What are promising directions for organizing income support? This dissertation compiles four empirical studies analyzing new and existing policies in three different income support schemes. Chapter 2 uses a field experiment to study the effects of giving social assistance claimants a greater say in reemployment decisions, for example, by exempting them from job search and other activity-related requirements. Chapter 3, using the same experiment, investigates the employment effects of allowing claimants to keep a larger fraction of income earned on top of benefits. Chapter 4 examines the effects of benefit sanctions and warnings in unemployment insurance, exploiting the quasi-random assignment of claimants to caseworkers, which systematically vary in their tendency to impose a certain measure. Finally, Chapter 5 uses a field experiment to study the labor supply effects of generous and unconditional cash support offered to economically vulnerable households.

Timo Verlaat (1989) holds a Bachelor's degree from Zeppelin University in Friedrichshafen, Germany, and a Research Master degree (cum laude) from Utrecht University School of Economics, the Netherlands, both in Economics. He carried out his PhD research at Utrecht University School of Economics from 2016 to 2021 and is currently working as a researcher at the Netherlands Bureau for Economic Policy Analysis (CPB).

ISBN 978-90-393-7476-4

Timo Verlaat **Carrot and Stick:** Experiments With Social Welfare Policies

U.S.E. 061

Utrecht University

School of Economics

Carrot and Stick:
Experiments With Social Welfare Policies

Timo Verlaat

U.S.E. 061

Carrot and Stick:
Experiments With Social Welfare Policies

Timo Verlaat

Manuscript committee: Prof. dr. P.T. de Beer
Prof. dr. M. Belot
Prof. dr. M. Goos
Prof. dr. J.J. de Laat
Prof. dr. A.M. Salomons

ISBN: 978-90-393-7476-4
U.S.E. Dissertation Series
USE 061
Printed by Ridderprint, www.ridderprint.nl

Carrot and Stick:
Experiments With Social Welfare Policies

Wortel en stok. Experimenten met sociale zekerheidsbeleid

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus,
prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 17 juni 2022 des middags te 2.15 uur

door

Timo Lee Luis Verlaat

geboren op 2 februari 1989
te Fürth, Duitsland

Promotor:	Prof. dr. S. Rosenkranz
Copromotor:	Dr. L.F.M. Groot

This dissertation was made possible with the financial support from a Research Talent grant (no. 406.16.538) from the Dutch Research Council (NWO).

Contents

Acknowledgements

My PhD journey started at a lecture on multidisciplinary macroeconomics in the fall semester of 2015, when Professor Mark Sanders had to leave our class 20 minutes early due to a meeting at the municipal office about "an interesting new project on unconditional welfare benefits." I remember sending an email the next day asking him if the project still needed a student assistant. Five days later, I was working my way through the first project proposal.

A little more than six years later, this journey has come to an end, and I am sincerely thankful for the support and advice that I have received on the way to finishing this dissertation. First, I would like to express my gratitude to my supervisors, Professor Stephanie Rosenkranz and Professor Loek Groot. Both jumped on board without any hesitation when I decided to apply for a PhD bursary of the Dutch Research Council. Stephanie and Loek, you both believed in me to lead this research project. I thank you for your encouragement, your guidance, and for always putting me first when there was the opportunity to share our research with policymakers and scholars abroad. I am also grateful to Professor Mark Sanders, not only for his advice and feedback on my work, but also for giving me the opportunity to join the project in the first place and for being the first to speak the words: "You should think about going for a PhD, because this project will need one."

I am very grateful to the members of the manuscript committee, Professor Paul de Beer, Professor Michèle Belot, Professor Maarten Goos, Professor Joost de Laat, and Professor Anna Salomons. Thank you for devoting some of your valuable time to reading my dissertation and providing feedback on my work. I am also thankful to the Dutch Research Council (NWO) for their generous funding, and to the European Social Fund (ESF), the program *Handhaving en Gedrag*, the Dutch Ministry of Social Affairs and Employment, and Utrecht University School of Economics for financially supporting different projects and research activities during my PhD.

Not far into my PhD, the "interesting new project on unconditional welfare benefits," later baptized *Weten wat werkt* (in English: What Works), grew into a major endeavor. To my great fortune, I was surrounded by a group of motivated and persistent people inside and outside of the university. I am very grateful to everyone at the Municipality of Utrecht who believed in the project and helped take it to the finish line. I counted more than 30 people in the municipal office who were involved at some point. I am afraid that I might miss someone once I start listing names, but I thank each and every one of you! Not to forget, I am also very grateful to all the participants of *Weten wat werkt*.

At the university, I could rely on the skilled project assistance of Katja van Dien. Katja, thank you for backing me up, offering crucial advice, and never losing your optimism. Marcel de Kruijk provided excellent research assistance. Marcel, you were the most independent and thorough RA one could wish for. I am also grateful to our data manager, Frans de Liagre Böhl, who helped us maneuver endless rounds of privacy impact assessments, and to our student assistant, Justine Miller, whose accuracy was exactly what we needed during the final stretch. Thanks also to Anne Kool, Myrthe Jansen, and Stijn Houweling for enriching our final policy report with qualitative insights gathered for their Master theses; and the university's Public Engagement Program for their (financial) support during various post-experiment activities. Last but not least, thanks again to Stephanie Rosenkranz, Loek Groot, and Mark Sanders for running this project with me.

While setting up our field experiment in Utrecht, similar experiments with welfare benefits were in the making in six other Dutch cities (Amsterdam, Deventer, Groningen, Nijmegen, Tilburg, and Wageningen). I want to thank the researchers involved—Paul de Beer, János Betkó, Kirsten Blom-Stam, Sandra Bos, Arjen Edzes, Peter Gramberg, Jac van der Klink, Ruud Muffels, Richard Rijnks, Jack de Swart, and Viktor Venhorst—for sharing their expertise and ideas. Special thanks go to a scientific supervisory committee installed by the Ministry of Social Affairs and Employment, which consisted of Lex Borghans, Lex Burdorf, Menno Fenger, Willemijn van Gastel, Jaap de Graaf, Saskia Keuzenkamp, and Els Sol, and to the Netherlands Bureau for Economic Policy Analysis (CPB), particularly Jonneke Bolhaar, Egbert Jongen, and Alice Zulkarnain. The comments and feedback provided by this group of people have proven invaluable.

Not long into my PhD, I was lucky to meet Arne Meeldijk, who then worked at the Dutch Unemployment Insurance Agency (UWV). Arne soon became an excellent partner for ping-ponging research ideas, one of which ended up filling Chapter 4 of this dissertation. I owe Arne a thank you for many reasons but, above all, for opening the door to doing research at and with UWV. Besides Arne, some other people were crucial to the above-mentioned chapter. I am thankful to the scientific supervisory committee of the program *Handhaving en Gedrag* for their sharp eyes and their many helpful suggestions. The committee comprised Miriam Adriaanse, Robert Dur, Sjoerd Goslinga, Amra Mustafić, Jacques Niehof, and Joyce Vliegen. I am also very grateful to UWV for facilitating the project and to the many UWV employees who took the time to answer my questions and requests. Special thanks go to Peter Berkhout, whose advice and data support was essential. Another thank you goes to Elena Fumagalli and Heike Vethaak for providing feedback at different stages of the project. Last but not least, I want to thank Erik Bijleveld from Radboud University for his guidance as a collaborator on this project.

In my third year, I had the great pleasure of joining the Catalan Institute of

Public Policy Evaluation (Ivàlua) in Barcelona as a guest researcher. I am incredibly grateful to Federico Todeschini for being enthusiastic about my unsolicited proposal and to Mireia Climent and Erika Pérez for arranging the collaboration and my visits. At Ivàlua, I received a warm welcome every single time. Many thanks to the Ivàlua team for showing me around and taking me to dinners, yoga classes, and beer tastings. The work at Ivàlua culminated in Chapter 5, and in addition to Federico I also owe a thank you to Xavier Ramos from the Autonomous University of Barcelona, the second collaborator on this project.

When Utrecht University School of Economics (U.S.E.) popped up in my Google search window almost a decade ago, I did not know it would become a place to stay. I spent a wonderful time at U.S.E. both during the Research Master and during my PhD, and many people deserve a word of thanks: my colleagues (that is, support staff and faculty), all the other fantastic PhDs at our department, and last but not least my students. I am grateful for the many things I learned from you and the many laughs we shared at seminars, workshops, lectures, lunches, *borrels*, and PhD dinners—first in person, later online.

To my paranymphs Lucia Rossel and Vincent Schippers, and the rest of office 2.11, Thomas Gomez, Bora Lancee, and Fujin Zhou—where to begin? Sharing an office, a cottage in Zeeland, and a box full of Christmas decorations with you were the best things that could have happened to me. I repeat: Doing a PhD without you would not have been half as fun, and you have enriched this journey in so many ways. I will miss you all dearly! Thanks as well to Catherine Blanchard, who I met on my first day at the university and who soon became a friend and later my PhD sister from another department. Thanks to Ernst-Jan de Bruijn, a valued companion and perfect sparring partner for everything related to field experiments, survey design, and welfare conditionality. And thanks to my friends, before Utrecht and since Utrecht—you all deserve a <3.

