is cooperative, tolerant, and inclusive, there is recent evidence pointing to prejudice and unjust treatment towards particular vulnerable groups. There is widespread literature indicating high and systematic discrimination against minority groups in the employment market and in access to mortgage loans (Andriessen et al., 2012; Derous, 2011; Derous et al., 2009; Aalbers, 2007). At the same time, the European Union Agency for Fundamental Rights found that, in the Netherlands, one third of Muslims feel discriminated because of their religion (FRA, 2017). Another survey conducted by Ipsos in 2014 states that 47% of Dutch residents find cultural mixing “a problem”, 55% think that the number of immigrants “is too high”, and 57% overestimate the presence of ethnic minorities living in the country (NU, 2014). These points establish an opportunity to test theoretical insights and instruments in two areas: *(i)* prosocial interactions between residents in an advanced multicultural society, and *(ii)* which new tools may promote minority group access to essential goods and services in The Netherlands.

The first area of research explores human cooperation in modern societies. The theoretical predictions of classical economic and evolutionary principles signal that individuals are only

interested in benefiting or assisting members in their network and marginalize and exclude outsiders -*strangers*- from cooperative benefits (and related advantages) in communal interactions. This vulnerability hampers the biological success and threatens the long-term survival of strangers -and outgroups- (Rand and Nowak, 2013; Capraro, 2013; Nowak, 2006; Bowles and Gintis, 2011). However, the validity of these classical assumptions has been revisited and subjected to empirical scrutiny in laboratory settings (Fehr and Gächter, 2002; Balafoutas et al., 2014a). Chapter 2 of this thesis builds on these laboratory-based experiments and designs a field experiment to examine the existence of prosocial preferences by analyzing interactions between strangers in Dutch society. The design uses actors that interact with strangers (participants) in the field. Each actor is requested to drop a glove (or litter) in a public park in order to record whether strangers passing by help (or scold) them. The research also signals which cognitive mechanisms may promote or discourage the enforcement of these behaviors by implementing a treatment based on the time that participants have to respond.

The second area examines marginalization and exclusion of ethnic minorities in European housing markets. Residential location determines access to better standards of education, health and employment (Ioannides, 2002; Buck, 2001; Cutler and Glaeser, 1997; Urban, 2009), and there is substantial research which shows the high and systematic prevalence of housing discrimination against ethnic minorities in countries neighboring The Netherlands (Van der Bracht et al., 2015; Auspurg et al., 2017; Bonnet et al., 2016; Ahmed and Hammarstedt, 2008). Insufficient knowledge of laws, language and local customs have also prevented minority groups from accessing housing (and submitting formal complaints) in The Netherlands (Will, 2003). More interestingly however, there is scarce research on ways to mitigate housing discrimination and/or to encourage minority access to the housing market. To contribute to this evidence, Chapter 3 examines two new instruments to promote access of Turkish and Moroccan minorities to the housing market in Amsterdam. The study tests whether the -less prominent- secondary website of Craigslist and whether a *(positive)* reference letter from a previous landowner can enhance access to housing for these two minority groups. The design employs two audit-based experiments to compare the rate of renters' responses from fictitious applicants. The first experiment compares responses to Turkish and Moroccans versus -majority- Dutch applicants to evaluate housing discrimination. The second experiment examines the impact of a positive reference versus a control group in the case of each ethnic minority.

1.2 Promoting the wellbeing of People Living with HIV in Burkina Faso

Burkina Faso is a low-income and land-locked country in West-Africa. Akin to other Sub-Saharan African countries, Burkina Faso faces challenges in safeguarding the wellbeing of People living with HIV (PLHIV) and their families (UNAIDS, 2019; Burger et al., 2019; Scanlon and Vreeman, 2013). The main vulnerability of this group arises from the need for life-long compliance to antiretroviral therapy to achieve survival, but patients often discontinue medical appointments within two years and adherence is frequently below target (Fox and Rosen, 2010; Rosen et al., 2007; Mills et al., 2006). Societal perceptions about the threat of mortality may also lead patients to experience marginalization and exclusion in areas such as: stigma (to themselves and their relatives), employment, higher medical expenses to manage the disease, and food insecurity (Scanlon and Vreeman, 2013). These vulnerabilities may deepen in a low-prevalence (0.8%) and low-income setting such as Burkina Faso, where the competition for resources and the conceptualization of PLHIV as "outcasts" may be intensified (UNAIDS, 2019).

Promoted by the World Health Organization, a growing area of research examines whether a system of mobile phone message reminders may enhance HIV-treatment compliance in low-resource settings. The reception of messages reminds patients that their lives are important and that they should comply with and follow treatment (WHO, 2011). However, a systematic review of these interventions indicates that only 60% of SMS programs improve retention and/or adherence in low- and middle-income countries (Demena et al., 2020). To contribute to this literature, the second section of this thesis generates several research outputs arising from a large-scale and long-term health intervention on PLHIV undergoing antiretroviral therapy in Burkina Faso. Chapter 4 outlines the experimental protocol of the intervention. The intervention comprises a randomized controlled trial of PLHIV undergoing ART. The evaluation tests the impact of a system of short message reminders (SMS) on bio-physical treatment-markers and psychosocial outcomes in 3,838 patients randomized across four treated groups and one control. It evaluates three primary outcomes (retention, adherence, and physical progression) and three general secondary outcomes (missed visits to health care center, patient health, and psychosocial measures). Treated individuals receive messages that vary per content (text, image, or their combination) and time (once or twice a week). The analysis is based on four rounds of surveys (baseline and three follow-ups) at 6, 12 and 24 months to assess the durability of the treatment effects. The protocol describes the design and characteristics of the intervention (Wagner et al., 2016).

In overall terms the intervention provides three major contributions to the literature. First, the intervention relies on a nationally representative sample of patients that include a richer and wider sample of participants. In contrast, most published work examines only a few geographical areas and/or includes less than 500 participants (Demena et al., 2020). Thus, the intervention improves the external validity of the results in terms of scope and geographical coverage of participants. Second, the study investigates the impact of mobile reminders on a host of psychosocial measures of wellbeing. This contribution expands the conventional approach of purely standard bio-physical and treatment-related outcomes. Third, the intervention uses patient-level panel data to examine the durability of treatment effects over a two-year timeframe. No previous intervention in West-Africa has successfully followed patients for such an extended period of time.

Before presenting the results of the main intervention, Chapter 5 examines the adaptation and biomedical transition of PLHIV as they manage antiretroviral therapy over time. The study investigates whether subjective and objective measures of health change as patients integrate ART into their lives. In a cohort study using information from the baseline survey, this work compares health measures of patients' in two categories: short-term (≤24 months) versus longer-term adherents (>24 months). The approach also examines underlying changes in the determinants of each health measure. To guide the work, the study advances a new framework that helps explain why patients' experiences vary over time, and how adaptation occurs as they manage their disease in the long run.

Chapter 6 presents the results of the impact evaluation of the main mHealth intervention. The chapter presents results for the pooled sample of patients and provides analysis for specific sub-groups over the two years. The description also highlights the durability of treatment effects in primary and secondary outcomes, and it examines potential heterogenous impacts between short- and long-term effects, and related categories.

Chapter 7 explores the association between reported experiences of marginalization and exclusion arising from HIV-stigma and subjective health in the sample of PLHIV in Burkina Faso. This article employs patient-level fixed effects models to minimize the influence of confounding factors over the four rounds of the intervention. The analysis also examines potential predictors of HIV-stigma. The examination of HIV-stigma is especially relevant in Burkina Faso as the low-prevalence of HIV combined with low access to resources may

intensify the experiences of stigmatization towards PLHIV and exacerbate vulnerability in their lives.

Part I: Promoting prosocial behaviour in The Netherlands

Chapter 2: Intuitive Help and Punishment in the Field [2]

2.1 Introduction

Prosocial behavior and cooperation are at the heart of modern social interactions (see for instance, Axelrod and Hamilton, 1981; Axelrod, 1984; Fehr and Gächter, 2002; Fehr and Fischbacher, 2003; Nowak and Sigmund, 2005; Nowak, 2006; van Veelen et al., 2012; Gächter, 2012; Capraro, 2013). Conventional theoretical explanations for the origins of cooperation in today's societies pertain to kin selection or direct and indirect reciprocity (Hamilton, 1964; Trivers, 1971; Axelrod and Hamilton, 1981; Nowak et al., 1995; Fehr and Gächter, 2002; Nowak and Sigmund, 2005; Nowak, 2006). However, consider for instance one-shot encounters between strangers in daily life. It is somewhat puzzling that strangers often help or act for others' benefit at a personal cost (Fehr and Gächter, 2002; Balafoutas et al., 2014a). If such interactions are not solely governed by self-interest as standard economic models or theories of evolution would predict (Nowak, 2006; Sigmund, 2010), then the natural question that arises is: why do strangers help each other? And which factors and mechanisms can promote or undermine prosocial behavior? Answering such questions is central to our understanding of human cooperation and the evolution of societies.

A more recent literature has focused on the cognitive mechanisms underlying human cooperation (Rand et al., 2012; Rand and Nowak, 2013; Rand and Epstein, 2014; Rand, 2016; Capraro and Cococcioni, 2016). This literature rests on the dual-process cognitive framework featuring two competing systems of decision making: (1) intuitive, automatic, emotional, unconscious and faster decisions based on prior experience and beliefs versus (2) deliberative, more controlled, rational, reflective, effortful and slow decisions (Sloman, 1996; Kahneman, 2003; Loewenstein and O'Donoghue, 2004; Frankish and Evans, 2009; Kahneman, 2012; Evans and Stanovich, 2013). The resulting line of reasoning which motivates current experimental work is that intuition promotes prosocial behavior, while deliberation overrides such behavior (Rand, 2016).

In order to shed light on intuition versus deliberation in cooperative decision making, a series of lab experiments has employed decision time manipulations. Time pressure is used to

[2] Artavia-Mora L, Bedi AS and Rieger M. 2017. Intuitive help and punishment in the field. *European Economic Review*, 92, pp.133-145.

trigger intuitive decisions and time delay is meant to elicit deliberative decisions.[3] A host of studies suggests that time delay indeed reduces prosocial behavior (Cone and Rand, 2014; Rand et al., 2012, 2013, 2014; as well as a meta study by Rand, 2016). A second group of studies has found no effect of such time manipulations (Tinghög et al., 2013; Verkoeijen and Bouwmeester, 2014), or, when there is an effect, it is limited to some groups of people and to specific social contexts (Rand et al., 2013; Capraro and Cococcioni, 2016). For instance, effects seem to disappear among experienced experimental subjects (Rand et al., 2013).[4] Effects can also depend on an experimental subject's general trust level (Rand and Kraft-Todd, 2014) and the overall trust environment (Capraro and Cococcioni, 2016).[5]

A related literature has also examined the impact of time manipulation on the decision to punish individuals. There is evidence that time delay moderates impulsive desires to punish offenders by decreasing negative emotions such as disagreement and resentment (Sutter et al., 2003; Grimm and Mengel, 2011; Wang et al., 2011; Neo et al., 2013).

To rationalize these patterns Rand et al. (2014, p.2) proposed the Social Heuristics Hypothesis. The hypothesis links learning from experience and daily interactions with decision outcomes in unusual contexts such as a lab. Put differently, it theorizes that "daily life typically involves factors such as repetition, reputation and the threat of sanctions, all of which can make cooperation in one's long term self-interest" which in turn shapes "generalized cooperative intuitions." And these internalized inclinations may or may not lead to intuitive cooperation in the lab or other atypical contexts (see also Jordan et al., 2014).

More recently, Bear and Rand (2016) have formalized these empirical patterns within a cognitive game theory model of cooperation. Individuals in the model can use intuition or deliberation when deciding how to interact with others. However, deliberate responses require individuals to make an additional effort as they need to reason through their decision to cooperate or defect. In other words, there is a "trade-off" between making costly but informed decisions and cheap intuitive ones. That said, individuals play two sorts of games: (i) one-shot, or (ii) reciprocal, repeated prisoner's dilemma games (occurring at probabilities

[3] Another strand of the literature has correlated response times with cooperation (Nielsen et al., 2014; Piovesan and Wengström, 2009; Recalde et al., 2014). However, recent work has shown that such an approach does not support a test of intuition versus deliberation but rather reflects decision conflicts (Krajbich, Bartling, Hare and Fehr, 2015; Evans, Dillon, and Rand, 2015). Therefore, we do not discuss these studies in more detail.

[4] Rand et al. (2013) report "vanishing" effect sizes associated with time manipulations in studies among increasingly experienced subjects recruited on Amazon Mechanical Turk.

[5] Capraro and Cococcioni (2016) recruited inexperienced subjects from India, a social environment with low quality institutions and levels of trust. They find that time manipulation has no effect on cooperation rates.

1-p and p, respectively). Only in repeated games do choices of individuals impact future encounters. The "social environment" then shapes the evolution and ultimately the nature of individuals. If the environment is sufficiently reciprocal (in other words, the probability p of repeated interactions is relatively high), individuals may follow a dual-process cognitive framework. That is, they will be intuitive cooperators who are, however, able to use costly deliberation. In the case of repeated games, deliberation may also lead to cooperation, while in one-shot games it favors defection. In sum, the model provides a rationale for why time delay (i.e. inducing deliberation) reduced cooperation in some of the previous experiments.

We contribute to the literature by examining whether time delay impacts the likelihood of helping and punishing strangers using a natural field experiment. While the concepts "cooperate" and "defect" do not translate directly to our experimental setting, the model by Bear and Rand (2016) does yield information on the expected impact of time delay. In our setting, we examine individually costly pro-social behavior in terms of helping and direct punishment and we experimentally induce deliberation through time delay (as outlined below). If the overall environment in the field is sufficiently reciprocal, and if helping behavior is correlated with the cooperative nature of individuals and driven by a similar dual-process cognitive framework, then time delay should decrease helping rates. Furthermore, if helping and punishment are driven by similar cognitive and psychological processes, we should see a similar pattern in both cases.

Our experiment has two notable features compared to previous studies in the lab: first, it yields a demographically diverse pool of subjects. Second, we can randomly and very naturally manipulate the time available to make a decision. Our design minimizes human mistakes and examines everyday human behavior (List, 2007, 2011; Balafoutas et al., 2014a). Rather than pushing buttons, the experiment is designed such that subjects may actually help a stranger at a personal cost. However, there is also one important design difference compared to papers which conduct lab experiments. While our paper is located in the emerging literature on intuitive cooperation, we examine pure helping behavior, and in an extension, direct and indirect punishment in one-shot interactions among strangers. Our design does not feature strategic elements nor create room for further interactions.

In order to test for the effect of time delay on helping and punishment rates in the field, we closely follow the methodology by Balafoutas and Nikiforakis (2012) and Balafoutas et al. (2014a). Specifically, we use actors that interact with strangers in the field. In the helping a

stranger experiment, we asked an actor to drop a glove in a public park. We then record if strangers (subjects) who are passing by help, that is, pick up or alert the actor about the glove. We also collect basic socio-demographic characteristics of subjects through a post-experimental survey.

The novelty of our paper compared to the original methodology used by Balafoutas and co-authors is that we are adding an exogenous time manipulation treatment to test for intuitive versus deliberative responses. The actor drops the glove when an approaching subject is at a distance of either 4.5 m or 13 m from him/her. In our experiment, the average subject has about 3.5 s or 10 s to decide whether to help or defect. The short distance is our time pressure or baseline case meant to trigger fast and intuitive help, while the longer distance is our time delay treatment that promotes slow and deliberative responses.

In an extension we also estimate the impact of time delay on direct and indirect punishment in the field by building on Balafoutas et al. (2014a) and (2012). Punishing a stranger who violates a social norm may benefit society, but also comes at a personal cost and involves the fear of retaliation or counter-punishment (Fehr and Fischbacher, 2004; Gächter et al., 2008; Nikiforakis, 2008; Balafoutas et al., 2014a; 2014b, 2016). In the case of direct punishment, our actors violate the non-littering norm by dropping an empty plastic bottle. Related, we also investigate whether time delay impacts fears or thoughts of retaliation among our subjects. We collect this information through the post-experimental survey. Based on previous experimental evidence, we hypothesize that time delay will reduce punishment rates. Note however that most of the previous evidence in this literature stems from bargaining settings with second-party punishment (Sutter et al., 2003; Grimm and Mengel, 2011; Neo et al., 2013).[6] That is to say, the participant (i.e. the punisher) is directly or personally affected by the behavior of other participants. Our littering experiment does not feature bargaining and is closer to (but not strictly) a third-party punishment experiment (see discussion in Balafoutas and Nikiforakis, 2012). In our setting participants are as much affected by the littering as other subjects, while in experiments on third-party punishment the punisher is affected by the norm violation in an indirect manner. In sum, our set-up and the norms governing punishment are conceptually different compared to previous studies, so the resulting impacts of time delay may differ as well.

[6] A study by Wang et al. (2011) features a third-party punishment experiment where longer time delay was found to lower punishment by observers.

As direct punishment is typically low in the field, we also examine indirect punishment defined as "withholding help from a norm violator" (Balafoutas and Nikiforakis, 2012; Balafoutas et al., 2014a). Indirect punishment is a thought-through form of punishment, so we expect that withholding help is more likely to be enforced in the time delay treatment.

We find that -on average- our subjects are intuitive helpers, but deliberation overrides this impulse. Time delay reduces helping rates from 74.6–56.5%, which amounts to a 24% reduction (p-value=0.03, chi2(1)=4.73, n=129).[7] In other words, time delay substantially increases defection rates or selfishness. This impact is robust to the inclusion of socio-demographic covariates. In our indirect punishment experiment on helping a norm violator, the impact of time delay is similarly large and important in terms of magnitude, amounting to a 27% reduction in helping rates. However the impact is only weakly statistically significant (p-value=0.12, chi2(1)=2.46, n=117; and p-value=0.083 for a one-sided Fisher's exact test given that our starting hypothesis was that time delay reduces helping rates). Further we find no evidence that time delay impacts rates of direct punishment (p-value=0.55, chi2(1)=0.37, n=236). However, in line with previous work, we find that 55% of our subjects feared retaliation (or at least considered it) as a potential response to punishment. More importantly, we find that time delay increases these fears by 12.7% points (p-value=0.08, chi2(1)=3.16, n=194). In sum, our results on helping a stranger are in line with studies showing a negative link between time delay and cooperation. In contrast, we do not find a significant impact of time delay on direct punishment. We document statistically weaker, negative impacts of time delay on helping a norm violator in our indirect punishment experiment. When interpreting these findings, it is important to note that helping or directly punishing a stranger are costly actions compared to indirect punishment (i.e. simply withholding help towards a norm violator).

Our paper is organized as follows. Section 2 details the experimental design and data. Section 3 presents the results. Section 4 compares findings to related studies and concludes.

2.2 The experiment

Our hypothesis is that time delay undermines helping and punishment rates. This section details the three experiments to test this hypothesis along with the choice of field location, treatments as well as the experimental procedures and subject characteristics.

[7] Unless indicated we report two-sided tests throughout the paper.

2.2.1 Three social dilemmas

We designed an experiment featuring three social dilemmas: (1) helping a stranger, (2) helping a norm violator and (3) direct punishment of a norm violator.

In the *first dilemma* we triggered help from subjects by asking an actor to drop one bicycle glove in a public park (as shown in Photos A2.1 and A2.2 in the Appendix). We interpret the decision to help or defect as the choice between benefiting others at a personal cost versus selfishness. We chose the glove drop for four reasons: First, gloves are complements. Losing one glove makes the second glove useless. Second, a glove falls noiselessly, and it is thus credible that the actor does not notice the loss of the glove and requires help. Third, gloves are big enough to be seen from a distance which is necessary for our decision time treatments as we explain below. Fourth, a glove is neither too cheap nor too expensive. We wanted to minimize distortions to cooperate or defect based on the moral justification and value of the object. Using an expensive object (i.e. jewelry) may lead to higher rates of help but also potentially to theft. Conversely, using a cheap object (i.e. pencil) may dissuade help and might be perceived as littering. Arguably, gloves are a good compromise.

The *second dilemma* extends the glove experiment by adding the violation of a social norm. Specifically, we asked the actor to litter just before dropping the glove. The actor litters (throws an empty plastic bottle) and then drops the glove (see Photo A2.3 in the Appendix). The idea is to investigate whether humans punish directly (costly punishment) or indirectly (withholding help) if an individual litters and to assess the impact of time delay on punishment. In this dilemma there is a possibility that indirect punishment crowds out direct punishment as discussed in Balafoutas et al. (2014a). And crowding out itself may be impacted by time delay. In other words, the impact of time delay on punishment in this experiment is no longer exogenous.

The *third dilemma* is a pure punishment experiment without potentially confounding effects of crowding out. The actor only litters and we randomly manipulate the time to decision as in the two dilemmas above. In other words, the subject decides whether to directly punish or not.

2.2.2 Location

The location of the experiment was a pedestrian path in Park Malieveld in The Hague in the Netherlands. The location is appropriate for at least three reasons: First, the path is straight

and bordered by trees. It is the only way to cross the park. Therefore, it is hard for subjects to avoid or dodge the decision to help or not (see Photos A2.1 and A2.5 in the Appendix). More importantly, there is little distraction and visibility is good (see Photos A2.6 and A2.9 in the Appendix). It is easy to see the glove drop. Second, based on prior observation we noticed that people on the path tend to walk alone and that there can be large distances between them. This fact is important since we only wanted to sample subjects walking alone. Confounds such as reciprocity and social pressure are thus minimized. Subjects can make "private" and "anonymous" decisions in a public space (Photos A2.1–9 in the Appendix). Third, the location is at the heart of the city and is surrounded by many stores, institutions, and workplaces such as government offices, learning institutes, shopping areas, university faculties, commercial businesses, and non-governmental organizations. Also, the central train station of The Hague is located at the end of the park. These contextual factors yield a relatively diverse pool of subjects.

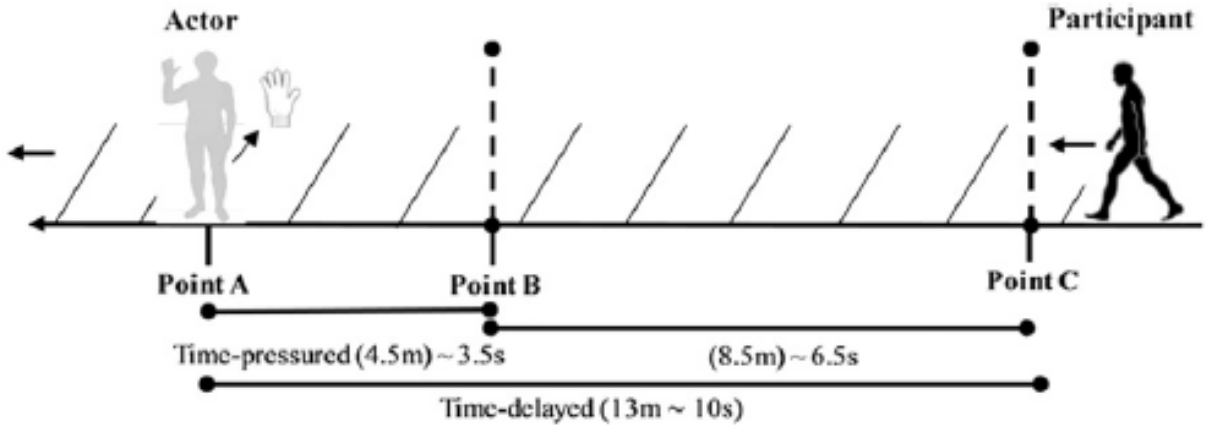
2.2.3 Experimental treatments

Our experiment tests the impact of time delay on the likelihood of helping and punishing strangers. We exogenously varied the time available to make a decision by changing the distance between subject and actor. Average human walking time is about 1.3 m/s (Mohler et al., 2007). We use two distance treatments, one short (baseline treatment) and one long. The short distance is 4.5 m between the subject and actor. This provides roughly 3.5 s for an individual to decide whether to help or defect. This treatment elicits decisions under time pressure. The longer distance is 13 m and subjects have 10 s to decide. The longer time period is designed to promote a slower or deliberate decision. We calibrated distances based on location visibility. If the glove is dropped at a distance closer than 4.5 m, the field of vision is too narrow and restricted. If the glove is dropped from a distance further than 13 m, visibility declines.

Figure 2.1 depicts the time manipulation treatments. The actor is depicted in grey and the subject in black. Point A indicates the location where the actor drops a glove and triggers the social dilemma. The actor drops the glove either when the subject is at point B (4.5 m) or at point C (13 m). Due to the arbitrary assignment of subjects to points, we can isolate the effect of distance on the likelihood of helping the actor. In the helping a norm violator experiment, the actor litters (violates the non-littering norm) before the participant reaches point B or C.

Then the actor drops a glove when points B or C are reached. In the direct punishment experiment, the actor litters when the subject reaches point B (4.5 m) or C (13 m).

Figure 2.1. Diagram of time pressure (baseline) and time delay treatments.



2.2.4 Subject selection and experimental procedure

Social dilemmas one and two (helping a stranger and helping a norm violator) were performed during 11 days in July 2015. The third dilemma (punishing a norm violator) was performed a year later during 14 days in May and June 2016.[8] The data collection took place on working days; unless weather conditions were not suitable (rain, storm, or low temperature). The treatments were assigned before the subjects arrived, and are thus independent of subject type. We used one female and one male actor. The first author of this paper was the male actor and the female actor was a research assistant. Both actors were aware of the hypothesis to be tested. The researcher recorded the data and was located at a distance to avoid social pressures (see Photo A2.2 in the Appendix).

Each trial began when the researcher selected a participant. The selection was based on two criteria: First, the subject needed to be alone with no other individual walking in the same or opposite direction. This criterion was imposed to eliminate social pressures. Second, the participant had to be in no visible hurry nor visibly distracted. Photos A2.4 and A2.5 in the Appendix show a typical participant in the experiment. In a few cases subjects were not selected since the actor or surveyor knew the subject personally.

[8] We thank an anonymous referee for suggesting the "pure" punishment experiment that was not included in the working paper version of the paper. The follow-up experiment took place a year later, at the same location, using the same procedures and actors.

Each experiment started with the actor sitting on a bench at Point A. Parked bicycles marked points A, B and C (see Figure 2.1). The actor held a pair of gloves (and if applicable the plastic bottle for the punishment experiment) and carried a bag (see Photos A2.1 and A2.2 in the Appendix). The actor then left the bench and started crossing the path, waiting for the participant to reach point B or point C. When the subject reached either point B or C, the actor "accidentally" dropped one bike glove without noticing while trying to put it in a bag for social dilemmas one and two. In the case of social dilemma three, the actor only littered at point B or C. The actor then stopped at his/her bicycle and pretended to be looking for the keys of the bicycle or to make a phone call (see Photos A2.1 and A2.3 in the Appendix). The actor waited until the participant revealed the decision at point A.

It is important to note that the actor ignored any voice alerts far from point A. Instead, the actor only responded at point A (see Photos A2.10–12 in the Appendix). This ensured that each participant had the same time to help/punish the actor.

In the last step of the experiment, that is once the participant had made a decision at point A, the researcher noted down the results, while the actor quickly interviewed the participant (Photos A2.13 and A2.14 in the Appendix). The survey collected demographic characteristics such as age, gender, time lived in The Netherlands, willingness to undertake risks in daily life and height as a proxy of physical strength and confidence (the short surveys can be found in the Appendix). We use these variables to investigate treatment balance, as well as impact heterogeneity.

To summarize, each participant's time to decision depended on individual walking speed, but most importantly on the distance to the dropped glove in dilemmas one and two, or littering in social dilemma three. The baseline scenario with the shorter available decision time was designed to elicit fast –intuitive- decisions, while the time delay treatment was meant to promote slower – deliberate- decisions.

2.5 Subject characteristics and treatment balance

2.5.1 Social dilemma 1 and 2

We ran 267 trials for social dilemmas one and two – 137 for helping a stranger and 130 for helping the norm violator, respectively. The sample size is sufficient to detect a medium sized

effect.[9] Table 2.1 shows that subject characteristics are balanced across treatments suggesting that randomization was achieved. These basic demographic statistics stem from the post-experiment survey.[10] The participants have an average age of 44 years and 61% are male. The average subject has lived for 38 years in the Netherlands. 76.5% of people in our sample have lived their entire lives in the Netherlands.

While we picked distances to ensure maximum visibility, people might not have noticed the drop of the glove. This is not a problem per se if visibility issues are not systematically related to the distance treatment. After the experiment we asked people if they had noticed the glove drop (see post-experimental survey in the Appendix). 92% of subjects acknowledged seeing it. There are no significant differences in this variable between the time and social dilemma treatments (see p-values stemming from two-sided t-tests in Table 2.1). Our main results exclude the subjects that did not acknowledge noticing the glove drop and littering.

Including or excluding these cases yields statistically identical results as we show in a robustness check below.

Table 2.1. Social dilemma 1 and 2 - characteristics of subjects and balance across treatments (pooled and full sample)

Variable	Obs	Mean	Std. Dev.	Min	Max	Randomization balance	
						Time treatment	Dilemma type
						P-values Δ	
Respond to survey	267	87.64	0.32	0	1	0.59	0.07
Age	234	43.78	13.97	15	76	0.56	0.69
Male (1=male, 0=female)	267	0.61	0.49	0	1	0.23	0.87
Height (in cm)	230	175.32	10.77	147	204	0.46	0.23
Years lived in The Netherlands	234	37.74	20.38	0	76	0.70	0.59
Native	234	0.76	0.42	0	1	0.32	0.62
Willingness to take risks (0=lowest, 10=highest)	233	5.74	1.85	0	10	0.44	0.23
Ack. seeing glove drop	267	0.92	0.27	0	1	0.26	0.21

Note: Full sample. For the analysis, we exclude subjects that did not acknowledge seeing the glove drop. Native is defined as having always lived in the Netherlands.

[9] Setting Cohen's d to 0.5, power to 80% and significance level to 5% in a power calculation for a test (two-tailed) of two proportions, the required sample size per social dilemma is 126. The required sample size was calculated using the "pwr" package in R. The relevant script is in the replication folder.

[10] Note that the response rate to the survey was 88% and non-response is unrelated to the time and social dilemma treatment at the 5% level of significance (see p-values stemming from two-sided t-tests in Table 2.1). However, our study uses participants' self-reports that are prone to limitations (see Schwarz, 1999; Sadana et al., 2002; Rand and Kraft-Todd, 2014).

2.5.2 Social dilemma 3

In the follow-up direct punishment experiment which was conducted a year later we ran 252 trials (see Table 2.2). The sample size of this follow-up experiment was chosen in order to be able to detect a small/medium effect size given that punishment rates are generally lower than helping rates.[11] The characteristics of the subjects in the follow-up experiment are similar to the subjects used to analyze social dilemmas 1 and 2 (compare Tables 1 and 2). More specifically, differences in subjects traits are statistically insignificant with the exception of height and risk taking. In social dilemma 3, the average subject is 2.4 cm taller (p-value=0.05, n=313, t(311)=2.01) and exhibits a 8.3% higher risk taking measure (p-value=0.03, n=310, t(308)=2.24) compared to the sample from a year earlier.[12] The time manipulation treatment with respect to subject characteristics is balanced. In the post-experiment survey, about 94% of subjects reported that they had seen the act of littering. We drop these observations from subsequent analysis, which does not affect our subsequent results. Sample sizes by treatment status and dilemma are presented in Table 2.3.

Table 2.2. Social dilemma 3 - characteristics of subjects and treatment balance.

	Obs	Mean	Std. Dev.	Min	Max	Randomization balance
Variable						Time treatment P-values Δ
Respond to survey	252	82.53	0.38	0	1	0.51
Age	208	43.67	13.32	16	76	0.61
Male (1=male, 0=female)	252	54.76	0.50	0	1	0.80
Height (in cm)	208	176.82	9.76	150	200	0.39
Years lived in The Netherlands	208	38.69	18.84	0	76	0.70
Native	208	78.85	0.41	0	1	0.38
Willingness to take risks (0=lowest, 10=highest)	202	6.05	1.68	0	10	0.46
Ack. seeing glove drop	252	93.65	0.24	0	1	0.30

Note: Full sample. This experiment took place a year later than the experiment presented in Table 2.1. For the analysis, we exclude subjects that did not acknowledge

Table 2.3. Sample sizes used in the analysis by social dilemma and time manipulation treatment.

	Dilemma 1 Helping a stranger	*Dilemma 2* Helping a norm violator	*Dilemma 3* Direct punishment
Time pressure	67	58	120
Time delay	62	59	116
N	129	236	236

Note: We exclude from the analysis subjects that did not acknowledge seeing the glove drop.

[11] Setting Cohen's d to 0.25, power to 80% and significance level to 5% in a power calculation for a test (two-tailed) of two proportions, the resulting sample size is 256.

[12] Reported results from these balance tests between the first and follow-up experiment are based on our final analysis sample and available on request.

2.3 Results

In what follows, we first present results on the impact of time delay on help (social dilemma 1 and 2), and then on punishment (social dilemmas 2 and 3). The results section closes with an investigation of impact heterogeneities along subject characteristics. Throughout the discussion, we refer to time delay as the treatment and take time pressure as the baseline scenario. In other words, we are looking at a (binary) increase in the available decision time and its impact on helping and punishment rates.

2.3.1 Helping strangers and norm violators

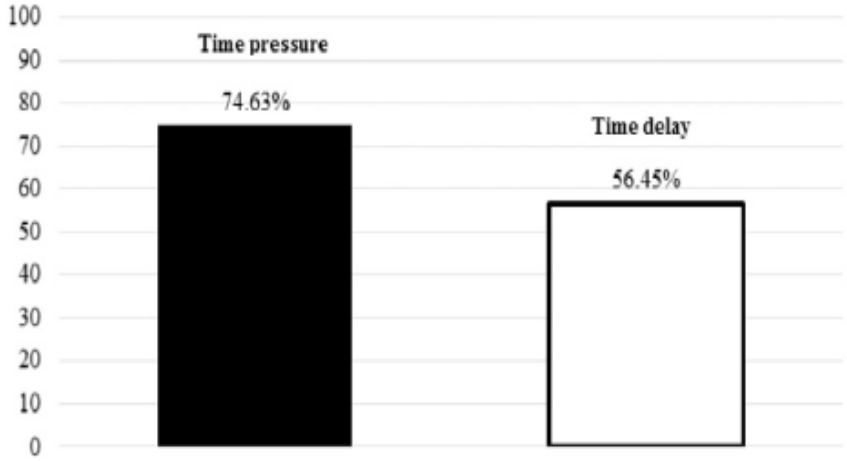
Time delay impacts helping rates in our experiment. The overall evidence indicates that our participants are naturally inclined to help, but behave more selfishly when they have more time to think.

2.3.1.1 Social dilemma 1: helping a stranger

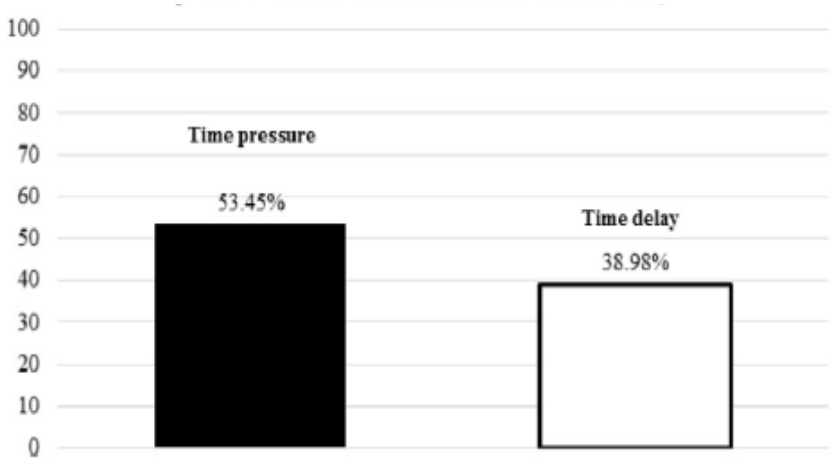
Time delay has a sizeable negative impact on helping rates in dilemma 1, as indicated by Panel A in Figure 2.2 The time delay treatment reduces the helping rate from 74.6–56.5%. This 18% point reduction is significant with a p-value of 0.03 (chi2(1)=4.73, n=129). Overall, across treatments, 66% of subjects helped. In comparison, Balafoutas et al. (2014a) find that only 39.7% of subjects helped an actor who had dropped a book at a German railway station. This difference is consistent with the fact that in our experiment the actor dropped the item "without noticing it." This could have plausibly boosted the motivation of our subjects to help.