My academic journey would not have been possible without the love and support of my family, my parents Susanne and Uwe, my brother Leon, and my grandparents Bärbel, Reinhold, and Garti. Liebe Familie, ich kann mich glücklich schätzen euch zu haben. Danke für eure Unterstützung über all die Jahre; ohne euch stünde ich nicht hier. Hidde, a final and special thank you goes to you. We met soon into the PhD and, what can I say, lucky me! Thank you for all the joy, and laughter, and excitement, and spot-on *taaladviezen* you bring into my life!

Timo Verlaat
Utrecht, April 2022

List of Tables

List of Figures

List of Abbreviations

2SLS	Two-stage least squares
AEA	American Economic Association
ATE	Average treatment effect
ATT	Average treatment effect on the treated
BRR	Benefit reduction rate
CAPI	Computer-assisted personal interviewing
CATI	Computer-assisted telephone interviewing
CAWI	Computer-assisted web interviewing
CBS	*Centraal Bureau voor de Statistiek* (Statistics Netherlands)
CPB	*Centraal Planbureau* (Netherlands Bureau for Economic Policy Analysis)
EITC	Earned Income Tax Credit
EU	European Union
EU-SILC	European Union Statistics on Income and Living Conditions
FE	Fixed effects
GDP	Gross domestic product
GPS	Global positioning system
IRB	Institutional review board
ISCED	International Standard Classification of Education
ITT	Intent-to-treat effect
IV	Instrumental variables
JIVE	Jackknife instrumental variables estimation
LATE	Local average treatment effect
NACE	Statistical Classification of Economic Activities in the European Community
NIT	Negative income tax
NWO	*Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (Dutch Research Council)
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares

OSF	Open Science Foundation
PAP	Pre-analysis plan
PES	Public employment services
PPP	Purchasing power parity
RCT	Randomized controlled trial
REC	Real Economy Currency
RF	Reduced form
RGC	*Renta garantizada de ciudadanía* (Guaranteed citizenship income)
RI	Randomization inference
SD	Standard deviation
SMI	*Suport municipal d'inclusió* (Municipal Inclusion Support Benefit)
UI	Unemployment insurance
UWV	*Uitvoeringsinstituut Werknemersverzekeringen* (Dutch Employee Insurance Agency)
WSNP	*Wet schuldsanering natuurlijke personen* (Natural Persons Debt Restructuring Act)
WTC	Working Tax Credit
WY	Westfall-Young

CHAPTER 1

Introduction

1.1 Motivation

In many countries around the world, citizens can rely on state-organized income support in situations of financial need. Quite often, this support has grown into extensive social welfare systems that offer various forms of targeted aid. Examples are *minimum income schemes*, which pay a subsistence income to alleviate poverty and enable participation in social life, *social insurance schemes*, which offer compensatory support in case of unforeseen income shocks, or *family benefits*, which target the economic stability of families. In many cases, income support is combined with *labor market programs*, which aim to promote job finding as a means to reach economic self-sufficiency. The scale of present-day income support schemes is shown in the government budgets needed to finance them. In 2017, OECD countries on average spent 0.6 percent of their GDP to compensate for unemployment, 0.5 percent on labor market programs, and 0.5 percent on minimum income schemes and other forms of income maintenance (OECD, 2021e). In the Netherlands and Spain, two countries included in this dissertation, expenditures in those three areas totaled 3.1 and 2.5 percent of GDP, respectively.

In the past three decades, income support schemes have come under pressure in several advanced economies. Many schemes underwent significant reforms, most of which aimed at cutting expenditures and making programs more effective in activating the unemployed. In the Netherlands, e.g., a 2015 reform introduced tighter regulations and quid pro quo arrangements for welfare benefits. In Germany, the 2004-2005 labor market reforms, also known as *Hartz* reforms, included a major restructuring of income support for the unemployed. And today, many countries are experiencing a lively public and political debate on how to improve the functioning of different support schemes. Take the Netherlands, a study site in most chapters of this dissertation, as an example. Only recently, in 2019, the country's government commissioned a comprehensive report on the future of the Dutch labor market (the so-called Borstlap Committee report; Commissie Regulering van Werk, 2020). Included in the report are clear-cut recommendations for reforming government support in case of unemployment. Benefit schemes, the report claims, should invest in stimulating the intrinsic motivation of job seekers, offer personal and individualized supervision, and strengthen financial incentives to work in addition to benefits.[1] Furthermore, schemes should maintain activating components, such as financial penalties, and provide a sufficient buffer in situations of need. As will become apparent later, some of these suggestions are subject to the studies included in this dissertation.

There is one aspect specifically that has received much attention in the policy discussions of recent years. This aspect concerns the conditionalities imposed by government support. As recipients should stand on their own feet again as soon as possible, benefit schemes need to balance the positive effects of providing insurance or protection and the potentially adverse incentive effects of that very support. Job seekers may find it optimal, e.g., to put less effort into a strenuous and time-consuming job search and claim unemployment benefits longer than strictly necessary. People receiving a minimum income may refrain from looking for a job or working more hours if benefits are reduced heavily for every additional euro earned.

Policymakers have responded to these challenges by making benefit receipt conditional on fulfilling different activity-related criteria (sometimes referred to as *compliance requirements*), among other things. These criteria may differ between schemes and countries, but they have in common that they aim to enforce behaviors conceivably discouraged by program participation (Venn, 2012). Broadly speaking, criteria may take the form of supportive ("carrots") or restrictive policies ("sticks") (Arni et al., 2022). Examples of the former type of policy are education and training programs, counseling, and job search assistance. Restrictive policies may include targets on effort provision, requirements to accept suitable work, or obligations to engage in unpaid activities. Commonly, adherence to different requirements is monitored, with sanctions following in case of infractions.

Although much of standard economic theory favors compliance requirements as a means to counter disincentives (see Fredriksson and Holmlund, 2006, for a review), there are also arguments calling this approach into question. These arguments stress the potentially adverse side effects of control, which they see as rooted in either economic incentives or psychological mechanisms. Arguments in the former tradition may point to impaired labor market outcomes when claimants accept lower wage offers in response to a financial sanction or drop out of the labor force altogether while trying to avoid supervision. Arguments relying on psychological explanations may stress the benefits of autonomy and self-determined behavior (Deci and Ryan, 1985; Frey and Jegen, 2001), point out reciprocal tendencies (Falk and Fischbacher, 2006), or allude to the additional cognitive tax that control places

[1] Similar advice can be found in reports of other organizations providing policy recommendations (see, e.g., OECD, 2015, 2021c; Wetenschappelijke Raad voor het Regeringsbeleid, 2020).

on financially constrained individuals (Mullainathan and Shafir, 2013).[2]

Perhaps unsurprisingly, questions of conditionality are not just subject to a lively public discussion but also see adoption in pilots, policy experiments, and scaled up programs around the world. In development aid, e.g., direct cash transfers without any strings attached have become an increasingly common instrument to reduce poverty and vulnerabilities (Handa et al., 2018). Experiments with unconditional unemployment benefits in Finland (Hämäläinen et al., 2021), an unconditional basic income in Germany (Bohmann et al., 2021), or a voluntary job guarantee in Austria (Kasy and Lehner, 2021) indicate a growing interest in the effects of reduced conditionality, also in advanced economies.

Despite these examples, there is still much to learn about how reduced conditionality can be incorporated into the design of effective income support. How do people react when provided with a subsistence base that puts no constraints on their behavior? What happens when programs try to stimulate intrinsic motivation among job seekers? Is help and support more effective when giving benefit recipients a greater say in *how* they want to be helped and supported? Are activating components, such as financial sanctions, effective means to change behavior? These are the questions that guide this dissertation and motivate the studies included in it. While it is inevitable that more work will be needed to reach robust conclusions, each chapter aims to expand our understanding of what could be the answers to these questions.

In what follows, I will discuss in more detail the objectives and contributions of this dissertation (1.2), describe the methods and data used (1.3), and provide an outline and summary of the individual chapters (1.4).

1.2 Objectives and Main Contributions

This dissertation addresses the following overarching research question: "What are promising directions for organizing income support?" I conduct four empirical investigations targeting existing and new approaches to organizing income support to answer this question. Instead of focusing on a single type of benefit, I compile work on three different schemes: a minimum income scheme (social assistance), a social insurance scheme (unemployment insurance (UI) benefits), and a cash transfer program operating outside the existing social protection system. All three schemes have in common that they provide temporary support; the goal is to secure the economic self-sufficiency of claimants in the shortest amount of time possible. Consequently, labor market-related outcomes, mainly (durable and high-quality) employment, play a prominent role in all four studies. Focusing on labor market-related outcomes alone would constitute a too-narrow focus, however. After all, government support may affect outcomes other than those related to paid work.