Figure 2.2. Average rates of helping a stranger (social dilemma 1) and helping a norm violator (social dilemma 2) by time manipulation treatment.

Panel A. Helping a stranger (Δ means p-value: 0.03, chi2(1):4.73, n:129)



Panel B. Helping a norm violator (Δ means p-value: 0.12, chi2(1):2.46, n:117; p-value;0.083, one-sided Fishes' exact test)



Note: P-values stem from chi-squared tests. Panel B also reports a one-sided Fisher's exact test.

2.3.1.2 Social dilemma 2: helping a norm violator

Similar patterns emerge in the experiment on helping the norm violator (see Panel B, Figure 2.2). To start with, it is not surprising that overall helping rates drop when the actor litters before losing a glove (compare panels A and B by treatment group in Figure 2.2). These uniform overall reductions due to the littering treatment are statistically significant at the 5% level. In other words, the initial littering causes subjects to reduce help. The impact of time delay has the same sign as in dilemma 1 and is large in magnitude, but is only weakly statistically significant. The impact amounts to a 14.47% point reduction with a p-value of

0.12 (chi2(1)=2.46, n=117[13]; and p-value=0.083 for a one-sided Fisher's exact test given that our hypothesis was that time delay reduces helping rates). When we average across treatments, 46% helped the norm violator. Note that Balafoutas et al. (2014a) find that 18.6% of subjects helped a norm violator.

2.3.1.3 Comparing social dilemmas 1 and 2

Time delay impacts are comparable in size across social dilemmas. Delay causes 24% (helping a stranger) and 27% (helping the norm violator) reductions in helping rates, respectively, although, the effect is only weakly statistically significant in dilemma 2.

Table 2.4 allows a direct comparison between the two experiments using a regression model. Both time delay and violation of the social norm decrease helping rates (columns 1 and 2). The two treatment effects are similar in magnitude and statistically significant (column 3). The adverse impact of norm violation amounts to −19.3% points (column 3). This reduction captures indirect punishment or refusal to help. This result is in line with Balafoutas et al. (2014a), who report that a norm violation causes a 21.1% point reduction in helping.

Does time delay magnify or decrease the effect associated with norm violation? Note that the interaction between the two treatments (time delay and norm violation) is positive but insignificant (column 4). The standard error is very large relative to the point estimate. In other words, there is no evidence that time delay reduces the negative effect of norm violation on the likelihood of helping a stranger (in absolute terms).

Table 2.4. Regression results pooling social dilemma 1 (helping stranger) and social dilemma 2 (helping norm violator).

Dep. Variable- Helping	(1)	(2)	(3)	(4)
Time delay	-0.169*		-0.164*	-0.182*
	(0.063)		(0.062)	(0.083)
Norm violator		-0.197*	-0.193*	-0.212*
		(0.062)	(0.062)	(0.085)
Time delay*Norm violator				0.037
				(0.124)
Constant	0.648*	0.659*	0.738*	0.746*
	(0.043)	(0.042)	(0.049)	(0.054)
N	246	246	246	246

Note: Linear regression model. Robust standard errors in parentheses. Symbols denote significance levels at +p < 0.1, *p < 0.05. Subjects who did not acknowledge

[13] Setting power to 80% and significance level to 5% in a power calculation for a test (two-sided) of two proportions, the resulting sample size would need to be 370. For a one-sided test, the sample size would need to be 292.

2.3.1.4 Robustness to the inclusion of co-variates

Table 2.5 presents regression models for the two binary response variables (helping, helping the norm violator) where we include subject covariates. Due to the randomization of treatment, point estimates associated with time delay are stable across models to this inclusion. Note also that, across the board, covariates are insignificant. Only male subjects are significantly less likely to help a norm violator in our setting.

Table 2.5. Helping in social dilemmas 1 and 2 and inclusion of covariates.

	Helping a stranger			Helping a norm violator		
	(1)	(2)	(3)	(4)	(5)	(6)
Time delay	-0.182*	-0.173*	-0.202*	-0.145	-0.157	-0.154
	(0.083)	(0.085)	(0.084)	(0.092)	(0.100)	(0.102)
Subject characteristics						
Age			0.005			0.003
			(0.004)			(0.004)
Male			0.089			-0.242*
			(0.116)			(0.117)
Native			0.126			-0.090
			(0.114)			(0.135)
Risk taking			0.021			0.016
			(0.025)			(0.027)
Height (in cm)			-0.000			0.005
			(0.006)			(0.006)
Constant	0.746*	0.742*	0.280	0.534*	0.553*	-0.289
	(0.054)	(0.054)	(0.927)	(0.066)	(0.073)	(1.021)
N	129	124	124	117	100	100

Note: Linear regression model. Robust standard errors in parentheses. Symbols denote significance levels at *p < 0.05.

2.3.2 Direct punishment of strangers

We now present the impact of time delay on direct punishment in social dilemma 2 (the actor first litters then loses a glove) and social dilemma 3 (the actor only litters). The former features a possible choice between direct and indirect punishment (withholding help), while the latter evokes only direct punishment in response to littering. To preview our findings, there is no systematic evidence that time delay affects punishment.

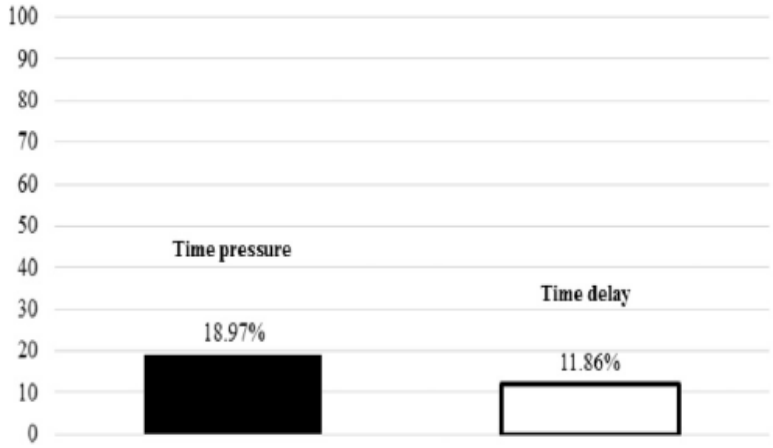
2.3.2.1 Social dilemma 2: direct punishment of the norm violator (with possible indirect punishment)

Impacts of time manipulations on direct punishment of littering in social dilemma 2 are presented in Panel A of Figure 2.3. Overall 15% of subjects punish directly. The impact of

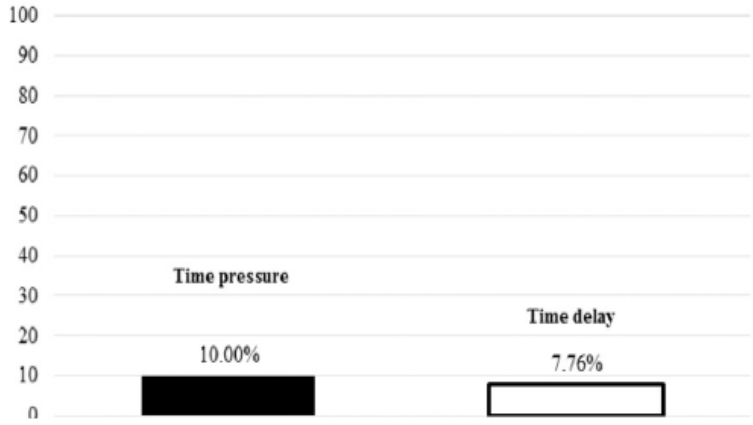
time delay amounts to −7.11% points and is insignificant[14] (p-value=0.29, chi2(1) =1.13, n=117).[15] In Table 2.6, we regressed punishment on the treatment dummy and subject characteristics. The point estimate associated with time delay remains insignificant. The only significant co-variate is age, which correlates positively with punitive behavior.

Figure 2.3. Average rates of direct punishment of norm violator (social dilemma 2 and 3) by time manipulation treatment.

Panel A. Social dilemma 2 (Δ means p-value: 0.29, chi2(1):1.13, n:117)



Panel B. Social dilemma 3 (Δ means p-value: 0.55, chi2(1):0. 37, n:236)



Note: P-values stem from chi-squared tests.

[14] Setting power to 80% and significance level to 5% in a power calculation for a test (two-sided) of two proportions, the resulting sample size would need to be 802.

[15] Note that indirect punishment may overlap with direct punishment. Figure A2.1 decomposes punishment rates into pure punishment and punishment with help. Bear in mind that sample sizes become small and the time delay treatment is no longer exogenous. However, time delay is uncorrelated with punishment and helping rates (compare left panel of Figure A2.1), while it is significantly and negatively associated with pure punishment rates (compare right panel of Figure A2.1).

Table 2.6. Direct punishment in social dilemmas 2 and 3 and inclusion of covariates.

Dep. Variable Direct punishment	Social dilemma 2			Social dilemma 3		
	(1)	(2)	(3)	(4)	(5)	(6)
Time delay	-0.071	-0.118	-0.099	-0.022	-0.046	-0.030
	(0.067)	(0.073)	(0.071)	(0.037)	(0.043)	(0.041)
Age			0.007*			0.004*
			(0.002)			(0.002)
Male			-0.104			-0.126*
			(0.073)			(0.047)
Native			0.064			-0.017
			(0.075)			(0.056)
Risk taking			0.028			0.043*
			(0.018)			(0.013)
Height (in cm)			0.001			0.003
			(0.004)			(0.003)
Constant	0.190*	0.213*	-0.361	0.100*	0.120*	-0.771+
	(0.052)	(0.060)	(0.625)	(0.028)	(0.033)	(0.456)
N	117	100	100	236	194	194

Note: Linear regression model. Robust standard errors in parentheses. Symbols denote significance levels at +p < 0.1, *p < 0.05.

2.3.2.2 Social dilemma 3: direct punishment of the norm violator

Likewise, there is no significant impact of time delay on punishment in social dilemma 2 (see Panel B of Figure 2.3). Overall 9% of subjects punished. The impact of time delay is small and amounts to −2.24% points with a p-value of 0.55[16] (chi2(1)=0.37, n=236). Table 2.6 presents a corresponding regression model with subject covariates. The impact of time delay remains insignificant. Unlike in the helping models, a series of covariates significantly predict punishment. Older, risk-taking and female subjects are more likely to punish.[17]

2.3.2.3 Comparison with previous findings on punishment

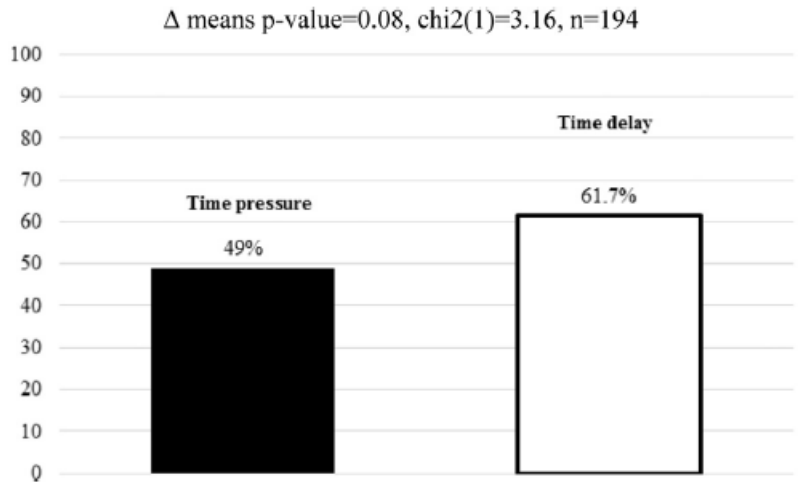
Note that Balafoutas and Nikiforakis (2012) and Balafoutas et al. (2014a) report punishment rates as low as 4% and 6.8% in littering experiments in train stations in Greece and Germany, respectively. These two studies argue that the low rates of punishment are due to fear of counter punishment (retaliation) by subjects; 71.9% and 57.8% of respondents in Greece and Germany offered this as a reason for not punishing. In another study by Balafoutas et al. (2016), 60% of respondents in a survey (in the subset of those who mentioned that they were bothered by littering) reported that they would not punish due to the fear of counter-punishment. And interestingly, such fears were more frequent for a severe compared to a small norm violation. These patterns highlight the key role of counter-punishment in the

[16] Setting power to 80% and significance level to 5% in a power calculation for a test (two-sided) of two proportions, the resulting sample size is 5038.

[17] Social dilemma 3 has more than double the amount of observations than dilemma 2, which may help explain the more precise estimates associated with covariates.

field. In our experiments, we followed Balafoutas and Nikiforakis (2012) and Balafoutas et al. (2014a) but added the decision time manipulation. In the time delay group, the punishment rates were 11.86% (social dilemma 2) and 7.76% (social dilemma 3), respectively. In the case of dilemma 3,[18] we also enquired if subjects were afraid of punishing and found that, on average across treatments, 55% of respondents reported fear of retaliation and as shown in Figure 2.4, time delay increases this fear by 12.7% points (p-value=0.08, chi2(1)=3.16, n=194). Likewise, Table 2.7 shows that this impact is robust to the inclusion of covariates. Moreover, we also found a negative and significant correlation between risk taking behavior and the likelihood of reporting this fear.

Figure 2.4. Fears or thoughts of retaliation/conflict (social dilemma 3) by time manipulation treatment.



Note: P-value stems from a chi-squared test.

Furthermore, Balafoutas et al. (2014a) find evidence that indirect punishment crowds out direct punishment. In other words, subjects would prefer not to help rather than directly tell a norm violator not to litter. We regressed the likelihood of punishing on a dummy for social dilemma 2 (capturing the possibility of indirect instead of direct punishment) and time delay (see Table A2.2). The point estimate associated with social dilemma 2 is negative (pointing to substitution between direct punishment and a refusal to help) but insignificant. While interpreting these results, it is important to note that our experiment was not designed to test crowding out. Specifically, we are pooling data from the first experiment with the experiment added a year later. Observable subject characteristics are well balanced across experiments

[18] We did not include this question in the first round of experiments. We thank an anonymous associate editor for this suggestion.

(as discussed in Section 2.5). However, unobservable experimental conditions may differ and confound estimates associated with the variable "social dilemma 2" in Table A2.2.

Table 2.7. Fears or thoughts of retaliation/conflict (social dilemma 3) and inclusion of covariates.

Dep. Variable. Fears of punishment	(1)	(2)	(3)
Time delay	0.127+	0.122+	0.123+
	(0.071)	(0.071)	(0.066)
Age			-0.001
			(0.003)
Make			-0.124
			(0.094)
Native			-0.056
			(0.084)
Risk taking			-0.100*
			(0.017)
Height (in cm)			-0.005
			(0.005)
Constant	0.490*	0.495*	2.064*
	(0.050)	(0.051)	(0.806)
N	194	193	193

Note: Linear regression model. Robust standard errors in parentheses. Symbols denote significance levels at +p < 0.1, *p < 0.05.

2.3.3 Treatment heterogeneity for help and punishment

Table A2.3 in the Appendix examines stability of estimates across sub-samples of actors as well as subject characteristics. While interpreting estimates and standard errors, it is important to note that the analysis suffers from a multiple comparison problem.[19]

First, we excluded subjects that did not notice the actor or the glove drop itself. Row 1 splits the sample into people that acknowledged witnessing the drop versus those that did not. Differences between point estimates are small and statistically insignificant. Second, we used two actors – one male and one female. One could be worried that the gender of the actor influences behavior, which may be problematic for the external validity of our results. Row 2 indicates that our point estimates are stable across actors. Of course, more actors are necessary to investigate actor-specific traits (such as height and age) that may influence the behavior of participants. Third, subject characteristics do not matter in a systematic way (see Rows 3–7). In sum, we can document little consistent heterogeneity in the treatment impacts.

[19] A larger sample would be needed to look at several interaction terms at the same time. In particular, the punishment regressions suffer from this caveat given that the majority of subjects defected and sample sizes become small across sample splits.

2.4 Discussion and conclusion

We contribute to the literature on the origins of human cooperation by studying how intuition and deliberation govern decisions to help or punish strangers at a personal cost. We ran a natural field experiment using actors that lost a personal item and/or litter as first done by Balafoutas and Nikiforakis (2012) and Balafoutas et al. (2014a). We extended this existing methodology by randomly manipulating decision times available to subjects. Time pressure was meant to trigger intuitive decisions and time delay was used to elicit deliberative ones. The aim was to test if time delay reduces helping and punishment rates.

At a more general level, one contribution of this paper is that we can alleviate the concern that one too easily conflates human error and intuition in a lab setting (Recalde et al., 2014). While of course we cannot fully rule out errors or confusion, a field experiment using an everyday situation offers a natural setting. "Bending down and picking up a glove" as many of our subjects did is plausibly less error-prone than pushing a button in the lab. We cannot, however, rule out that defectors made an error under time pressure. Fortunately, such errors would work against us finding an anti-social effect of time delay.

We document that time delay decreases the likelihood of helping strangers in the field. These results are in line with evidence from lab experiments showing that time delay lowers rates of cooperation (see Rand et al., 2012; Rand et al., 2014), as well as recent theoretical work by Bear and Rand (2016). One important difference compared to many lab experiments is that our setting assesses helping behavior without strategic interactions.

How do rates of cooperation and time delay effects compare with previous studies? Rand et al. (2014) find that under time delay, contributions to public good games decrease by 21%. A meta study of 67 studies by Rand (2016) suggests that inducing deliberation decreases cooperation by 17.3%. In our set-up, time delay leads to a 24% reduction in the likelihood of helping a stranger (p-value=0.030, chi2(1)=4.73, n=129). The corresponding impact on helping a norm violator is as similarly important in magnitude (27%) but only weakly statistically significant (p-value=0.12, chi2(1)=2.46, n=117; p-value=0.083, one-sided Fisher's exact test).

Similar to previous studies and on the basis of our findings we argue that, on average, humans are naturally predisposed to help strangers. To examine this more explicitly we asked our subjects if they found it difficult to make the decision to help or not (see post-

experimental survey in the Appendix). The responses indicate that time delay made it significantly harder to make a decision. The proportion of people that reported “quite a lot” or “a lot“ of difficulty was 26% points higher for those in the time-delayed treatment in the case of helping a stranger (p-value=0.00, chi2(1)=11.32, n=122) and 15% points in the case of helping a norm violator (p-value=0.14, chi2(1)=2.22, n=101).

Our second main hypothesis - that time delay lowers direct punishment rates - was not supported by the data. The treatment effect was small and insignificant (−2.24% points, p-value=0.55, chi2(1)=0.37, n=236). This non-result is somewhat surprising, given that we do find that people fear retaliation in response to punishment and given that time delay magnifies this fear by 12.7% points (p-value=0.08, chi2(1) =3.16, n=194). In addition, recent work has underlined the importance of counter-punishment in the field (Balafoutas et al., 2016). So, ceteris paribus, time delay should also lower punishment.

What can this null-finding tell about direct punishment more generally? While a careful investigation of this issue requires additional work, three remarks are in order. First, Peysakhovich et al. (2014) report that a subject's experimental behavior is highly persistent across different cooperation tasks and also stable across time. However, choices in cooperation and punishment games are not related. In other words, cooperation and punishment seem to have different psychological underpinnings. It is thus plausible that time delay might affect only helping but not punishment in the field. Second, experimental manipulations of punishment behavior seem to work differently in lab and field experiments (Balafoutas et al., 2016). Also contrary to previous lab evidence, Balafoutas et al. (2016) find that a manipulation of the severity of the norm violation has no effect on punishment rates. Finally, overall punishment rates are low in our data (only 10% of subject punished) and socio-demographic characteristics explain only 10% of the variation. At the same time, the fear of counter-punishment is high and increases with time delay. The denouement of this discussion is that cooperation and punishment appear to be driven by different processes and future experiments need to examine punishment, the fear of counter-punishment and the related cognitive processes in more detail.

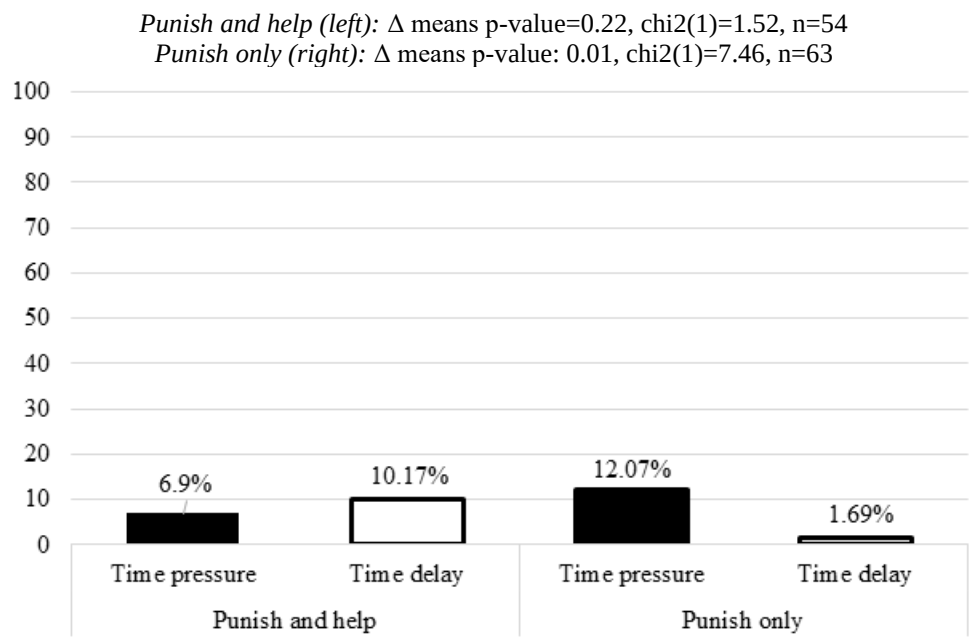
Our study can inform future research on human cooperation and punishment, both in the lab and the field: First, participants have different abilities to digest information and make decisions. While randomization ensures balance across ability types, the average effects documented in our paper may conceal heterogeneity in terms of individual processing speeds

(Rubinstein, 2007). The speed, complexity, and importance of the dilemma itself may influence human behavior (see related studies by Fehr and Rangel,2011). Second, the impact of time delay may be specific to the helping and punishment task and context. Future work may investigate the stability and universality of our estimates across space, time, and task (Rieger and Mata, 2015). In other words, our methodology could change along various dimensions such as the type of time manipulation, different social dilemmas, research team and country setting. Third, it may be worth investigating heterogenous effects specific to actor traits such as gender, height, country origins and age. Fourth, laboratory studies should continue uncovering the underlying mechanisms driving intuitive decision making and how they evolve over time (for recent research along these lines see Peysakhovich and Rand, 2016). Related, it would be interesting to examine intuitive helping and punishment behavior among children as they age and engage in new experiences. Finally, our null result on punishment behavior warrants additional work.

2.5 Annex

2.5.1 Figures

Figure A2.1: Direct punishment **and** helping a stranger (social dilemma 2) by decision time treatment



Note: P-values stem from chi-squared tests. Figure 3, Panel A presents overall rates of punishment. The figure above decomposes these rates into subjects that punished and helped, and those that only punished. When interpreting the time manipulation patterns in the sub-samples, it is important to bear in mind that this is an endogenous sample split.

2.5.2 Tables

Table A2.1. Helping behaviors by treatments (in %)

Stage	**Behavior**	**Helping a stranger**		**Helping a norm violator**	
		Time pressure	*Time delay*	*Time pressure*	*Time delay*
Behavior before decision	Looks around	12	5.7	6.5	13
	Hesitates	14	20	29	13
	Voice alert	28	28.6	22.6	17.4
	No reaction	46	45.7	41.9	56.6
Helping behavior	Physical contact & points	42	40	35.5	34.8
	Bends down & picks up glove	32	25.7	22.6	30.4
	Voice alert & points	26	34.3	41.9	34.8

Table A2.2. Crowding-out test (indirect vs. direct punishment) pooling social dilemma 2 and social dilemma 3

Dep. var. Direct punishment	(1)	(2)	(3)
Dilemma 2	-0.020	-0.035	-0.017
(crowding out)	(0.030)	(0.032)	(0.031)
Time Delay	-0.049+	-0.073*	-0.062+
	(0.029)	(0.033)	(0.032)
Age			0.004*
			(0.001)
Male			-0.070*
			(0.034)
Native			-0.006
			(0.040)
Risk taking			0.024*
			(0.009)
Height (in cm)			0.002
			(0.002)
Constant	0.113*	0.134*	-0.542
	(0.025)	(0.029)	(0.352)
N	353	294	294

Notes: Linear regression model. Unlike in the main analysis, subjects that punished *and* helped are coded as zero for the purpose of the crowding-out test. Robust standard errors in parentheses. Symbols denote significance levels at +p<0.1, *p<0.05.

Table A2.3: Time delay impact heterogeneity on helping (social dilemmas 1 and 2) and direct punishment (social dilemmas 2 and 3) in sub-samples

	Sample	(1) Helping a stranger (Dilemma 1)	(2) Helping a norm violator (Dilemma 2)	(3) Direct Punishment (Dilemma 2)	(4) Direct Punishment (Dilemma 3)
(1)	Full sample	-0.19*	-0.15+	-0.07	-0.02
		(0.08)	(0.09)	(0.06)	(0.04)
	Acknowledged seeing	-0.18*	-0.14	-0.07	-0.02
	glove drop/littering(col. 4)	(0.08)	(0.09)	(0.07)	(0.04)
	P-value Δ	*0.70*	*0.87*	*0.93*	*0.70*
(2)	Male Actor	-0.22+	-0.10	-0.13	-0.05
		(0.12)	(0.13)	(0.1)	(0.05)
	Female Actor	-0.15	-0.19	-0.00	0.00
		(0.12)	(0.13)	(0.09)	(0.06)
	P-value Δ	*0.67*	*0.62*	*0.33*	*0.45*
(3)	Male subject	-0.16	-0.12	-0.12+	-0.02
		(0.10)	(0.11)	(0.07)	(0.04)
	Female subject	-0.20	-0.23	-0.00	-0.03
		(0.13)	(0.15)	(0.13)	(0.06)
	P-value Δ	*0.81*	*0.57*	*0.41*	*0.89*
(4)	Age above median	-0.27*	-0.14	-0.10	-0.06
		(0.1)	(0.13)	(0.11)	(0.07)
	Age below median	-0.14	-0.19	-0.09	-0.04
		(0.13)	(0.14)	(0.06)	(0.04)
	P-value Δ	*0.42*	*0.81*	*0.92*	*0.82*
(5)	Native	-0.20*	-0.20+	-0.15+	-0.07
		(0.09)	(0.11)	(0.08)	(0.05)
	Other	-0.22	-0.04	0.08	0.03
		(0.19)	(0.21)	(0.07)	(0.09)
	P-value Δ	*0.91*	*0.51*	*0.04*	*0.37*
(6)	Height above median	-0.14	-0.13	-0.12	-0.05
		(0.10)	(0.14)	(0.10)	(0.05)
	Height below median	-0.22	-0.21	-0.1	-0.04
		(0.13)	(0.14)	(0.10)	(0.06)
	P-value Δ	*0.68*	*0.67*	*0.89*	*0.92*
(7)	Risk taking above median	-0.04	-0.22+	-0.09	-0.07
		(0.11)	(0.13)	(0.10)	(0.06)
	Risk taking below median	-0.38*	-0.07	-0.13	0.03
		(0.13)	(0.14)	(0.10)	(0.03)
	P-value Δ	*0.05*	*0.44*	*0.74*	*0.15*

Note: Linear regression models. P-values below estimates stem from tests of the equality of coefficients. We use full sample medians to investigate heterogeneities in terms of subject's age, years lived in the Netherlands, height, and risk preferences. Native refers to those who have always lived in the Netherlands. Models on the sub-samples are jointly estimated. Robust standard errors in parentheses. Symbols denote significance levels at +p<0.1, *p<0.05.

2.5.3 Photos [20]

Photo A2.1: Helping a stranger in the time pressure treatment (baseline)

Photo A2.2: Helping a stranger in the time delay treatment

Photo A2.3: Helping a norm violator in the time delay treatment

Note: Notice the empty plastic bottle and the glove in the scene.

[20] All photos were made after the experiment and scenes were re-enacted for illustration.

Photo A2.4: Characteristics of location and position of researcher

Note: The location and the position of the researcher permits private and anonymous decisions.

Photo A2.5: Characteristics of participants

Note: A subject must be alone and in no visible hurry nor visibly distracted. There is also no other subject coming from the opposite direction.

Photo A2.6: Location, front view

Photo A2.7: Location, left view

Photo A2.8: Location, right view

Photo A2.9: Location, back view

Photo A2.10: Example of helping behavior: participant bends down, picks up the glove and gives it back to the actor

Photo A2.11: Example of helping behavior: voice alert and pointing

Photo A2.12: Participant defects

Photo A2.13: Post-experimental survey

.

Photo A2.14: Close-up of the actor surveying the participant after the experiment

2.5.4 Surveys

A2.1 Post-experiment survey (Social Dilemma 1 and 2)

After each participant crossed Point A, the actor/actress followed the participant and asked: “Excuse me, I am a researcher of Erasmus University and I just (littered and) dropped my glove as an experiment. Could I ask you a few quick questions? We can walk together if you want.”

1. Did you see the (littering and the) drop of the glove ? 0. No _____. 1. Yes _____.
2. How willing are you to take risks in general? From 0 to 10 where max. is 10: _____.
3. What is your height in cm? _____ cm.
4. What is your age? _____ years.
5. How long have you lived in The Netherlands? _____ (in years / months).
6. How difficult was to make the decision of what to do? 0. Not at all _____. 1. Just a little _____. 2. Quite _____. 3. A lot _____.

Note: The survey was administered in English, since The Hague is an international city and a large majority of people speak English.

A2.2 Post-experiment survey (Social Dilemma 3)

The survey was administered in the same way as the previous survey for Social Dilemma 1 and 2 during the follow-up experiment a year later. Ordering of questions was changed to better accommodate a new question: Question 3 below was added. Again, the survey was administered in English, since The Hague is an international city and a large majority of people speak English.

After each participant crossed Point A, the actor/actress followed the participant and asked: "Excuse me, I am a researcher of Erasmus University and I just littered as an experiment. Could I ask you a few quick questions? We can walk together if you want."

1. Did you see me littering? 0. No _____. 1. Yes _____.
2. How difficult was to make the decision to punish? 0. Not at all _____. 1. Just a little _____. 2. Quite _____. 3. A lot _____.
3. Where you afraid to punish for fear of retaliation (or thoughts about it)? 0. No _____. 1. Yes _____.
4. What is your age? _____ years.
5. How long have you lived in The Netherlands? _____ (in years / months).
6. From 0 to 10, how willing are you to take risks in general? (max. is 10): _____.
7. What is your height? _____ cm.

Chapter 3: Averting housing discrimination? Evidence from Craigslist and references in Amsterdam

3.1 Introduction

Housing discrimination continues to hinder the wellbeing and living standards of ethnic minorities. The literature indicates that residential location influences access to better education, employment, credit, security and healthcare even in advanced societies (Cohen-Cole, E., 2011; Brevoort, 2011; Urban, 2009; Aalber, 2007; Buck, 2001; Cutler and Glaeser, 1997). One area of research examines the high prevalence of exclusion experienced by Arab/Muslims minorities in private housing markets in Europe (for a detailed review see Table 3.1). However, there remains scarce evidence about instruments that may avert discrimination and/or enhance access in this market. Combining these notions, the present study asks: Does a less prominent -secondary- platform avert discrimination against (male) ethnic minorities in the housing market? And likewise, does a *(positive)* reference letter from a previous landlord enhance access? To answer these questions, this article conducts two separate experiments on an auxiliary platform, namely Craigslist in Amsterdam (The Netherlands). The first trial evaluates discrimination against Turks and Moroccans, and the second tests the impact of a reference letter on the likelihood of accessing housing. [21]

The historical prevalence of housing discrimination against minority groups motivates the search for settings that are less prejudicial and stimulate access. This study first examines whether a less prominent and secondary website is less likely to be associated with discrimination, in contrast to past evidence which has been based on the leading and most prominent housing platforms in various European markets (review Table 1).[22] There are two main reasons why a secondary platform could lead to less prejudicial outcomes for marginalized groups. First, in secondary settings each rental offer is likely to receive a lower quantity of applications as the platform receives less visits from candidates looking for housing. This lower demand for housing is likely to increase interest and attention for each applicant and each potential applicant will have a higher likelihood of selection given the limited competition. This aspect would thus increase the chances of accepting candidates

[21] The Netherlands has a long history of receiving migrants from Turkey and Morocco (Zorlu and Hartog, 2001). In 2019, residents with (first or second generation) Turkish background accounted for 10% of residents with a migrant background and 2.7% of total residents. Likewise, residents with Moroccan ancestry accounted for 9.8% of migrants and 2.3% of the total residents (CBS, 2020).

[22] Table 3.1 provides a detailed review on the prevalence, size and countries of housing discrimination in past research. All studies show evidence of discrimination towards ethnic minorities in the leading and most prominent website platforms in European countries (except from those done via telephone).

with minority profiles even in cases where renters would prefer otherwise. Second, an auxiliary platform may be viewed as a website intended for minorities in a country as opposed to those who advertise and access more prominent websites. In cases where agents using secondary platforms are sufficiently different from the majority, their decisions may be less prejudicial towards minority groups as compared to agents participating in leading platforms. In short, agents using secondary platforms may be from minority groups, and thus also more accepting of other members who have a minority status.[23]

The study explores the first question using the Internet-based platform Craigslist in Amsterdam for two reasons. First, Craigslist is a secondary website that is active every day in The Netherlands, but to a lower extent than most prominent platforms.[24] Craigslist is free of charge, it is widely known inside and outside the country, and it displays about 100 new rental offers per month (during the period of this study). In comparison, the most popular websites: funda.nl has about 43 million visits per month, and Pararius.com has 1 million visitors per month (Funda.nl, 2020; Pararius, 2020). Moreover, the posts indicate that Craigslist is mainly used by local expats and foreigners looking to relocate. This trait suggests that Craigslist offers an environment where advertisers have very different characteristics as compared to leading Dutch platforms. Local expats and foreigners may perhaps be more tolerant, open-minded and as a collective inherently more diverse (coming from diverse backgrounds) and have attitudes that are more accepting of other cultures. This may translate into reduced discrimination against ethnic minorities. Second, Craigslist is only available in Amsterdam. This sharpens the focus of the research although it prevents a country-wise intervention. Of perhaps greater relevance, Amsterdam is the capital and most populated city in the country. It faces permanent pressures from residents and foreigners searching for home, and historically it has amongst the highest number and proportion of immigrations and relocations in the country. For instance, in 2018 the Greater Amsterdam region had 109,000 new residents comprising 11% of total moves (CBS, 2020). Based on these considerations, Craigslist offers a useful platform to study the first question of this research.

[23] For instance, non-Dutch speaking Western minorities in Amsterdam maybe more willing to engage with non-Dutch speaking non-Western minority groups.

[24] Craigslist.com did not respond to numerous messages from the author requesting information about the number of visitors and visibility of the website inside The Netherlands.

The second question of this research examines the impact of a *(positive)* reference letter on access to housing. Two central motives explain why a reference letter from a previous landlord may foster access in housing markets. First, traditional network structures in modern societies favor the selection of more privileged and connected groups to find successful rental accommodation. This system of informal referrals exacerbates inefficiency and inequality in renters' choices as it leads to favor friends, relatives, or close members (Beamen and Magruder, 2012; Loury, 2006; Montgomery, 1991). Instead, the inclusion of a formal reference letter in the first step of an application process may help to counterbalance the *a priori* negative preferences and/or perceptions about the characteristics of an unknown candidate (given inherent qualities such as ethnicity).[25] Second, landlords may improve the credibility and knowledge about an applicants' hard and soft skills by examining a formal referral from a past valid informant. This action offers an opportunity to match the most qualified and ideal candidate in terms of character and dependability to the best rental option. The information in the reference thus surpasses the typical and plain application criteria commonly used (i.e. name, education, employment, and/or disposable income). Hence, this instrument may alleviate the systematic and intrinsic information asymmetry between renters' and house owners that often causes adverse effects in the housing market (Abel et al., 2017).

The present study contributes to the literature in three ways. First, this study presents evidence on whether the use of a secondary website platform is associated with less discrimination against ethnic minorities in the housing market of a European capital. This evidence adds to the conventional literature concerned with measuring the prevalence of discrimination in the leading and most prominent country-specific websites. Second, it provides evidence on whether a reference letter enhances access of ethnic minorities to housing. This work is relevant as there is scarce research on instruments that stimulate access to the housing market for ethnic minorities. Standard work only examines the benefits of applicants' information in terms of which specific socio-economic predictors lead to better or worse rates of discrimination (i.e. Andersson et al., 2012; Bosch et al., 2010). There is also one article that has examined the impact of external policy messages on lessors' attitudes (Fang et al., 2014). Two other labor market studies have found that references stressing candidates' personality and skills can alleviate information frictions and improve accessibility of minority groups in South Africa and Germany (Abel et al., 2017; Kaas and Manger, 2011). Third, in contrast to existing studies which rely on a few profiles, this study uses a large pool

[25] This situation closely resembles the case of discrimination in labor markets (Abel et al., 2017).

of ethnic profiles. The study employs 48 fictitious profiles across three identities. This enables us to examine whether there are specific commonly used ethnic names (for instance, Mohammed) which elicit adverse responses (or not) or is the phenomenon more widespread.

Overall, the study tests three hypotheses in two experiments. The first experiment tests whether seeking housing through Craigslist mitigates discrimination against Turkish *(Hypothesis #1)* and Moroccan candidates *(Hypothesis #2)*. Separately, the second experiment tests whether including a positive reference letter from a previous Dutch landlord improves the likelihood of renters' responses towards ethnic minorities *(Hypothesis #3)*.

The research examines these three hypotheses by implementing two audit-based experiments. The identification strategy employs an audit-system method to compare the rate of renters' responses from fictitious candidates. In Experiment #1 the study compares responses to male Turks and Moroccans versus -*majority*- Dutch male candidates to examine housing discrimination. The literature has shown that men experience stronger discrimination than women candidates (Carlsson and Eriksson, 2014). Thus, if there is no discrimination against men on Craigslist then it may be expected that there will be no discrimination against women from ethnic minorities as well. Experiment #2 investigates the impact of a positive reference letter from a previous fictitious landlord, included in the application, versus a control group (same ethnicity with and without reference). The inclusion of the reference letter is expected to enhance the tenancy prospects of the ethnic minority applicant.

The study confirms that there is no housing discrimination against Turks and Moroccans on Craigslist, and that the reference letter has no effect on access. In Experiment #1, the results show that as compared to Dutch candidates, applications from Turkish candidates have a lower call-back of -1.8%-points (p-value=0.68, n=414) while Moroccans experience a lower call back of -2.1%-points (p-value=0.60, n=400). In both cases the effects are statistically insignificant. In Experiment #2, the reference letter has a negative and insignificant effect in the pooled sample of applications of -3.6%-points (p-value=0.34, n=636). The study concludes that a reference letter has no impact on improving housing access for ethnic minorities (*rejecting Hypothesis #3*).

The remainder of this article is structured as follows: Section 2 provides background details on housing discrimination in The Netherlands. Section 3 details the experimental design of

the two trials. Section 4 discusses the data characteristics. Section 5 provides results, and Section 6 discusses the results and concludes.

3.2 Background on housing discrimination in The Netherlands

A study of housing discrimination against Arabs/Muslims minorities in The Netherlands is interesting for three reasons. First, although the Netherlands has a long tradition of migration, cultural mixture and acceptance (Zorlu and Hartog, 2001), in the past decade Dutch and European society has experienced three dramatic events that may have intensified prejudice, discrimination and negative attitudes towards Arabs/Muslims: *(i)* the International Economic Crisis that started in 2008 has exacerbated discourses and policies favoring Europeans before immigrants, *(ii)* the intense influx of refugees from the Middle East since 2015, and *(iii)* the increase in the number and strength of terrorist attacks in Europe from individuals with Arab/Muslim backgrounds (Auspurg et al., 2017; Baldini and Federeci, 2011). An analysis from Ipsos in 2014 reports that 47% of Dutch residents say that cultural mixing is "a problem", 55% consider the number of immigrants "too high", and 57% overcalculate the presence of ethnic minorities in The Netherlands (NU, 2014). Hence, the current context challenges the international and historical perception of The Netherlands as a society with high cross-cultural trust, cooperation, and tolerance.

Second, there is strong and persistent evidence of labor discrimination against Turkish and Moroccan residents in The Netherlands. Past research has shown that candidates with Turkish and Moroccan profiles have lower chances of obtaining an invitation for a job interview (Blommaert et al., 2014; Andriessen et al., 2012; Derous, 2011; Derous et al., 2009; Bovenkerk et al., 1995; Büyükbozkoyum, 1991).[26] Recent studies have found that these ethnic minorities may experience penalties that can reach up to 25 percentage points (Bloommaert et al., 2013; Andriessen et al., 2012). Residents with Turkish and Moroccan backgrounds are also more likely to have temporary contracts as compared to Dutch workers, even after controlling for age, educational skills, and gender (Andriessen and Dagevos, 2007; Dagevos, 2001). Moreover, an index of ethnic hierarchy gives a low rank to Moroccans and Turks in terms of ethnic similarities with native Dutch people and their overall capacity to integrate in The Netherlands (Dagevos and Gijsberts, 2007; Verkuyten and Zaremba, 2005; Hagendoorn and Hraba, 1989). A related study finds that Moroccans are in the second lowest

[26] Results indicate that 50% of candidates with non-Western backgrounds have positions where they use lower skills than their education, whereas the figure halves to 25% in residents from the -*majority*- Dutch labor force (Andriessen et al., 2012; Andriessen, 2010).

position in employers' likability across four ethnic minorities (together with Turkish, Surinamese and Antillean residents). Turkish residents obtained the highest position while Antillean residents were placed fourth (Nievers, 2010; Kruisbergen and Veld, 2002).

Third, there is historical indication of substantial housing discrimination against residents with ethnic minority profiles across European countries such as Belgium, Germany, Sweden, Finland, France, Spain, Italy, and Greece (review Table 3.1). For instance, in Sweden males with the name Mohammed are 24.8 percentage points (p.p) less like to receive a response from a landowner than candidates with Swedish-sounding names such as Erik (Ahmed and Hammarstedt. 2008). In Germany candidates with Turkish names are 9.1 p.p less likely to be able to access their desired housing as compared to applicants with German names (Auspurg et al., 2017); while in Italy, Arab/Muslim applicants are discriminated by 23.1 p.p across the country (Baldini and Federici, 2011). In Belgium, ethnic minorities with (without) accent are discriminated by 15.5 p.p (12.4 p.p) as compared to candidates with a native accent (Van der Bracht et al., 2015). Notably, to the best of my knowledge, despite this vast and persistent evidence, there is no academic publication examining this subject in the (private) Dutch housing market.[27] However, there are reports in the popular press. In a newspaper article, a fictitious Moroccan candidate - Rachid El Haddaoui, experienced discrimination of 28 p.p compared to Jaap van de Ven (Elibol and Tielbeke, 2018). In addition, previous studies have found that race harms access to Dutch mortgage loans (Aalbers, 2007), and that language restrictions and limited knowledge of housing laws prevent immigrants from making legal complaints against local rental managers (Will, 2003).

[27] Amongst the main reasons for the limited evidence is that real state agencies dominate the administration of the Dutch private housing market (for instance refer to: funda.nl or pararius.com). First, this factor is *apriori* assumed to reduce the potential discrimination from landowners. Second, the large presence of specialized rental agencies means that a few agents may manage numerous rental offers increasing the probability of detecting fictitious candidates that would affect the research.

Table 3.1: Review of literature on housing discrimination in European markets

Authors (journal, year)	**Country (city, location)**	**Treatments**	**Design**	**Desc. Results** (for men)	**Impact** (with controls)
Baldini and Federici (*J. of Housing Economics*, 2011)	Italy (41 cities) in Website: Subito.it	Rental applications of Italian, Arab/Muslim and East-European men and women (with and without information on quality of individual)	(RAP) Random Assignment Procedure (one per offer)	Italians: 62% Arabs: 39% -23.1%-points (p<1%, n=3,676)	-28.3%-points (p<1%, n=1,438)
Auspurg et al. (*J. of Housing Economics*, 2017)	Germany (Munich in Newspaper: *Süddeutsche Zeitung*, and website: *immoscout24.de*)	Rental applications varying with ethnicity (German or Turkish), occupation (high or low), and marital status (married or single).	(RMPM) Randomly Matched Paired Method (two per offer)	Germans: 70%, Turkish: 61.5% -9.1%-points (p<1%, n=1,274)	n.a
Andersson et al. (*Land Economics*, 2012)	Norway (all available in website: finn.no)	Rental applications varying with ethnicity (Norwegian or Arab), and employment (bank vs. warehouse)	RAP	N. men: 58.5%, Arab men: 46.4 (12.1%-points, p-value<1%)	-12.3%-points (p<10%, n=950)
Bosch et al. (*Regional Sc. And Urban Economics*, 2010)	Spain (20 of the largest cities website-based) in Website: loquo.com	Spanish vs Moroccan names with "standard vs high-quality" applicants (about current employment and occupation)	RMPM	-20.8%-points (p<1%, n=2,240)	-23%-points (p<1%,n=1,997)
Acolin et al. (*J. of Urban Economics*, 2016)	France (in website: non-disclosed)	Rental applications varying with ethnicity (French vs Northern Africa, Sub-saharan Africa, Turkey, Poland, and Portugal-Spain)	RMPM	France: 72.3%, Turkey: 63.4% (p-value<5%)	n.a
Ahmed and Hammarstedt (*J. of Urban Economics*, 2008)	Sweden (in website: blocket.se)	Rental applications from Swedish male, Swedish female, and Arabic/Muslim male	RMPM (three per offer)	Swedish male: 41%, Arab male: 18% (24.8%-points, p<1%, 1,000)	-21.3%-points (p<5%,n=1,056)
Öblom and Antfolk (*PLoS One*, 2017)	Finland (in website: Tori.fi, in Helsinki, Turku and Tampere)	Rental applications varying with ethnicity (Finish, Swedish and Arab), and gender (male vs female)	RMPM	Males, Arabic: 18% Finish: 34% Swedish: 38%	-16.8 (p<1%)
Bonnet et al. (*Urban Studies*, 2015)	Paris (in website: seloger.com leading internet platform at the time)	Phone-calls varying with ethnicity (French vs Arab), and gender (male vs female)	RMPM (varying location and ethnicity in conversations)	When ethnicity goes first: -9.10% (p=0.32, n=77)	Ethnicity first: 11.2% (p>10%) -no impact-
Drydakis (*J. of Population Economics*, 2011)	Athens (diverse local newspapers)	Phone-calls from women varying with ethnicity (Greek and Albanian), and rent (low, middle and upper)	RAP	-25.6% (p<1%, n=4,884)	n.a
Carlsson & Eriksson (*J. of Ethnic and Migration Studies*, 2015)	Greater London Area (local website, not found in article)	Rental applications from male varying with ethnicity (British, Eastern European, Indian, African and Arabic), and skill job (low vs. high)	RAP	British: 54% Arabic: 34% (p<1%, n=5,143)	-20%-points (p<1%)
Bengtsson et al. (*Applied Economic Letter*, 2012)	Stockholm (in website: blocket.se)	Rental applications from male varying with ethnicity (Swedish vs Arabic)	RMPM	n.a	-8.6%-points (p<1%,n=1,110)
Ahmed et al. (*Land Economics*, 2010)	Stockholm (in website: blocket.se)	Rental applications from males varying with ethnicity (Swedish vs Arabic)	RAP	n.a.	-16.8%-points (p<1%, n=841)
Sacherová (bachelor thesis, 2016)	Slovakia (in website: mojeanketa.cz)	Rental applications from males varying with ethnicity (Slovak vs Roma)	RMPM	Slovak: 47.5% Roma: 39.4%	-9.5%-points (n=336, p<10%)

3.3 Experimental design

This section provides details on the experimental design including the location, the two experiments, the justification of the ethnic names used, the application procedure, and the econometric approach.

3.3.1 Craigslist as housing website platform in Amsterdam

For all communications and data collection the study employed the online website Craigslist.org in Amsterdam (The Netherlands).[28] Craigslist is a useful Internet-based platform that provides opportunities to buy, sell, and rent all types of goods and services including housing and employment. With regard to housing, private individuals, landowners, and real estate agents may upload offers to lease houses, apartments, and rooms, and tenants may submit applications when interested in an option. All participants can interact privately and anonymously for free, and without restriction to size or price of the housing unit. This feature minimizes social and reciprocity pressures in the decisions of lessors (as explained by Hanson and Hawley, 2011). To have full access to the services of the website every agent must register and provide an email address. The experimental design applies to housing offers made by rental managers and combines offers from landlords and real estate agents (but excludes the same real estate agencies to avoid detection and repetition).

3.3.2 The two experiments

The study implements two field experiments to evaluate rental discrimination (*Experiment #1*) and the impact of a reference on housing requests (*Experiment #2*). The main approach in both studies is an audit-system method whereby two messages are submitted to a single rental offer using written applications via e-mail. One message includes the treatment to evaluate, and the other message is the control group. This approach reduces potential biases in results caused by direct-physical or voice contact with the renter. Experiment #1 evaluates housing discrimination against ethnic minorities by comparing applications from Turkish or Moroccan candidates versus requests from -*majority*- Dutch applicants. Experiment #2 examines the impact of a positive reference letter from a previous landowner versus a control group (without letter). Table 3.2 provides an overall view of the experiments and treatment groups, and the following two sub-sections provide further details on the design of each experiment.

[28] For more information refer to the website: amsterdam.craigslist.org.

Table 3.2: The two experiments: treatments and ethnic groups

Experiment #1: Ethnic discrimination	**Experiment #2: Impact of reference letter**
Treatments #1 and #2	Treatment #3 (pooled sample)
(a) Turkish vs Dutch candidate (b) Moroccan vs Dutch candidate	(a) Turkish candidate with and without reference letter using a Dutch name (b) Moroccan candidate with and without reference letter using a Dutch name

3.3.2.1 Experiment #1: Testing housing discrimination against ethnic minorities

The first experiment investigates prejudice against candidates with Turkish and Moroccan identities as shown in Table 3.2.[29] The design examines the response rates from rental managers to housing requests made from each of the three ethnic identities: *(i)* Turkish, *(ii)* Moroccan or *(iii)* -*majority*- Dutch. The first group of messages investigates discrimination against Turkish candidates by comparing managers' responses to applicants from Turkish versus Dutch candidates. The second group of messages is from Moroccan applicants and compares requests from Moroccan candidates versus requests sent by -*majority*- Dutch candidates. Thus, the study sent applications with Turkish or Moroccan identities separately, and in every case, forwarded a second message using a Dutch identity.[30] In both cases, the design uses messages from the ethnic minority as treatment and compares responses to messages sent by Dutch candidates who serve as a control. Panel A of Annex Table A3.1 displays the message scripts sent in Dutch (and translated to English) for Experiment #1.

3.3.2.2 Experiment #2: Impact of a positive reference letter using a Dutch name

The second experiment tests the impact of a positive reference from a previous landlord (with a Dutch name) on both ethnic minorities using the audit-system as outlined in Table 3.2.[31] The design evaluates the rate of responses from rental managers to housing requests from Turkish or Moroccans candidates with a reference letter (treatment) versus similar candidates without references (control group). Thus, the study submitted applications with Turkish or Moroccan identities separately that included or did not include a positive reference letter. Experiment #2 examines the impact of the reference letter on the pooled sample. The

[29] The study submitted 1,136 applications to a total of 568 offers to detect a discrimination of 12 percentage points for each ethnic minority. The power calculation was done in STATA with a power of 80% and statistical significance of 5%.

[30] The order of the messages was chosen at random before start of fieldwork. This procedure is explained in Section 2.4 on "Procedure in applications".

[31] This second experiment aimed at a total of 1,650 applications to 825 requests to identify an impact of 7 percentage points with a power of 80% and statistical significance of 5%.

additional message scripts and specific reference letter for Experiment #2 are shown in Panel B and Panel C of Annex Table 1, respectively.

3.3.3 Generation of ethnic identities: Dutch, Turkish and Moroccan male candidates

The design generated a combined total of 48 fictitious male candidates across the three ethnic groups. The field experiment thus employed 16 profiles using names with -*majority*- Dutch identity, 16 profiles using Turkish names, and 16 profiles with Moroccan identity as shown in Table 3.3. The comparison between Dutch and Turkish or Moroccan names is convenient given the historical context and since the names of both minority groups are very different from the Dutch population (Zorlu and Hartog, 2001). Notably, the research focuses on male applicants given that past evidence in European housing markets show that male ethnic minorities experience stronger discrimination than female candidates (Carlsson and Eriksson, 2014).

Table 3.3 Applicants names: Dutch, Turkish and Moroccan candidates

Profiles	Dutch	Turkish	Moroccan
First names	Willem, Jan, Peter, and Dirk	Mehmet, Mustafa, Murat, and Hasan	Yilamz, Sahin, Kaya, and Öztürk
Family names	De Jong, Jansen, De Vries, and Van den Berg	Mohammed, Ahmed, Said, and Rachid	Loukili, Moussaoui, Idrissi, and Tahiri

Note: The Meertens Institute (2018) provided the information.

The experimental design reviewed official government statistics stored by the Meertens Institute to select credible and realistic names in the generation of the fictitious candidates' profiles (The Meertens Institute, 2018). The approach searched for the four most popular (male) first and last names of residents in The Netherlands with -*majority*- Dutch heritage, and the four most common (first and last) names for residents with Turkish and Moroccan heritage (at least one parent). Thus, the fictitious profiles signal ethnic minorities with second-or-higher-generation immigrant origin or first-generation origin with legal residency. Hence, the study combined the (4) first and (4) last names from each ethnic group to create a pool of 16 profiles with full names for a complete pool of 48 identities.

3.3.4 Procedure: Generation of profiles, data collection and message applications

A total of 48 fictitious e-mail accounts in *gmail.com* were created to apply to each rental offer selected from Craigslist.org. These 48 accounts correspond to each of the 48 fictitious names generated to signal Turkish (n=16), Moroccan (n=16) and Dutch identities (n=16) as outlined in Table 3.3. The Gmail electronic accounts were generated with the structure:

name.lastname.756. On a few occasions the addresses were expanded to four digit-numbers because the initial option was already taken. An Excel file was used to record details about each rental offer selected at extraction. This file recorded the main information about the type of housing (apartment, house or room) and other characteristics (price in euros, area in square meters, location, number of rooms, furnished or not, available parking space, presence of balcony/terrace or garden, and luxurious design). The file also included specific identifiers of each advertisement where possible (title of advertising, names, and email accounts), as well as the particular features of the experimental treatments and results that were evaluated in each option (ethnic origins and names, times of applications, and managers' responses). A second folder stored a digital copy of the original website announcements in PNG or PDF format.

The study collected data for the two experiments in separate phases of fieldwork over a period of two years. The collection of data for Experiment #1 ran between July 2018 and April 2019.[32] This fieldwork gathered data to test Hypothesis #1 and #2. The fieldwork for Experiment #2 started in July 2019 and it finished on March 2020 (with the onset of the Corona crisis) to test Hypothesis #3. Importantly, the two phases of fieldwork employed the same combination of messages, and data collection took place only during working days at least three times a week. The audit-design sent each-paired set of messages with a delay of one hour and no more than two-hours of difference. The order and selection in the application of messages and treatments was set via STATA computer-randomization before the start of each fieldwork phase. This method supports randomization in treatments, in name of applicants by ethnic groups, the sort of ethnic minority, and order and type of message submitted. However, the actual number of housing applications mailed per working day varied according to available offers per day and week. All messages were submitted in Dutch as shown in Annex Table 1. In those cases where landlords responded, the researcher politely declined the offer (in some cases a reply was ignored).

3.3.5 Empirical approach

The empirical analysis is based on one outcome indicator and two treatment indicators. The outcome indicator is a renter's positive reply (*R)* to a housing application as 1, and zero otherwise. Ethnicity (*Eth*) is a binary indicator which takes on a value 1 when the applicant is

[32] The main part of the study finished on April 2019, but one observation was double-counted and re-executed on July 2019.

from a minority group (Turkish or Moroccan), and zero when the applicant has a Dutch name. The second treatment variable indicates whether an application includes a reference letter (1) or not (zero).

In the case of both experiments, the analysis proceeds along similar lines. It starts with a comparison of means and bivariate analysis of the responses to the applications within the two separate experimental designs. Subsequently, multivariate linear probability models are estimated to: *(i)* identify the impact of Turkish and Moroccan identities on responses from renters, and *(ii)* identify the impact of the reference letter on applications from ethnic minorities. Explicitly, in the first experiment, a renter's response (R) is treated as a function of an applicant's ethnic identity (Dutch, Turkish or Moroccan), and various sets of characteristics as shown in (*e1*):

$$R_{ih} = \beta_1 Eth_{ih} + \beta_2 Type_{ih} + \beta_3 Features_{ih} + \beta_4 D_{ih} + u_{ih} . \quad (e1)$$

In (*e1*), R_{ih} refers to a binary indicator which takes a value of 1 when candidate i obtains a positive response from the renter with regard to housing offer h, and zero otherwise. [33] The coefficient β_1 denotes the impact of ethnicity (Eth_{ih}). Four additional sets of variables are included. These include the type of housing (β_2), features of the offer (β_3), and a variable controlling for offers displayed more than three times to capture potential concerns such as: special preferences from the renter, particular difficulties in renting the location, or problems with the website (β_4). Lastly, the model includes an idiosyncratic error term (u_{it}). Separate models are estimated for Turkish and Moroccan profiles.

Following the literature on housing discrimination (review articles in Table 1), this analysis includes $Type_{ih}$ which distinguishes between three types of dwellings - apartment, room, or a house. Regarding the features of the dwelling ($Features_{ih}$), the specification includes the price (in euros), area of the dwelling on offer (in square meters), the number of rooms in the dwelling (continuously), a series of dummy variables indicating whether a location is furnished, has parking space, has balcony or terrace, and/or has a garden (see literature in Table 1). Two additional dichotomous variables indicate whether the offer includes all utilities (i.e. gas, electricity, and water) and whether the location has luxurious qualities in terms of expensive constructions materials and stylish appearance/design.

[33] In practice the results use total responses as there is no statistically significant difference with only positive responses as shown in Table A2.

Four variants of (*e1*) are estimated. The first model provides bivariate estimates of the ethnicity effect based on the full sample. The second specification recalculates the bivariate analysis restricting the sample to the total number of observations available in the most complete model (fourth specification). The third model includes the type of housing on offer ($Type_{ih}$). The fourth model adds the full set of housing characteristics. All regressions include the dummy indicating on whether an offer is shown more than three times on the website (D_{ih}).

A second linear probability model estimates the impact of the reference letter. The model employs the same econometric approach and group of variables as in (*e1*) with three differences. First, the renters' responses (outcome variable) are now limited to applications from ethnic minorities (Turkish and Moroccans). Second, the specification switches the regressor of ethnic identity (Eth_{ih}) for the indicator of the reference letter ($Reference_{ih}$), and third, the specification includes ethnic-profile fixed effects ($Eth\text{-}Profile_{ih}$) to control for potential differences which may arise across profiles.. The following model presents the full specification with these two adjustments:

$$R_{ih} = \beta_1 Reference_{ih} + \beta_2 Type_{ih} + \beta_3 Features_{ih} + \beta_4 D_{ih} + \beta_5 Eth\text{-}Profile_{ih} + u_{ih} \quad (e2)$$

Four variants of (*e2*) are estimated. The first model provides bivariate estimates of the ethnicity effect for the full sample, while the second provides bivariate estimates restricting the sample to complete observations. The third specification adds type of location ($Type_{ih}$), and the fourth controls for the complete set of characteristics. All variables in (e2) are defined in the same was as in (e1), and all estimates include ethnic-profiles fixed effects.

3.4 Characteristics of rental offers

This section outlines the characteristics of the two samples. A total of 2,366 housing applications were sent for 1,183 vacancies in two rounds of fieldwork between the period June 2018 to March 2020).[34] The following two sub-sections describe the characteristics of the housing offers in each sample. Since the study uses an audit-system in which a pair of applications is sent to the same housing offer, there are no differences in traits within Experiment #1 and Experiment #2. However, traits of the dwelling on offer may differ across ethnicities and across the two experiments.

[34] The study aimed to collect a total of 2,790 housing applications for 1,395 vacancies but fieldwork was suddenly interrupted on March 13th given the lockdown measures put forward to deal with the Coronavirus crisis. Fieldwork was completed for Experiment #1, and for Experiment #2, date was collected for 75% of the target sample (615 of 825 observations).

3.4.1 Experiment #1: Testing discrimination against Turks and Moroccans

The advertised characteristics of the rental offers in the samples with Turkish versus Moroccans (n=568) names are statistically balanced at 5% of significance with three exceptions (see Table #3.4). These exceptions (from a total of 13 tests[35]) show a small and positive difference in the case of Turkish applications in the area of the housing unit (0.18-points, p-value=0.03) the number of rooms (5%-points, p-value=0.00), and a small and negative difference in apartments (-5%-points, p-value=0.02).[36] On average, the rental cost is Euro 1,465 and has an area of 83 square meters. The average dwelling has two bedrooms, and most are furnished (90%). Half the dwellings include a balcony/terrace (49%) and only 10% provide luxurious accommodation. About 53% of adverts include utilities in the price, 67% offer parking space (including street), and 13% have gardens. Table 3.4 displays these characteristics for both Turkish and Moroccans applications.

Table 3.4. Summary statistics in Experiment #1

Variable	N	Mean	Std. Dev.	Min	Max	Mean		Comparison (1) vs. (2)
						Turkish (1)	Moroccan (2)	
Price	1,124	1,465.40	789.51	70	4,500	1,454.02	1,476.78	0.63
Square meters	890	83.44	46.99	8	360	85.64	81.29	0.17
Apartment	1,134	0.82	0.38	0	1	0.80	0.85	0.02
Housing	1,134	0.09	0.28	0	1	0.09	0.09	0.66
Room	1,134	0.09	0.28	0	1	0.11	0.06	0.00
Utilities included	1,076	0.53	0.50	0	1	0.52	0.54	0.40
Bedrooms	1,122	1.65	0.92	1	6	1.67	1.63	0.48
Number of spaces/rooms	1,106	3.18	1.35	1	9	3.27	3.09	0.03
Has parking	1,112	0.67	0.47	0	1	0.69	0.65	0.19
Furnished	1,118	0.90	0.30	0	1	0.91	0.90	0.73
Balcony/Terrace	1,110	0.49	0.50	0	1	0.48	0.50	0.43
Garden	1,106	0.13	0.33	0	1	0.14	0.12	0.27
Luxury	1,104	0.10	0.29	0	1	0.09	0.10	0.35

3.4.2 Experiment #2: Testing the impact of a positive reference

Table 3.5 shows that the features of the housing are balanced at 5% statistical significance in the case of all variables, except one. The Turkish applications are slightly less likely to have been sent to furnished apartments (-4%-points, p-value=0.02)[37]. The average dwelling is on offer at a price of Euro 1,220 euros and offers an area of 77 square meters. Locations have at

[35] At the 5% level the study expects on average at least 1 difference to be statistically different by chance.

[36] These variables are included for robustness in multivariate analysis (shown in results section). More importantly, the inclusion or exclusion of these variables does not change the main results as shown in complementary robustness checks in Annex Table A3.2.

[37] The inclusion or exclusion of these variables does not change main results as shown in robustness checks in Annex Table A3.3.

least two rooms and include a minimum of one bedroom. Almost all offers are furnished (93%) and about a third include a balcony/terrace (37%). In contrast, only 9% are luxurious and 16% include utilities in the advertised price, 65% have parking spaces, and 11% offer access to garden.

Table 3.5. Summary statistics in Experiment #2

Variable	N	Mean	Std. Dev.	Min	Max	Mean		Comparison (1) vs. (2)
						Turkish (1)	Moroccan (2)	
Price of offer	1,178	1219.60	772.67	85	6,000	1207.12	1231.78	0.58
Square meters	834	76.94	85.67	3	1,600	73.10	80.38	0.22
Apartment	1,040	0.55	0.51	0	3	0.82	0.78	0.07
Housing	1,226	0.80	0.40	0	1	0.03	0.05	0.17
Room	1,226	0.04	0.20	0	1	0.15	0.17	0.28
Utilities included	1,226	0.16	0.36	0	1	0.55	0.55	0.99
Bedrooms	1,074	1.35	0.63	1	4	1.35	1.34	0.61
Number of spaces/rooms	1,028	2.76	0.93	1	7	2.76	2.76	0.95
Has parking	1,056	0.65	0.48	0	1	0.64	0.67	0.31
Furnished	1,112	0.93	0.26	0	1	0.91	0.95	0.02
Balcony/Terrace	1,044	0.37	0.48	0	1	0.40	0.34	0.05
Garden	1,028	0.11	0.32	0	1	0.10	0.12	0.26
Luxury	998	0.09	0.29	0	1	0.10	0.08	0.45

3.5 Results

This section presents the results of the two experiments. The first study examines housing discrimination against Turks and Moroccans, while the second assesses the impact of a positive reference from a previous Dutch landowner in applications submitted by ethnic minorities. The main analysis focuses on the likelihood of receiving a response to an application versus an application being ignored. In some cases, the response was a rejection, but the results are not sensitive to the exclusion of such responses. For instance, the overall response rate was 32.8% when pooling together both experiments. This response rate was composed of 23.3% who asked for more information about the candidate, 4.4% offered a viewing or asked for a phone conversation (1.8%). In 2.3% of the cases, renters rejected the application stating that the unit was already taken, and 1.1% stated that the residence was about to be rented. Analysis excluding the responses with outright rejections is provided in Tables A2 and A3.

3.5.1 Experiment #1 – No discrimination against Turkish nor Moroccans

The descriptive results show that 33.80% of renters responded to the pooled sample of applications from both minority groups while 35.92% to requests from Dutch members as

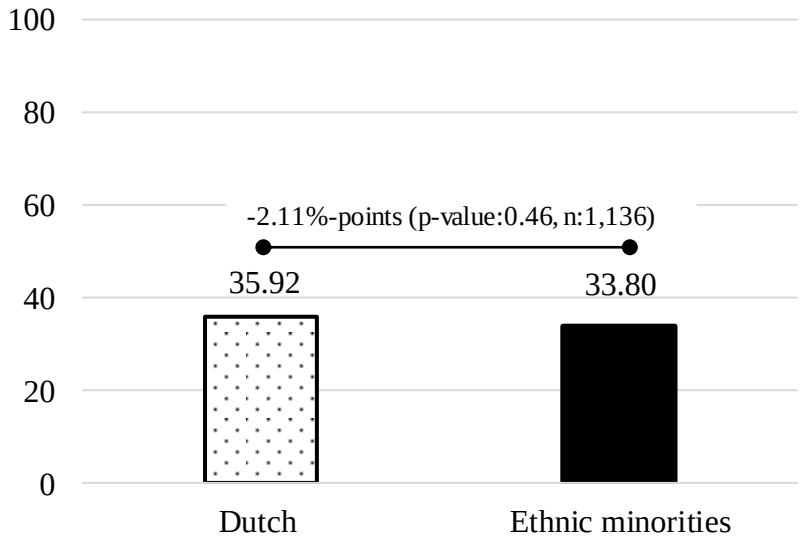
illustrated in Panel A of Figure 3.1. This difference represents a negative and statistically insignificant effect against ethnic minorities of about -2.11%-points (p-value=0.46, n=1,136).[38] Panel B of Figure 3.1 breaks down this general result to analyze the two-initial hypotheses of this paper. There is evidence of a small albeit statistically insignificant discrimination against Turkish applicants of -1.76%-points (p-value=0.67, n=568). The corresponding figure for Moroccan applicants is -2.46%-points (p-value=0.53, n=568) and also statistically insignificant.[39] These results confirm Hypothesis #1 and Hypothesis #2 of no housing discrimination against minorities on Craigslist.

The main results of Experiment #1 persist after controlling for confounding factors. In Table 3.6 (columns 1 to 4), the coefficients on ethnicity remain negative and statistically insignificant across specifications. Model (4) adjusted for the full set of confounding variables shows a negative and statistically insignificant coefficient of -2%-points (p-value=0.51, n=814). The same pattern appears in Table 3.7 which provides separate estimates for Turks and Moroccans. In column 4, Turkish candidates experience a negative and insignificant effect of -1.8%-points (p-value=0.68, n=414). In the case of Moroccans as well, the effect is negative and there is a statistically insignificant impact of -2.1%-points (p-value=0.60, n=400) as shown in column 8.

[38] These results imply that ethnic minorities have a negative effect in renter's response that requires a very large sample of observations to detect. The study must submit at least to 8,025 double-applications in total.
[39] The study would need to apply about 12,034 times to find Turkish discrimination and nearly 3,869 times to identify Moroccan discrimination.