[2]Other strands of criticism may come from a political-philosophical perspective and refer to, e.g., questions of social rights and justice (see, e.g., Eleveld et al., 2020). The evaluation of the policy interventions featured in this dissertation in this regard is beyond the scope of this dissertation.

Examples are health or education outcomes. Studying outcomes more broadly is crucial to understanding the comprehensive impacts of income support policies and the trade-offs they may entail. For that reason, in most of my studies, I consider outcomes from different domains. To collect outcome data, I use both administrative data sources and surveys. All studies included employ (quasi-) experimental research designs. In the following section, I will describe the data and empirical frameworks that I use in more detail.

With this dissertation, I contribute to the understanding of which policies improve (and do not improve) the working of income support. In doing so, my work complements a rich empirical literature studying policy designs and instruments related to income support. This literature draws from evaluations of single program components, such as active labor market policies (see, e.g., Card et al., 2017; Filges et al., 2015; Kluve, 2010, for an overview), benefit sanctions (see McVicar, 2020, for an overview), and different compliance requirements (see, e.g., Arni and Schiprowski, 2019; Cairo and Mahlstedt, 2021), to studies of entire programs, such as the 1970s negative income tax experiments in the United States (see Burtless, 1986, for a description of the experiments). The investigations included in this dissertation expand this literature in different directions. Therefore, I will now highlight some main objectives and contributions per individual chapter. While this section provides an introductory overview, a more detailed discussion of the relevant literature and contributions can be found in the respective chapters.

In the second and third chapters, I study the effectiveness of three alternative regulatory regimes in social assistance. Whereas Chapter 2 concerns two interventions that alter the way claimants are supervised and supported, Chapter 3 focuses on changing financial work incentives. Initially, all three interventions were part of the same study, a randomized controlled trial (RCT) in Utrecht (the Netherlands). I evaluate the interventions in two separate chapters due to their different nature and distinct contributions to the existing knowledge base.

Chapter 2 aims to answer the following research question: “What are the effects of implementing autonomy-enhancing regimes in social assistance?” I use the term *autonomy-enhancing regimes* to describe interventions that give benefit claimants more agency over their reemployment decisions. The research question is of great relevance, as evidence from the fields of psychology and behavioral economics suggests potential gains from self-determined behavior (see, e.g., Koen et al., 2016). The question is also relevant from a policy perspective, considering that arrangements promoting autonomy instead of control may reduce administrative effort and costs. Importantly, little is known about the effects of such regimes in the context of income support. Although there are a few studies that examine the effects of scaling down requirements or assistance for benefit recipients (see, e.g., Bolhaar et al., 2020; Johnson and Klepinger, 1994; Klepinger et al., 2002; McVicar, 2008, 2010), none of the treatments evaluated in earlier work actively emphasizes autonomy and choice. In addition to understanding whether autonomy-enhancing regimes produce favored results on average, it is also important to know if certain groups of claimants profit more or less from such regimes. Therefore, I also examine to what extent claimants with different gender, education levels, or ages respond differently to the interventions tested. The main contribution of Chapter 2 is to show that autonomy-enhancing regimes in social assistance can improve labor market outcomes, particularly among certain subgroups.

Chapter 3 addresses another important question related to the design of income support: "Do generous earnings exemptions stimulate welfare claimants to work?" Earnings exemptions (also earnings disregards) refer to arrangements under which benefits are not reduced one-on-one in the case of other income. Such policies aim to stimulate part-time work in addition to benefits, which may serve as a stepping stone toward full-time employment and independence from benefits. Whether earnings exemptions are effective is of great interest, given that such policies are easy to implement, requiring no more than changes in administrative systems. The working of earnings exemptions is also relevant from a cost-benefit perspective, considering that reducing benefits less than one-on-one is costly, but may pay off if work incentives are strong enough. From previous work, we know that earnings exemptions can effectively stimulate part-time work among particular subgroups of welfare claimants, e.g., single mothers (Knoef and van Ours, 2016). Less is known about their effects among a broader group of claimants. The main contribution of Chapter 3 is to show that slight adaptations to financial work incentives may stimulate part-time but do not increase chances of full-time exit from benefits.

In Chapter 4, I direct my attention to a specific instrument commonly used to enforce the compliance of UI benefit claimants with rules and regulations: financial sanctions. The chapter aims to answer the following research question: "What are the effects of imposing benefit sanctions and reprimands on job seekers?" Benefit sanctions are a temporary (full or partial) withdrawal of benefits in case of infractions, e.g., if claimants violate minimum job search requirements, refuse to accept suitable work, or misbehave toward a caseworker. A better understanding of how sanctions translate into claimant behavior and economic outcomes is paramount to designing optimal support schemes. Although existing empirical evidence consistently shows positive impacts of imposing sanctions on job finding and unemployment duration (see, e.g., Abbring et al., 2005; van den Berg et al., 2004; Svarer, 2011), less is known about their effects on other outcomes, such as quality of work indicators or health. Furthermore, there is ample evidence on the effectiveness of softer disciplinary measures, such as reprimands. The main contribution of Chapter 4 is to reveal trade-offs between sanctions and reprimands as policy options. Although sanctions seem a valuable tool to enforce future compliance, the evidence speaks in favor of imposing reprimands when outcomes are considered in a broader sense. In addition, the chapter makes a methodological contribution. I apply an instrumental variable approach, a technique frequently used in other areas of study, to solve endogeneity problems arising from the selective imposition of sanctions and reprimands.

In Chapter 5, I leave the Dutch setting to study a municipal cash transfer program in Spain. The research question is: "What are the labor supply effects of offering generous and unconditional income support?" In recent years, unconditional income support has become a popular tool to combat poverty and vulnerabilities in low- and middle-income countries. Evaluations of such programs report positive impacts on economic, health, and quality of life outcomes (see, e.g., Haushofer and Shapiro, 2016; Pega et al., 2017), while effects on labor supply are absent or only moderately negative (see, e.g., Banerjee et al., 2017; Bastagli et al., 2016). However, there are only a few examples of comparable programs being tested in a high-income country (see Marinescu, 2017, for a review of a few programs tested in the US). In even fewer cases, this concerns programs offering generous support, i.e., transfers which allow for a living at or above subsistence level. Understanding the effects of unconditional transfers in a setting different from their typical application is of essential importance for policymakers designing poverty relief. The program I study is a municipal program that targets vulnerable households in deprived neighborhoods of Barcelona. Households received a monthly cash transfer, without strings attached, capped at twice the poverty level. The main contribution of Chapter 5 is to show that unconditional transfers seem to involve a trade-off. While leading to lower labor supply on average, such effects appear to be limited to households with care responsibilities.

1.3 Data and Methodology

In this section, I address the empirical frameworks and data used in this dissertation. I begin by discussing the former. I rely on two different strategies to identify the causal effects of the programs and instruments studied. Chapter 2, 3, and 5 build on research designs that randomly assign units of observation to intervention groups, subject to a given treatment, and control groups, subject to the care-as-usual regime—in other words, RCTs. In Chapter 4, I exploit a research setting in which exogenous variation in treatment occurs naturally. More specifically, subjects run different risks of receiving the treatment of interest because they are quasi-randomly assigned to individuals with different propensities of imposing the treatment. I will now discuss both approaches and the estimation strategies involved in more detail.

Estimating causal effects in the former setting is straightforward, as it involves comparing groups assigned to be treated with groups assigned to remain untreated. Still, some challenges remain, such as, when encountering noncompliance, as is the case in different chapters of this dissertation. Noncompliance describes a situation in which some subjects do not *actually* receive the treatment they were originally *assigned* to receive. As actually treated and untreated subjects may form non-random subsets of the original treatment and control groups, identifying the average treatment effect (ATE) is no longer feasible. Analyzing the trials included in this dissertation, my co-authors and I deal with this challenge by estimating, where applicable, two alternative types of causal effects: intent-to-treat (ITT) effects and local average treatment effects (LATE). While the ITT describes the effect of *implementing* a program, the LATE is the treatment effect for a subgroup of subjects referred to as *compliers* (Angrist and Imbens, 1994). These are subjects that receive the treatment if and only if assigned to do so.

I use data from two RCTs, one conducted with social assistance claimants in

Utrecht and one with economically vulnerable households in Barcelona. The first trial, subject to Chapter 2 and 3, is based on a collaboration with the Municipality of Utrecht. In this collaboration, my co-authors and I were responsible for scientific supervision, including experimental design, data collection, and analysis. The trial in Utrecht received IRB approval by the Ethical Review Committee of the Faculty of Law, Economics and Governance at Utrecht University and was pre-registered—together with a pre-analysis plan (PAP)—at the RCT Registry of the American Economic Association. To increase transparency, the respective chapter includes an appendix motivating deviations from the PAP. The second trial, subject to Chapter 5, originates from a research project initiated by the City Council of Barcelona. The work included in the chapter stems from a research visit at the Catalan Institute of Public Policy Evaluation (*Ivàlua*), the organization entrusted with conducting the empirical analysis. Due to joining the project at a later stage, a PAP remains absent for this study.