Figure 3.1. Descriptive results in Experiment #1: Likelihood of manager's response to ethnic group

Panel A. Pooled sample



Panel B. Within ethnic minority *(Hypothesis #1 and Hypothesis #2)*

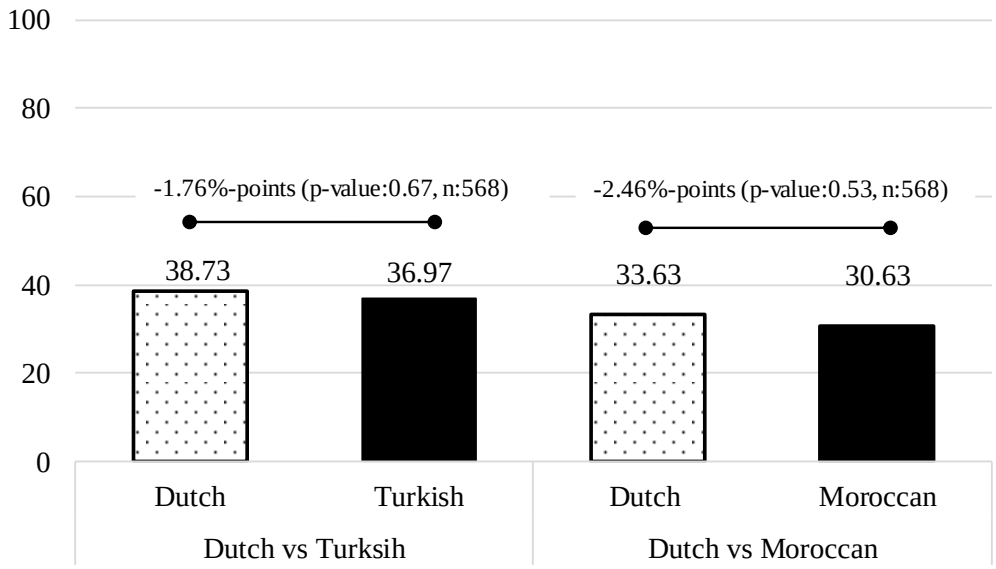


Table 3.6. Pooled regression results in Experiment #1

	Likelihood of renters' response (mean response to Dutch candidates: 36%)			
	(1)	(2)	(3)	(4)
Ethnic minority (=1, Dutch=0)	-0.021	-0.021	-0.021	-0.020
	(0.441)	(0.521)	(0.521)	(0.507)
Apartment (=1)			0.112***	-0.069
			(0.009)	(0.231)
Room (=1)			0.142+	-0.243*
			(0.077)	(0.014)
Price of location				-0.000***
				(0.001)
Square meters				-0.000
				(0.563)
Number of spaces/rooms				0.001
				(0.975)
Furnished				0.002
				(0.960)
Parking available				-0.067+
				(0.096)
Balcony/Terrace				-0.019
				(0.577)
Garden				-0.064
				(0.231)
Utilities included				0.202***
				(0.000)
Luxury space				0.137***
				(0.003)
Constant	0.405***	0.419***	0.313***	0.595***
	(0.000)	(0.000)	(0.000)	(0.000)
N	1,136	814	814	814

Note: Linear regression models. Models include a control variable for messages that appeared at least three times during fieldwork and robust standard errors. Symbols denote significance levels at *p<10%, **p<5%, ***p<1%

Table 3.7. Within ethnic group regression results in Experiment #1

	Likelihood of renters' response							
Mean response to Dutch candidates:	Turkish (39%)				Moroccan (33%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic minority (=1, Dutch=0)	-0.015	-0.017	-0.017	-0.018	-0.027	-0.023	-0.023	-0.021
	(0.707)	(0.704)	(0.705)	(0.679)	(0.484)	(0.602)	(0.609)	(0.602)
Apartment (=1)			0.014	-0.294***			0.233***	0.141***
			(0.836)	(0.000)			(0.000)	(0.008)
Room (=1)			-0.104	-0.546***			0.362***	0.028
			(0.428)	(0.000)			(0.000)	(0.812)
Price				-0.000				-0.000***
				(0.822)				(0.000)
Squate meters				-0.001				0.001
				(0.335)				(0.206)
Number of spaces/rooms				-0.037				0.030
				(0.179)				(0.157)
Furnished				-0.021				0.068
				(0.747)				(0.267)
Parking available				-0.120*				0.007
				(0.028)				(0.903)
Balcony/Terrace				0.008				-0.036
				(0.862)				(0.442)
Garden				-0.203*				0.019
				(0.012)				(0.779)
Utilities included				0.174***				0.213***
				(0.002)				(0.000)
Luxury space				0.095				0.216***
				(0.158)				(0.000)
Constant	0.437***	0.453***	0.446***	0.936***	0.372***	0.383***	0.149***	0.196
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.122)
N	568	414	414	414	568	400	400	400

Note: Linear regression models. Models include a control variable for messages that appeared at least three times during fieldwork and robust standard errors. Symbols denote significance levels at *p<10%, **p<5%, ***p<1%

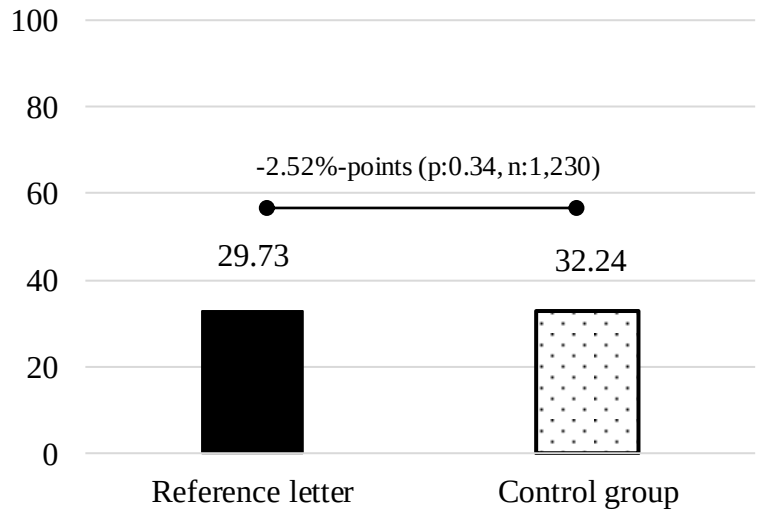
3.5.2 Experiment #2– No impact of reference letter to ameliorate access

The study finds that 29.73% of requests with a reference letter received a response as opposed to 32.24% for those in the control group. As shown in Panel A of Figure 3.2, this difference represents a negative and statistically insignificant pooled effect of the reference in housing applications of -2.52%-points (p-value=0.34, n=1,230).[40] The effect of the reference letter remains negative and insignificant within ethnic minorities as shown in Panel B of Figure 3.2. In the case of Turkish candidates the reference letter has an effect of -1.86%-points (p-value=0.61, n=622) and in the case of Moroccans of -3.2%-points (p-value=0.40, n=608).[41] Hence, the study rejects Hypothesis #3 and concludes that at least on Craigslist a reference letter does not impact responses from rental administrators.

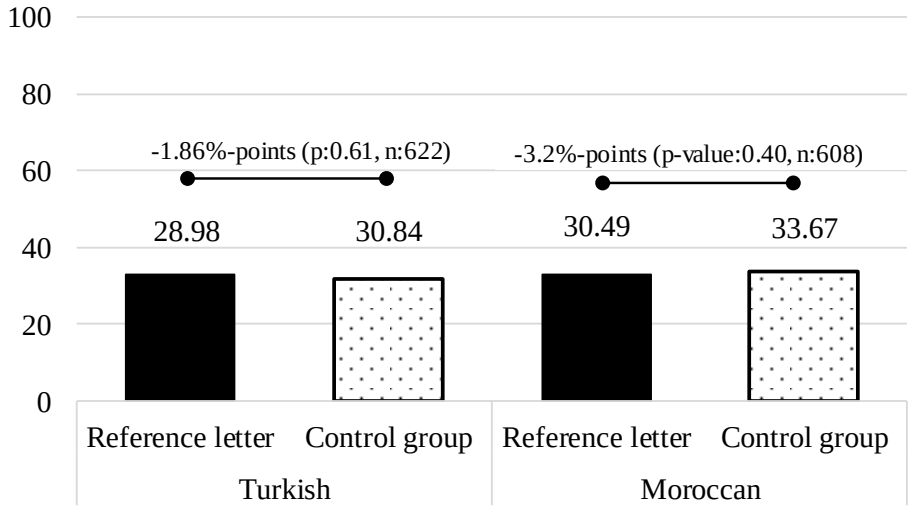
[40] The experiment would need to apply to 5,407 housing offers to identify the impact in the pooled sample.
[41] These coefficients indicate that a reference letter has a negative effect that requires a very large sample to identify. The sample needed for Turkish candidates reaches 9,619 and for Moroccan is 3,444.

Figure 3.2. Descriptive results in Experiment #2: Likelihood of manager's response to reference letter

Panel A. Pooled sample *(Hypothesis #3)*



Panel B. Within ethnic minorities



The impact of the reference letter remains negative and statistically insignificant in multivariate analysis. The pooled sample results of the reference letter remain negative and statistically insignificant across the four specifications in Table 3.8. In Model (4) the effect corresponds to an effect of -3.6%-points (p-value=0.34, n=636). This finding rejects Hypothesis #3 in the present study. The reference letter does not improve access of the pooled ethnic minorities in this market. Moreover, this effect persists within each ethnicity (Table 3.9). In the case of Turkish candidates, the inclusion of the reference has a negative and insignificant effect on the likelihood of getting a response of -4.2%-points (p-value=0.34, n=338) in model (4). In applications from Moroccans, the reference letter has a negative and

insignificant effect of -2.7%-points (p-value=0.60, n=298). In sum, the reference does not show different impacts within each ethnicity neither which leads to reject Hypothesis #3.

Table 3.8. Pooled regression results in Experiment #2

	Likelihood of renters' response (mean response to control: 32%)			
	(1)	(2)	(3)	(4)
Reference letter (=1)	-0.026	-0.033	-0.033	-0.036
	(0.337)	(0.358)	(0.355)	(0.281)
Apartment (=1)			0.281***	0.148***
			(0.000)	(0.005)
Room (=1)			0.446***	0.223*
			(0.000)	(0.018)
Price				-0.000***
				(0.000)
Squate meters				0.001***
				(0.000)
Number of spaces/rooms				0.018
				(0.443)
Furnished				-0.034
				(0.534)
Parking available				-0.104*
				(0.013)
Balcony/Terrace				-0.017
				(0.652)
Garden				0.027
				(0.643)
Utilities included				0.162***
				(0.000)
Luxury space				0.160*
				(0.025)
Constant	0.315***	0.339***	0.073	0.220
	(0.000)	(0.003)	(0.497)	(0.131)
N	1230	636	636	636

Note: Linear regression models. Models include ethnic profile fixed effects, a control variable for messages that appeared at least three times during fieldwork, and robust standard errors. Symbols denote significance levels at *p<10%, **p<5%, ***p<1%

Table 3.9. Within ethnic group regression results in Experiment #2

	Likelihood of renters' response							
	Turkish				Moroccan			
Mean response to control:	(31%)				(34%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reference letter (=1)	-0.024	-0.037	-0.038	-0.042	-0.028	-0.027	-0.025	-0.027
	(0.529)	(0.438)	(0.434)	(0.340)	(0.468)	(0.609)	(0.627)	(0.597)
Apartment (=1)			0.273***	0.116*			0.297***	0.238*
			(0.000)	(0.044)			(0.000)	(0.032)
Room (=1)			0.373***	0.134			0.584***	0.413*
			(0.000)	(0.184)			(0.000)	(0.023)
Price				-0.000***				-0.000+
				(0.000)				(0.072)
Squate meters				0.001***				0.001
				(0.000)				(0.302)
Number of space/rooms				0.023				0.004
				(0.436)				(0.923)
Furnished				-0.101				0.053
				(0.148)				(0.543)
Parking available				-0.181***				-0.016
				(0.002)				(0.801)
Balcony/Terrace				-0.032				0.009
				(0.547)				(0.871)
Garden				0.050				0.034
				(0.446)				(0.752)
Utilities included				0.183***				0.132*
				(0.000)				(0.032)
Luxury space				0.180+				0.147
				(0.078)				(0.188)
Constant	0.319***	0.204*	-0.055	0.302*	0.316***	0.346***	0.049	0.012
	(0.000)	(0.043)	(0.587)	(0.045)	(0.000)	(0.004)	(0.621)	(0.960)
N	622	338	338	338	608	298	298	298

Note: Linear regression models. Models include ethnic profile fixed effects, a control variable for messages that appeared at least three times during fieldwork, and robust standard errors. Symbols denote significance levels at *p<10%, **p<5%, ***p<1%

3.6 Discussion and Conclusion

This paper contributes to the literature on discrimination against vulnerable groups in modern societies. The study focused on ways and instruments to avert discrimination and/or to improve access of ethnic minorities to desired housing. It provided evidence on whether the use of a secondary -auxiliary- platform averts discrimination, and on whether a reference letter from a previous landlord improves the likelihood of accessing housing. The study employed an audit-system and it used the website Craigslist in Amsterdam to compare renters' responses to applications from ethnic minorities versus -*majority*- Dutch members. To manipulate the identity of applicants the study used male first and last names from Dutch versus minority identities with Turkish and Moroccan names. Turkish and Moroccan identities are chosen considering their historical and current presence in Dutch society. This design was used to evaluate three hypotheses: (*H1*) Turkish and Dutch applicants are equally

likely to receive a response from renters, (*H2*) Moroccan and Dutch applicants are equally likely to receive a response from renters, and (*H3*) a positive reference letter improves the likelihood of receiving a reply from rental managers as compared to a control group without reference.

Experiment #1 confirms the hypothesis that there is no discrimination against Turkish and Moroccans candidates on housing offered through Craigslist in Amsterdam. The coefficients on ethnic identity are negative but statistically insignificant in the case of both groups *(accepting Hypothesis #1 and #2).* In contrast, Experiment #2 refutes the hypothesis that a *(positive)* reference letter from a previous *(Dutch)* landowner enhances the likelihood of receiving a response from managers *(rejecting Hypothesis #3)*.

These results are in marked contrast to the existing literature and suggest that searching for housing on secondary platforms is likely to yield greater success for minorities. Craigslist encompasses two essential features of settings expected to lessen discrimination: *(i)* lower market prominence with only about 100 new rental offers per month, and *(ii)* tailor-designed by and for minorities looking to relocate. The first point ensures that renters are more encouraged to accept minority candidates given the lower market visibility and competition for each advertised offer. The second point reflects that Craigslist targets local expats and foreigners instead of the average -*majority*- Dutch agent. While Craigslist is a site catering to Western, English speaking minorities it is likely that given the minority status of such a group it is more welcoming of other minorities, albeit, non-Western. Second, consistent with the results from the first experiment, the positive reference letter does not impact the response from rental managers contrary to evidence from labor markets in South Africa and Germany (Abel et al., 2017; Kaas and Manger, 2012). In fact, the evidence is suggestive of a negative effect of the reference letter although it is statistically insignificant.

The study has one main limitation. While it does evaluate the initial phase of housing applications where discrimination is expected to have the largest impact (Auspurg et al., 2017; Riach and Rich, 2002), it does not evaluate every stage of potential housing discrimination. It is possible that despite the encouraging initial responses, the outcome for minorities is not as sanguine after a meeting or a viewing.

This study motivates future work in three areas. First, future work could examine whether minority-led platforms (by minorities and for minorities) could achieve similar non-

discriminatory results in other European markets. Second, despite the lack of effectiveness of a reference letter in this case, future work could implement the reference letter in countries (and websites) with historical evidence of housing discrimination such as Germany, France and Sweden (Auspurg et al., 2017; Bonnet et al., 2016; Ahmed and Hammarstedt, 2008). Third, another interesting approach for researchers in the field may be to examine non-monetary incentives to stimulate access to goods and services. A prominent line of economic research employs moral incentives to avert the negative behavior of non-repayment of credit-card debts by simply reminding customers about the "injustice of non-repaying their debts" (Bursztyn et al., 2019). Another study uses a normative statement to remind students not to discriminate and a descriptive norm on peer behavior to avert gender discrimination in student evaluations of teaching (Boring and Philippe, 2017). Second,

This study is related to work on ingroup favoritism in labor, housing and other market-based interactions which examines instruments which may be used to promote prosocial behavior towards vulnerable groups (Weichselbaumer, 2016; Bartoš et al., 2017; List, 2004; Chuah et al., 2014; Cadsby et al., 2016; Chen and Li, 2009; Becker, 1957). The present work suggested instruments to stimulate access of vulnerable groups to essential markets. Secondary websites may help ethnic minorities access housing markets, however, given the limited housing offers on such platforms their use may be limited. To conclude, it is imperative to continue research on instruments which may help achieve greater equality for marginalized populations and to enhance living standards in modern societies.

3.7 Annex

3.7.1 Tables

Table A3.1. Messages employed in study

Panel A. Messages used in Experiment #1

Type 1	Type 2
(Example using Turkish identity) "Geachte heer, Mijn naam is Ahmed Idrissi en ik ben geïnteresseerd in het huren van het door u aangeboden appartement. Graag ontvang ik meer informatie en maak ik een afspraak voor een bezichtig. In afwachting van uw bericht. Dank u, Ahmed Idrissi" - *(Translation)* Dear Landlord, My name is Ahmed Idrissi and I am interested in the apartment you are offering. I would like to have more information and to fix a showing. I hope that you will write back to me. Awaiting your message. Thank you, Ahmed Idrissi	**(Example using Dutch identity)** "Geachte heer, Ik ben Jan de Jong. Graag ontvang ik informatie over uw appartement. Tevens zou ik graag een afspraak maken om het appartement te komen bekijken. U kunt mij bereiken op dit email adres. Met vriendelijke groeten, Jan de Jong" - *(Translation)* Dear Landlord, I am Jan de Jong. I would like to have more information about your advertised apartment. We could schedule a viewing. I hope that you will reply to my message. You can reach me at this email address. Kind regards, Jan de Jong

Panel B. Messages in Experiment #2 with reference letter

Type 1	Type 2
(Example using Turkish identity) "Geachte heer, Mijn naam is Ahmed Idrissi en ik ben geïnteresseerd in het huren van het door u aangeboden appartement. Graag ontvang ik meer informatie en maak ik een afspraak voor een bezichtig. **Ik aanhechten een referentie van mijn vorige verhuurder, Jan de Jong.** In afwachting van uw bericht. Dank u, Ahmed Idrissi" - *(Translation)*Dear Landlord, My name is Jan de Jong and I am interested in the apartment you are offering. I would like to have more information and to fix a showing. I hope that you will write back to me. **I attach a reference from my previous landlord, Jan de Jong.** Awaiting your message, Thank you, Ahmed Idrissi	**(Example using Moroccan identity)** "Geachte heer, Ik ben Said Loukili. Graag ontvang ik informatie over uw appartement. Tevens zou ik graag een afspraak maken om het appartement te komen bekijken. **Ik aanhechten een referentie van mijn vorige verhuurder, Jan de Jong.** U kunt mij bereiken op dit email adres. Met vriendelijke groeten, Said Loukili - *(Translation)* Dear Landlord, I'm Said Loukili. I would like to receive information about your apartment / house. I would also like to make an appointment to come and see the apartment. **I attach a reference from my previous landlord, Jan de Jong.** You can reach me at this email address. With best regards, Said Loukili

Panel C. Example of reference letter in Experiment #3

(Example using with Peter Jansen)

2019

Geachte heer,

Via deze brief beveel ik u de mevrouw Said Idrissi aan als huurder van uw woonruimte. Ik was zeer tevreden met hem als bewoner van mijn woonruimte: schoon, stil en een model huurder. Ik hoop dat u wilt overwegen uw woonruimte aan hem te verhuren.

Met vriendelijke groeten,

Peter Jansen
peterjansen756@gmail.com

- - - - -

- - - - - - - - - - - - - - - - -

(Translation) Dear Sir, Through this letter I recommend Mrs. Said Idrissi as a tenant of your living space. I was very satisfied with him as a resident of my living space: clean, quiet and a model tenant. I hope you want to consider renting out your living space to him.
With best regards,

(signature)

Peter Jansen
(email)

Table A3.2. Comparison in likelihood of total vs positive responses in Experiment #1

		Total		Positive		(1) vs. (3) P-value Δ	(2) vs. (4) P-value Δ
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Full sample	Ethnic minority (=1)	-0.021	-0.020	-0.026	-0.025	0.57	0.56
		(0.028)	(0.030)	(0.028)	(0.031)		
	Includes covariates	No	Yes	No	Yes		
	N	1136	814	1089	773		
Panel B. Turkish	Turkish (=1)	-0.018	-0.018	-0.021	-0.022	0.76	0.68
		(0.041)	(0.043)	(0.041)	(0.044)		
	Includes covariates	No	Yes	No	Yes		
	N	568	414	549	396		
Panel C. Moroccan	Moroccan (=1)	-0.025	-0.021	-0.031	-0.024	0.62	0.84
		(0.039)	(0.041)	(0.039)	(0.041)		
	Includes covariates	No	Yes	No	Yes		
	N	568	400	540	377		

Note: Seemingly unrelated linear probability models. Column 1 comprises the pooled sample of results. Column 2 evaluates the pooled sample including control variables. Column 3 only uses positive responses, and Column 4 examines positive responses with control variables. The last two Columns (5 and 6) tests for differences in coefficients. Symbols denote the level of statistical significance at *p<10%, **p<5%, and ***p<1%.

Table A3.3 Comparison in likelihood of total vs positive responses in Experiment #2

		Total		Positive		(1) vs. (3) P-value Δ	(2) vs. (4) P-value Δ
		(1)	(2)	(3)	(4)		
Panel A. Full sample	Reference (=1)	-0.025	-0.036	-0.026	-0.033	0.95	0.77
		(0.026)	(0.033)	(0.026)	(0.033)		
	Includes covariates	No	Yes	No	Yes		
	N	1,230	636	1,196	621		
Panel B. Turkish	Reference (=1)	-0.019	-0.042	-0.016	-0.041	0.79	0.91
		(0.037)	(0.043)	(0.037)	(0.043)		
	Includes covariates	No	Yes	No	Yes		
	N	622	338	609	332		
Panel C. Moroccan	Reference (=1)	-0.032	-0.026	-0.035	-0.24	0.75	0.85
		(0.038)	(0.051)	(0.038)	(0.051)		
	Includes covariates	No	Yes	No	Yes		
	N	698	298	587	289		

Note: Seemingly unrelated linear probability models. Column 1 comprises the pooled sample of results. Column 2 evaluates the pooled sample including control variables. Column 3 only uses positive responses, and Column 4 examines positive responses with control variables. The last two Columns (5 and 6) tests for differences in coefficients. Symbols denote the level of statistical significance at *p<10%, **p<5%, and ***p<1%.

Part II: Promoting wellbeing in People Living with HIV in Burkina Faso

Chapter 4: Protocol for a Randomized Controlled Trial Evaluating Mobile Text Messaging to Promote Retention and Adherence to Antiretroviral Therapy for People Living with HIV in Burkina Faso [42]

4.1 Introduction

4.1.1 Background

Several factors inhibit retention in care and adherence to antiretroviral therapy (ART) among people living with human immunodeficiency virus (PLHIV). Typical reasons include individual and social obstacles, such as lack of information on treatment procedures, social stigma, discrimination, and competing priorities that prevent patients from considering antiretroviral treatment as a worthwhile investment of time, energy, or resources (Lester et al., 2006; Mills et al., 2006). In developing countries, economic and contextual barriers may further amplify these challenges because patients are resource-constrained and health facilities may not be easily accessible leading to substantial costs in terms of wage losses and travel expenses (Scanlon and Vreeman, 2013).

Discontinuation of ART regimens is particularly prevalent in Sub-Saharan Africa. Systematic reviews have found that approximately 22.5% of patients discontinue ART within 10 months and 56% are lost to follow-up or death during the first 2 years of the treatment (Fox and Rosen, 2010; Rosen et al., 2007). HIV patients are perhaps not fully aware of the long-term consequences of dropping out from ART treatment. Lack of retention in care and adherence to antiretroviral therapy increases HIV viral loads and the probability of transmission, reduces the number of T-lymphocytes cells in the blood (CD4 count)-an indicator of how well the immune system is working and a strong predictor of HIV progression-leads to deterioration in the quality of life and can be responsible for creating virus strains that are resistant to current HIV medication (Eaton et al., 2012; Cohen et al., 2011; Wang et al., 2009; Gill et al., 2005). Favorable health outcomes for PLHIV require lifelong compliance with ART programs. At the macro-level the negative side effects of poor compliance with ART can deteriorate the efficiency and efficacy of public health care systems by increasing the burden of the disease and the potential costs of care in the future (Granich et al., 2012; Institute of Medicine, 2011; Loubiere et al., 2010).

[42] Wagner N, Ouedraogo D, Artavia-Mora L, Bedi A and Thiombiano BA. 2016. Protocol for a randomized controlled trial evaluating mobile text messaging to promote retention and adherence to antiretroviral therapy for people living with HIV in Burkina Faso. *JMIR Research Protocols*, 5(3), p.e170.

The World Health Organization (WHO) promotes the use of innovative mobile technologies to overcome barriers that undermine access to health care and the quality of care delivery in resource-poor countries (WHO, 2011). mHealth is an important element of this approach as it can help alleviate some of the existing obstacles in the delivery of quality care (Jongbloed et al., 2015; Free et al., 2013; Gurol-Urganci et al., 2013). Mobile technologies may help patients undergoing ART to maintain the treatment routine as they provide instant communication unrestricted to location (Chaiyachati et al., 2014; Coomes et al., 2012). Specifically, the use of text messaging (short message service, SMS) reminders may support PLHIV to take their pills every day, schedule refills of their prescriptions, and assist them through common side-effects. Such reminders are considered to be a low-cost, low-barrier intervention (Lester et al., 2008). Especially in resource-constrained developing countries where some patients live far from the health centers and a system of regular home visits by health care providers is not in place (Scanlon and Vreeman, 2013; Lester et al., 2006; Mills et al., 2006), regular text messages may help patients to remain in care and adhere to their antiretroviral regimens (Govindasamy et al., 2014; Finitsis et al., 2014; Guy et al., 2012). These advantages have promoted the rapid expansion of mHealth projects that aim to improve health outcomes in patients living with HIV and other diseases across the developing world (Global Digital Health Network, 2016).

Despite the possibility of cost-efficient, easy outreach through SMS, recent studies have shown contrasting evidence. While several studies have demonstrated that mobile text message reminders are effective in enhancing adherence to ART programs others do not find any effects (Shet et al., 2014; Mbuagbaw et al., 2012; da Costa et al., 2012; Horvath et al., 2012; Pop-Eleches et al., 2011; Kunutsor et al., 2010; Lester et al., 2010). Research in Kenya demonstrates that 53% of the participants who received weekly SMS reminders achieved adherence of at least 90% during the 12 months of the study (Pop-Eleches et al., 2011). In the control group, only 40% of the participants achieved similar adherence levels. Evidence from a small study in Brazil also suggests that adherence to ART increased due to text message reminders at least during the 4-month study period (da Costa et al., 2012). In contrast, a 6-month study in Yaoundé, Cameroon, found that standardized motivational mobile text messages did not increase adherence (Mbuagbaw et al., 2012) Likewise, research conducted in some states in India did not find a statistically significant impact of mobile phone reminders on time to virological failure or ART adherence at the end of a 2-year study period (Shet et al., 2014). Moreover, we are not aware of any long-term study that has been

conducted in a Francophone African country where perceptions, preferences, and health systems are considerably different as compared with Anglophone African countries (Meessen et al., 2014).

4.1.2 Justification

The background provided above motivates continued research on the impact of mHealth interventions in developing countries. The majority of the existing studies are limited to geographically circumscribed areas such as capital centers, and are based on small sample sizes or on short-time horizons (Shet et al., 2014; Mbuagbaw et al., 2012; da Costa et al., 2012; Horvath et al., 2012; Pop-Eleches et al., 2011; Kunutsor et al., 2010; Lester et al., 2010). For instance, the Brazilian study carried out a 4-month study with as few as 21 Brazilian female PLHIV (da Costa et al., 2012). Similarly, the 2 studies in Kenya followed less than 550 participants and included no more than 3 health facilities for a period of 12 months (Horvath et al., 2013; Pop-Eleches et al., 2011; Kunutsor et al., 2010; Lester et al., 2010). The intervention in Cameroon included 1 hospital and 198 participants for a period of 6 months (Mbuagbaw et al., 2012). While the intervention in India had the longest time horizon of 2 years and a relatively large sample of 631 participants it was based on data from 3 health centers located in only 2 states (Shet et al., 2014). Thus, research based on a sample across an entire nation with a longer-term horizon is warranted. The present study is timely as it is based on a wide-spread intervention across Burkina Faso with a 2-year horizon. These design features improve the study's external validity. Furthermore, the study will enhance our understanding of the extent to which mHealth interventions promote healthy behaviors and support psychosocial wellbeing. Therefore, the study will contribute to an improved understanding of when, why, and for whom mHealth interventions work (Madhvani et al., 2015; Walshe, 2007)

4.1.3 Objectives

The main objective of this trial is to determine the impact of four different packages of SMS message reminders to promote HIV patients' retention and adherence to ART as well as their health outcomes in a large-scale randomized controlled trial (RCT) in a Francophone country, namely Burkina Faso. We hypothesize that patients who receive text messages are encouraged to take their pills and reminded of the importance of ART for their health so that they remain in care longer than those who do not receive text messages. We also anticipate that enhanced retention and adherence will lead to positive health outcomes. Our objectives

are (1) to inform best practices for enrolling PLHIV into ART programs and supporting them throughout care, (2) to provide insights on the key obstacles confronting patient retention and adherence to ART, (3) to advise on the most effective application of mobile technology for health interventions, including the short-, medium-, and long-term benefits that can be anticipated, and (4) to encourage long-term patient success with ART by promoting feasible and efficient strategies that may be adopted in resource-constrained settings. Concerning the effectiveness of SMS text messaging reminders, our intervention includes two complementary objectives. First, the study evaluates whether patients may experience fatigue from the SMS text messaging reminders. By carrying out 3 surveys at 6, 12, and 24 months after the launch of the intervention, we aim at determining the optimal period for such a type of mHealth intervention to be efficient. Second, the trial will evaluate the effects of message type (text vs ASCII image) and frequency (weekly versus semi- weekly), i.e., the differences in the four treatment arms, on patient outcomes.

4.2 Methods

4.2.1 Trial Setting

The study is implemented in the Francophone African country of Burkina Faso. The Joint United Nations Programme on HIV/AIDS (UNAIDS) reports that this country suffered a HIV prevalence of 0.9 in 2014 and the total number of PLHIV is estimated to be around 110,000 of which 94,000 are adults and 18,000 are children younger than 15 years (UNAIDS, 2015). The feminization of HIV is also observed in Burkina Faso since 59 percent of the adult PLHIV are women according to the National Council for the Fight against AIDS and Sexually Transmitted Infections (CNLS-IST) -the national committee in charge of the surveillance and fight against HIV/AIDS. The country provides free anti-retroviral treatment and its provision of HIV care follows WHO's international guidelines and strategies (Burkina Faso Ministry of Health, 2016; WHO, 2013). Burkina Faso's National Plan to combat HIV includes decentralization policies and multisector participation with the aim to increase the number of PLHIV who enroll at health centers that provide ART therapy and care (Burkina F, 2016). Despite these efforts the fight against HIV and for ART adherence and retention remain a concern. Although the actual number of PLHIV enrolled in official files increased from 70,230 in 2013 to 76,342 in 2014, the actual number of patients undergoing active antiretroviral treatment was only 60% (42,145) in 2013 and 61% (46,623) in 2014. Similarly, despite the fact that the number of patients that becomes lost to follow up decreased almost

by half from 2013 to 2014 (833 to 443), they are all attributed to fatalities (UNAIDS, 2015). These figures provide a first indication that adherence and retention to ART cannot be assured in Burkina Faso. And indeed, a survey carried out by UNAIDS in 2012 among 2,800 Burkinabe PLHIV revealed that adherence is perceived as a challenge due to negative side effects from treatment, the time and resources needed to regularly refill the stock of drugs and stigma (UNAIDS, 2012). The study we aim to implement allows us to assess whether and how retention and adherence may be improved.

4.2.2 Study Design

The study design rests on a five-arm, randomized, controlled trial with four treatment arms and one control group. With the support of the Burkinabe Ministry of Health and local health centers, HIV patients have been screened and recruited at 80 health care facilities that provide antiretroviral therapy across the 13 regions of Burkina Faso. Using a 7:7:7:7:10 allocation ratio, patients are randomized to one of the five-arms with the control group being slightly oversampled as we introduce two types of control groups; a control group of patients drawn from health centers which do not receive any of the interventions and a spillover-prone control group drawn from health centers which are visited by both treated and untreated patients. A member of the international survey team carried out the randomization in Stata. The team member did not have any interactions with patients. Once patients were randomly allocated, the intervention was launched, and the SMS messages are sent. Patients in each of the five groups receive the public standard care but only four groups receive an SMS reminder that varies by the type of message (text versus American standard code for information interchange (ASCII) image) and its frequency (weekly vs semiweekly). The study collaborates with the health care personnel attending to the patients as well as the associations and self-help groups that provide psychosocial support to PLHIV. The former provides medical information and the latter carry out the survey interviews. The health personnel and the enumerators are not informed about the outcome of the random allocation of the participants. Because they do not know which patient is in which group, we do not expect bias. The study will keep track of the patients 6, 12, and 24 months after the intervention is launched.

4.2.3 Facilities Selection

The investigation will be implemented in health facilities across the 13 regions of Burkina Faso. The health centers are selected from a total of 100 registered facilities that are recorded

in the documents of the CNLS-IST as providers of antiretroviral therapy across the 13 health care regions of Burkina Faso (UNAIDS, 2015). To be eligible to participate in this study, health centers need to fulfill two conditions. First, they need to exist and they need to be functioning. While this sounds like a straightforward requirement it is possible that despite being registered, facilities might have never become operational or have shut down. Second, the health centers need to be willing to collaborate. Specifically, we enrolled a total of 80 health centers with an intended average of 40 patients being recruited per center depending on the eligible population of PLHIV who frequent the health center. To avoid under or overrepresentation of a particular health center, eligible facilities must meet a minimum number of 15 HIV patients and are allowed to enroll a maximum number of 150.

4.2.4 Participants

4.2.4.1 Eligibility and Informed Consent

The participants need to fulfill five conditions. As an initial step for eligibility, participants must be enrolled in an ART program in 1 of the health care centers collaborating in the study. Second, each patient must provide written informed consent confirming their participation in the study. The informed consent includes signed permission to consult their medical records over the duration of the study. Third, participants must be older than 15 years because the focus of the study is on adult PLHIV. Fourth, participants who have been under ART for less than 4 years are preferred, although, experienced patients are also included. We aim at assessing differential effects for individuals initiating ART versus individuals who have been under treatment for a long time. Because the existing literature suggests that drop-out is highest among those initiating treatment, we place more emphasis on individuals who have recently started ART. Therefore, we aim that at least two-thirds of the sample consists of patients that have been under ART for less than 4 years. We trained and informed the health centers and enumerators about this sampling feature. We requested to be informed (by phone) in the case that patients were included, that have been longer on ART. Thus, from the central level we closely monitored the oversampling of patients, who have been under ART for less than 4 years. Fifth, patients must have reliable access to a mobile phone including a stable network connection. Access to a mobile phone is not a major bottleneck: according to the World Factbook there were 12.5 million subscribers in Burkina Faso in 2014, 68 of 100 inhabitants have a mobile phone, and there are 3 major mobile networks that reach out to the entire country (Central IA, 2016; Africa & Middle East Telecom-Week, 2016).

4.2.4.2 Recruitment

The recruitment period lasted for 4 months because patients return to the health centers to refill their stock of antiretroviral medication at different intervals. This time period has been sufficient to reach the target sample size of 3800 individuals.

4.2.5 Randomization

Randomization was undertaken after the collection of the baseline data. As a first step we randomly identified 8 pure control health centers using the random number generator in Excel. These pure control centers allow us to assess whether there are within health center spillovers from those who receive reminders and those who do not. Across the 72 remaining health centers, we randomly assign individuals to 1 of 5 groups (i.e., 1 of 4 treatment packages or the control group of patients who do not receive any text message). Due to the potentially large heterogeneity of patient socioeconomic profiles and variations across health centers, the study applies covariate balancing across multiple baseline covariates to minimize imbalance between treatment groups. To increase statistical power and the precision of the results, participants are ordered along key characteristics and randomized within these ordered blocks. The following characteristics are included in the ordering: gender, age and weight of the participant, duration under antiretroviral treatment, health center identifiers, the reported distance to the health center, CD4 counts, and subjective health rating.

Following this 2-step procedure, participants are randomly assigned to 1 of 5 groups (4 of which receive various packages of SMS reminders and a control group, which does not). Thus, there are 2 control groups. One control group consists of individuals who do not receive messages but are affiliated to health centers where others receive messages and a pure control group, which consists of individuals who do not receive messages and are affiliated to the 8 pure control health centers where no one receives a message. The introduction of 2 types of control groups is motivated by a desire to address potential spillovers/contamination between treated persons (those who receive a SMS text message) and patients in the control group (those who don’t receive the SMS text message) but frequent the same health center. We know that medical interventions such as deworming and cancer screenings have spillovers, similarly voter awareness and cash transfer programs have spillovers to untreated but close populations (Pruitt et al., 2014; Gine et al., 2012; Angelucci and De Giorgi, 2009; Miguel et al., 2004). Having 2 control groups allows us to address whether both treated and untreated patients in the treatment health centers are covered by the intervention (Gine et al.,

2012). In the analysis, we will introduce a dichotomous variable for control patients from mixed-treatment control sites to assess whether their outcomes differ compared with patients from pure control sites who are unlikely to experience spillovers.

4.2.6 Intervention

The intervention comprises 4 treatment groups and 1 control group to investigate the impact of different types of SMS text message reminders sent out at different frequencies on patient outcomes. For groups 1 to 4, messages will be sent on a weekly basis. Patients of each of the 4 treatment groups will receive at least 1 text or image message per week. Image messages are used to ensure that the intervention reaches out to people with limited literacy skills and they will be sent as ASCII pictures. Because we do not expect the participants to have smartphones, we use ASCII images that can be displayed on basic mobile phones. The content and frequency of the messages will vary for each of the 4 treatment groups. While the first treatment group will receive only 1 text message per week (low frequency), the second treatment group receives a total of 2 text messages per week (high frequency). Meanwhile, the third group will get 1 text and 1 image message per week and, the fourth group will receive 2 image messages per week. All messages will be sent at 8 am and patients will not be prompted to respond. The impact of the treatments will be assessed in follow-up surveys at 6, 12, and 24 months after the start of the intervention. A detailed identification of the 5 groups, type of interventions, and timings are presented in Table 4.1.

Table 4.1. Identification of the five intervention groups

Treatment arms		Intervention (per week)	Timing
Control group			
	Eight pure control health centers	No text nor image messages	
	Individuals from the remaining 72 centers randomly allocated to the control group.		
Treatment 1: text-only, low frequency		One text message	Every Monday at 8 am
Treatment 2: text-only, high frequency		Two text messages	Every Monday and Friday at 8 am
Treatment 3: text and image		One text message; one image message	Text: every Monday at 8 am; image: every Friday at 8 am
Treatment 4: image only		One message	Every Monday at 8 am

Furthermore, text messages are sent in French or 1 of the local languages (Moore, Jula, Gulmancema, Fulfulde, or Dagara). The baseline survey determined the main language of

each participant to be used in the text messages. The content of the text messages will vary to ensure that the participants remain curious about the messages over the study period. We are concerned that individuals stop looking at an identical message after some time as they already know the content of the standardized message. Furthermore, we opted for a mix of messages instead of standardized messages due to concerns about exclusively "negative" framing (Textbox 4.1). However, we acknowledge that this approach does not allow us to systematically compare the implications of standardized messages across different treatment arms. Figure 4.1 shows two examples of the ASCII images that are sent.

Textbox 4.1. Example of text messages that are sent

Hello. Do not forget to take your pills.

Your health is important. Take your pills.

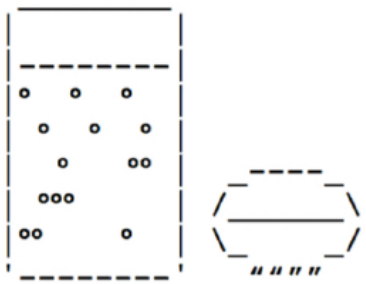
You are very important. Do not play with your health. Take your pills.

Are you very busy? This is why I remind you to take your pills.

Don't forget that you are strong, unique, funny and blessed. You are needed. This is why I would like to remind you to regularly take your pills.

Figure 4.1. Examples of ASCII images used in the intervention

Example #1: Glass of water and pill



Example #2: Pill and thumb-up

Patients in the control group do not receive any text messages but only the standard care. Along with periodic clinical check-ups and treatment counseling, this includes routine monitoring of patient CD4 cells as measure of disease progression. Adherence support and/or additional treatment counseling may also be provided at the community level. This will be assessed during follow-up surveys.

4.2.7 Sample Size

Power calculations need to account for the stratified nature of the sample. There are 80 treatment centers. We randomly preserve 8 pure control centers to assess spillovers. Within each of the remaining 72 centers, treated and untreated participants are sampled. Thus, for the majority of the centers treatment allocation is not at the level of the center but at the level of the individual. The sample size is calculated with the clustersampsi command of Stata, Version 13 assuming a power of 80% and a significance level of 5%. Furthermore, we impose an intraclass correlation coefficient of 0.015 to quantify the degree to which patients within the same health center are related (Kranzer et al., 2012). We allow for cluster sizes to vary because we know that the size of the eligible population varies across clusters. Thus, we impose a coefficient of variation of cluster sizes of 0.5 (ratio of the standard deviation of cluster sizes to the mean cluster size).

Because we collect baseline information we can control for observable characteristics and their correlation with the outcome. We impose a correlation of 0.85. Lastly, because we expect that 35% of the patients will forget to take antiretroviral (ARV) medication from time to time, a target sample size of 3800 PLHIV is needed to obtain a "minimum detectable effect" of an increase in adherence from 65% to 70%. These adherence figures are based on a reported 67% adherence among female PLHIV in Burkina and the government target to bring adherence up to at least 80% (Burkina F, 2014; UNAIDS, 2012). As we are conservative about the possibilities of mHealth to raise adherence by 15% the sample was set up in such a way to also identify small gains.

4.2.8 Duration and Follow-Up Surveys

The intervention will run for a period of 2 years. We will conduct four outcome assessments during baseline as well as 6, 12, and 24 months into the intervention.

4.2.9 Ethical Concerns

The study faces two ethical concerns. First, we need to access patient medical records to monitor health indicators. We need to gain participant consent and have to ensure that confidentiality is protected during the period of the study. Only the health personnel and the local enumerators will know the individuals. In the dataset, all patient information will be anonymous and only linked to the participant's identification number that is given in the context of the study. No identity information will be disclosed. Second, as evident in the examples provided above, in order to prevent negative social stigma, the reminders do not

disclose a participant's seropositive status. Participants will be informed of the possibility that texts could be read by other individuals who have access to their phone and they will have to consent to this risk in order to participate in the study. All relevant ethical clearance from the national ethics committee has been obtained (N° 2014-12-144).

4.2.10 Implementation of the Text Message Reminder System

The local telecommunication company EVOLE provides a computerized platform for sending the SMS text message reminders, re-launching text messages if delivery fails, and monitoring message receipt. The messages are sent in bulk and a record is kept of all communication. The computerized platform allows us to verify whether the text and image messages have been sent and that the intervention has been properly implemented. We will also use the monitoring -data to analyze the quality of the SMS text message communication. Details about the study design, facility selection, and patient eligibility including the random allocation mechanism, the intervention and the outcome variables are graphically presented in Figure 4.2.

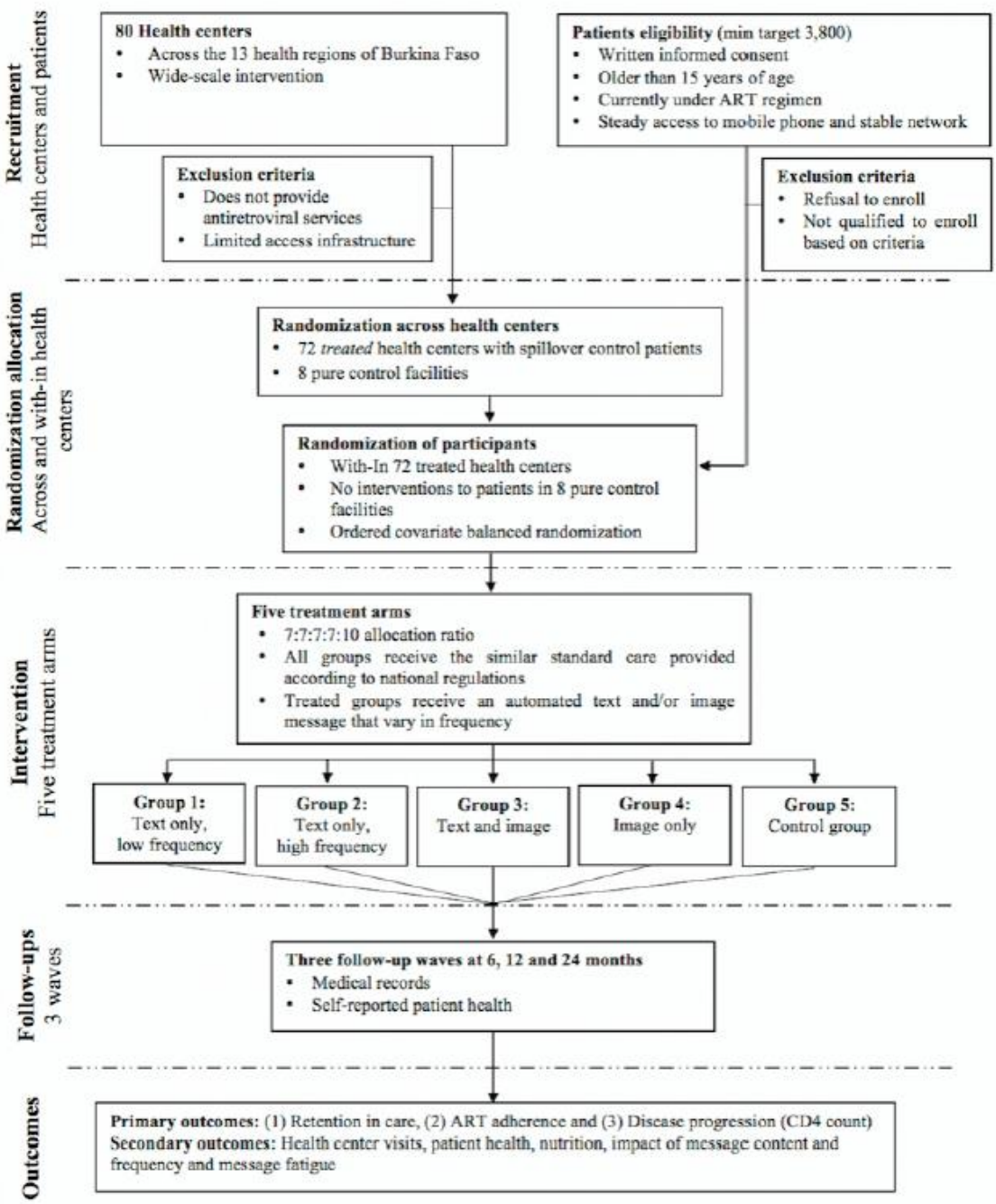
4.2.11 Study Measures

4.2.11.1 Primary Outcomes

The study focuses on the measurement of three primary outcomes: retention in care, adherence to ART, and disease progression. Retention is measured by a dichotomous variable-whether a patient remains on ART 6, 12, and 24 months after the intervention has started as compared with baseline (Rasschaert et al., 2012; Achieng et al., 2012). Retention could have been conceptualized also as the incidence or number of missed visits during a given reference period (Mugavero et al., 2012). However, remaining in care is in itself an important challenge in Burkina Faso. This is in stark contrast to the situation of PLHIV in developed countries. Therefore, we decided to choose remaining on an ART regimen as our retention measure. Adherence or rather lack of adherence to ART is evaluated as the number of doses missed in the past 7 days, which is a self-reported measure (Chkhartishvili et al., 2014; Oyugi et al., 2004; Chesney, 2006). It may have been better to use actual pharmacy dispensing data to calculate a medication possession ratio. However, our field experience suggested that it was not always feasible to obtain detailed and correct information about pharmacy dispensing. Contrary to developed countries where standards across hospitals may be similar, in Burkina Faso we observed considerable differences in the standards and routines across health centers both between regions and within them. Because we assess a

wide-scale intervention our study also includes rural and remote areas, which are perhaps even less capable of providing this information. Therefore, we opted for inclusiveness knowing that this implies adjustments in the type of comparable indicators that can be collected and analyzed across health centers (Chesney, 2006). Lastly, disease progression is analyzed using patients' CD4 counts.

Figure 4.2 Study flow



Information on these primary outcomes will be collected using individual questionnaires and patient medical records. Table 4.2 presents details about the indicators with information about the date of data collection.

4.2.11.2 Secondary Outcomes

Information on 5 groups of secondary outcomes will be collected. First, we will obtain information about the incidence and number of missed health center visits. Second, because we are also interested in more general physical and psychosocial health aspects, we will gather information about the body mass index, the incidence of coinfections, mortality, as well as subjective and mental health ratings. In addition, we will collect information on the patients' levels of risk preference and their subjective discount factors. Third, we will also measure nutritional outcomes. Fourth, to assess design effects and the duration of effectiveness of the mHealth intervention, we will gather data on the message type and frequency. The sensitivity to message type (text vs ASCII image) and the frequency (weekly vs semiweekly) of receiving the messages will be measured by comparing the effects of the 4 different interventions on the primary outcomes (Shet et al., 2014; Mbuagbaw et al., 2012; da Costa et al., 2012; Horvath et al., 2012; Pop-Eleches et al., 2011; Kunutsor et al., 2010; Lester et al., 2010). Finally, message fatigue will be measured using patients' perceptions and possible changes in the impact on the primary outcome indicators across follow-up survey rounds. The secondary outcomes will be measured 6, 12, and 24 months after commencement of the intervention.

Table 4.2. Primary and secondary outcome variables

Classification		Indicators/measures	Baseline	Follow-ups (months after)		
				6	12	24
Primary outcomes						
	Retention in care	Dichotomous measure whether a patient remains on ART	X	X	X	X
	ART adherence	Number of doses missed in the past 7 days	X	X	X	X
	Disease progression	CD4 count	X	X	X	X
Secondary outcomes						
	Health center visits	Incidence and number of missed visits		X	X	X
	Patient health	Biomarker: body mass index	X	X	X	X
		Incidence of coinfection	X	X	X	X
		Mortality	X	X	X	X
		Subjective health rating	X	X	X	X
		Measures of mental health		X		X
	Preferences	Risk preference		X		X
		Subjective discount factor		X		X
	Nutrition		X		X	
	Message type and frequency	Comparison of the four interventions		X	X	X
	Message fatigue	Patients' perceptions of the intervention		X	X	X

4.2.12 Analysis Plan

Our analysis will exploit the randomized nature of the intervention to attribute treatment effects. In addition to a simple comparison between the pooled treatments and the control group we will also assess the differential impact of the 4 treatment arms. We will employ multivariate regression models to assess the impact of the interventions on each of the primary outcome variables. For the continuous outcome measure (CD4 count) we plan to employ an Ordinary Least Squares model. For the count data (pill doses missed) and if the continuous outcome measure is skewed, we propose to make use of a Poisson model; for the dichotomous outcome (remaining on ARV regimen), we propose to employ a Logit model. In all models we will control for the clustering of participants within health centers. We will make use of the coefficient estimates from the 4 treatment arms to establish a preference ordering of the effectiveness of the different types (text vs ASCII image) and frequencies (weekly vs semiweekly) of the SMS text message reminders. We will also estimate quantile regressions to identify which group(s) of participants are most (least) likely to have gained from the interventions. Comprehensive data on patient characteristics collected at baseline will allow us to control for confounding factors and patient heterogeneity.

These patient characteristics include age, gender, ethnicity, education, whether the patient is the head of the household, income, and whether the patient works. The simple comparison of outcome variables across groups will be complemented by a difference-in-difference identification strategy where we jointly employ the data from the three follow-up surveys. This analysis will permit us to tease out the differences in health outcomes attributable to each of the 4 interventions while controlling for time-fixed effects.

The data will be collected and coded by trained enumerators and analyzed using the statistical and data analysis program STATA. We will use conventional levels of significance at 1%, 5%, and 10%. Because we identified three primary outcome variables, we will make use of an inflated alpha when determining significance levels for each and every outcome separately. We apply this correction because the more significance tests we conduct at $\alpha=0.05$, the more likely we are to claim that we have a statistically significant result. For a significance level of 5% the Bonferroni inflated alpha is $0.05/3 = 0.015$.

We expect missing data to be below 5% and plan to address it using imputation techniques. However, the technique that we do use will depend on the characteristics of the missing information. Our goal is to avoid reducing the sample size without compromising statistical validity.

4.3 Results

Our project finished the recruitment of patients in May 2015. We have recruited the targeted 3800 patients across the 80 health centers and even oversampled by 38 patients. We kept 8 pure control health centers and randomized participants in the remaining 72 treated health centers across the treatment arms and the control group. The intervention started in October 2015 and follow-up data collection has started in April 2016. Analysis of the intervention has not yet started. The project was challenged by the sociopolitical instability in Burkina Faso during 2014, 2015, and early 2016. The initial launch of the study coincided with a coup d'état in Burkina Faso in November 2014. Despite the political turmoil in the country, the study received ethical clearance from the Burkinabe Ethics Committee for Research in Health in December 2014. A second coup d'état in September 2015 further challenged the project. But we could continue with the project and launch the intervention in October 2015. In January 2016, the terrorist attacks in Burkina Faso's capital Ouagadougou added another layer of uncertainty to the project. Throughout the sociopolitical instabilities in the country the local and the international team managed to keep the project running. We therefore expect

that we can successfully complete the project and conduct the intended analysis as outlined in this protocol.

This intervention is funded by 3ie-International Initiative for Impact Evaluation and conducted by the International Institute of Social Studies of Erasmus University of Rotterdam in collaboration with Univsersité Polytechnique de Bobo-Dioulasso.

4.4 Discussion

4.4.1 Summary

To date, very few rigorous evaluations have examined the impact of SMS text messaging reminders on retention and adherence of PLHIV. Existing studies are based on small samples not delivered widely at a national scale and tend to focus on short time horizons (Shet et al., 2014; Mbuagbaw et al., 2012; da Costa et al., 2012; Horvath et al., 2012; Pop-Eleches et al., 2011; Kunutsor et al., 2010; Lester et al., 2010). To the best of our knowledge, this study will be the first long-term RCT to assess the effects of a mHealth intervention using an intervention that is delivered at national scale. The study will also assess the effects of message type (text vs ASCII image) and frequency (weekly vs semiweekly), as well as, whether patients experience message fatigue over the course of the 2-year period of the intervention. To assess the impact of this intervention in a multivariate fashion, we will collect information on sociodemographic traits and several diseases-related outcome indicators. The findings of this study will enhance understanding of how interventions using mobile technology can influence HIV care and treatment. The conclusions from this study will contribute to more informed recommendations for mHealth in health care provision and the development of national and international guidelines for mHealth usage. We expect to define when, why, and for whom mHealth interventions may work. The findings may also be applicable to other African countries, which have similar epidemiological HIV profiles and social conditions as in Burkina Faso.

4.4.2 Limitations

We are not able to control for all possible external factors that may influence the results. These factors include, among others, appropriate supply of medication, political or civil struggles that may influence the functioning of the health system, and other community-level changes during the 2-year intervention period. All these contextual confounders will be documented and considered while interpreting the findings.

4.4.3 Conclusions

This study aims to contribute to the evidence on mHealth usage to support HIV care delivery in resource-poor settings. We will evaluate the impact of mHealth interventions for care and treatment of PLHIV, and thereby expect to inform strategies to improve health-related outcomes among specific populations. The study expects to advance the research on how long mHealth interventions remain effective in the context of wide-scale interventions in developing countries. We will determine the impact of mHealth in the short-, medium-, and long-term, and thus advance understanding of how mHealth interventions may complement other social and behavioral health interventions in developing countries.

Chapter 5: Adaptation and biomedical transition of people living with HIV to antiretroviral treatment in Burkina Faso [43]

5.1 Introduction

People living with HIV (PLHIV) can now adapt to the disease and enjoy a long and healthy life by adhering to antiretroviral treatment (ART). Recent studies show that patients transit from an early period of disruptive psychological and physical trauma to a stable long-term state of good health (Baranov, Bennett, & Kohler, 2015; Thornton, 2012). This process involves changes in priorities, values and objectives in life between the early phase of uncertainty and fear about survival, and the mature phase after securing survival and reincorporating into society (Russell & Seeley, 2010; Russell et al., 2007). While confronting a life-threatening disease is unsettling and transformative, the literature on adaptation explains that people first react strongly to negative health shocks but revise to conventional levels of wellbeing over time (Albrecht & Devlieger, 1999). To date, there is scarce evidence on how subjective and objective measures of health are shaped throughout the course of the transition phase, and there is even less evidence about the role of socio-economic correlates in the adaptation process.

This paper investigates how patients cope and adapt to the life-threating diagnosis of HIV in a resource-poor setting. We assess whether there are shifts in subjective and objective measures of health and their associated socio-economic determinants. Changes in both health measures are contrasted since the adjustment processes are likely to be different as the indicators capture different aspects of wellbeing

Three general reasons guide the research on subjective and objective measures of health. First, both measures have become an important goal of public policies and a non-negligible criterion for resource allocation in health care (Dolan & Kahneman, 2008). Second, existing evidence has demonstrated that measures of health and wellbeing are credible predictors of economic activity and biological outcomes. Indicators of subjective health and wellbeing have enhanced our understanding of unemployment, inflation, gross domestic product, social capital, migration, social relationships, and related areas (Di Tella, MacCulloch, & Oswald, 2001; Dolan & Kahneman, 2008). Similar studies have found a strong association between subjective and objective measures with health conditions such as mortality, severity of mental

[43] Artavia-Mora L, Wagner N, Thiombiano BA and Bedi A. 2020. Adaptation and biomedical transition of people living with HIV to antiretroviral treatment in Burkina Faso. *Global Public Health*, 15(5), pp.638-653.

disorders, disability, intensity and prevalence of chronic diseases, medical expenses and utilization, stress, and physical functioning (Benjamins, Hummer, Eberstein, & Nam, 2004; Doiron, Fiebig, Johar, & Suziedelyte, 2015; Jylhä, 2009; Wu et al., 2013). Yet, the findings have to be interpreted knowing that the ratings of subjective health may vary across patients' understanding, believes and sociocultural context even within one country. Third, the success in predicting health outcomes has encouraged research on the underlying determinants that predict subjective health (Addai & Adjei, 2014; Koelmeyer, English, Smith, & Grierson, 2014).

The leading theories on hedonic adaptation and response shift provide a rationale why PLHIV may adapt to the disease over time. First, hedonic adaptation involves the reaction and judgment of individuals to negative or positive stimuli according to their own histories that act as reference points of comparison (Frederick & Loewenstein, 1999). The phenomenon conceptualizes that a person who undergoes a negative (health) shock could adapt to the situation so that the new poorer state does not translate into lower subjective health in the long-term (Dolan & Kahneman, 2008). A host of research confirms the presence of hedonic adaptation in diverse areas of wellbeing demonstrating that life shocks can have small impacts on happiness (Oswald & Powdthavee, 2008). There is also evidence of hedonic adaptation in many areas of health such as disability, chronic illnesses, and long-term treatments. In a longitudinal study, patients with prior heart histories were less likely to report worse subjective health when they experienced a new heart problem compared to first-time patients (Wu, 2001). In a similar vein, a study on patients of hemodialysis shows that they are happier than the predictions of healthy individuals (Riis et al., 2005). Moreover, people with a new sudden physical disability initially report aversion to work, but over time, they return to their initial level of wellbeing at a rate of about 30–50% (Schwartz, Andresen, Nosek, & Krahn, 2007). And more recently, it has been shown that women adapt to breast cancer in terms of health-quality (Tessier, Blanchin, & Sébille, 2017).

Second, response shift is a phenomenon closely related to hedonic adaptation (Schwartz & Sprangers, 2000). Response shift conceptualizes that life priorities and expectations are updated and recalibrated after new life circumstances (Postulart & Adang, 2000; Schwartz & Sprangers, 1999; Sharpe & Curran, 2006). Three factors may trigger this redefinition: (i) recalibration, (ii) reprioritization, and (iii) reconceptualization. Recalibration involves the alteration in an individual's internal standards after a new life-changing situation.

Reprioritization refers to changes in personal priorities and values such as valuing health more than income after the announcement of a deadly disease. Reconceptualization comprises the redefinition of an individual's standard of health over their life (Pierret, 2007). Thus, response shift implies a change in the reference point of what good health means when confronted with an unfortunate diagnosis. Previous research has documented response shift in patients of asthma, diabetes, epilepsy, hemophilia, breast cancer, and kidney disease (Sharpe, Butow, Smith, McConnell, & Clarke, 2005; Tessier et al., 2017).

The present study contributes to the literature in four ways presenting an important feature compared to past research. First, we compare short-term versus longer-term adherents to ART to assess whether patients of HIV reach a steady level of subjective health in the course of treatment adherence. We also approximate the response shift phenomenon by investigating whether subjective health is higher among PLHIV compared to the healthy population. Second, we contrast subjective and objective measures of health since they evaluate different aspects of wellbeing and their adjustment processes are likely to differ. Third, our analysis strengthens the existing literature by using a nationally representative sample of 3625 PLHIV in a resource-poor setting, namely Burkina Faso. Fourth and most importantly, although there is ample research about hedonic adaptation and response shift for a range of medical conditions, there is scant evidence about patients of HIV under ART, especially in resource-poor contexts where income constraints and perceived social stigma may be more intense. The study that is closest to ours analyses 1700 Australian PLHIV showing that income, employment, presence of social support, self-reports about recent sexual activity and participation in support groups are strong predictors of subjective health (Koelmeyer et al., 2014).

To test the adaptation and biomedical transition in health we investigate subjective and objective measures of health analyzing a database with 3625 observations in a cohort study design. The subjective health measure that we employ uses a five-point Likert scale to provide a qualitative self-report about patients' perception of their health status. Rankings range from (1) very bad to (5) very good health. The objective measure is a physical biomarker and consists of patients' CD4 counts, which is the measure of disease progression that is most readily available in the Burkinabe health system. According to WHO (2013), the CD4 count provides an impartial indication about the state of the immune system of PLHIV. In the empirical analysis of adaptation and biomedical transition in health measures we

employ two steps: First, we assess changes in subjective health and its socio-economic determinants. Second, we compare the findings for subjective health with the predictors of the objective indicator of health.

5.1.1 Adaptation and biomedical transition in PLHIV: From short-term to long-term adherents

PLHIV experience a severe and intense process of adjustment between the early phase of shock after discovering the disease, and the mature phase of management and acceptance (Kralik, Koch, Price, & Howard, 2004). The early experience of starting ART is disruptive for a person's life (Russell & Seeley, 2010). Treatment initiators struggle to maintain control and a positive outlook on life since they experience stigma, fear, isolation, distress, and uncertainty about survival (Rhodes, Bernays, & Terzić, 2009). Incorporating ART in daily routines requires major non-linear adjustments in activities and living standards that involve many challenges along the way (Sharpe & Curran, 2006). Accommodating the treatment scheme and the new health condition require drastic adjustments in activities, relationships, parenting, employment and consumption patterns, leisure, and many other areas of life (Kralik et al., 2004; Pierret, 2007; Rhodes et al., 2009; Russell & Seeley, 2010; Russell et al., 2007).

Instead, research demonstrates that patients can return to an improved and healthy status when adhering to ART in the long-term. Several studies describe how PLHIV experience transformative and empowering effects after maturing in the management and acceptance of the treatment (Neuman, Obermeyer, & MATCH Study Group, 2013; WHO, 2013). By securing survival and recovering physical strength, long-term adherents take agency and reactivate their life as independent workers and achievers (Russell & Seeley, 2010). Long-term adherents can even attain the leading status of experts in their communities becoming valuable resources for new patients and the broader community (Pierret, 2007; Russell & Seeley, 2010).

5.1.2 Conceptual framework: Predictions about adaptation and biomedical transition

Assuming the presence of response shift we expect that as PLHIV integrate antiretroviral treatment into their daily routines they demonstrate higher ratings of subjective health compared to the general population (Figure 5.1). PLHIV revalue their state of health since loss aversion changes their perception of good health. This effect increases their perception of subjective good health. We do not anticipate hedonic adaptation as patients experience

permanent costs and side effects of the disease. Thus, response shift will not revert or disappear so that subjective health will stay high and stable over time (or as long as patients adhere to ART). We hypothesize that the subjective report could increase among long-term adherents as they continuously feel physically better. Comparatively, with regard to objective health, we expect that short-term adherents show lower health than the general population since they have only recently initiated ART as a response to a deterioration of their health. In contrast, those patients that remain under ART will improve their objective health status in the long-term. Long-term adherence to ART can suppress HIV leading to (a close to) normal life.

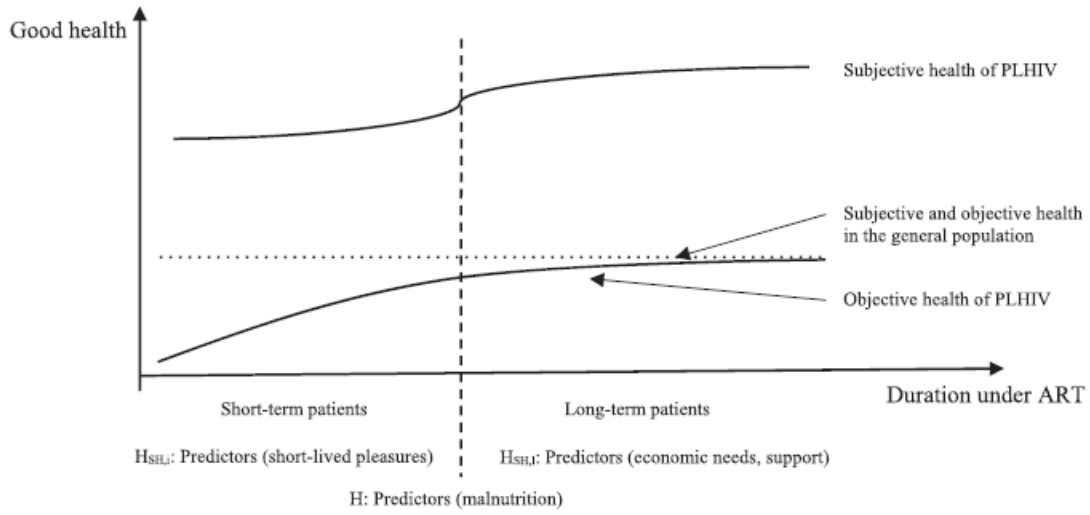
Furthermore, we also expect adaptation in the determinants of health. Since the early process of adaptation shifts patients' priorities, we expect that short-lived pleasures that are sources of immediate happiness will be positively correlated with the subjective health measure (Pierret, 2007; Russell & Seeley, 2010). The focus on short-lived pleasures is plausible since patients are aware that ART does not necessarily work for everybody (Russell & Seeley, 2010; Russell et al., 2007; WHO, 2013). This is what we refer to as hypothesis HSH,i in Figure 5.1. In contrast, long-term adherents can return to 'normality' once they have adapted to the disease and have managed to integrate ART in their lives (Pierret, 2007; Russell et al., 2007). In terms of the determinants of subjective health we anticipate that short-lived pleasures lose their importance in the long-term by shifting personal values towards economic needs and social success. This is denoted as hypothesis HSH,l in Figure 5.1. Thus, we expect the reversal to pre-disease levels in the determinants.

Independent of phase of adherence, we expect that patients of HIV with better nutrition also report better subjective health (see hypothesis H in Figure 5.1) (Masa, Chowa, & Nyirenda, 2018).

For the measure of objective health, we expect a sharp and strong biomedical transition from a very low level in early adherents to a higher level in long-term patients. The existing literature provides little guidance on determinants of health biomarkers among PLHIV. Nutrition is the only exception that has systematically shown to support the development of resilience against HIV (Anema, Vogenthaler, Frongillo, Kadiyala, & Weiser, 2009; Houtzager, 2009; Koethe et al., 2010). The literature has identified three pathways: (i) there is an overlap between HIV prevalence and malnutrition (Anema et al., 2009), (ii) PLHIV are more vulnerable to malnutrition than the general population (Houtzager, 2009), and (iii)

malnutrition is a predictor of mortality among PLHIV (Koethe et al., 2010). Thus, we consider malnutrition as the link between subjective and objective health, and we expect malnutrition to be negatively associated with both indicators.

Figure 5.1: Theoretical predictions.



Notes: The vertical axis displays the level of health and the horizontal axis the duration under ART. SH abbreviates subjective health. The hypothesis HSH,i is formulated for short-term adherents with respect to subjective health; the hypothesis HSH,l is formulated for long-term adherents with respect to subjective health; hypothesis H is for all PLHIV and refers to both subjective and objective health.

5.2 Materials and methods

5.2.1 Study setting

Burkina Faso is a land-locked West-African country. According to the Human Development Index in 2018 the country scores (very) low in human development by ranking 183 out of 189 nations. The country is ethnically diverse and multilingual where the 'Mossi' ethnicity constitutes 50% of the population.

In 2016, the adult (15–49 years) HIV prevalence rate was 0.8% (UNAIDS-a, 2016). Since 2005, the prevalence rate has been stable, and by 2010 new HIV infections had decreased by 45% (UNAIDS, 2012). Nevertheless, there are considerable disparities in prevalence rates by place of residence and region: Prevalence is lower in rural (0.6%) compared to urban areas (2.3%) (UNAIDS, 2012). The feminization of HIV is also a phenomenon observed in Burkina Faso with two-thirds of the registered HIV cases being women (UNAIDS-b, 2016).

Burkina Faso provides medical HIV support in 100 public, private and community health centers located in 13 regions of the country (UNAIDS-a, 2016). Since 2010 antiretroviral medication is free, and in 2015, Burkina Faso achieved the continuous provision of ART to all patients under treatment. There are an estimated 95,000 PLHIV in the country, of whom between 46,000 and 70,300 are on ART (UNAIDS, 2012). By 2016, the coverage rate of ART was 70% (UNAIDS-a, 2016). However, the challenges of transportation, stigma and time frequently prevent PLHIV to adhere to ART. In practice, some centers also require patients to pay for specialized medical tests such as the CD4 count. For that reason, not all PLHIV have up-to-date test results in the study.

Another important feature of the Burkinabe context is the presence of HIV/AIDS self-help groups/associations. These voluntary associations serve as support groups offering regular meetings, counselling, and occasional financial funds for nutrition and medical tests. The main objective of the self-help groups is to prevent PLHIV from social isolation, leading them to a functional life as individuals, family members and members of their communities.

5.2.2 Study design

We analyze a nationally representative sample of patients of HIV who undergo ART in Burkina Faso in 2015. The total dataset comprises 3625 patients randomly selected from 80 health centers across the 13 health regions of the country. The patients were interviewed between February and May 2015. Eligibility consisted in being older than 15 years and providing written informed consent. Moreover, trained enumerators privately administered the survey to ensure confidentiality (Wagner, Ouedraogo, Artavia-Mora, Bedi, & Thiombiano, 2016). The survey contains information on patients' subjective health, CD4 counts, as well as demographic and socio-economic characteristics.

5.3 Data analysis

5.3.1 Outcome indicators

The outcome indicator for subjective health is based on responses to the survey question 'How do you rate your health?'. Responses are given on a five-point Likert scale ranging from (1) very bad, (2) bad, (3) fair, (4) good, to (5) very good. We employ the responses in two ways. First, we directly assess the given health ratings. Second, we build a dichotomous variable combining response categories (4) and (5) as being in good health, denoted by one

and zero otherwise. We analyze both subjective health variables to assess the sensitivity of our findings to the definition of subjective health.

The measure of objective health is the CD4 count as it provides an indication of the health of the immune system and no other biomarker is regularly assessed by the Burkinabe health system. The CD4 count of a healthy person ranges between 500 and 1500 CD4 cells per cubic millimeter of blood. PLHIV who have a CD4 count over 500 are in good health whereas those below 200 are at a high risk of death (WHO, 2013). Similar to subjective health, we analyze the CD4 count directly and by building a dichotomous variable with patients with CD4 counts above 500 (in good health), coded as one, and zero, otherwise.

5.3.2 Control variables

As determinants of good health we consider an individual's nutritional status, demographic, educational and economic characteristics, and two related variables: (i) whether the individual participates in an HIV/AIDS self-help group, and (ii) whether the individual had sexual intercourse in the week prior to the survey. We capture nutritional status employing weight. Ideally, we would have used BMI to proxy nutrition but we do not have height measurements for the majority of the sampled individuals. The individual-level demographics are gender, age, marital status, religion and ethnicity.

Additionally, we control for household demographics. The household demographics capture household size using: (i) whether the individual has children, and (ii) the share of male and female individuals above 15 years. The latter two variables show whether the household has female adults and capture the possibility of labor sharing. We also use four educational categories and a binary variable for fluency in French. We compare patients with primary, basic secondary and higher education to no formal education (excluded category).

Three variables reflect the economic situation of the patient: (i) an indicator evaluates active work; (ii) there is information on individual income; and (iii) we include whether the individual comes from an agricultural household. Moreover, we use a binary variable to indicate whether the patient participates in an HIV/AIDS self-help group. Finally, we control for sexual intercourse in the last seven days. Because sexual transmission is one of the major transmission channels of HIV, there have been considerable sensitization efforts around HIV and sex (Kalichman, Carey, & Johnson, 1996). Therefore, we consider information about sexual activity a potential determinant of the wellbeing of PLHIV.