Turning to the second empirical approach, in Chapter 4 I use a natural experiment instead of a controlled study design. I exploit a research setting in which subjects (benefit claimants) are quasi-randomly assigned to individuals (caseworkers) who systematically vary in their tendency to impose a particular treatment (sanctions and reprimands). I estimate causal effects employing an instrumental variable (IV) design. More specifically, I use the tendency of caseworkers to impose a certain treatment as an instrument for actual treatment receipt, an approach also referred to as *leniency design* (Cunningham, 2021). Naturally, this approach relies on more assumptions than identification in a randomized controlled setting. First, the instrument should be correlated with the endogenous regressor—in my study, the probability of actual treatment (instrument relevance). Second, the instrument must not correlate with the error term and should not directly affect the outcome of interest (instrument exogeneity). I will discuss the plausibility of these assumptions in detail in the respective chapter. Using an IV design, the estimated effects in Chapter 4 are, again, local average treatment effects. Here, the subgroup of compliers comprises claimants that could have received a different treatment had they been assigned to a different caseworker deciding on their treatment. Similarly to the RCT my co-authors and I conducted in Utrecht, I pre-registered the study subject to Chapter 4. Pre-registration took place at the registry of the Open Science Foundation. Again, I include deviations from the pre-registration in a separate appendix.

I will now focus on the data used. As mentioned before, the outcomes studied in the different chapters span various domains, including labor market and job search-related outcomes, quality of life measures, activities related to social participation and human capital formation, psychological outcomes, and compliance with rules and regulations. Labor market-related outcomes play a prominent role in all four chapters. My co-authors and I collect data on these outcomes mostly using administrative data sources. In Chapter 2–4, all located in a Dutch setting, these sources are social security records available at Statistics Netherlands (CBS). Dutch social security records are comprehensive, covering the entire population of employed workers in the Netherlands, and entail, among other things, information on earnings, hours worked, and the type of employment contract entered. CBS data is non-public but accessible for statistical and scientific research under certain conditions.[3] Chapter 5 uses non-public social security data from the Spanish Ministry of Labor and Social Economy. These records are less extensive in terms of information covered, as they only list an individual's employment status. Both data sources allow me to follow individuals over time, either monthly (CBS) or in ten-day intervals (Spanish Ministry of Labor and Social Economy).

For other outcome variables, I mostly turn to surveys. Where possible, I use validated and widely used survey instruments, e.g., questions from the 36-Item Short-Form Health Survey (Ware and Sherbourne, 1992) or the European Union Statistics on Income and Living Conditions (EU-SILC). In cases where I have information from both surveys and administrative sources, I compare outcomes to preclude bias due to socially desirable responding or misreporting. In all cases, I measure survey outcomes at both follow-up and baseline, that is, before randomizing units and implementing treatments. As is common practice, I use the information on baseline outcomes to verify the validity of randomization mechanisms and increase the precision of effect estimates. A link to the surveys used can be found in the respective chapters.

It is only in Chapter 4 that I also rely on proprietary data. This data concerns benefit recipients' compliance and job search behavior and stem from administrative data sources at the Dutch Employee Insurance Agency (UWV).

1.4 Outline and Summary of Chapters

In what follows, I briefly summarize each of the four chapters that form the core of this dissertation. I organize these four chapters as individual articles, including separate introductions and conclusions. This structure should allow the reader to zoom in on a specific topic or read the chapters in a different order. Naturally, this approach leads to some overlap and repetition, which I hope the reader will excuse. I close this dissertation with the sixth and last chapter. In this chapter, I discuss my main findings in light of the overarching research question, review some of the main limitations, and suggest directions for further research.

1.4.1 Requirements Versus Autonomy: What Works in Social Assistance?

In this chapter, I study the effects of two alternative regulatory regimes in social assistance. Both regimes build on reducing welfare conditionality and strengthening claimant agency in decisions related to reemployment. Under the first regime, social assistance claimants were exempted from all compliance requirements related to benefit receipt. These requirements included active job search, accepting paid work, regular meetings with a caseworker, and following training and activation programs. The regime also suspended monitoring by the welfare office. As

[3] For further information: microdata@cbs.nl.

claimants still had to prove eligibility, the benefit was not wholly unconditional. Under the second regime, claimants received intensive one-on-one counseling by a permanent caseworker. However, the intensive counseling program consulted the wishes of claimants regarding the kind of support and assistance that would be provided.

The two alternative regimes were evaluated in an RCT with 752 social assistance claimants sampled in Utrecht. I find that the exemption treatment significantly improved labor market-related outcomes, which contrasts with the predictions offered by standard economic theory. Roughly 1.5 years after the start of the trial, exempted claimants were on average 75 percent more likely to have work that allows for exiting benefits than their counterparts in the control group. Furthermore, exempted claimants had higher chances of working under a permanent contract, i.e., a contract with no end date. Subgroup analyses suggest that the exemption treatment worked particularly well for female, lower educated, and younger claimants, while effects are negative for higher educated claimants. By contrast, estimates for the counseling treatment—although positive—are less pronounced and largely statistically insignificant. Using survey data, I find no evidence of effects on job search behavior, social participation, health, and well-being for both treatments. In the exemption group, respondents reported higher levels of experienced autonomy.

Chapter 2 is a joint work with Stephanie Rosenkranz, Loek Groot, and Mark Sanders, all of whom were affiliated with Utrecht University School of Economics at the time of conducting the study. I thankfully acknowledge financial support from the European Social Fund (no. 2018EUSF2011696) and the Dutch Ministry of Social Affairs and Employment. The results of this chapter (together with the results of Chapter 3) are also summarized in a policy report (Verlaat et al., 2020a).[4]

1.4.2 Do Earnings Exemptions Stimulate Paid Work Among Welfare Claimants?

In Chapter 3, I investigate the effects of another alternative policy regime in social assistance. In contrast to the previous chapter, which concerned alternative approaches related to monitoring and support, I now study financial incentives to work and earn income in addition to receiving benefits. The intervention tested increased financial work incentives in two ways. First, it lowered the rate at which the welfare office reduces benefits as earned income is received. In other words, social assistance claimants faced a lower implicit tax rate on additional earnings. Second, the intervention eliminated a six-month time limit in which claimants can keep a share of additional earnings. The intervention was part of the RCT in Utrecht that also forms the basis of the previous chapter.

I find tentative evidence that the more generous earnings exemption had a positive effect on employment rates. Nearly 1.5 years after implementing the treatment, the chances of being employed were roughly 40 percent higher among treated

[4]The report, which is only available in Dutch, and other supplementary material can be accessed at `https://dspace.library.uu.nl/handle/1874/395951`.

subjects than control subjects. Employment effects appear to be driven by part-time work, while exit to full-time employment remained pretty much unaffected. I also find improvements in the income situation of treated claimants. Looking at benefit expenditures, I find no evidence for budgetary effects. I label the evidence as tentative as the estimates in Chapter 3 are surrounded by greater uncertainty than would be desirable. Like the previous chapter, Chapter 3 is a joint work with Stephanie Rosenkranz, Loek Groot, and Mark Sanders.

1.4.3 The Effects of Sanctions and Reprimands in Unemployment Insurance

The goal of the first two chapters was to study alternatives to the status quo regime in social assistance. In Chapter 4, I deviate from this path in two respects. First, instead of studying a policy regime, I now focus on a specific policy instrument. Second, instead of examining a new program, I now evaluate an already existing program feature. Specifically, I study the causal effects of imposing disciplinary measures when job seekers receiving UI benefits violate job search requirements. Disciplinary measures include benefit sanctions, which have a notable financial impact, and reprimands, which merely entail a warning and do not involve any financial consequences. Today, such measures constitute an integral feature of UI benefit schemes in many countries around the world. I study the effects of imposing sanctions and reprimands in the Dutch UI benefits system. Dutch UI benefit recipients are required to conduct and register four job search activities every four weeks. Violating this requirement may result in a reprimand or a sanction, most commonly a 25 percent cut in benefits for four months. As mentioned above, I use an IV approach (leniency design) to solve endogeneity problems stemming from the selective imposition of sanctions and reprimands.

In contrast to previous work, I find no evidence that imposing benefit sanctions promotes job finding or reduces benefit dependency. The same holds for reprimands. However, I find tentative evidence for negative side effects, namely that sanctions may persistently harm future earnings. Another finding relates to the compliance behavior of claimants. While sanctions generally lower the probability of re-offense, reprimands seem to result in a stronger behavioral response when looking at reported job search activities. An important side note concerns the potentially deterring effects of sanctions. Not only imposing a sanction but the mere fact of having a sanctioning threat in place may already affect outcomes. Studying these effects was beyond the scope of this chapter.