5.3.3 Empirical specification

We employ descriptive and multivariate analyses to examine our predictions. First, we carry out an independent sample t-test for a comparison of means of the outcome variables. Second, we implement a linear regression model for the multivariate analysis. As a robustness check we make use of an ordered probit model for the subjective health ratings. We specify the following model for patient i who has attended health center c to assess the determinants of health across the transition path:

$$Y_{ic} = \beta_0 + \beta_1 Malnutr_{ic} + \beta_2 Demo_{ic} + \beta_3 Edu\_Eco_{ic} + \beta_4 Group_{ic} + \beta_5 Sex_{ic} + \gamma_c + \varepsilon_{ic} \quad (1)$$

where Y_{ic} represents either subjective or objective health. We control for malnutrition using weight ($Malnutr_{ic}$) and the demographic variables are denoted by $Demo_{ic}$. To adjust for non-linearities we include weight and its squared term.[44] In a second step we add the educational and economic controls represented by Edu_Eco_{ic} and, in a third step we add two related variables: $Group_{ic}$ capturing the membership in an HIV/AIDS self-help group, and Sex_{ic} comprising active sexual intercourse in the last seven days. In all specifications we include health center-specific effects (γ_c) to capture differences in infrastructure, local traditions, local policies, and geography. The fixed effects also address concerns about context-specific perceptions of the subjective health rations. The standard errors are clustered at the health centre level to account for correlations in the unobservable characteristics (ε_{ic}).

We split our sample into two groups to allow for a comparison of cohorts within our sample: short-term and long-term adherents. We divide the sample at 24 months of adhering to ART. Three criteria guide this choice. First, the recent medical literature on HIV demonstrates that PLHIV experience the strongest psychological and biological changes during the first two years on ART (Russell et al., 2007). Second, current qualitative evidence explains how the early challenges that PLHIV face in quality of life and objectives become more mature and complete after two years (Kralik et al., 2004; Rhodes et al., 2009; Russell et al., 2007; WHO, 2013). And third, drop-out rates are higher during the initial two years of adherence to ART (Fox & Rosen, 2010; Rosen, Fox, & Gill, 2007). But there is limited evidence on the psychological and physical evolution of PLHIV after two years of adherence to ART. By splitting the sample at 24 months we capture the most prominent differences between shorter

[44] According to WHO only 5.2% of the total Burkinabe population is obese (WHO, 2016). As comparison, in the United States adult obesity rates exceed 35% in five states, the lowest rate is 22.3% (SOB, 2017).

and longer-term ART adherents. Hence, we estimate Equation (1) separately for the two cohorts of HIV patients in our database to compare short-term (≤24 months) versus long-term adherents (>24 months).

5.4 Results

This section contains four parts. First, we introduce the characteristics of our sample. Second, we provide descriptive statistics for the health outcomes. The third part outlines adaptation in the predictors of subjective health and the fourth part shows the results for the transition in the determinants of the objective health indicator.

5.4.1 Descriptive statistics

5.4.1.1 Descriptive statistics: Determinants

Descriptive statistics are presented in Table 5.1. The average age of a patient in our sample is 41.3 years and the majority are female (69%). The high rate of female PLHIV is not unusual and is commonly referred to as the feminization of HIV (Quinn & Overbaugh, 2005). But the large share of women may simply reflect a higher detection rate as in Burkina Faso HIV testing is part of antenatal care. Likewise, sex differences in health-seeking behaviors between Burkinabe women and men may also explain the lower level of ART enrolment by men (Auld et al., 2015).

The average household size is 7 which is similar to the average household size of 7.4 reported in the 2014 Burkina Faso Continuous Multisector Survey (Enquête Multisectorielle continue) (World Bank, 2017). The majority of respondents have children (89%) and are married or live in cohabitation (57%). Among short-term adherents the share of married individuals is larger, namely 59.6% versus 55.8% among long-term adherents (p-value = 0.066). A considerable share of the sample loses their partner in the course of the disease. We observe nearly 23.7% of widows/widowers among long-term adherents versus 18.1% among short-term patients (p-value = 0.000). We also report the share of working age (fe)males in the households of the respondents to capture intrahousehold demographics. The average household has a share of 26.4% (35.9%) (wo)men. Muslims account for roughly half the sample (50.7%), Christians for 44.9%. Mossi represent the main ethnicity with 52.6% of respondents, which is similar to official country statistics.

More than half the sample has no formal education with no differences between short-term and long-term patients. Concerning economic characteristics, we observe that long-term

adherents are qualitatively speaking 3.3 percentage points more likely to report that they are actively working. This is reasonable as survival encourages patients to continue their life and work.

Almost 40% of the patients are members of an HIV/AIDS self-help group. This figure highlights that Burkina Faso is a country with very active associations. However, participation in these groups is not very prominent among short-term adherents (24.9%). It increases among long-term adherents to almost 50% (p-value = 0.000). The literature confirms this finding as social isolation and stigma are an initial reaction in the process of adjustment to the disease (Russell & Seeley, 2010; Russell et al., 2007). Lastly, about 37% of individuals had sexual intercourse in the week prior to the survey.

5.4.1.2 Descriptive statistics: Outcome indicators

The self-assessed health rating is very high with an average of 4.36 out of 5 and 92% (n = 3625) of patients report that they are in good or very good health. This is a substantially higher figure compared to the general population. According to the latest World Values Survey for Burkina Faso in 2007, the general population reports an average rating of 2.96 out of 4 (after transforming the scale for comparison), and only 72% (n = 1519) report good subjective health. This finding is consistent with previous evidence on the response shift. Amongst short-term patients we observe an average health rating of 4.3, whereas longer-term adherents have an average of 4.4. This difference is small albeit statistically significant (p-value = 0.000) showing that while response shift continuous after 24 months, the main force driving the shift occurs in the initial 24-months.

The response shift phenomenon is further highlighted when considering the share of individuals who report being in (very) good health. Among the short-term patients this share amounts to 89.6% (n = 1408) while among the long-term adherents the share is 93.7% (n = 2217). The difference of 4.1 percentage points corresponds to a small positive and statistically significant difference between the two cohorts (p-value = 0.003). Thus, we document a response shift in subjective health which occurs quickly and does not change greatly over time.

Table 5.1. Descriptive Statistics

Variable	Full Sample		Short-term patients		Long-term patients		Diff. in means p-value
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
Weight	59.929	10.646	59.106	10.729	60.451	10.562	0.012
Female	0.690	0.463	0.678	0.468	0.697	0.460	0.257
Age	41.274	9.438	39.376	9.706	42.479	9.063	0.000
Household size	6.807	5.191	6.803	5.285	6.810	5.132	0.980
Participant has children	0.888	0.315	0.879	0.326	0.894	0.308	0.228
Share of women above 15	0.359	0.212	0.361	0.225	0.357	0.204	0.777
Share of men above 15	0.264	0.199	0.252	0.189	0.272	0.204	0.011
Marital status (Excluded category: Divorced separated)							
Married	0.572	0.495	0.596	0.491	0.558	0.497	0.066
Widow	0.215	0.411	0.181	0.385	0.237	0.425	0.001
Single	0.153	0.360	0.159	0.366	0.148	0.356	0.477
Religion (Excluded category: Other and none)							
Muslim	0.507	0.500	0.561	0.496	0.473	0.499	0.007
Christian	0.449	0.497	0.398	0.490	0.482	0.500	0.005
Main ethnicity: Mossi	0.526	0.499	0.529	0.499	0.524	0.500	0.905
Education (Excluded category: No formal education)							
Primary	0.290	0.454	0.277	0.448	0.299	0.458	0.305
Secondary and higher	0.198	0.398	0.192	0.394	0.201	0.401	0.625
French is main language	0.101	0.301	0.112	0.316	0.093	0.291	0.213
Economically active	0.357	0.479	0.337	0.473	0.370	0.483	0.261
Income in 1,000 FCFA	14.532	59.783	13.987	59.834	14.878	59.761	0.717
Lives in agricultural household	0.546	0.498	0.513	0.500	0.568	0.495	0.253
Member of self-help group	0.398	0.490	0.249	0.433	0.493	0.500	0.000
Sexual intercourse in last 7 days	0.366	0.482	0.364	0.481	0.367	0.482	0.920
Outcome indicators							
Subjective health rating (1-5)	4.358	0.649	4.298	0.686	4.396	0.622	0.035
Share of good health	0.921	0.270	0.896	0.306	0.937	0.243	0.003
CD4 counts	454.297	229.338	388.381	214.086	489.601	229.509	0.000
Ind. of good health (CD4 count > 500)	0.368	0.482	0.266	0.442	0.423	0.494	0.000

Notes: The total number of observations is 3625. The subsample of antiretroviral short-term adherents consists of 1408 observations, the subsample of long-term patients consists of 2217 observations. For the CD4 count only 835 (1559) observations are available for the subsample of antiretroviral short-term (long-term) patients. The comparison of means between short-term and long-term adherents accounts for clustering of the error term at the health center level.

Furthermore, we find strong patterns of biomedical transition between short-term and long-term adherents.[45] Short-term patients have an average CD4 count of 388 while for long-term adherents it is 490 (p-value = 0.000). Similarly, the share of short-term patients in good health is 26.6% whereas it is 42.3% among long-term adherents. The increase in the CD4 count is consistent with medical and economic research demonstrating the efficacy and usefulness of ART to boost CD4 counts in PLHIV (WHO, 2013). Notably, the contrasting evolution between the sharp change in CD4 counts and the slow change in the subjective health measure suggests that while both indicators are useful and informative, they measure

[45] The average CD4 count is 454.3. The share of individuals with objective good health is 36.8% (CD4 count of 500 or above).

different dimensions of wellbeing. A pairwise correlation between subjective health and the CD4 count yields a small (4%) albeit statistically significant (p-value = 0.059) correlation.

In sum, subjective health is an omnibus measure, contrarily to the CD4 count, and includes both physical and psychological aspects. The rapid adjustment in subjective health may occur as reaction to obtaining ART and the sense of relief that comes with acquiring a new lease of life. Instead, psychological factors influence less the objective measure, which is mainly driven by the use of antiretroviral medication. Therefore, both health measures should not be used as substitutes in this context.

5.4.2 Adaptation in subjective health

In Table 5.2 we present the determinants of subjective health for short-term versus long-term patients. The outcome variable is a five-point Likert scale rating of subjective health. Three results stand out that are similar across cohorts: First, individuals with better nutritional status also report higher subjective health. The coefficient is small albeit statistically significant. A one standard deviation increases in weight (10.65) increases subjective health by 0.23 to 0.28. This increase is small given that the average subjective health rating is 4.4. Although small, the effect of one sample standard deviation increase in weight (10.65) amounts to 35% of the sample standard deviation of subjective health (0.65).[46]

Second, being economically active is a positive determinant of subjective health for short-term and long-term adherents. We observed in the field and in the literature that for the former, having a paid task gives them purpose and meaning in a period where they are experiencing emotional distress. Suspending work promotes feelings of solitude, low self-esteem and poor outlook (Neuman, Obermeyer, & MATCH Study Group, 2013; Rhodes et al., 2009; Russell & Seeley, 2010; Russell et al., 2007). For long-term adherents, having economic activity and income secures household wellbeing. This implies that general economic wellbeing regains importance in long-term patients once they have secured their survival (Kralik et al., 2004; Russell & Seeley, 2010; Russell et al., 2007; WHO, 2013).

Third, adherents from households with a larger share of women above 15 years report higher subjective health. In the field, we found that women are more likely to be supportive in work share arrangements. Since two-thirds of our sample consists of women, this finding is not

[46] Note that for the sample at hand, the negative impact of the squared weight term is negligible since it is almost zero. It only affects individuals with a weight above 170 kilograms. None of the respondents weighs that much.

particularly surprising but only highlights the collectivist culture of Burkina Faso. Note that the relationship becomes more pronounced for long-term ART patients showing that help from family members might be considered more important in the long-run.

Next, we identify two main determinants of subjective health among short-term patients. First, we observe that individuals who have children report higher subjective health. The effect is 16% to 18% of a standard deviation in subjective health. Second, enjoying the short-lived pleasure of sexual intercourse is positively associated with subjective health among short-term patients. The estimate of 0.12 explains as much as 18% of the standard deviation in subjective health. This finding shows that individuals who are in the early phase of disruption and adjustment place a relatively larger importance for living now and enjoying the current happiness of life.

In contrast, for long-term adherents the (short-term) effects of having children and sexual intercourse fully vanish. We find a shift in determinants away from short-lived and subjective sources of joy towards purely economic and practical considerations. Income is positively and statistically significantly associated with the subjective health of long-term patients.[47] While the economic effect is small –a one standard deviation increase in income (59.78) increases subjective health by 0.02– it indicates that economic concerns gain importance in the long-run. A second finding shows that members of agricultural households report lower subjective health in the long-term. This is not surprising since the agricultural households in the sample are poor and engage in labor-intensive activities. Thus, this result supports the notion that economic concerns regain importance among long-term adherents. In addition to economic factors, we find that married individuals report higher subjective health. Being married increases subjective health by
17% of a standard deviation and supports the idea of a positive effect of stable family arrangements for long-term antiretroviral patients. Finally, we find a small negative effect stemming from age.

To assess the robustness of our findings we employ two additional models. First, we estimate a model employing a dichotomous subjective health variable coding (very) good health as one and zero otherwise. The results confirm the patterns discussed above suggesting that the findings are not sensitive to the scaling or the definition of the outcome variable. Second, we

[47] For short-term patients, income is negatively and insignificantly associated with subjective health.

employ an ordered probit model for the subjective health ratings and the findings are again similar to those discussed above (results not shown).[48]

Table 5.2. Determinants of subjective wellbeing for ART short-term versus long-term patients (Outcome variable: Likert scale rating of subjective wellbeing).

	Short-term adherent			Long-term adherent		
	(1)	(2)	(3)	(4)	(5)	(6)
Weight	0.028**	0.026**	0.025**	0.024***	0.023***	0.023***
	(0.012)	(0.012)	(0.012)	(0.006)	(0.006)	(0.006)
Weight2	-0.000*	-0.000*	-0.000a	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.034	0.038	0.050	-0.005	0.003	0.005
	(0.052)	(0.052)	(0.052)	(0.036)	(0.037)	(0.039)
Age	-0.003	-0.003	-0.003	-0.003**	-0.003**	-0.003**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Household size	-0.003	-0.003	-0.003	0.002	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Participant has children	0.114**	0.106**	0.095*	0.043	0.038	0.039
	(0.052)	(0.051)	(0.052)	(0.049)	(0.050)	(0.049)
Share of women above 15	0.133*	0.130*	0.117	0.154**	0.138**	0.134**
	(0.077)	(0.077)	(0.074)	(0.063)	(0.063)	(0.063)
Share of men above 15	0.028	0.041	0.045	-0.042	-0.034	-0.038
	(0.095)	(0.097)	(0.096)	(0.068)	(0.069)	(0.070)
Marital status (Excluded category: Divorced, separated)						
Married	0.142	0.139a	0.107	0.112**	0.109*	0.102*
	(0.086)	(0.087)	(0.086)	(0.055)	(0.056)	(0.055)
Widow	0.115	0.099	0.100	0.063	0.054	0.056
	(0.095)	(0.095)	(0.096)	(0.057)	(0.057)	(0.056)
Single	0.076	0.074	0.061	0.093a	0.093a	0.092*
Education (Excluded category: No formal education						
Primary		-0.046	-0.043		-0.026	-0.028
		(0.048)	(0.048)		(0.027)	(0.027)
Secondary and higher		-0.097	-0.094a		-0.058	-0.058
		(0.060)	(0.058)		(0.046)	(0.046)
Lives in agricultural household		-0.024	-0.028		-0.059*	-0.065*
		(0.059)	(0.058)		(0.032)	(0.033)
Economically active		0.128**	0.110**		0.082**	0.078**
		(0.051)	(0.049)		(0.034)	(0.035)
Income in 1,000 FCFA		-0.000a	-0.000		0.000**	0.000**
		(0.000)	(0.000)		(0.000)	(0.000)
Member of self-help group			0.000			0.051
			(0.076)			(0.033)
Sexual intercourse in last 7 days			0.120***			0.047
			(0.031)			(0.035)

Notes: OLS results. The total number of observations for antiretroviral short-term (long-term) patients is 1408 (2217). The number of health centers, i.e. fixed effects, is 74 (75) for antiretroviral short-term (long-term) patients. All regressions include control variables for religion (Christian, Muslim), ethnicity (Mossi) and French being the main language but coefficients are not shown. Standard errors are clustered at the health center level. ***/**/* indicate significance at the 1/5/10% level, respectively.

[48] Detailed results are omitted for the sake of brevity but available from the authors upon request.

5.4.3 Biomedical transition in objective health

Table 5.3 presents the results associated with the determinants of objective health.[49] Three patterns persist throughout the adaptation process. First and similar to subjective health, weight is positively associated with the objective health measure. The effect is larger for short-term patients but persists for long-term patients. An increase in weight of one standard deviation (10.65) increases the CD4 count of short-term patients by about 90. This finding suggests that weight is the link function between the two health measures.

Second, while there is no gender difference in subjective health, we find strong gender differences in the objective indicator of health. Across treatment stages, women have a CD4 count that is at least 50 counts higher on average. This is consistent with the pattern found in the data that women have spent a longer duration on ART as compared to men. Third, individuals who are members of self-help groups have substantially higher CD4 counts.

Moreover, we also identify treatment stage-specific determinants. We find an interesting sign reversal with respect to being married. Short-term adherents who are married report a CD4 count that is roughly 25 counts lower but statistically insignificant. Since early PLHIV experience feelings of stigma, fear, and isolation, it might be that the closeness of living as a married couple makes it harder for short-term patients to (secretly) take their medication (Pierret, 2007; Rosen et al., 2007; Russell & Seeley, 2010). However, for long-run adherents, marriage increases CD4 counts and the coefficient is precisely identified. The effect is similar to an increase of at least 60 counts. A similar sign reversal is present in patients with secondary or higher education. Initially, patients that are more educated have lower CD4 counts compared to non-educated individuals but the effect reverts for long-term patients. While we do not find any direct income or economic effects, we observe that long-term adherents who live in large households are worse off. Reducing household size by one sample standard deviation (5.19) increases their CD4 count by roughly 10 counts.

[49] Information on CD4 counts was collected directly from health facility records and is not based on patients' self reported recollections. Trained hospital staff provided the information. For confidentiality reasons we did not have access to the health facility records. The CD4 count is available for 66% of the patients (2,394 of 3,625) and as discussed earlier, on average, it increased from 388 among short-term patients to 490 among long-term adherents.

Table 5.3. Determinants of CD4 counts for ART short-term versus long-term patients.

	Short-term adherent			Long-term adherent		
	(1)	(2)	(3)	(4)	(5)	(6)
Weight	9.160**	9.637***	9.040**	4.566	4.505	4.876*
	(3.626)	(3.616)	(3.607)	(2.839)	(2.867)	(2.884)
Weight2	-0.048*	-0.050*	-0.046	-0.018	-0.019	-0.020
	(0.028)	(0.028)	(0.028)	(0.019)	(0.019)	(0.019)
Female	57.466***	55.070**	54.666**	51.719***	54.272***	53.584***
	(21.068)	(23.290)	(24.018)	(14.390)	(14.496)	(14.538)
Age	-1.842**	-2.117**	-2.143**	-0.292	-0.036	-0.124
	(0.826)	(0.820)	(0.840)	(0.810)	(0.827)	(0.798)
Household size	-0.493	-0.946	-0.838	-2.083**	-1.905**	-1.833*
	(1.533)	(1.377)	(1.413)	(0.930)	(0.954)	(0.946)
Participant has children	-3.387	-3.788	-5.763	-11.796	-12.474	-11.107
	(23.089)	(23.346)	(23.408)	(20.692)	(20.964)	(20.963)
Share of women above 15	-18.446	-15.414	-15.806	14.122	10.723	9.275
	(43.266)	(43.079)	(43.377)	(26.891)	(27.046)	(27.443)
Share of men above 15	5.413	18.179	18.710	-5.769	-4.526	-6.589
	(33.469)	(33.016)	(32.369)	(32.991)	(32.328)	(33.279)
Marital status (Excluded category: Divorced, separated)						
Married	-25.954	-27.157	-25.209	62.762**	64.877**	67.259**
	(27.049)	(28.009)	(26.757)	(26.927)	(27.050)	(26.508)
Widow	-23.927	-28.610	-27.410	50.439**	53.879**	54.593**
	(31.160)	(31.343)	(30.752)	(24.684)	(24.912)	(24.721)
Single	3.191	11.418	15.605	72.332**	69.899**	71.633**
	(33.971)	(33.528)	(33.084)	(29.393)	(28.976)	(28.871)
Education (Excluded category: No formal education						
Primary		-32.832	-31.393		8.967	8.192
		(20.234)	(19.714)		(12.571)	(12.495)
Secondary and higher		-40.085	-41.162*		54.838**	53.417**
		(24.148)	(23.809)		(21.509)	(21.519)
Lives in agricultural household		37.454	34.771		-18.043	-20.436
		(24.118)	(24.158)		(19.710)	(19.593)
Economically active		1.670	0.166		9.163	6.003
		(19.484)	(20.814)		(14.955)	(14.957)
Income in 1,000 FCFA		0.002	0.002		-0.017	-0.012
		(0.238)	(0.242)		(0.069)	(0.070)
Member of self-help group			38.341**			38.160***
			(17.483)			(13.961)
Sexual intercourse in last 7 days			7.443			3.794
			(20.172)			(15.614)

Notes: OLS results. The total number of observations for antiretroviral short-term (long-term) patients is 835 (1559). The number of health centers, i.e. fixed effects, is 67 (67) for antiretroviral short-term (long-term) patients. All regressions include control variables for religion (Christian, Muslim), ethnicity (Mossi) and French being the main language but coefficients are not shown. Standard errors are clustered at the health center level. ***/**/* indicate significance at the 1/5/10% level, respectively.

To assess the robustness of our results we also employ an indicator variable of objective health. Individuals are coded as healthy who have a CD4 count of 500 or above. We identify the same determinants and dynamics as for the model with the counts.[50]

Overall, the biomedical transition is more pronounced than changes in subjective health. Moreover, the changes occur at different speeds indicating that subjective and objective health measures capture different aspects of patient wellbeing.

5.5 Discussion

We contribute to the literature on response shift by examining temporal changes in subjective and objective health amongst PLHIV. We also examine the socio-economic determinants of these health outcomes between a cohort of short-term (24 months or less) and long-term patients (more than 24 months). To examine the adaptation process, we employ a nationally representative sample of PLHIV undergoing ART in Burkina Faso. The subjective health measure employs a qualitative 5-point Likert scale from very poor (1) to very good (5) health, while the objective measure uses the CD4 count.

We find a small and statistically significant increase in self-reported good health from short-term to longer-term patients (90% to 94%, p-value = 0.000). We also find that there is response shift in subjective health resulting in a higher proportion of patients rating their health as good (92%) as compared to the general population (72%). A comparison of determinants of subjective health between short-term and longer-term patients shows that patients have different priorities, objectives and values between the two ART phases. In the cohort of short-term patients, we find that the presence of children and recent sexual intercourse increases their subjective health. In longer-term chronic patients we find that economic and practical considerations determine their subjective health. Thus, we observe a considerable shift in the predictors of subjective health over time.

We contrast changes in subjective health with the biomedical transition using a measure of objective health. Contrary to subjective health we find a gradual process of adjustment in the CD4 count. The CD4 count of short-term antiretroviral patients is about 100 counts lower compared to long-term adherents. The transitional dynamics in the determinants of the objective measure show that in the short-term more educated individuals have lower CD4 counts. However, in the long-term, highly educated individuals have better biomarkers.

[50] Detailed results are omitted for the sake of brevity but available from the authors upon request.

Another interesting finding is that while across disease stages men and women do not differ in subjective health, females have considerably higher health biomarkers. This raises the need for gender-specific responses to HIV/AIDS.

Overall, the findings indicate that subjective and objective measures of health capture different aspects of wellbeing. The subjective health measure provides an overly optimistic indication while the objective measure underestimates the actual beneficial effects of ART (in particular at ART initiation). Nutrition is the only parameter that consistently predicts higher levels of subjective and objective health across disease stages. Thus, our research provides further evidence for integrating food support and nutritional training within the existing system of HIV care. This fact is particularly important in the current situation of funding uncertainties and competing budget priorities (Dolan & Kahneman, 2008; Loewenstein & Ubel, 2008).

The study is not free from limitations: First, the research suffers from drawbacks that are inherent to the analysis of self-reported data. The results may comprise reporting errors and biases if the participants have systematically misinterpreted scales and/or questions. To the extent possible we aimed at limiting this source of bias by carefully training our enumerators. In addition, the response ratings may vary across patients' understandings, believes and sociocultural contexts. Since all patients are from Burkina Faso we consider the results valid for the country but it needs to be empirically verified whether our results are generalizable to other African countries. Second, patients might report good health in response to survey questions as signaling mechanism of good adherence. However, this hypothesis has shown no real effects in other chronic diseases (Parkinson) (Dolan & Kahneman, 2008; Skogar et al., 2012). Moreover, even if there is over-reporting about the quality of health, our findings would hold although the underlying mechanisms would change. Response shift would not be due to a change in the internal reference point but vis-à-vis the outside world. Third, the differences in sample sizes across short-term versus longer-term patients is likely to affect precision and comparability of our estimates. Note that the overall sample of 3625 observations is a nationally representative, random sample of PLHIV in Burkina Faso. The sample split was introduced with respect to the duration under ART and the resulting imbalance in the sample size between short-term and longer-term patients is merely a representation of the study context, i.e. patients can prolong their life substantially by adhering to ART. Importantly, both sub-samples include more than 1400 patients and are

similar in the number of observations to other international surveys that are often used for micro and macro cross-country comparisons such as the Worlds Values Survey (n = 1500) and Afro-Barometer Survey (n = 1200). Therefore, we expect the estimates to be precise and the comparison to be credible.

In sum, this research highlights the variations in patients' health outcomes and predictors across the different stages of the disease. Treating all individuals under ART in the same way, independently of how long they have been taking medication, is likely to miss out the variability in living situations and constraints of short-term versus long-term patients. Thus, the study advances our understanding of adaptation processes of PLHIV. We need to acknowledge adaptation to the disease and design health programs away from one-size-fits-all solutions to stage of disease-specific support and counseling. Incorporating adaptation processes in the design of health programs allows us to develop better tailored and ultimately more cost-efficient support strategies (Loewenstein & Ubel, 2008; WHO, 2013). In conceptualizing policies, we should focus on ameliorating treatment and resilience management by accounting for situational and context-specific factors and the possibility that these are likely to change in the course of the disease. A recent systematic review about mobile phone interventions (SMS/IVR/calls) to support PLHIV on ART in low and middle-income countries suggests that effective tools still need to be found to accompany PLHIV in the long-run (Demena, Artavia-Mora, Ouedraogo, Thiombiano, & Wagner, 2020). The research at hand suggests that it is particularly important to diminish obstacles to ART continuation and to recognize the needs of PLHIV depending on their specific demographic, economic, social and physical characteristics, especially given that HIV is nowadays a life-long chronic condition that societies must bear. Related, additional research should examine the generalizability of the presented results for other African countries and for other life-threating diseases (i.e. Tuberculosis).

Chapter 6: The impact of mobile reminders on retention, adherence, and psychosocial wellbeing in PLHIV- Evidence from a randomized controlled trial in Burkina Faso [51]

Embargo version

[51] In collaboration with Natascha Wagner, Denis Ouedraogo, Boundia A. Thiombiano, and Arjun Bedi.

Chapter 7: Experienced stigma as a determinant of subjective health in a large-scale and panel-data sample of PLHIV in Burkina Faso

Embargo version

Chapter 8: Conclusion

Based on two small- and one large-scale intervention, this thesis investigated the challenges experienced by marginalized and excluded groups and presented ways to improve their wellbeing in two countries. Part 1 of the thesis was based on experiments conducted in The Netherlands (Chapters 2 and 3) while Part II of the thesis was based on a nation-wide intervention in Burkina Faso (Chapters 4 to 7). Chapter 2 revisited classical theoretical principles on human cooperation and explored cognitive mechanisms to prompt prosocial behaviors. Chapter 3 tested instruments to enhance minority access to the housing market. Chapters 4 through 7 presented various research outputs drawn from a large-scale and long-term health intervention on PLHIV undergoing antiretroviral therapy in Burkina Faso.

8.1 Promoting human cooperation and access to housing in The Netherlands

The multicultural and tolerant characteristics of Dutch society provided an opportunity to test theoretical principles about the evolution of human prosociality and to put forward instruments to improve minority access to an essential market in The Netherlands. From a theoretical standpoint, the first chapter showed that people engage in costs to themselves to enforce prosocial choices even towards strangers in their communities (outside their network). Strangers are (for the most part) not marginalized or excluded from cooperative benefits and advantages. The chapter identified the existence of prosocial preferences in real-life helping and punitive behaviors in a natural task. The helping task was conducted in a field setting and avoided noise from social and reciprocity pressures hampering previous field research (Balafoutas et al., 2014; 2012). The findings showed that a shorter time span to decide encouraged helping behavior while a longer time span reduced the probability of helping strangers. This evidence is comparable to laboratory studies which tend to show that shorter times to make decisions favor cooperation, although the overall findings remain inconclusive across tasks, populations and time (Merkel and Lohse, 2019; Bouwmeester et al., 2017; Myrseth and Vollbrant, 2016; Rand, 2016; Capraro and Cococcioni, 2016; Rand et al., 2012;). At its core, this work challenges classical principles which assume that humans are mainly driven by self-interest, and it provides indicative evidence that human beings are intuitively helpful as a shorter time to make a decision encourages prosocial preferences.

The second chapter implemented two experiments to investigate new instruments to avert housing discrimination against ethnic minorities. The research found that on a secondary housing platform (Craigslist) there is no discrimination against Turkish and/or Moroccan

candidates compared to -*majority*- Dutch applicants. The findings indicated no evidence of prejudice against ethnic minorities, unlike previous literature which has shown high and systematic prevalence of discrimination in Dutch labor markets and leading housing markets in other European countries (Andriessen et al., 2012; Bloommaert et al., 2013; Auspurg et al., 2017; Van der Bracht et al., 2015). Thus, the study puts forward the idea that the use of auxiliary website platforms as opposed to the use of the most prominent housing platforms in a country is less likely to be associated with discrimination against ethnic minorities. Moreover, the article also discovered that a reference letter from a previous landlord does not impact the response from renters contrary to previous evidence from Germany and South Africa (Abel et al., 2017; Kaas and Manger, 2012).

8.2 Promoting wellbeing in PLHIV in Burkina Faso

The second half of the thesis consisted of various articles which drew on information from a randomized controlled trial to improve health outcomes in PLHIV in Burkina Faso. The intervention was launched in 2015 to explore the impact of a system of message reminders on a set of bio-physical and psychosocial outcomes. The design of the intervention examined five treatment-arms with four treated groups that received messages that varied in content (text versus image) and frequency (low versus high). The intervention was based on a nationally representative sample of 3,838 patients selected across the country for a period of two years and it collected patient-level information through four survey rounds (baseline and three follow-ups). The analysis of the program tested the durability of treatment effects (at 6, 12, and 24 months) and potential heterogeneities within sub-groups (i.e. between short- and long-term).

Using information from the baseline survey, Chapter 4 examined the process of adaptation and biomedical transition in measures of subjective and objective health as patients undergo antiretroviral treatment. The chapter contrasted outcomes for short versus long-term ART adherents. The results showed that these two measures of health capture different dimensions of the adaptation process as patients manage their treatment in the long-term. With regard to subjective health, there is a small and statistically significant increase (90 to 94%) in self-reported good health as per short-term versus longer-term patients. In other words, subjective reports of good health remain high and improve only slightly with time of adherence to ART. It is argued that with access to ART, PLHIV may rapidly rescale their perceptions of what is good health. In contrast to the limited change in the subjective health measures, the objective

measure of good health was substantially higher among longer-term (42%) compared to short-term patients (27%).The findings indicate that subjective and objective measures of health capture different aspects of wellbeing. The broader subjective health measure provides an overly optimistic picture while the narrower objective measure underestimates the beneficial effects of access to ART.

In Chapter 5 the thesis presented the results of the mHealth intervention. The program tested four message reminders that varied in content (text or image) and frequency (once or twice a week) to promote bio-physical, treatment-related, and psychosocial outcomes in PLHIV undergoing antiretroviral therapy over two years. The pooled sample results showed no global impact on primary outcomes (retention, adherence, and physical health) nor within follow-up surveys. This finding is comparable to previous evidence which has shown mixed impacts of mHealth interventions on bio-physical and treatment measures. A recent systematic review mentions that 40% of mHealth interventions have no impact on retention or adherence in low- and middle-income countries (Demena et al., 2020).

More interestingly, the intervention provided evidence of a large and positive impact of the intervention on a wide range of psychosocial measures (secondary outcome). Treated patients improved their reports of current and future happiness, faith in life, subjective health, and subjective health since previous survey. The image-reminders produced the strongest impacts across most measures, while the specific indicators of current and future happiness systematically improved in all four types of message reminders. Moreover, the psychosocial impacts in treated patients were stable and durable over the two years of the intervention. Message reminders generated the strongest effects in the psychosocial outcomes at six months which started to decay at twelve months, and disappeared at 24 months. However, these impacts endured for the entire 24 months in the sub-group of patients that confirmed the reception and comprehension of reminders. These impacts were also durable amongst longer-term adherents. Hence, the intervention extends the discussion about the cost-effectiveness of mHealth to other unexplored dimensions of health. The bulk of the literature focuses on a narrow set of standard bio-physical and treatment-related indicators without recognizing the deeper and important psychosocial benefits.

Chapter 7 explored the association between HIV-related stigma and subjective health of PLHIV. Employing patient-level fixed effects models, the study showed that stigma has a negative and statistically significant association with subjective health (2.3%-points, p-

value=0.090). Results indicated that income, household size and sexual activity were also positive predictors of subjective health. The analysis showed that retention in care reduced experienced stigma while regular participation in PLHIV-support groups led to increased awareness of stigma.

8.3 Conclusion

This thesis puts forward strategies aimed at improving the living standards and wellbeing of marginalized and excluded populations in modern societies. In two separate sections, the thesis tested theoretical, methodological, and policy-oriented instruments designed to assist vulnerable social groups in two different countries. For vulnerable groups in The Netherlands, this thesis presented mechanisms to promote human prosociality towards strangers and to improve access of ethnic minorities to housing. The thesis demonstrated that humans engage in costly prosocial (helping and punishment) actions towards members outside their network. The analysis demonstrated that designing programs and structures that prompt fast and/or automatic decisions may encourage prosocial choices in daily tasks. In terms of discriminating against vulnerable groups, the analysis showed that secondary housing platform-settings (similar to Craigslist in Amsterdam) may avert the high and systematic prevalence of ethnic discrimination found in many European markets. In contrast, there remains a need for instruments that improve access to housing given the lack of impact of the reference letter.

The second half of the thesis generated ideas and approaches to improve the wellbeing of PLHIV in Burkina Faso. The results presented in the thesis indicate that current public health programs need to recognize the process of patients' adaptation in the management of their treatment to achieve holistic resilience strategies in care support. To continue designing policies that assume all patients experience the same difficulties and progression misses the variability in perceptions and constraints between short-term and long-term cases. Especially in developing countries, health systems should expand the implementation of one-size-fits-all solutions to identify stage-of-disease specific support. This new view in policy may be particularly helpful to adopt in PLHIV in Burkina Faso, but also when individuals are exposed to other life-threatening diseases with different context-specific concerns.

This thesis also expands the scope of previous mHealth interventions in resource poor settings. It informs that systems of mobile reminders enhance patients' psychosocial health,

even when there is no impact on conventional indicators. While the benefits of mental health are larger in the early stage, the positive impacts continue over two years. Thus, this work suggests that past mHealth trials may have missed substantial health benefits by examining only a narrow set of standard indicators.

Finally, this thesis highlights the relevance of HIV-stigmatization and puts forwards strategies for its mitigation in Burkina Faso. From a health policy view, this thesis indicates that HIV-stigma is associated with poor subjective health and it argues that retention in treatment can mitigate social stigma besides improving physical progression. In contrast, the role of support group programs may need to be revisited as participation in such programs is associated with improvements in physical recovery but also prompts stigma. Future strategies should take these factors into consideration to develop cost-efficient policies and locally-adjusted programs in Burkina Faso.