Chapter 4 is a joint work with Erik Bijleveld from Radboud University Nijmegen. I gratefully acknowledge financial support from the Dutch program *Handhaving en Gedrag* (in English: Law Enforcement and Behavior), an interdepartmental partnership of six government bodies. I am very grateful to the Dutch Employee Insurance Agency (UWV) for facilitating and supporting this project. The results of this research are also summarized in a policy report (Verlaat et al., 2021).[5]

[5]The report, which is only available in Dutch, can be accessed at

1.4.4 The Labor Supply Effects of Generous and Unconditional Cash Support

In Chapter 5, I study the labor supply effects of a generous and unconditional cash transfer program in Spain. The program, initiated by the City Council of Barcelona, targeted economically vulnerable households in Barcelona's most deprived neighborhoods. I call the program generous, as cash transfers raised a family's income to the subsistence level. On average, households received a monthly payment of approximately half the statutory minimum wage. I label the program as unconditional as payments did not depend on recipients' actions, such as job search, paid work, healthcare use, or investments in education. In contrast to a fully unconditional cash transfer, eligibility depended on a means test. The program was trialed in an RCT with 1,288 households. In addition to studying the overall impacts of the program, I compare the effects in different (randomized) treatment arms. First, I contrast households assigned to a social activation plan with households receiving the cash transfer alone. Second, I compare households that faced a 25–35 percent benefit reduction rate with households facing a benefit reduction rate of 100 percent. Studying these modalities allows for a better understanding of how different design features affect outcomes.

I find that assignment to the program resulted in strong negative labor supply effects. Nearly two years after starting the program, recipients assigned to receive the cash transfer were roughly 20 percent less likely to work on average. This effect is persistent, at least for six months post-treatment. Combining financial support with a social activation plan seems to have hampered employment chances even more. So did a 100 percent benefit withdrawal rate. Studying effect heterogeneity, I find indications that negative labor supply effects were almost entirely driven by households with care responsibilities. This result might warrant a re-assessment of the impact of the program. If reductions in labor supply result from the substitution of labor for care tasks, the broader welfare effects of the program may be positive, after all.

Chapter 5 is a joint work with Federico Todeschini from the Catalan Institute of Public Policy Evaluation (*Ivàlua*) and Xavier Ramos Morilla from the Autonomous University of Barcelona. I thankfully acknowledge financial support from a Utrecht University School of Economics mobility grant.

https://www.handhavingengedrag.nl/onderzoeken/sancties-voor-burgers.

CHAPTER 2

Requirements Versus Autonomy: What Works in Social Assistance?*

2.1 Introduction

When designing benefit schemes for the unemployed, a crucial dilemma is maintaining incentives to search for a job. Policymakers in different countries have responded to that dilemma by making benefits receipt conditional on claimants' (job search) behavior (Venn, 2012). Accordingly, claimants are required to actively look for work, accept suitable jobs, or participate in training and activation programs. At the same time, to ensure compliance, welfare authorities monitor claimants' behavior and sanction or fine them in case of a breach. If anything, the trend over the past three decades shows governments tightening conditions for benefit recipients (Knotz, 2018). This chapter reports on a policy experiment that had a benefit scheme change in the opposite direction. More specifically, we study the case of a welfare office softening the principle of conditionality, instead actively promoting claimant autonomy.

The policy experiment included a scheme that suspended all compliance requirements tied to benefits receipt. Claimants were given full autonomy, and the welfare office would no longer monitor their behavior or oblige them to meet caseworkers and engage in employment services. Note that claimants still had to prove eligibility, which means that benefits were not wholly unconditional. We will refer to this first scheme as the *exemption* treatment. A second scheme involved more intensive counseling; however, it was under the tenet of following the claimant's lead

*This research was pre-registered at the AEA RCT Registry under no. 0003592 (Verlaat et al., 2020c). A pre-analysis plan (PAP) is available at the registry. We report discrepancies with the PAP in Table 2.E.1 in Appendix 2.E. The study received IRB approval by the Ethical Review Committee of the Faculty of Law, Economics and Governance at Utrecht University under file no. 2018-002.

when providing assistance and guidance. We will refer to this second scheme as the *counseling* treatment. We partnered with a Dutch municipality—responsible for the policy experiment—to conduct an empirical evaluation of these two alternative approaches.[1]

Benefit schemes for the unemployed are intended as temporary safety nets. Therefore, we are primarily interested in the following research question: What are the effects of the two treatments on various labor market outcomes, that is, employment, independence from benefits (self-sufficiency), and the quality of reemployment? In addition, we study the effects on experienced autonomy, job search behavior, social participation, health, and well-being.

While the configuration of the two alternative schemes largely originates from the policy sphere, we can still formulate some stylized theoretical underpinnings which guided our thinking about potential effects on labor market outcomes.[2] For the counseling treatment, economic theory leads us to expect higher chances of job finding and self-sufficiency to the extent that counseling reduces market frictions and increases search efficiency. For the exemption treatment, standard theory suggests effects in the opposite direction. First, suspending job search requirements, monitoring and sanctions should lead to lower search incentives and higher reservation wages, resulting in lower and slower transitions from welfare to work. Similarly, dropping out of counseling and employment services could lead to lower search effort and less efficient search, both of which may negatively affect job finding and exit from welfare. Regarding post-unemployment outcomes, exemption may lead to higher quality job matches (due to higher reservation wages) and higher employment stability (due to fewer referrals to temporary work). For counseling, we would not necessarily expect to see such effects.

Importantly, both treatments give autonomy to claimants, i.e., opportunities of choice and self-direction. This may constitute another channel through which treatment effects materialize. Autonomy plays a central role in both Self-Determination Theory (Deci and Ryan, 1985) and its counterpart in economics, Motivation Crowding Theory (Frey and Jegen, 2001). Both theories claim that motivation does not only vary in terms of levels (i.e., how much motivation?), but also in terms of type (i.e., controlled versus intrinsic motivation). People experience controlled motivation when they feel pressured or obliged to do something and intrinsic motivation when acting out of choice and volition. Motivation Crowding Theory highlights the

[1]The policy experiment was possible due to a waiver issued by the Dutch Ministry of Social Affairs and Employment. This waiver allowed municipalities, which in the Netherlands administer the national social assistance scheme, to temporarily deviate from the treatment of benefit claimants prescribed by the law. The waiver also allowed for a third treatment, which—instead of changing requirements and supervision—allowed for a more generous earnings disregard. Our partner municipality implemented this third treatment alongside the other two treatments. We decided to evaluate this third treatment separately due to its different character. While an earnings disregard only changes financial work incentives, we are interested in the effects of autonomy versus control in this chapter. In summary, we find that a more generous earnings disregard promotes part-time work among benefit claimants (for detailed results, see Chapter 3).

[2]For formal accounts, see, e.g., van den Berg and van der Klaauw (2006) or Fredriksson and Holmlund (2006).

interplay between these two types of motivation: external interventions, if perceived as controlling, may undermine intrinsic motivation and lower overall performance (crowding-out effect). In contrast, interventions may foster intrinsic motivation and increase overall performance when perceived as supportive (crowding-in effect). Following the reasoning of the two theories, we may expect that exempting claimants reduces crowding-out effects. In that case, motivational effects may compensate for the negative consequences of reduced regulation. In the case of counseling, opportunities of choice and self-direction may give room to motivation crowding-in and lead to better outcomes.

We evaluate the two alternative approaches in a randomized controlled trial (RCT) with 565 social assistance recipients sampled in the city of Utrecht. We use administrative data from social security records at Statistics Netherlands to obtain information on labor market outcomes. Additionally, we employ three waves of surveys to collect information on experienced autonomy, job search behavior, social participation, health, and well-being.

We find that exempting claimants has a significantly positive impact on labor market outcomes. Roughly 1.5 years after the start of treatment, exempted claimants are on average 75 percent more likely to be self-sufficient due to paid work as compared to the control group. They are also more likely to have found work under a permanent contract. Furthermore, survey results suggest gains in experienced autonomy among that group. Our findings for the counseling group are also generally positive but less pronounced and not statistically significant in most months of observation. Whereas we see effects getting stronger over time for the exemption group, the effects of counseling appear sooner but level off in the longer term. Overall, we do not find evidence of effects on job search behavior, social participation, health, and well-being. On a cautionary note, our limited sample size does not allow us to very confidently reject small treatment effects or false positives due to performing multiple comparisons. Still, we deem our findings worth reporting, as they provide strong evidence against *negative* labor market effects, a prediction offered by standard economic theory.