References

1. Aalbers MB. 2007. Place-based and race-based exclusion from mortgage loans: evidence from three cities in the Netherlands. *Journal of Urban Affairs, 29*, pp.1–29.
2. Abel M, Burger R and Piraino P. 2017. The value of reference letters. Policy Research Working Paper #8266.
3. Achieng L, Musangi H, Ong'uti S, Ombegoh E, Bryant L, Mwiindi J, et al. 2012. An observational cohort comparison of facilitators of retention in care and adherence to anti-retroviral therapy at an HIV treatment center in Kenya. *PLoS ONE.* 7:e32727
4. Acolin A, Bostic R and Painter G. 2016. A field study of rental market discrimination across origins in France. *Journal of Urban Economics*, *95*, pp.49-63.
5. Addai I, and Adjei J. 2014. Predictors of self-appraised health status in sub-Saharan Africa: The case of Ghana. *Applied Research in Quality of Life*, 9(2), 233-253.
6. Adjei C. 2013. The Socio-Economic impacts of Stigmatization on young women (16-24) infected with HIV in Accra, Ghana (Ridge hospital).
7. Africa and Middle East Telecom-Week. 2016. Burkina Faso Mobile Fixed Network Operators list URL: http://www.africantelecomsnews.com/Operators_Regulators/ List_of_African_fixed_and_mobile_telcos_Burkina_Faso.html [accessed 2016-03-29] [WebCite Cache]
8. Ahmed AM and Hammarstedt M. 2008. Discrimination in the rental housing market: A field experiment on the Internet. *Journal of Urban Economics*, *64*(2), pp.362-372.
9. Ahmed AM, Andersson L and Hammarstedt M. 2010. Can discrimination in the housing market be reduced by increasing the information about the applicants? *Land Economics*, 86(1), pp.79-90.
10. Albrecht GL, and Devlieger PJ. 1999. The disability paradox: high quality of life against all odds. *Social Science and Medicine*, 48(8), 977-988.
11. Akerlof GA, Kranton RE. Economics and identity. *The Quarterly Journal of Economics*. 2000 Aug 1;115(3):715-53.
12. Amuri M, Mitchell S, Cockcroft A, and Andersson N. 2011. Socio-economic status and HIV/AIDS stigma in Tanzania. *AIDS Care*, 23(3), 378-382.
13. Andersson L, Jakobsson N and Kotsadam A. 2012. A field experiment of discrimination in the Norwegian housing market: Gender, class, and ethnicity. *Land Economics*, *88*(2), pp.233-240.
14. Anema A, Vogenthaler N, Frongillo EA, Kadiyala S, and Weiser SD. 2009. Food insecurity and HIV/AIDS: current knowledge, gaps, and research priorities. *Current HIV/AIDS Reports*, 6(4), 224-231.
15. Andriessen I, and Dagevos J. 2007. Discriminatie op de arbeidsmarkt: Kwantitatieve analyses [Discrimination on the labor market: quantitative analyses]. In Andriessen, I., Dagevos, J., Nievers, E. J., Boog, I. (Eds.), Discriminatiemonitorniet-westerse allochtonen op de arbeidsmarkt 2007 [Monitor on discrimination of non-western migrants on the labor market] (pp.101-141). Den Haag, Netherlands: Sociaal en Cultureel Planbureau.
16. Andriessen I. 2010. Ontwikkelingen in kansenongelijkheid: Meritocratisering, conjunctuur of sociale acceptatie? [Meritocracy, economic cycle or social acceptation?]. In E. Nievers, I. Andriessen, L. Faulk, & J. Dagevos (Eds.), *Discriminatiemonitor niet-westerse migranten op de arbeidsmarkt 2010* [Monitor on discrimination of non-western migrants on the labor market] (pp. 30-51). Den Haag, Netherlands: Sociaal en Cultureel Planbureau.
17. Andriessen I, Nievers E, Dagevos J, and Faulk L. 2012. Ethnic discrimination in the Dutch labor market: Its relationship with job characteristics and multiple group membership. *Work and Occupations*, *39*(3), pp.237-269.
18. Angelucci M, and De Giorgi G. 2009. Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption? *American Economic Review.* 99
19. Artavia-Mora L, Wagner N, Thiombiano BA and Bedi A. 2020. Adaptation and biomedical transition of people living with HIV to antiretroviral treatment in Burkina Faso. *Global Public Health*, pp.1-16.
20. Auld AF, Shiraishi RW, Mbofana F, Couto A, Fetogang EB, El-Halabi S, Lebedonyane R, Tlhagise P, and Colebunders R. 2015. Lower levels of antiretroviral therapy enrollment among men with HIV compared with women: 12 countries, 2002-2013. Morbidity and mortality weekly report/USA: Department of Health and Human Sciences. Centers for Disease Control. 64(46):1281-303.
21. Auspurg K, Hinz T, and Schmid L. 2017. Contexts and Conditions of Ethnic Discrimination: Evidence from a Field Experiment in a German Housing Market. *Journal of Housing Economics*, *35*, pp.26-36.

22. Axelrod R and Hamilton WD. 2981. The evolution of cooperation. *Science*. Mar 27;211(4489):1390-6
23. Axelrod R. The Evolution of Cooperation. 1984. Basic Books
24. Balafoutas L and Nikiforakis N. 2012. Norm enforcement in the city: A natural field experiment. *European Economic Review*. Nov 1;56(8):1773-85.
25. Balafoutas L, Nikiforakis N, Rockenbach B. 2014. Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences*. Nov 11;111(45):15924-7.
26. Balafoutas L, Grechenig K, Nikiforakis N. 2014. Third-party punishment and counter-punishment in one-shot interactions. *Economics letters*. Feb 1;122(2):308-10.
27. Balafoutas L, Nikiforakis N, Rockenbach B. 2016. Altruistic punishment does not increase with the severity of norm violations in the field. *Nature Communications*. Nov 1;7(1):1-6.
28. Baldini M, and Federici M. 2011. Ethnic discrimination in the Italian rental housing market. *Journal of Housing Economics*, *20(1)*, pp.1-14.
29. Baranov V, Bennett D, and Kohler HP. 2015. The indirect impact of antiretroviral therapy: Mortality risk, mental health, and HIV-negative labor supply. *Journal of Health Economics*, 44, 195-211.
30. Bartoš V, Bauer M, Chytilová J, and Matějka F. 2016. Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition. *American Economic Review*, *106(6)*, 1437-1475.
31. Barr A, Dekker M, Janssens W, Kebede B, and Kramer B. 2019. "Cooperation in Polygynous Households." *American Economic Journal: Applied Economics*, 11 (2): 266-83.
32. Barr A, Dekker M, Mwansa F, and Zuze TL. 2020. Financial decision-making, gender and social norms in Zambia: preliminary report on the quantitative data generation, analysis and results. *CeDEx discussion paper series*. ASC Working Paper 147 / 2020.
33. Beaman L, and Magruder J. 2012. Who gets the job referral? Evidence from a social networks experiment. *American Economic Review*, 102(7), pp.3574-93.
34. Bear A, and Rand DG. 2016. Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences*. Jan 26;113(4):936-41.
35. Beaulieu M, Adrien A, Potvin L, and Dassa C. 2014. Stigmatizing attitudes towards people living with HIV/AIDS: Validation of a measurement scale. *BMC Public Health*, *14*(1), 1246.
36. Becker G. 1957. The Economics of Discrimination. University of Chicago Press, Chicago
37. Benjamins MR, Hummer RA, Eberstein IW, and Nam CB. 2004. Self-reported health and adult mortality risk: an analysis of cause-specific mortality. *Social Science and Medicine*, 59(6), 1297-1306.
38. Bengtsson R, Iverman E, and Hinnerich BT. 2012. Gender and ethnic discrimination in the rental housing market. *Applied Economics Letters*, *19*(1), pp.1-5.
39. Bicchieri C. 2016. *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.
40. Blommaert L, Coenders M, and van Tubergen F. 2014. Ethnic discrimination in recruitment and decision makers' features: Evidence from laboratory experiment and survey data using a student sample. *Social Indicators Research*, *116*(3), pp.731-754.
41. Bonnet F, Lalé E, Safi M, and Wasmer E. 2016. Better residential than ethnic discrimination! Reconciling audit and interview findings in the Parisian housing market. *Urban Studies*, *53*(13), pp.2815-2833.
42. Boring A, and Philippe A. 2017. *Reducing discrimination through norms or information: Evidence from a field experiment on student evaluations of teaching* (No. 17-72). Institute for Advanced Study in Toulouse (IAST).
43. Bosch M, Carnero MA, and Farre L. 2010. Information and discrimination in the rental housing market: Evidence from a field experiment. *Regional Science and Urban Economics*, *40*(1), pp.11-19.
44. Bovenkerk F, Gras MJ, and Ramsoedh D. 1995. *Discrimination against migrant workers and ethnic minorities in access to employment in the Netherlands* (Vol. 4). Employment Department, Internat. Labour Office
45. Brevoort KP. 2011. Credit card redlining revisited. *Review of Economics and Statistics*, 93(2), pp.714-724.
46. Brickman P, Coates D, and Janoff-Bulman R. 1978. Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, 36(8), 917.
47. Bouwmeester S, Verkoeijen PP, Aczel B, Barbosa F, Bègue L, Brañas-Garza P, Chmura TG, Cornelissen G, Døssing FS, Espín AM, and Evans AM. 2017. Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, 12(3), pp.527-542.
48. Bowles S and Gintis H. 2013. *A Cooperative Species*. Princeton University Press.
49. Buck N. 2001. Identifying neighborhood effects on social exclusion. *Urban Studies*, 38, 2251–2275.

50. Burger C, Burger R and van Doorslaer E. 2019. The Health Impact of Free Access to Antiretroviral Therapy in South Africa. *Available at SSRN 3506209*.
51. Burkina Faso Ministry of Health. Burkina Faso Ministry of Health. 2016. Standards Protocols of Medical Care for People Living With HIV URL: http://www.remed.org/recommandations_officielles_du_ministere_de_la_sante_nov2008.pdf [accessed 2016-03-25] [WebCite Cache]
52. Bursztyn L, Fiorin S, Gottlieb D, and Kanz M. 2019. Moral Incentives in Credit Card Debt Repayment: Evidence from a Field Experiment. *Journal of Political Economy* 127, no. 4: 1641-1683.
53. Büyükbozkoyum Ö, Stamatiou M, and Stolk M. 1991. Turkse HTS'ers zoeken werk. Verslag van een sollicitatie-experiment. *Sociologische Gids*, *38*(3), pp.187-192.
54. Cadsby CB, Du N, and Song F. 2016. In-group favoritism and moral decision-making. *Journal of Economic Behavior & Organization*, *128*, pp.59-71.
55. Cahill K and Perera R. Competitions and incentives for smoking cessation. *Cochrane Database of Systematic Reviews*. 2011(4).
56. Capraro V. 2013. A model of human cooperation in social dilemmas. *PloS ONE*. 8(8).
57. Capraro V and Cococcioni G. 2016. Rethinking spontaneous giving: Extreme time pressure and ego-depletion favor self-regarding reactions. *Scientific Reports*. Jun 2;6:27219.
58. Capital Value. 2019. Housing shortage in The Netherlands rises to 263,000 dwellings. Capital Value Residential Real Estate Services. Retrieved from: https://www.capitalvalue.nl/en/news/housing-shortage-in-the-netherlands-rises-to-263000-dwellings
59. Carlsson M and Eriksson S. 2015. Ethnic discrimination in the London market for shared housing. *Journal of ethnic and migration studies*, *41*(8), pp.1276-1301.
60. Carlsson M and Eriksson S. 2014. Discrimination in the rental market for apartments. *Journal of Housing Economics*, *23*, pp.41-54
61. Carpenter C J. 2010. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. *Health Communication.* 25(8): 661–669.
62. CBS. 2016. Population growth fueled by immigration. Centraal Bureau voor de Statistiek NL. Reviewed: November 1st, 2016. Accessible at: https://www.cbs.nl/en-gb/news/2016/04/population-growth-fuelled-by-immigration
63. CBS. 2020. General statistics on The Netherlands'. Centraal Bureau voor de Statistiek NL. Accessible at: https://opendata.cbs.nl/statline/#/CBS/en/navigatieScherm/thema?themaNr=5350
64. Central IA. 2016. The World Factbook. Telephones - Mobile Cellular - Country Comparison to the World. URL: https://www.cia.gov/library/publications/the-world-factbook/fields/2151.html [accessed 2016-08-11] [WebCite Cache]
65. Chaiyachati KH, Ogbuoji O, Price M, Suthar AB, Negussie EK, Bärnighausen T. 2014. Interventions to improve adherence to antiretroviral therapy: a rapid systematic review. *AIDS;* Suppl 2:S187-S204.
66. Chan B, and Tsai A. 2015. Trends in HIV-related stigma in the general population during the era of antiretroviral treatment expansion: An analysis of 31 Sub-Saharan African Countries.. *Infectious Diseases Society of America*; Vol. 2, No. suppl_1, p. 404
67. Charness G and Gneezy U. 2009. Incentives to exercise. *Econometrica*. 77(3):909-931.
68. Chen Y and Li SX. 2009. Group identity and social preferences. *American Economic Review*, *99*(1), pp.431-57.
69. Chesney MA and Smith AW. 1999. Critical delays in HIV testing and care. *American Behavioral Scientist.*42(7):1162–74.
70. Chesney MA. 2006. The elusive gold standard. Future perspectives for HIV adherence assessment and intervention. *Journal of Acquire Immune Deficiency Syndrom*e; Suppl 1:S149-S155.
71. Chkhartishvili N, Rukhadze N, Svanidze M, Sharvadze L, Dehovitz JA, Tsertsvadze T, et al. 2014. Evaluation of multiple measures of antiretroviral adherence in the Eastern European country of Georgia. *Journal of International AIDS Society.* 17:18885
72. Chuah SH, Hoffmann R, Ramasamy B, and Tan JH. 2014. Religion, ethnicity and cooperation: An experimental study. *Journal of Economic Psychology*, *45*, pp.33-43.
73. CL-IST. 2005. Strategic Framework to Combat HIV / AIDS and STIs (CSLS) 2006-2010. National Council against AIDS and STIs. Ouagadougou: CNLS-IST; p.101 [Visited on the 15/04/2014]. Online: https://hivhealthclearinghouse.unesco.org/library/documents/cadre-strategique-de-lutte-contre-le-vihsida-et-les-ist-2006-2010

74. Clingingsmith D, Khwaja AI, and Kremer M. 2009. Estimating the impact of the Hajj: religion and tolerance in Islam's global gathering. *Quarterly Journal of Economics*. 124(3):1133-1170.
75. CNLC-IST. 2016. Plan national multisectoriel de lutte contre le VIH. Conseil national de lutte contre le SIDA et les IST. Burkina Faso. URL: http://hivhealthclearinghouse.unesco.org/library/documents/plan-national-multisectoriel-de-lutte-contre-le-vih-le-sida-et-les-ist-pnm-annee-0 [accessed 2016-03-29] [WebCite Cache]
76. Cohen-Cole E. 2011. Credit card redlining. *Review of Economics and Statistics*, 93(2), pp.700-713.
77. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. 2011. Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine*; 365:493-505
78. Cone J and Rand DG. 2014. Time pressure increases cooperation in competitively framed social dilemmas. *PLoS ONE*. 9(12).
79. Coomes CM, Lewis MA, Uhrig JD, Furberg RD, Harris JL, and Bann CM. 2012. Beyond reminders: a conceptual framework for using short message service to promote prevention and improve healthcare quality and clinical outcomes for people living with HIV. *AIDS Care*. 24:348-357.
80. Craig GM, Daftary A, Engel N, O'Driscoll S, and Ioannaki A. 2017. Tuberculosis stigma as a social determinant of health: a systematic mapping review of research in low incidence countries. *International Journal of Infectious Diseases, 56*, 90-100.
81. Currie C, Zanotti C, Morgan A, Currie D, De Looze M, Roberts C, Smith OR and Barnekow V. 2009. Social determinants of health and well-being among young people. Health Behaviour in School-aged Children (HBSC) study: international report from the 2009/2010 Survey: World Health Organization; 2009;2010:271. ISBN 978 92 890 1423 6.
82. Curtis E, Comiskey C, and Dempsey O. 2015. Correlational research: Importance and use in nursing and health research. *Nurse Researcher*, 6(1), pp.20-25
83. Cutler DM and Glaeser EL. 1997. Are Ghettos good or bad? *Quarterly Journal of Economics*, 112, 827–872 .
84. Da Costa TM, Peres BJ, Gomes de Costa DA, Sigulem D, de Fátima MH, Filho AC, et al. 2012. Results of a randomized controlled trial to assess the effects of a mobile SMS-based intervention on treatment adherence in HIV/AIDS-infected Brazilian women and impressions and satisfaction with respect to incoming messages. *International Journal of Medical Informatics*. 81:257-269
85. Da Costa TM, Barbosa BJP, Gomes e Costa DA, Sigulem D, de Fátima Marin H, Castelo Filho A, and Pisa IT. 2012. Results of a randomized controlled trial to assess the effects of a mobile SMS-based intervention on treatment adherence in HIV/AIDS-infected Brazilian women and impressions and satisfaction with respect to incoming messages. *International Journal of Medical Informatics*. 81(4): 257-269.
86. Dagevos J. 2001. Rapportage minderheden 2001: Meer werk [Ethnic minorities report 2001: More jobs]. Den Haag, Netherlands: Sociaal en Cultureel Planbureau.
87. Dagevos J and Gijsberts M. 2007. Jaarrapport integratie 2007. *The Hague: Sociaal en Cultureel Planbureau SCP-publicatie, 27*.
88. Dahlui M, Azahar N, Bulgiba A, Zaki R, Oche OM, Adekunjo FO, and Chinna K. 2015. HIV/AIDS related stigma and discrimination against PLWHA in Nigerian population. *PloS One, 10*(12), p.e0143749.
89. DellaVigna S and Gentzkow M. 2010. Persuasion: Empirical Evidence. *Annual Review of Economics*. 2(1):643-669.
90. Dawson-Amoah CG. 2015. Determinants of HIV stigma among healthcare workers in Ghana. Walden University, Walden University ScholarWorks.
91. de Groot C and Vrieselaar N. 2019. Housing shortage continues to drive up house prices. RaboResearch - Economic Research. Dutch Housing Market Quarterly. Review article at: https://economics.rabobank.com/publications/2019/november/housing-shortage-continues-to-drive-up-house-prices/
92. Deacon H, Uys L and Mohlahlane R. 2009. HIV and stigma in South Africa. In *HIV/AIDS in South Africa 25 years on* (pp. 105-120). Springer, New York, NY.
93. Demena B, Artavia-Mora L, Ouedraogo D, Thiombiano BA and Wagner N. 2020. A systematic review of mobile phone interventions (SMS/IVR/calls) to improve adherence and retention to antiretroviral treatment in low and middle-income countries. *AIDS Patient Care and STDS*, 34(2), pp. 59-71.
94. Derous E. 2011. Geen baan voor een Marokkaan? Discriminatie bij cv-screening nader bekeken [No job for a Moroccan? Discrimination in screening job resumes]. *Gedrag and Organisatie,* 24(2), 126-137.

95. Derous E, Nguyen H, and Ryan A. 2009. Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance*, 22, 297-320.
96. Di Tella R, MacCulloch RJ, and Oswald AJ. 2001. Preferences over inflation and unemployment: Evidence from surveys of happiness. *American Economic Review*, 91(1), 335-341.
97. Di Tella R, MacCulloch RJ, and Oswald AJ. 2001. Preferences over inflation and unemployment: Evidence from surveys of happiness. *American Economic Review*, 91(1), 335-341.
98. Dijker AJ, Koomen W, and Kok G. 1997. Interpersonal determinants of fear of people with AIDS: The moderating role of predictable behavior. *Basic and Applied Social Psychology*, *19*(1), pp.61-79.
99. Dlamini PS, Wantland D, Makoae LN, et al. 2009. HIV stigma and missed medications in HIV-positive people in five African countries. *AIDS Patient Care and STDS*. 23(5):377–87.
100. Doiron D, Fiebig DG, Johar M, and Suziedelyte A. 2015. Does self-assessed health measure health?. *Applied Economics*, 47(2), 180-194.
101. Dolan P and Metcalfe R. 2012. Valuing health: a brief report on subjective well-being versus preferences. *Medical Decision Making*, 32(4), 578-582.
102. Dolan P and Kahneman D. 2008. Interpretations of utility and their implications for the valuation of health. *The Economic Journal*, 118(525), 215-234.
103. Dolan P, Peasgood T, and White M. 2008. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, 29(1), 94-122.
104. Drake AL, Unger JA, Ronen K, Matemo D, Perrier T, DeRenzi B, Richardson BA, Kinuthia J, and John-Stewart G. 2017. Evaluation of mHealth strategies to optimize adherence and efficacy of Option B+ prevention of mother-to-child HIV transmission: Rationale, design and methods of a 3-armed randomized controlled trial. *Contemporary Clinical Trials*. *57*:44-50.
105. Drydakis N. 2011. Ethnic discrimination in the Greek housing market. *Journal of Population Economics*, 24(4), pp.1235-1255.
106. Earnshaw VA, Smith LR, Chaudoir SR, Amico KR, and Copenhaver MM. 2013. HIV stigma mechanisms and well-being among PLWH: a test of the HIV stigma framework. *AIDS Behavior*.17:1785–95.
107. Eagly AH and Chaiken S. 1984. Cognitive theories of persuasion. *Advances in Experimental Social Psychology*. 17:267-359.
108. Eaton JW, Johnson LF, Salomon JA, Bärnighausen T, Bendavid E, Bershteyn A, et al. 2012. HIV treatment as prevention: systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa. *PLoS Medicine.* 9:e1001245.
109. Egrot M. 2007. Renaître d'une mort sociale annoncée: recomposition du lien social des personnes vivant avec le VIH en Afrique de l'Ouest (Burkina Faso, Sénégal). *Cultures et Sociétés*, (1), pp.49-56.
110. Elibol R and Tielbeke J. 2018. Investigation Discrimination on the housing market: Rachid is also just a nice boy. De Groene Amsterdammer. Retrieved from: https://www.groene.nl/artikel/rachid-is-ook-gewoon-een-nette-jongen
111. Ellis RP, Fiebig DG, Johar M, Jones G and Savage E. 2013. Explaining health care expenditure variation: Large-sample evidence using linked survey and health administrative data. *Health Economics*, 22(9), 1093-1110.
112. Elul B, Lamb MR, Lahuerta M, Abacassamo F, Ahoua L, Kujawski SA, Tomo M, and Jani I. 2017. A combination intervention strategy to improve linkage to and retention in HIV care following diagnosis in Mozambique: A cluster-randomized study. *PLOS Medicine*. *14*(11): e1002433.
113. Evans AM, Dillon KD, and Rand DG. 2015. Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General*. Oct;144(5):951.
114. Evans JS and Stanovich KE. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science.* May;8(3):223-41.
115. Fang AH, Guess AM and Humphreys M. 2014. Messaging Strategies to Combat Housing Discrimination in New York City: Evidence from a Field Experiment. Evaluation Report 2.0. Columbia University for the New York City Commission on Human Rights. Fair Housing 2014.
116. Farmer MM and Ferraro KF. 1997. Distress and perceived health: mechanisms of health decline. *Journal of Health and Social Behavior*, 298-311.
117. Fehr E and Rangel A. 2011. Neuroeconomic Foundations of Economic Choice--Recent Advances. *Journal of Economic Perspectives.* Dec;25(4):3-0.
118. Fehr E and Gächter S. 2002. Altruistic punishment in humans. *Nature*. Jan;415(6868):137-40.

119. Fehr E and Fischbacher U. 2003. The nature of human altruism. *Nature*. Oct;425(6960):785-91.
120. Fehr E and Fischbacher U. 2004. Third-party punishment and social norms. *Evolution and Human Behavior*. Mar 1;25(2):63-87.
121. Finitsis DJ, Pellowski JA, and Johnson BT. 2014. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. *PLoS ONE*. 9:e88166
122. Fox MP and Rosen S. 2010. Patient retention in antiretroviral therapy programs up to three years on treatment in sub-Saharan Africa, 2007-2009: systematic review. *Tropical Medicine and International Health.* Suppl 1:1-15.
123. FRA. 2017. Second European Union Minorities and Discrimination Survey: Muslim- Selected Findings. European Union Agency for Fundamental Rights. EU-MIDIS II. Accessible at: http://fra.europa.eu/en/publication/2017/eumidis-ii-muslims-selected-findings
124. Frederick S and Loewenstein G. 1999. Hedonic Adaptation. Well-Being. The foundations of Hedonic Psychology/Eds. D. Kahneman, E. Diener, N. Schwarz. NY: Russell Sage, 302-329.
125. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. 2013. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. *PLoS Medicine*. 10:e1001363
126. Frey BS and Stutzer A. 2002. What can economists learn from happiness research? *Journal of Economic Literature*, *40*(2), pp.402-435.
127. Fox MP and Rosen S. 2010. Patient retention in antiretroviral therapy programs up to three years on treatment in sub-Saharan Africa, 2007-2009: systematic review. *Tropical Medicine and International Health*. Jun;Suppl 1:1-15
128. Frankish K and Evans JS. 2009. The duality of mind: An historical perspective. In two minds: Dual processes and beyond. In J Evans and K Frankish (eds.) In Two Minds: Dual Processes and Beyond, Oxford University Press, pp.1-29
129. Funda, 2020. About Funda. 2000 - 2020 funda - Powered by NVM. Website available at: https://content.funda.nl/over/funda/
130. Gächter S. 2012. A cooperative instinct. *Nature*. Sep;489(7416):374-5.
131. Gächter S, Renner E, and Sefton M. 2008. The long-run benefits of punishment. *Science*. Dec 5;322(5907):1510.
132. Gausset Q, Mogensen HO, Yameogo WME, Berthé A, and Konaté B. 2012. The ambivalence of stigma and the double-edged sword of HIV/AIDS intervention in Burkina Faso. *Social Science and Medicine.*74(7) :1037-44.
133. Geldof M, Thiombiano BA, and Wagner N. 2020. When I receive the SMS, it is a sign of love: Symbolic Connotations of SMS messages for People living with HIV in Burkina Faso. *AIDS Care. 1-8*.
134. Genberg BL, Hlavka Z, Konda KA, Maman S, Chariyalertsak S, Chingono A, et al. 2009. A comparison of HIV/AIDS-related stigma in four countries: negative attitudes and perceived acts of discrimination towards people living with HIV/AIDS. *Social Science and Medicine*, 68, 2279e2287.
135. Gill CJ, Hamer DH, Simon JL, Thea DM, and Sabin LL. 2005. No room for complacency about adherence to antiretroviral therapy in sub-Saharan Africa. *AIDS*. 19:1243-1249.
136. Gine X and Ghazala M. 2012. Together We Will Experimental Evidence on Female Voting Behavior in Pakistan. URL: http://ssrn.com/abstract=1871588 [accessed 2016-08-11] [WebCite Cache]
137. Global Digital Health Network2016. mHealth Working Group Inventory of Projects. URL: https://www.mhealthworkinggroup.org/projects/mhealth-working-group-inventory-projects [accessed 2016-03-29] [WebCite Cache]
138. Goffman E. 1963. Stigma: Notes on the Management of Spoiled Identities. Prentice-Hall, Englewood Cliffs, NJ.
139. Govindasamy D, Meghij J, Kebede NE, Clare BR, Ford N, and Kranzer K. 2014. Interventions to improve or facilitate linkage to or retention in pre-ART (HIV) care and initiation of ART in low- and middle-income settings--a systematic review. *Journal of International AIDS Society*. 17:19032
140. Granich R, Kahn JG, Bennett R, Holmes CB, Garg N, Serenata C, et al. 2012. Expanding ART for treatment and prevention of HIV in South Africa: estimated cost and cost-effectiveness 2011-2050. *PLoS ONE*. 7:e30216
141. Grimm V, Mengel F. 2011. Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters.* May 1;111(2):113-5.

142. Gruenais ME and Ouattara F. 2008. “Patients like the others" or how caregivers treat AIDS patients and the issue of stigmatization in Burkina Faso. *Science and Technology, Health Sciences*, Special issue no. 1: 93-100.
143. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, and Car J. 2013. Mobile phone messaging reminders for attendance at healthcare appointments. *Cochrane Database Systematic Review*. 12:CD007458.
144. Guy R, Hocking J, Wand H, Stott S, Ali H, and Kaldor J. 2012. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. *Health Service Research.* 47:614-632
145. Haberer JE, Musiimenta A, Atukunda EC, Musinguzi N, Wyatt MA, Ware NC, and Bangsberg DR. 2016. Short message service (SMS) reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. *AIDS*. *30*(8):1295-1300.
146. Hagendoorn L, Hraba J and Hagendoorn R. 1989. The ethnic hierarchy in the Netherlands: Social distance and social representation. *British Journal of Social Psychology*, *28*(1), pp.57-69.
147. Haile KK, Nillesen E and Tirivayi N. 2020. Impact of formal climate risk transfer mechanisms on risk-aversion: Empirical evidence from rural Ethiopia. *World Development*, 130, p.104930.
148. Hamilton WD. 1964. The genetical evolution of social behavior. II. *Journal of Theoretical Biology*. Jul 1;7(1):17-52.
149. Hanson A and Hawley Z. 2011. Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *Journal of Urban Economics*, *70*(2-3), pp.99-114.
150. Hatzenbuehler ML, Phelan JC and Link BG. 2013. Stigma as a fundamental cause of population health inequalities. *American Journal of Public Health*, *103*(5), pp.813-821.
151. Heijnders M and Van Der Meij S. 2006. The fight against stigma: an overview of stigma reduction strategies and interventions. *Psychological Health Medicine*;11:353–63.
152. Helliwell JF. 2006. Well-being, social capital and public policy: what's new? *The Economic Journal*, 116(510), C34-C45.
153. Herek GM. 1999. ‘AIDS and Stigma’, *American Behavioral Scientist*, 42(7), pp. 1106–1116.
154. Horvath T, Azman H, Kennedy GE, and Rutherford GW. 2012. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. *Cochrane Database Systematic Review.* 3:CD009756.
155. Houtzager LM. 2009. Nutrition in HIV: a review. *Benin Journal of Postgraduate Medicine*, 11(1).
156. Hughto JMW, Reisner SL and Pachankis JE. 2015. Transgender stigma and health: A critical review of stigma determinants, mechanisms, and interventions. *Social Science & Medicine*, *147*, 222-231.
157. Institute of Medicine. 2011. Preparing for the Future of HIV/AIDS in Africa: A Shared Responsibility. Washington, DC: The National Academies Press.
158. INSD. 2011. Enquête démographique et de santé et à indicateurs multiples (EDSBF-MICS IV), Rapport préliminaire, Burkina Faso, 2010. *Ouagadougou: INSD*, *40*. Online: https://www.unicef.org/bfa/french/bf_eds_2010.pdf#xd_co_f=NzA1MjZlNTMtNWU2OS00NzI0LWJkNGUtZjI5N2E1NWQ0MWMz~
159. Janz NK, and Becker MH. 1984. The Health Belief Model: A Decade Later. *Health Education Quarterly*. *11*(1):1–47.
160. John LK, Loewenstein G, Troxel AB, Norton L, Fassbender JE, and Volpp KG. 2011. Financial incentives for extended weight loss: a randomized, controlled trial. *Journal of General Internal Medicine*. 26(6):621-626.
161. Jongbloed K, Parmar S, van der Kop M, Spittal PM, and Lester RT. 2015. Recent evidence for emerging digital technologies to support global HIV engagement in care. *Current HIV/AIDS Reports*. 12:451-461.
162. Jordan J, Peysakhovich A, and Rand DG. 2015. Why We Cooperate. In The moral brain: A multidisciplinary perspective. MIT Press, Cambridge
163. Joseph DD, Nhavoto JA, Augusto O, Ponce W, Traca D, Nguimfack A, and de Sousa Palha C. 2016. SMSaúde: Evaluating Mobile Phone Text Reminders to Improve Retention in HIV Care for Patients on Antiretroviral Therapy in Mozambique. *Journal of Acquired Immune-Deficiency Syndromes*. *73*(2): e23-30.
164. Jylhä M. 2009. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science and Medicine*, 69(3), 307-316.
165. Kaas L and Manger C. 2012. Ethnic discrimination in Germany's labour market: a field experiment. *German Economic Review*, *13*(1), pp.1-20.

166. Kahneman D. 2003. A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*. Sep; 58(9):697.
167. Kahneman D. 2012. Thinking, fast and slow. Penguin, London, United Kingdom
168. Kahneman D and Tversky A. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(2), 263.
169. Kalichman SC, Carey MP, and Johnson BT. 1996. Prevention of sexually transmitted HIV infection: A meta-analytic review of the behavioral outcome literature. *Annals of Behavioral Medicine*, 18(1), 6-15.
170. Kalichman SC and Simbayi LC. 2003. HIV testing attitudes, AIDS stigma, and voluntary HIV counselling and testing in a black township in Cape Town, South Africa. *Sexually Transmitted Infections*.79(6):442–7.
171. Kazianga H and Wahhaj Z. 2020. Will urban migrants formally insure their rural relatives? Family networks and rainfall index insurance in Burkina Faso. *World Development*, *128*, p.104764.
172. Kazianga H, De Walque D and Alderman H. 2012. Educational and child labour impacts of two food-for-education schemes: Evidence from a randomized trial in rural Burkina Faso. *Journal of African Economies*, *21*(5), pp.723-760.
173. Koelmeyer R, English DR, Smith A, and Grierson J. 2014. Association of social determinants of health with self-rated health among Australian gay and bisexual men living with HIV. *AIDS Care*, 26(1), 65-74.
174. Koethe JR, Lukusa A, Giganti MJ, Chi BH, Nyirenda CK, Limbada MI, Banda Y, and Stringer JS. 2010. Association between weight gain and clinical outcomes among malnourished adults initiating antiretroviral therapy in Lusaka, Zambia. *Journal of Acquired Immune Deficiency Syndromes*; 53(4), 507.
175. Koomen LEM, Burger R and van Doorslaer E. 2019. Effects and determinants of tuberculosis drug stockouts in South Africa. *BMC Health Services Research*, *19*(1), p.213.
176. Krajbich I, Bartling B, Hare T, and Fehr E. 2015. Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*. Jul 2;6(1):1-9.
177. Kralik D. 2002. The quest for ordinariness: transition experienced by midlife women living with chronic illness. *Journal of Advanced Nursing*, 39(2), 146-154.
178. Kralik D, Koch T, Price K, and Howard N. 2004. Chronic illness self-management: taking action to create order. *Journal of Clinical Nursing*, 13(2), 259-267.
179. Kranzer K, Govindasamy D, Ford N, Johnston V, and Lawn SD. 2012. Quantifying and addressing losses along the continuum of care for people living with HIV infection in sub-Saharan Africa: a systematic review. *Journal of International AIDS Society*. 15:17383
180. Kruisbergen EW and Veld T. 2002. Een gekleurd beeld. Over beelden, beoordeling en selectie van jonge allochtone werknemers [A coloured image. Study about images, evaluations and selection of young migrant employees]. Assen, Netherlands: Koninklijke Van Gorcum.
181. Kunutsor S, Walley J, Katabira E, Muchuro S, Balidawa H, Namagala E, et al. 2010. Using mobile phones to improve clinic attendance amongst an antiretroviral treatment cohort in rural Uganda: a cross-sectional and prospective study. *AIDS Behavior.* 14:1347-1352.
182. Ky-Zerbo O, Desclaux A, Somé JF, El Asmar K, Msellati P and Obermeyer CM. 2014. La stigmatisation des PVVIH en Afrique: analyse de ses formes et manifestations au Burkina Faso. *Santé Publique*, *26*(3), pp.375-384.
183. Lee RS, Kochman A and Sikkema KJ. 2002. Experienced stigma among people living with HIV-AIDS. *AIDS and Behavior*, *6*(4), pp.309-319.
184. L'Engle KL, Green K, Succop SM, Laar A, and Wambugu S. 2015. Scaled-Up mobile phone intervention for HIV care and treatment: Protocol for a facility randomized controlled trial. *JMIR Research Protocols*. *4*(1):e11.
185. Lester RT, Gelmon L, and Plummer FA. 2006. Cell phones: tightening the communication gap in resource-limited antiretroviral programmes? *AIDS.* 20:2242-2244.
186. Lester R and Karanja S. 2008. Mobile phones: exceptional tools for HIV/AIDS, health, and crisis management. *The Lancet Infectious Diseases*. 8:738-739.
187. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. 2010. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomized trial. *Lancet;* 376:1838-1845.
188. Li J, Mo PK, Wu AM and Lau JT. 2017. Roles of self-stigma, social support, and positive and negative effects as determinants of depressive symptoms among HIV infected men who have sex with men in China. *AIDS and Behavior*, *21*(1), 261-273.