Our findings contribute to understanding how autonomy-enhancing regimes in unemployment benefit schemes might work. In doing so, our results add to a large and expanding literature evaluating activation policies for the unemployed.[3] Comparable to the exemption treatment are three field experiments and three quasi-

[3] A large part of this literature evaluates particular activation instruments (or active labor market policies), such as subsidized employment, skills training, or job search assistance. Systematic reviews of that literature (see, e.g., Card et al., 2017; Kluve, 2010) find that the effectiveness of such programs is highly related to the program type. While public sector employment programs appear to be largely detrimental, wage subsidies and programs including job search assistance appear to have stable positive effects. The effects of training are small and sometimes negative in the short term, but they show larger impacts over a longer time horizon. Another strand of the literature studies the effects of imposing benefit sanctions (see, e.g., Abbring et al., 2005; van den Berg et al., 2004; van der Klaauw and van Ours, 2013; Lalive et al., 2005). These studies generally find that sanctions reduce unemployment duration and—at least in the short term—increase employment chances, but can be detrimental for reemployment quality and longer-term outcomes (see McVicar, 2020, for a recent review).

experiments, all of which have in some way studied the effects of scaling down requirements or assistance. The first experimental studies looked at eliminating job search requirements and monitoring in the U.S. unemployment insurance (UI) system in the 1990s (Johnson and Klepinger, 1994; Klepinger et al., 2002). Both studies find that, consistent with standard economic theory, eliminating requirements and monitoring delay exit from UI while increasing post-unemployment earnings in the latter study.[4] McVicar (2008, 2010) exploit the temporary suspension of monitoring claimants' job search efforts due to welfare office refurbishments in the United Kingdom and Northern Ireland. They also find that periods of zero monitoring lead to lower exit rates from UI and longer spells on average. To the best of our knowledge, the only evidence for effects in the opposite direction comes from a non-experimental study in Australia. Using a matching design, Gerards and Welters (2021) find that UI benefit recipients subjected to requirements take longer to find work and stay in employment shorter than similar recipients without such requirements. A study that, instead of eliminating requirements and monitoring, removes employment services and counseling comes from Bolhaar et al. (2020). The authors study the effectiveness of different activation policies using a field experiment with social assistance claimants in Amsterdam. To form a control group, some subjects are excluded from receiving any assistance and employment services whatsoever (while leaving formal compliance requirements intact). The authors find mixed results, suggesting that no guidance can lead to better, equal, or worse labor market outcomes, depending on which form of active guidance they take as the counterfactual.

The exemption treatment in our study is comparable to the first set of studies mentioned above because it fully eliminates job search requirements and monitoring. In contrast to the study from Bolhaar et al. (2020), however, exemption does not mean exclusion from assistance and employment services but instead makes their use voluntary. Importantly, in none of the earlier work was scaling down requirements or assistance combined with actively emphasizing autonomy and choice, which is a distinguishing feature of the interventions we study. Following behavioral theories, this aspect may alter the effectiveness of the treatment. To the best of our knowledge, we are the first to study this type of intervention in the context of activating the unemployed.

With regards to the counseling treatment, there exist more related experimental studies. Most concern mandatory counseling or activation programs (often in combination with increased monitoring) and provide evidence for positive labor market effects (see, e.g., Dolton and O'Neill, 1996, 2002; Gorter and Kalb, 1996; Maibom et al., 2017; Meyer, 1995). An exception is van den Berg and van der Klaauw (2006), who find no effect of counseling and monitoring on exit to work for Dutch UI recipients. Previous evidence on the effectiveness of rather *intensive* activation and counseling programs comes from Graversen and van Ours (2008) and

[4] Studying the long-term employment outcomes (up to nine years later) of the Johnson and Klepinger (1994) experiment, Lachowska et al. (2016) find negligible overall effects of eliminating requirements.

Markussen and Røed (2016). The former study evaluates a high-intensity early-stage job search assistance, training, and counseling program targeted at Danish UI recipients. The latter study is on a Norwegian activation program for hard-to-employ social assistance recipients. Both programs appear to have successfully increased employment prospects.[5] What distinguishes these earlier interventions from our treatment is that they are mostly mandatory, pre-defined programs. To the best of our knowledge, we are the first to test a program that gives claimants a strong say in how they want to be assisted. Again, this additional aspect of autonomy may significantly change the impact of the treatment.

Two further contributions of our work concern the population studied and the method applied. Much of the existing work on activating the unemployed has focused on recipients of UI benefits. These are individuals who are reasonably close to the labor market. Less is known about what works for individuals with long unemployment histories and less favorable labor market prospects. Moreover, only a minor share of evidence on activation policies stems from experimental studies. Exploiting a randomized research design, we expand the evidence base in that direction.

The remainder of this chapter is organized as follows. The following section briefly describes how social assistance is organized in the Netherlands and introduces the local experimental setting. Section 2.3 covers the experimental design and methods. Section 2.4 introduces our data sources, while Section 2.5 discusses implementation and Section 2.6 our empirical strategy. Section 2.7 presents our results. In Section 2.8 we discuss potential mechanisms, while Section 2.9 concludes.

2.2 Policy Context

2.2.1 Social Assistance in the Netherlands

Social assistance in the Netherlands is a non-contributory guaranteed minimum income scheme that provides a monthly benefit payment to households with income below the subsistence level. The scheme is designed as a safety net for those who are, in principle, fit for work. Nonetheless, many claimants face problems related to health, finances, or family that may impede reemployment. Unemployed job-seekers usually claim social assistance because they have exhausted other benefits (e.g., UI benefits) or have never been eligible for other benefits in the first place. As such, the scheme is quite comparable to last-resort assistance in other European countries and not so much to the United States, where welfare typically targets single-parents with low income.

Eligibility for the scheme depends on a means test that examines income (including the income of household partners) and wealth.[6] People that are deemed

[5] In a study following up Graversen and van Ours (2008), Rosholm (2008) finds that overall effects seem to be driven by a threat-effect of the mandatory program. In contrast, the effects for the single program components (job search assistance, training, counseling) are largely insignificant or negative.

[6] In addition to that, claimants have to be legal residents.

unable to work or have reached the state pension age are eligible for other schemes. Benefits are paid for an unlimited duration and benefit levels vary between 70 and 190 percent of the net minimum wage, depending on household size.[7] Earnings on top of benefits are subject to a marginal tax rate of 100 percent, except for the first six months in which income is earned.[8] In addition to social assistance, recipients may collect means-tested housing and healthcare allowances, and parents may receive further allowances for the cost of raising children.[9] Importantly, the scheme is decentralized, which means that municipalities and regional councils are in charge of processing applications, paying out benefits, counseling and monitoring recipients, and detecting benefit fraud.

Benefits receipt is tied to various compliance requirements, which primarily target labor market behavior and are comparable to those of other OECD countries (Immervoll and Knotz, 2018). In principle, recipients are obliged to actively look for paid work and accept any suitable job offered. They might also be requested to participate in activation or training programs, register with a temporary employment agency, or undertake volunteer work. In addition, recipients have to cooperate with the local welfare office. This includes showing up for caseworker meetings and providing all the information needed to assess eligibility and determine the benefit level. Recipients' behavior is monitored, and noncompliance may result in a benefit sanction—a cut in one's benefits between 30 and 100 percent for a maximum of three months, depending on the severity and recurrence of the infringement. In practice, however, benefit sanctions are rarely imposed.[10]

Some recipients are formally and temporarily exempted from certain requirements, usually job search requirements, for medical reasons or because they provide informal care for next-of-kin. Caseworkers may also give informal exemptions. In practice, compliance requirements often differ, depending on which target group a recipient belongs to. Elderly recipients, e.g., usually face less stringent requirements than younger recipients. The same holds for recipients with a great distance from the labor market compared to recipients with good labor market prospects.

[7] At the start of our study, benefit payments amounted to €992 ($1,279 PPP) per month for single-person households (70 percent of the net minimum wage) and €1,417 ($1,827 PPP) per month for two-person households (100 percent of the net minimum wage). The benefit level is capped at 190 percent of the minimum wage for households consisting of five or more adults. During the study, benefit levels had increased by 4 percent. Please note that in the Netherlands, taxes and other contributions are deducted from benefit payments beforehand. Thus, the mentioned figures refer to net income. Note that we use the 2018 OECD purchasing power parity (PPP) exchange rate to convert euros into U.S. dollars (OECD, 2021d).

[8] For a total of six months, recipients are allowed to keep 25 percent of their monthly net earnings up to a maximum of €202 ($260 PPP) per month.

[9] Taken together, the maximum housing and healthcare allowance for a single-person household amount to an approximately additional €450 ($580 PPP) per month.

[10] In the three years before our experiment, sanctioning rates averaged at roughly 4 percent on the national level (Divosa, 2021).

2.2.2 Local Setting

Our experiment took place in the city of Utrecht, which is the fourth largest city in the Netherlands.[11] Municipalities and regional councils usually organize the execution of the social assistance scheme through local welfare departments. In Utrecht, that is the *Work & Income Department*, which operates a single welfare office. As of January 1st, 2018, the share of households in Utrecht receiving social assistance benefits was slightly above the national average, at 5.9 percent (11,110 individuals) compared to 5.5 percent. The study took place in a period of sustained national economic growth (2.7 percent in 2018) with a relatively tight labor market (80 vacancies per 100 unemployed in Q4/2018).[12]

Although many social assistance claimants have long unemployment histories, there still exists quite some heterogeneity concerning labor market prospects. To better target their employment services, the welfare office in Utrecht classifies claimants into four broad categories, which reflect distance to the labor market. Class I includes claimants, which are directly employable and expected to find paid work within one to three months. Claimants in class II are expected to find paid work within three to 12 months. Class III comprises claimants that are currently unfit for work due to personal circumstances. A fourth class concerns claimants with an occupational disability. Classification takes place during the application process, and it is automated, taking into account claimant characteristics, such as age, education level, language skills, and health status. Caseworkers have discretionary power to adjust the classification. Formally, compliance requirements apply to claimants of all four classes. However, in practice, claimants in classes III and IV are treated with more leniency than claimants in the other two classes.