189. Linnemayr S, Huang H, Luoto J, Kambugu A, Thirumurthy H, Haberer JE, Wagner G, and Mukasa B. 2017. Text messaging for improving antiretroviral therapy adherence: no effects after 1 year in a randomized controlled trial among adolescents and young adults. *American Journal of Public Health. 107*(12):1944-1950.
190. List JA. 2004. The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics*, *119*(1), pp.49-89.
191. List JA. 2007. Field experiments: a bridge between lab and naturally occurring data. *The B.E. Journal of Economic Analysis & Policy*. Mar 22;5(2).
192. List JA. 2011. Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives*. Sep;25(3):3-16.
193. Logie C and Gadalla TM. 2009. Meta-analysis of health and demographic correlates of stigma towards people living with HIV. *AIDS Care* 21: 742–753.
194. Loewenstein G and O'Donoghue T. 2004. Animal spirits: Affective and deliberative processes in economic behavior. SSRN 539843.
195. Loewenstein G and Ubel PA. 2008. Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics*. 92(8-9): 1795-810.
196. Logie CH, Wang Y, Lacombe-Duncan A, Wagner AC, Kaida A, Conway T, Webster K, de Pokomandy A and Loutfy MR. 2018. HIV-related stigma, racial discrimination, and gender discrimination: Pathways to physical and mental health-related quality of life among a national cohort of women living with HIV. *Preventive Medicine*, *107*, pp.36-44.
197. Loubiere S, Meiners C, Sloan C, Freedberg KA, and Yazdanpanah Y. 2010. Economic evaluation of ART in resource-limited countries. *Current Opinion HIV AIDS*. 5:225-231
198. Lucas RE, Clark AE, Georgellis Y and Diener, E. 2003. Reexamining adaptation and the set point model of happiness: reactions to changes in marital status. *Journal of Personality and Social Psychology, 84(3),* 527.
199. Loury LD. 2006. Some contacts are more equal than others: Informal networks, job tenure, and wages. *Journal of Labor Economics*, 24(2), pp.299-318
200. Lyons A, Heywood W and Rozbroj T. 2016. Psychosocial factors associated with resilience in a national community-based cohort of Australian gay men living with HIV. *AIDS and Behavior*, *20*(8), pp.1658-1666.
201. Ma HI, Saint-Hilaire M, Thomas CA and Tickle-Degnen L. 2016. Stigma as a key determinant of health-related quality of life in Parkinson's disease. *Quality of Life Research*, *25*(12), 3037-3045.
202. Macé S and Le Lec F. 2011. On fatalistic long-term health behavior. *Journal of Economic Psychology*, 32(3), 434-439.
203. Madhvani N, Longinetti E, Santacatterina M, Forsberg BC, El-Khatib Z. 2015. Correlates of mobile phone use in HIV care: Results from a cross-sectional study in South Africa. *Preventive Medicine Reports.* 2:512-516.
204. Mahajan AP, Sayles JN, Patel VA, et al. 2008. Stigma in the HIV/AIDS epidemic: a review of the literature and recommendations for the way forward. *AIDS*. 22(Suppl 2):S67–79.
205. Mak WW, Cheung RY, Law RW, Woo J, Li PC and Chung RW. 2007. Examining attribution model of self-stigma on social support and subjective well-being among people with HIV+/AIDS. *Social Science & Medicine*, *64*(8), pp.1549-1559.
206. Malaba LC, Ward BJ, and Nathoo KJ. 2006. Mortality and morbidity among postpartum HIV-positive and HIV-negative women in Zimbabwe: risk factors, causes, and impact of single-dose postpartum vitamin A supplementation. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 43(1), 107-116.
207. Masa R, Chowa G, and Nyirenda V. 2018. Socioeconomic correlates of dietary diversity and its association with adherence and psychosocial functioning of people living with HIV in rural Zambia. *Nutrition and Health*, 0260106018761282.
208. Maughan-Brown B. 2010. Stigma rises despite antiretroviral roll-out: a longitudinal analysis in South-Africa. *Social Science and Medicine*, 70, 368e374.
209. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester RT, Mills EJ, Smieja M, et al. 2012. The Cameroon Mobile Phone SMS (CAMPS) trial: a randomized trial of text messaging versus usual care for adherence to antiretroviral therapy. *PLoS ONE.* 7:e46909
210. Meessen B, Hercot D, Noirhomme M, Ridde V, Tibouti A, Tashobya CK, et al. 2011. Removing user fees in the health sector: a review of policy processes in six sub-Saharan African countries. *Health Policy Plan*; Suppl 2:ii16-ii29.
211. Meier S and Sprenger CD. 2015. Temporal stability of time preferences. *Review of Economics and Statistics*, 97(2), 273-286.

212. Merkel AL and Lohse J. 2019. Is fairness intuitive? An experiment accounting for subjective utility differences under time pressure. *Experimental Economics*, 22(1), pp.24-50.
213. Miguel E and Kremer M. 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica.* 72:159-217
214. Mills EJ, Nachega JB, Buchan I, Orbinski J, Attaran A, Singh S, et al. 2006. Adherence to antiretroviral therapy in sub-Saharan Africa and North America: a meta-analysis. *JAMA.* 296:679-690.
215. Mohler BJ, Thompson WB, Creem-Regehr SH, Pick HL, and Warren WH. 2007. Visual flow influences gait transition speed and preferred walking speed. *Experimental Brain Research*. Aug 1;181(2):221-8.
216. Montgomery JD. 1991. Social networks and labor-market outcomes: Toward an economic analysis. The *American Economic Review*, 81(5), pp.1408-1418.
217. Myrseth KO and Wollbrant CE. 2016. Commentary: Fairness is intuitive. *Frontiers in Psychology*, 7, p.654.
218. Mugavero MJ, Westfall AO, Zinski A, Davila J, Drainoni ML, Gardner LI, and Retention in Care (RIC) Study Group. 2012. Measuring retention in HIV care: the elusive gold standard. *Journal of Acquire Immune Deficiency Syndromes.* 61:574-580
219. Mulcahy K and Kollamparambil U. 2016. The impact of rural-urban migration on subjective well-being in South Africa. *The Journal of Development Studies*, 52(9), 1357-1371.
220. Nam SL, Fielding K, Avalos A, Dickinson D, Gaolathe T, and Geissler PW. 2008. The relationship of acceptance or denial of HIV-status to antiretroviral adherence among adult HIV patients in urban Botswana. *Social Science and Medicine*, 67(2), 301-310.
221. Neuman M, Obermeyer CM, and MATCH Study Group. 2013. Experiences of stigma, discrimination, care and support among people living with HIV: a four country study. *AIDS and Behavior*, 17(5), 1796-1808.
222. Neo WS, Yu M, Weber RA, and Gonzalez C. 2013. The effects of time delay in reciprocity games. *Journal of Economic Psychology.* Feb 1;34:20-35.
223. Nielsen UH, Tyran JR, and Wengström E. 2014. Second thoughts on free riding. *Economics Letters*. Feb 1;122(2):136-9.
224. Nievers E. 2010. Personeelsselecteurs over niet-westerse migranten en discriminatie [Personeelsselecteurs over niet-westerse migranten en discriminatie]. In Nievers, E., Andriessen, I. (Eds.), Discriminatiemonitor niet-westerse migranten op de arbeidsmarkt 2010 [Monitor on discrimination of non-western migrants on the labor market] (pp. 52-85). Den Haag, Netherlands: Sociaal en Cultureel Planbureau.
225. Nikiema, PR. 2019. Impact of school feeding programmes on educational outcomes: Evidence from Burkina Faso. *Journal of African Economies*, 28 (3), 323-34.
226. Nikiforakis N. 2008. Punishment and counter-punishment in public good games: Can we really govern ourselves?. *Journal of Public Economics*. Feb 1;92(1-2):91-112.
227. Nillesen EE, Grimm M, Goedhuys M, Reitmann AK, and Meysonnat A. 2019. On the Malleability of Implicit Attitudes Towards Women Empowerment: Evidence from Tunisia. *IZA Discussion Paper No. 12471*
228. Nixon SA, Hanass-Hancock J, Whiteside A, and Barnett T. 2011. The increasing chronicity of HIV in sub-Saharan Africa: Re-thinking" HIV as a long-wave event" in the era of widespread access to ART. *Globalization and Health*, 7(1), 41.
229. Nyblade L, Pande R, Mathur S, MacQuarrie K, Kidd R, et al. 2003. Disentangling HIV and AIDS stigma in Ethiopia, Tanzania and Zambia. Washington, DC: International Center for Research on Women. p53.
230. Nyblade L. 2006. Measuring HIV stigma: existing knowledge and gaps. *Psychology, Health and Medicine*, 11, 335e345
231. Nobel-Media AB, 2020. The Sveriges Riksank Prize in Economic Sciences in Memory of Alfred Nobel. The Nobel Prize. Access at: https://www.nobelprize.org/prizes/economic-sciences
232. Nowak MA, May RM, and Sigmund K. 1995. The arithmetic of mutual help. *Scientific American*. Jun 1;272(6):76-81.
233. Nowak MA and Sigmund K. 1998. Evolution of indirect reciprocity by image scoring. *Nature*. Jun;393(6685):573-7.
234. Nowak MA. 2006. Five rules for the evolution of cooperation. *Science*. Dec 8;314(5805):1560-3.
235. NU.NL. 2014. Meerderheid vindt aantal allochtonen in Nederland 'te hoog'. Sanoma Digital The Netherlands B.V. NU. Accessible at: http://www.nu.nl/politiek/3928656/meerderheid-vindt-aantal-allochtonen-in-nederland-te-hoog.html?redirect=1.
236. Öblom A and Antfolk J. 2017. Ethnic and gender discrimination in the private rental housing market in Finland: A field experiment. *PloS ONE*, *12*(8), p.e0183344.

237. Orrell C, Cohen K, Mauff K, Bangsberg DR, Maartens G, and Wood R. 2015. A randomized controlled trial of real-time electronic adherence monitoring with text message dosing reminders in people starting first-line antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes*. *70*(5): 495-502.
238. Oster E. 2019. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), pp.187-204.
239. Oswald AJ and Powdthavee N. 2008. Does happiness adapt? A longitudinal study of disability with implications for economists and judges. *Journal of Public Economics*, 92(5-6), 1061-1077.
240. Ouattara F, ME Gruenais, Huygens P, and Traore A. 2004. Sida, stigma and exclusion. Anthropological study in Burkina Faso (Bobo- Dioulasso, Banfora, Niangoloko, Orodara). Paris: National Agency for AIDS Research; Marseille: Institute for Research for Development; Bobo-Dioulasso (Burkina Faso): SHADEI; Dec. ; 127 p. [Visited on the 15/04/2014]. Online: http://www.sfls.aei.fr/ckfinder/userfiles/files/Actus/doc/2006/vih_stigmat_exlu_burkina.pdf .
241. Oyugi JH, Byakika-Tusiime J, Charlebois ED, Kityo C, Mugerwa R, Mugyenyi P, et al. 2004. Multiple validated measures of adherence indicate high levels of adherence to generic HIV antiretroviral therapy in a resource-limited setting. *Journal of Acquire Immune Deficiency Syndrome*;36:1100-1102.
242. Pararius, 2020. About Pararius - rental properties in Holland. 2006 - 2020 Pararius B.V. Website available at: https://www.pararius.com/info/about-pararius
243. Park CL and Folkman S. 1997. Meaning in the context of stress and coping. *Review of General Psychology*, 1(2), 115.
244. Peysakhovich A, Nowak MA, and Rand DG. 2014. Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications*. Sep 16;5(1):1-8.
245. Peysakhovich A and Rand DG. 2016. Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*. Mar;62(3):631-47.
246. Pierret J. 2007. An analysis over time (1990–2000) of the experiences of living with HIV. *Social Science and Medicine*, 65(8), 1595-1605.
247. Pieter J. 2018. Netherlands housing shortage to last through 2020: Report. NL times. Retrieved from: https://nltimes.nl/2018/02/05/netherlands-housing-shortage-last-2020-report
248. Piovesan M and Wengström E. 2009. Fast or fair? A study of response times. *Economics Letters*. Nov 1;105(2):193-6.
249. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de WD, et al. 2011. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. *AIDS*. 25:825-834
250. Postulart D and Adang EM. 2000. Response shift and adaptation in chronically III patients. *Medical Decision Making*, 20(2), 186-193.
251. Pruitt SL, Leonard T, Murdoch J, Hughes A, McQueen A, and Gupta S. 2014. Neighborhood effects in a behavioral randomized controlled trial. *Health Place*. 30:293-300
252. Prus SG. 2011. Comparing social determinants of self-rated health across the United States and Canada. *Social Science and Medicine*, 73(1), 50-59.
253. Quinn TC and Overbaugh J. 2005. HIV/AIDS in women: an expanding epidemic. *Science*. 308(5728): 1582-3.
254. Rand DG, Greene JD, and Nowak MA. 2012. Spontaneous giving and calculated greed. *Nature*. Sep;489(7416):427-30.
255. Rand DG, Greene JD, and Nowak MA. 2013. Rand et al. reply. *Nature*. Jun;498(7452):E2-3.
256. Rand DG and Nowak MA. 2013. Human cooperation. *Trends in Cognitive Sciences*. Aug 1;17(8):413-25.
257. Rand DG, Peysakhovich A, Kraft-Todd GT, Newman GE, Wurzbacher O, Nowak MA, and Greene JD. 2014. Social heuristics shape intuitive cooperation. *Nature Communications*. Apr 22;5(1):1-2.
258. Rand DG and Epstein ZG. 2014. Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PloS ONE*. 9(10).
259. Rand DG and Kraft-Todd GT. 2014. Reflection does not undermine self-interested prosociality. *Frontiers in Behavioral Neuroscience*. Sep 3;8:300.
260. Rand DG. 2016. Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*. Sep;27(9):1192-206.
261. Rankin WW, Brennan S, Schell E, Laviwa J, and Rankin SH. 2005. The stigma of being HIV-positive in Africa. *PLoS Medicine*, 2(8), e247.

262. Rasschaert F, Koole O, Zachariah R, Lynen L, Manzi M, and Van Damme W. 2012. Short and long term retention in antiretroviral care in health facilities in rural Malawi and Zimbabwe. *BMC Health Services Research.* 12:444
263. Rawat R, Kadiyala S, and McNamara PE. 2010. The impact of food assistance on weight gain and disease progression among HIV-infected individuals accessing AIDS care and treatment services in Uganda. *BMC Public Health*, 10(1), 316.
264. Recalde MP, Riedl A, and Vesterlund L. 2014. Error prone inference from response time: The case of intuitive generosity. CESifo Working Paper Series, No. 4987
265. Resource Tracking Working Group. 2018. HIV Prevention Research & Development Investments, 2017 - Investing to End the Epidemic. Retrieved at: https://www.avac.org/sites/default/files/resource-files/HIV_resourceTracking2017.pdf. Last accessed: June 26, 2020.
266. Rhodes T, Bernays S, and Terzić KJ. 2009. Medical promise and the recalibration of expectation: hope and HIV treatment engagement in a transitional setting. *Social Science and Medicine*, 68(6), 1050-1059.
267. Riach P and Rich J. 2002. Field experiments of discrimination in the marketplace. *The Economic Journal, 112*, pp.480-518.
268. Rieger M and Mata R. 2015. On the generality of age differences in social and nonsocial decision making. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. Mar 1;70(2):200-12.
269. Riis J, Loewenstein G, Baron J, Jepson C, Fagerlin A, and Ubel PA. 2005. Ignorance of hedonic adaptation to hemodialysis: a study using ecological momentary assessment. *Journal of Experimental Psychology*: General, 134(1), 3.
270. Rintamaki LS, Davis TC, Skripkauskas S, Bennett CL, and Wolf MS. 2006. Social stigma concerns and HIV medication adherence. *AIDS Patient Care and STDs*, *20*(5), 359-368.
271. Robins S. 2005. Rights passages from" near death" to" new life": AIDS activism and treatment testimonies in South Africa. IDS Working Paper 251. Brighton: Institute of Development Studies.
272. Rosen S, Fox MP, and Gill CJ. 2007. Patient retention in antiretroviral therapy programs in sub-Saharan Africa: a systematic review. *PLoS Medicine*. 4(10): e298.
273. Rubinstein A. 2007. Instinctive and cognitive reasoning: A study of response times. *The Economic Journal*. Oct;117(523):1243-59.
274. Ruggeri, K. 2018. Behavioral Insights for Public Policy: Concepts and Cases. Routledge.
275. Russell S and Seeley J. 2010. The transition to living with HIV as a chronic condition in rural Uganda: Working to create order and control when on antiretroviral therapy. *Social Science and Medicine*, 70(3), 375-382.
276. Russell S, Seeley J, Ezati E, Wamai N, Were W and Bunnell R. 2007. Coming back from the dead: living with HIV as a chronic condition in rural Africa. *Health Policy and Planning*, 22(5), 344-347.
277. Sacherová K. 2016. How discriminatory is the housing market in Slovakia: experimental investigation? Charles University. Digital Repository.
278. Sadana R, Mathers CD, Lopez AD, Murray CJ, and Iburg K. 2002. Comparative analyses of more than 50 household surveys on health status. Summary measures of population health: concepts, ethics, measurement, and applications. Geneva: World Health Organization. Jan 1:369-86.
279. Sanjuán P, Molero F, Fuster MJ, and Nouvilas E. 2013. Coping with HIV related stigma and well-being. *Journal of Happiness Studies*, *14*(2), pp.709-722.
280. Scanlon ML and Vreeman RC. 2013. Current strategies for improving access and adherence to antiretroviral therapies in resource-limited settings. *HIV AIDS.* 5:1-17
281. Schwarz N. 1999. Self-reports: how the questions shape the answers. *American Psychologist*. Feb;54(2):93.
282. Schwartz CE and Sprangers, M.A. 1999. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Social Science and Medicine*, 48(11), 1531-1548.
283. Schwartz CE and Sprangers MA. 2000. Adaptation to changing health: Response shift in quality-of-life research. In American Psychological Association. Washington, DC.
284. Schwartz CE, Andresen EM, Nosek MA, Krahn GL, and RRTC Expert Panel on Health Status Measurement. 2007. Response shift theory: important implications for measuring quality of life in people with disability. *Archives of Physical Medicine and Rehabilitation*, 88(4), 529-536.
285. Sharpe L and Curran L. 2006. Understanding the process of adjustment to illness. *Social Science and Medicine*, 62(5), 1153-1166.
286. Sharpe L, Butow P, Smith C, McConnell D, and Clarke S. 2005. Changes in quality of life in patients with advanced cancer: evidence of response shift and response restriction. *Journal of Psychosomatic Research*, 58(6), 497-504.

287. Shet A, Costa AD, Kumarasamy N, Rodrigues R, Rewari BB, Ashorn P, et al. 2014. Effect of mobile telephone reminders on treatment outcome in HIV: evidence from a randomized controlled trial in India. *BMJ*. 347:g5978-g5978.
288. Sigmund K. 2010. The calculus of selfishness. Princeton University Press. Jan 4.
289. Sinha SR, and Barry M. 2011. Health technologies and innovation in the global health arena. *New England Journal of Medicine. 365:779–782.*
290. Simons GF and Fennig CD (eds.). 2018. *Ethnologue: Languages of the World, Twenty-first edition.* Dallas, Texas: SIL International. Online version: www.ethnologue.com.
291. Skogar Ö, Fall PA, Hallgren G, Bringer B, Carlsson M, Lennartsson U, Sandbjörk H, Törnhage CJ, and Lökk J. 2012. Parkinson's disease patients' subjective descriptions of characteristics of chronic pain, sleeping patterns and health-related quality of life. *Neuropsychiatric Disease and Treatment*. 8: 435.
292. Sloman SA. 1996. The empirical case for two systems of reasoning. *Psychological Bulletin*. Jan;119(1):3.
293. SOB. 2017. Adult Obesity in the United States [Internet]. The Trust for America's Health and the Robert Wood Johnson Foundation. The State of Obesity; 2004-2018 (Updated August 2017).
294. Sprangers MA and Schwartz CE. 1999. Integrating response shift into health-related quality of life research: a theoretical model. *Social Science and Medicine*, 48(11), 1507-1515.
295. Steward WT, Herek GM, Ramakrishna J, Bharat S, Chandy S, Wrubel J, and Ekstrand ML. 2008. HIV-related stigma: adapting a theoretical framework for use in India. *Social Science & Medicine*, 67(8), pp.1225-1235.
296. Stringer KL, Marotta P, Baker E, Turan B, Kempf MC, Drentea P, Stepanikova I, and Turan JM. 2019. Substance Use Stigma and Antiretroviral Therapy Adherence Among a Drug-Using Population Living with HIV. *AIDS Patient Care and STDs*, *33*(6), pp.282-293.
297. Sutter M, Kocher M, and Strauß S. 2003. Bargaining under time pressure in an experimental ultimatum game. *Economics Letters*. Dec 1;81(3):341-7.
298. Tajfel H and Billic M. Familiarity and categorization in intergroup behavior. Journal of experimental social psychology. 1974 Mar 1;10(2):159-70.
299. Tessier P, Blanchin M, and Sébille V. 2017. Does the relationship between health-related quality of life and subjective well-being change over time? An exploratory study among breast cancer patients. *Social Science and Medicine*, 174, 96-103.
300. The Meerstens Institute. 2017. Database obtained from The Meerstens Institute. Web page: https://www.knaw.nl/en/institutes/meertens-instituut.
301. Thompson SC and Janigian AS. 1988. Life schemes: A framework for understanding the search for meaning. *Journal of Social and Clinical Psychology*, 7(2-3), 260-280.
302. Thompson B, Diamond KE, McWilliam R, Snyder P, and Snyder SW. 2005. Evaluating the Quality of Evidence from Correlational Research for Evidence-Based Practice. *Exceptional Children* 71 (2), 181-194.
303. Thornton RL. 2012. HIV testing, subjective beliefs, and economic behavior. *Journal of Development Economics*, 99(2), 300-313.
304. Tinghög G, Andersson D, Bonn C, Böttiger H, Josephson C, Lundgren G, Västfjäll D, Kirchler M, and Johannesson M. 2013. Intuition and cooperation reconsidered. *Nature*. Jun;498(7452):E1-2.
305. Trivers RL. The evolution of reciprocal altruism. 1971. *The Quarterly Review of Biology*. Mar 1;46(1):35-57.
306. Tsai AC and Venkataramani AS. 2015. The causal effect of education on HIV stigma in Uganda: evidence from a natural experiment. *Social Science & Medicine*, *142*, pp.37-46.
307. UNAIDS. 2002. Situational analysis of discrimination and stigma against people living with HIV / AIDS in West and Central Africa: ethical and legal considerations. Abidjan: UNAIDS for West and Central Africa. Inter-Country Team for West and Central Africa, African Network of People Living with HIV. p.42. http://data.unaids.org/pub/report/2002/20020701_analysis_discrimination_en.pdf
308. UNAIDS. 2005. *Resource needs for an expanded response to AIDS in low-and middle-income countries.* The Joint United Nations Programme on HIV/AIDS. Geneva, Switzerland.
309. UNAIDS. 2012. Rapport d'activités sur la riposte au sida du Burkina Faso 2012. The Joint United Nations Programme on HIV/AIDS.
310. UNAIDS. 2015. Joint United Nations Programme on HIV/AIDS (UNAIDS). Global AIDS Response Progress Reporting (GARPR) URL: http://www.unaids.org/sites/default/files/country/documents/BFA_narrative_report_2015.pdf.Accessed: [accessed 2016-08-11] [WebCite Cache]

311. UNAIDS. 2012. Joint United Nations Programme on HIV/AIDS (UNAIDS). 2012. Global AIDS Response Progress Reporting. URL: http://www.unaids.org/sites/default/files/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/ce_BF_Narrative_Report[1].pdf [accessed 2016-03-29] [WebCite Cache]
312. UNAIDS (a). 2016. Rapport d'activités sur la riposte au sida du Burkina Faso 2016. The Joint United Nations Programme on HIV/AIDS
313. UNAIDS (b). 2016. Country fact sheets. The Joint United Nations Programme on HIV/AIDS.
314. UNAIDS. 2019. Country: Burkina Faso. Retrieved at: https://www.unaids.org/en/regionscountries/countries/burkinafaso
315. UNDP. 2019. Human Development Indices and Indicators: 2018 Statistical Update: Burkina Faso. United National Development Programme, Briefing note for countries on the 2018 Statistical Update. Retrieved at: http://hdr.undp.org/sites/all/themes/hdr_theme/country-notes/BFA.pdf
316. United Nations. 2020. Sustainable Development Goals: Goal 10: Reduce inequality within and among countries. United Nations. Access at: https://www.un.org/sustainabledevelopment/inequality/
317. Urban S. 2009. Is the neighborhood effect an economic or an immigrant issue? A study of the importance of the childhood neighborhood for future integration into the labour market. *Urban Studies*, 46, pp.583–603.
318. Van der Bracht K, Coenen A, and Van de Putte B. 2015. The Not-in-My-property syndrome: The occurrence of ethnic discrimination in the rental housing market in Belgium. *Journal of Ethnic and Migration Studies*, *41*(1), pp.158-175.
319. Van Veelen M, García J, Rand DG, and Nowak MA. 2012. Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences.* Jun 19;109(25):9929-34.
320. Verkoeijen PP and Bouwmeester S. 2014. Does intuition cause cooperation?. *PloS ONE*. 9(5).
321. Verkuyten M and Zaremba K. 2005. Interethnic relations in a changing political context. *Social Psychology Quarterly*, *68*(4), pp.375-386.
322. Vie TL, Hufthammer KO, Holmen TL, Meland E, and Breidablik HJ. 2014. Is self-rated health a stable and predictive factor for allostatic load in early adulthood? Findings from the Nord Trøndelag Health Study (HUNT). *Social Science and Medicine*, 117, 1-9.
323. Visser MJ, Makin JD, Vandormael A, Sikkema RJ, and Forsyth BWC. 2009. HIV/AIDS stigma in a South African community. *AIDS Care*, 21(2), 197-206.
324. Volpp KG, Troxel AB, Pauly MV, Glick HA, Puig A, Asch DA, Galvin R, Zhu J, Wan F, DeGuzman J, and Corbett E. 2009. A randomized, controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*. 360:699-709.
325. Voors MJ, Nillesen EE, Verwimp P, Bulte EH, Lensink R, and Van Soest DP. 2012. Violent conflict and behavior: a field experiment in Burundi. *American Economic Review*, *102*(2), pp.941-64.
326. Wagner N, Ouedraogo D, Artavia-Mora L, Bedi A, and Thiombiano BA. 2016. Protocol for a randomized controlled trial evaluating mobile text messaging to promote retention and adherence to antiretroviral therapy for people living with HIV in Burkina Faso. *JMIR Research Protocols*, 5(3).
327. Wagstaff A, van Doorslaer E, and Burger R. 2019. SMS nudges as a tool to reduce tuberculosis treatment delay and pretreatment loss to follow-up. A randomized controlled trial. *PloS ONE*, *14*(6), p.e0218527.
328. Wang H, Zhou J, He G, Luo Y, Li X, Yang A, Fennie K, and Williams AB. 2009. Consistent ART adherence is associated with improved quality of Life, CD4 counts, and reduced hospital costs in central China. *AIDS Research and Human Retroviruses*. 25(8):757-63.
329. Walshe K. 2007. Understanding what works—and why—in quality improvement: the need for theory-driven evaluation. *International Journal for Quality in Health Care*. Apr 1;19(2):57-9.
330. Wang H, Zhou J, He G, Luo Y, Li X, Yang A, et al. 2009. Consistent ART adherence is associated with improved quality of Life, CD4 counts, and reduced hospital costs in central China. *AIDS Resources Human Retroviruses.* 25:757-763.
331. Wang CS, Sivanathan N, Narayanan J, Ganegoda DB, Bauer M, Bodenhausen GV, Murnighan K. 2011. Retribution and emotional regulation: The effects of time delay in angry economic interactions. *Organizational Behavior and Human Decision Processes*. Sep 1;116(1):46-54.
332. Weichselbaume D. 2016. Discrimination against female migrants wearing headscarves. *SSRN Discussion Paper*.
333. Will G. 2003. RAXEN 4 Report to the European Monitoring Centre on Racism and Xenophobia (EUMC). European Forum for Migration Studies, Bamberg

334. WHO. 2002. WHOQOL-HIV instrument: The 120 QUESTIONS WITH RESPONSE SCALES & 38 IMPORTANCE ITEMS. Mental Health: Evidence and Research Department of Mental and Substance Dependence.
335. WHO. 2013. Consolidated guidelines on general HIV care and the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach (p. 269). Geneva: World Health Organization. ISBN 978 92 4 150572 7.
336. WHO. 2011. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzeland: World Health Organization.
337. WHO. 2007. Toward universal access: scaling up HIV/AIDS interventions in the health sector. *Geneva: World Health Organization.*
338. WHO. 2016. *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach*. World Health Organization. ISBN 978 92 4 154968 4.
339. Winskell K, Hill E, and Obyerodhyambo O. 2011. Comparing HIV-related symbolic stigma in six African countries: Social representations in young people's narratives. *Social Science and Medicine*, *73*(8), pp.1257-1265.
340. World Bank. 2014. LSMS Dataset Finder. Living Standard Measurement Study.
341. World Factbook. 2014. Burkina Faso. Washington, DC: Central Intelligence Agency.
342. World Factbook. 2018. Burkina Faso. Washington, DC: Central Intelligence Agency.
343. Yan H, Li X, Li J, Wang W, Yang Y, Yao X, Yang N, and Li S. 2019. Association between perceived HIV stigma, social support, resilience, self-esteem, and depressive symptoms among HIV-positive men who have sex with men (MSM) in Nanjing, China. *AIDS Care*, pp.1-8.
344. Wolfe WR, Weiser SD, Bangsberg DR, Thior I, Makhema JM, Dickinson DB, et al. 2006. Effects of HIV-related stigma among an early sample of patients receiving antiretroviral therapy in Botswana. *AIDS Care*, 18(8), 931e933.
345. Zorlu A and Hartog J. 2001. Migration and immigrants: the case of the Netherlands (No. 01-042/3). Tinbergen Institute Discussion Paper. No. 01-042/3

Curriculum vitae (short version)

Luis Artavia-Mora, a citizen from Costa Rica and the United States of America, obtained a BA degree with honours in Economics at the National University of Costa Rica in 2012, and a MA degree at Erasmus University of Rotterdam in 2015. He has published in economic and health journals, and he combines six years of academic and professional experience developing projects in international development. His interests cover the application of behavioural and experimental instruments to improve development goals across countries. He has worked in small-scale research on prosociality and social norms, and a large-scale intervention in Burkina Faso sending mobile messages to improve bio-physical markers and psychosocial health in patients of HIV with help from Dr. Natascha Wagner and Professor Arjun Bedi. During his doctoral studies, he represented the PhD community at the University's Council, and lead the design and teaching of various post-graduate courses in human behaviour and mathematics.

PhD Portfolio

LUIS ARTAVIA-MORA

Rotterdam, The Netherlands
(+31) 644450230 ⋄ lartavia90@gmail.com
Citizenships: United States of America and Costa Rica

EDUCATION

- **PhD Candidate in Behavioral and Development Economics** *Jan-2017 to Present*
  Erasmus University of Rotterdam
  - *Dissertation title:* 'Understanding and Overcoming Biases Against Marginalized Groups: Behavioral and Experimental Evidence from The Netherlands and Burkina Faso'
- **Master in Development Economics (distinction)** *Aug-2014 to 2015*
  Erasmus University of Rotterdam
  - *Major in econometric analysis of development policies*
- **Bachelors in Economics (distinction)** *2008 to 2012*
  National University of Costa Rica

HONORS

- Master degree and thesis (distinction) *Dec-2015*
- MA Scholarship Fund for Excellent Students *Aug-2014*
- BA Honor Scholarships *2009, 2011, 2012*

PUBLICATIONS AND WORK IN PROGRESS

- Wagner N, et al. Durability of mobile message reminders to improve retention, adherence and psychosocial wellbeing in PLHIV: A 2-year large-scale study in Burkina Faso. Work in progress.
- Artavia-Mora L. Averting housing discrimination? Evidence from Craigslist and reference letters in Amsterdam. Work in progress.
- Artavia-Mora L, et al. Stigma as social determinant of health in a large-scale and panel data sample of PLHIV in Burkina Faso. Work in progress.
- Demena B, et al. 2020. A Systematic Review of Mobile Phone Interventions (SMS/IVR/calls) to improve Adherence and Retention to Antiretroviral Treatment in Low and Middle-Income Countries. *AIDS Patient Care and STDS.*
- Artavia-Mora L, et al. 2020. Adaptation and Biomedical Transition of People Living with HIV to Antiretroviral Treatment in Burkina Faso. *Global Public Health*
- Artavia-Mora L, et al. 2017. Intuitive Help and Punishment in the Field. *European Economic Review* (92) 133–145
- Wagner N, et al. 2016. Protocol for a randomized controlled trial evaluating mobile short message services (SMS) to promote retention and adherence to anti-retroviral therapy for people living with human immunodeficiency virus (HIV) in Burkina Faso. *Journal of Medical Internet Research: Research Protocols* 5(3): e170

SKILLS

Softwares	STATA, R and Phyton (basic), SPSS, Overleaf-Latex, Qualtrics, M. Project (basic)
Languages	Spanish (Native), English (Fluent)

PROFESSIONAL EXPERIENCE

- **PhD Candidate** (eur.nl) *Jan-2016 to Present*
  Behavioral experiments on social norms in The Netherlands, and on the impact of mobile reminders to improve the well-being of PLHIV in Burkina Faso. Supervisory fieldwork visit for data collection surveillance in October 2017.
- **Consulting research analyst** (sanigest.com) *2013 to 2014*
  Development of projects and proposals related to health, poverty and education for public and private institutions in India, Kazakhstan, Moldova, Mexico, El Salvador and Bahamas among others.
- **Research assistant at CRn State of The Nation Report** (estadonacion.or.cr) *2012*
  Research and development of specific sections of the report. The sections included: the behavior of the housing market, the Collective Bonus Program and advancement in minimum wages.
- **Researcher at the Institute for Sustainability Policies** (ips.or.cr) *2011*
  -*Study*: "Financial Accessibility Research: Santa Cruz-Cartagena, Costa Rica" The project evaluated the main factors causing socioeconomic disparities between the two underdeveloped regions of Costa Rica. The tasks involved designing a methodological instrument (i.e. surveys) and to organize and conduct fieldwork.
- **Manager assistant at Industrias POLYAM** (family business) *2008 to 2012*

TEACHING

- Erasmus University of Rotterdam
  - STATA: Practical Applications for Statistical and Data Management *2019*
  - Refresher on Mathematics for Economists *2018-2020*
  - Human Behavior and Experiments in Development *2017-2019*
  - Foundations of Economic Development *2015*

CERTIFICATES

- Machine Learning A-Z: Hands-On Python and R in Data Science (DST) *Expected, October-2020*
- Behavioral Economics & Consumer Psychology in Marketing (MWA) *August-2020*
- Programming in R: Introduction to Data Science and Graphs (ITAM) *May-2020*
- Behavioral science for better public policies (IADB) *April-2020*
- Human Relations, Effective Communications and Interpersonal Competence at (DC) *Dec-2011*
- Etiquette and Protocol (NUCR) *July-2011*

LEAD POSITIONS

- PhD Representative at Association and University Council (EUR) *2018-2020*
- Master Class Representative (EUR) *2014-2016*