In addition to processing applications, paying out benefits, and monitoring claimants, the welfare office provides employment services. Those services include, among other things, caseworker counseling, job search assistance, job placements, or programs aimed at increasing claimants' employability. As the approaches evaluated in this study make such services voluntary (exemption) or possibly change the intensity at which they are provided (counseling), it is instructive to describe in more detail which role employment services play under the status quo regime.

The type and intensity of services claimants receive largely depend on their assigned class. Supervision of claimants in class I is primarily directed at stimulating job search. Hence, employment services usually prioritize job placements, job search assistance, and frequent counseling by caseworkers. Supervision of claimants in class II aims at increasing employability. Therefore, services for that group mainly consist of programs and instruments targeting skill development. Claimants in class III receive little employment services. Together with social workers and neighborhood centers, support for this group aims at promoting social participation and working towards employability.

Finally, it is important to note that claimants are usually assigned to teams of

[11] As of January 1st, 2018, Utrecht counted 347,483 inhabitants in 178,186 households.

[12] Information on economic growth and labor market tightness are retrieved from Statistics Netherlands.

caseworkers instead of having a permanent caseworker. Within a team, the team's workload usually determines which caseworker will contact a claimant on a particular day. This configuration will change under the counseling approach, as the next section will explain in more detail.

2.3 Experimental Design and Methods

2.3.1 Treatment Groups

For the evaluation, we randomly allocated social assistance claimants to two treatment groups and one control group. In what follows, we will describe all three groups in more detail.

Treatment 1 (Exemption). Treatment 1 replaced the status quo regime with a scheme under which claimants were given full autonomy in making decisions concerning reemployment. We emphasize four particular aspects of this treatment. Firstly, subjects were formally exempted from all compliance requirements that concern labor market behavior. Most importantly, that included obligations to look for paid work, accept any suitable job, and follow activation and training programs. They were also exempted from regular meetings with a caseworker. Subjects were only required to provide the welfare office with all the information necessary to assess eligibility and determine the benefit level. Secondly, when informing subjects about the new rules, the welfare office pointed out that claimants were *expected* to work on their reemployment but now had full autonomy to decide in which way.[13] Thirdly, employment services offered by the welfare office were still available to those who wished to receive them, yet, they were not actively offered. At the start of treatment, the welfare office informed about the possibility of opting in and out of employment services at any stage. Lastly, the welfare office contacted subjects after roughly eight months to inquire about their situation and reinforce the offer of providing employment services and support.

Treatment 2 (Counseling). Treatment 2 left the rules and regulations of the status quo regime unchanged. Yet, subjects in that group were subject to an intensive counseling scheme, which deviated from the status quo approach in four regards. Firstly, welfare recipients were assigned to an individual caseworker instead of being supervised by a caseworker team. Secondly, caseworkers' caseloads were roughly halved.[14] Thirdly, claimant and caseworker jointly decided on additional services to be provided. This could include extra meetings, coaching, programs, or training. In offering extra services, the question *"What do you need?"* was introduced as a leading principle. Lastly, the frequency of contact between caseworkers and claimants was increased. The goal was to be

[13] All communication material, which is only available in Dutch, can be accessed at `https://dspace.library.uu.nl/handle/1874/395951`.

[14] Caseworker teams usually deal with a caseload of 120 claimants per caseworker. Caseloads for individual caseworkers in treatment 2 amounted to approximately 40-50 claimants per caseworker.

in contact at least twice as much as under the status quo regime. Contact may refer to personal meetings and remote contact (phone calls, emails, letters).

Control group. The control group consisted of claimants that had consented to participate in the experiment but were assigned to the control condition. Subjects in the control group received the care-as-usual treatment according to the status quo regime.[15] We acknowledge that discretion in the status quo regime (as described in Section 2.2) may lead to significant heterogeneity in the treatment of control group subjects. As a result, our exemption treatment may only entail a slight deviation from the status quo for some participants. We assume that—if anything—this biases our results toward zero and makes our tests for treatment effects more conservative.

2.3.2 Sample Selection, Randomization, and Timing

Figure 2.1 depicts a flow chart describing the process and timing of sample selection and randomization. Our sample is a stock sample of existing social assistance recipients.[16] Specifically, the target population for our experiment included all claimants in Utrecht that had received benefits for a minimum of ten weeks on two cutoff dates and did not fall under any of seven exclusion criteria. For instance, recipients younger than 27 are subject to rules and regulations excluded from the government waiver (see Table 2.A.1 in Appendix 2.A for a complete list of exclusion criteria). Eventually, we counted 8,338 social assistance recipients eligible for participation, making up for 68 percent of social assistance recipients in Utrecht.

The government waiver required that claimants give consent before including them in an experiment. Accordingly, we first recruited subjects and randomly assigned those who agreed to participate to treatment and control groups. Due to this setup, our subjects knew about the experiment and its design features, such as random assignment and duration, and what the treatments entailed. Recruitment for the experiment took place during a large-scale campaign, running from February to April 2018. The campaign, administered by the Municipality of Utrecht, aimed at recruiting a sample as representative as possible and included information events in all city districts, information desks at the welfare office, letters, e-mails, and text messages, as well as posters, brochures, videos, and other informational material distributed at targeted locations across the city. Enrollment for the study consisted of two steps: signing an informed consent form and filling in a baseline survey. On the date of randomization, 752 claimants (around 9 percent of all claimants approached) had at least completed the first step and were included in the randomization.

[15] We note that control subjects received a 40-euro gift card for completing three surveys.

[16] We acknowledge that it would be preferable to test the interventions in both a stock sample and a flow sample of welfare claimants, considering that a policy change would affect both populations and assuming that the two populations differ in composition and exposure to the status quo regime. However, including a flow sample was not feasible as the government waiver allowing for different treatments was in place for 19 months at one stretch, and weekly inflow into the scheme would not have sufficed in reaching a large enough sample size in due time.

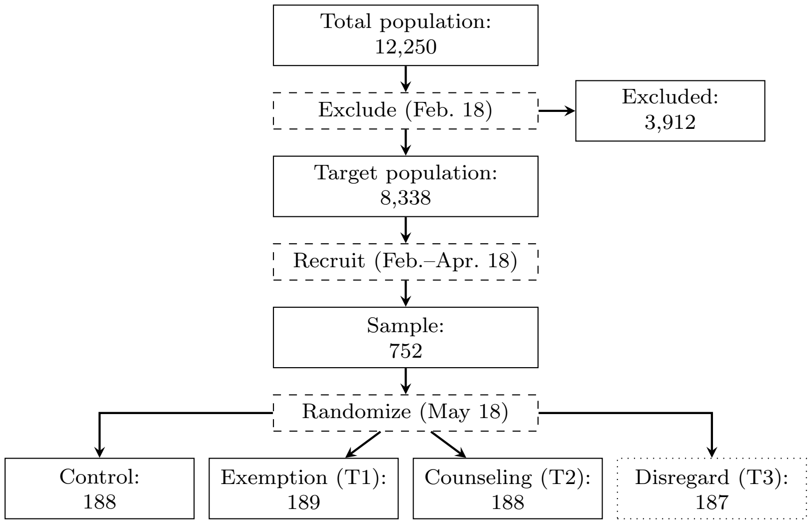


Figure 2.1 Process of Sample Selection and Randomization.

Randomization took place at the individual level. We randomized treatments over claimants instead of caseworkers as the welfare office does not randomly assign claimants to caseworkers.[17] To increase precision, we allocated subjects to experimental groups using a stratified randomization design. We used two variables to form strata: *living situation* (two categories), indicating if claimants receive benefits as a single-person household or not, and *class* (four categories), as determined by the welfare office during the application process (see Section 2.2.2).[18] In each stratum, participants were assigned to experimental groups with equal probability, using an algorithm for stratified randomization.[19] As a result, 189 (188) subjects were allocated to the exemption (counseling) treatment and 188 subjects to the control group. Thus, our final sample for this study consists of 565 claimants. In Section 2.4.3 we will compare this sample to our target population. The remaining 187 subjects belong to a third treatment group, subject to a more generous earnings disregard. We evaluate the third treatment in Chapter 3.

The treatments ran for 19 months, from June 1st, 2018, until December 31st, 2019. The treatment period was the same for all subjects. Subjects were informed about their group assignment via letters two weeks before the start of treatment. From month 17 onward, the welfare office contacted all subjects still receiving

[17]Some of the previous literature exploits (conditional) random assignment of claimants to caseworkers to identify treatment effects (see, e.g., Arni and Schiprowski, 2019; Bolhaar et al., 2019, 2020).

[18]At the time of randomization, we only had limited access to claimant background data available at the registries of the welfare office. We did not have access to the comprehensive data sources available at Statistics Netherlands, yet. We chose the two variables we expected to be most predictive for labor market outcomes among the available data.

[19]We used the algorithm provided by the user-written Stata package *randtreat* (Carril, 2017). The command provides different options for dealing with misfit observations, which are observations in excess after dividing the number of observations per stratum by the number of experimental groups. In our case, misfits were randomly allocated to a group within each stratum. The randomization syntax is available upon request.

benefits for a debrief, informing them about the situation post-treatment. Note that subjects continued in the same experimental group when exiting and re-entering social assistance during the experimentation period. Importantly, the treatment period ended before the COVID-19 pandemic started.

2.3.3 Compliance and Spillovers

The voluntary character of the experiment allowed subjects to withdraw from the study at any time. In case of withdrawal, subjects directly returned to the status quo regime, introducing one-sided noncompliance (failure-to-treat) into our experiment.[20] To account for noncompliance, we later estimate both intent-to-treat (ITT) and local average treatment effects (LATE). Figure 2.C.1 in Appendix 2.C shows cumulative withdrawal rates for all three groups during the 19-month treatment period. By the last month of the trial, withdrawal rates amounted to 9, 11, and 14 percent for the control, exemption, and counseling group, respectively.[21] Differences in withdrawal between the exemption and the control group ($p = 0.495$) and the two treatment groups ($p = 0.327$) are statistically indistinguishable. The difference in withdrawal between the counseling and the control group is significant at the 10 percent level ($p = 0.098$). It seems plausible that the counseling group reports higher withdrawal rates given the more intensive nature of the treatment.

We assume that the risk of spillovers is low. Our 565 subjects make up for, at most, 5 percent of the total population of social assistance claimants in Utrecht, and our sample is spread across all ten city districts. To assess the risk of spillover effects, we asked subjects halfway through the trial whether they knew any person outside their household participating in the experiment. Roughly 9 percent of all respondents and 8 percent of control group respondents answered that question with a *yes*. This leads us to believe that spillovers are not seriously affecting our results.

2.4 Data

2.4.1 Administrative Data Sources

Data on labor market outcomes come from social security records at Statistics Netherlands. These records contain monthly information on individuals' employment situation, including the start and end dates of employment, working hours, earnings, and information on the type of employment contract (temporary versus permanent). The nature of the records allows us to construct retrospective control variables, e.g., cumulative earnings before taking part in the study. The data also

[20]Monitoring the experiment and performing several implementation checks, we did not find any evidence that any subject assigned control received treatment. Therefore, we rule out the case of two-sided noncompliance.

[21]The withdrawal rates we observe are substantially lower than the treatment group dropout rates of 30-50 percent that Heckman et al. (1999) report for a variety of field experiments with labor market policies.

allows us to compare outcomes between groups before the start of treatment. We use the records to construct a longitudinal data set with monthly observations from October 2016 (19 months before randomization) to December 2021 (12 months after the treatments stopped). As the records include the whole population of the Netherlands, there is no missing data (attrition). Unfortunately, the employment data from Statistics Netherlands do not (yet) include information on self-employed individuals, i.e., own-account workers, due to lags in reporting of up to two years.

We also collect information on background characteristics at the time of invitation. From the civil registry at Statistics Netherlands, we obtain information on subjects' gender, age, the highest level of education, and the migration background. From the benefit registry at Statistics Netherlands, we collect information on the household composition (which determines benefit levels). Lastly, we obtain data on claimant classification, contacts with claimants, and employment services from the registries of the welfare office in Utrecht.

2.4.2 Survey Data and Attrition

Our administrative data sources lack information on other outcomes of interest, such as job search behavior or health and well-being. We used surveys to collect (self-reported) data on these outcomes and followed up twice to track outcomes over time. In total, we administered three waves of surveys (baseline, midline, and endline), each time using computer-assisted personal interviewing (CAPI) and computer-assisted web interviewing (CAWI).[22] In the case of CAPI, subjects were visited by trained interviewers, interview schedules were registered, and the location of the interview could be verified using GPS tracking. We provided questionnaires in the four most spoken languages in the Netherlands, and the field team consisted of interviewers fluent in those languages.[23] Our questionnaires were bench tested and piloted upfront. Respondents were thoroughly informed about the anonymity of the survey, among other things, to minimize the risk of social desirability bias. We can link our survey data to administrative information using unique identifiers for each subject.

As our administrative data sources provide us with complete records, we only face risk of attrition-related bias for survey outcomes. We can think of three main reasons for survey attrition in our study: non-response, active refusal to fill in a survey, and, as we were not allowed to contact subjects that had withdrawn from the study, unavailability due to withdrawal. In what follows, we will use the term *attrits* to describe subjects that did not fill in a particular survey, no matter which reason.

Table 2.B.1 in Appendix 2.B reports response rates relative to both the full sample and those approached during a particular wave. In general, response rates

[22]The baseline survey took place in the two months before randomization, the midline survey took place in treatment months 8 and 9, and the endline survey in months 15 and 16. Questionnaires can be accessed at `https://dspace.library.uu.nl/handle/1874/395951.`

[23]The languages were Dutch, English, Turkish, and Modern Standard Arabic. A professional bureau translated the questionnaires and made sure to include a cross-check by two translators.

are rather high. Even at endline, we were able to collect survey data for 70–80 percent of our subjects. While attrition rates do not differ notably between the control and the exemption group, we find a somewhat larger share of missing data in the counseling group. However, response rates relative to those approached are quite similar across groups (see Panel B of Table 2.B.1). This suggests that higher attrition rates in the counseling group mainly stem from higher withdrawal rates in that group.

To further investigate survey attrition, we assess whether attrits are different in terms of background characteristics (see Table 2.B.2) and whether their characteristics differ across experimental groups (see Table 2.B.3–2.B.4). While we report detailed results in Appendix 2.B, our findings can be summarized as follows. First, attrits in both follow-up waves are, on average, lower educated, further away from the labor market, more likely to have a non-Dutch background, and more likely to live in a multi-person household. To account for differences between attrits and non-attrits, we apply inverse probability weighting, based on the background characteristics listed in Table 2.1, when analyzing survey outcomes and report weighted results as a robustness check. Second, we find some evidence that attrits may differ across experimental groups. Therefore, as a final adjustment, we also report bounded treatment effects based on the bounding methods described by Horowitz and Manski (2000) and Lee (2009).[24]

2.4.3 Descriptive Statistics and Balancing

Table 2.1 reports background characteristics at the time of invitation for our full sample of 565 subjects.[25] Half of our sample are women, and 63 percent receive welfare benefits as a single-person household. On average, subjects are 47 years old. The largest share of participants (39 percent) is in class III, i.e., claimants with a great distance to the labor market. In addition, 26 percent are registered as directly employable (class I) and 32 percent as employable within 12 months (class II). On average, subjects receive benefits for 75 months at the start of the trial. The distribution of spell duration is right-skewed and has a long tail (median: 55 months; 95th percentile: 241 months), which explains the high average. Subjects may have worked full-time or part-time in addition to benefits in the pre-treatment period. To proxy employment history, we calculate cumulative earnings in the 24 months before recruiting, which average €1,620 ($2,088 PPP). As for education level, 23 percent completed higher education, and 25 percent completed intermediate education. Accordingly, 49 percent have a low level of education, i.e., no more than lower secondary education.[26]

[24]Horowitz-Manski bounds bracket the true treatment effect and are obtained by replacing missing values with extreme values (maximum or minimum). Lee bounds bracket the treatment effect for a subgroup of subjects, namely *always-reporters*. This approach involves throwing away the highest/lowest values in the group with the lower attrition rate (also referred to as *trimming*).

[25]One subject assigned to the counseling group passed away during the study. We exclude that subject from our analyses, which leaves us with an estimation sample of 564 observations.

[26]We classify education levels according to the International Standard Classification of Education (ISCED) 2011.

Table 2.1 Sample Descriptive Statistics and Balancing.

	Full sample	Group			p-value	
		Control	T1	T2	T1 vs. control	T2 vs. control
	(1)	(2)	(3)	(4)	(5)	(6)
Age (in years)	47.0	47.0	47.2	46.9	0.91	0.89
Female	0.50	0.49	0.47	0.55	0.74	0.24
Lower education	0.49	0.54	0.43	0.49	0.04	0.36
Intermediate education	0.25	0.23	0.25	0.26	0.58	0.46
Higher education	0.23	0.19	0.27	0.22	0.07	0.43
Education unknown	0.04	0.04	0.04	0.02	0.98	0.24
Dutch background	0.37	0.35	0.40	0.37	0.32	0.62
Western background	0.11	0.11	0.13	0.09	0.45	0.48
Non-western background	0.52	0.54	0.47	0.54	0.14	0.96
Current spell (in months)	74.5	79.1	72.4	71.9	0.35	0.29
Earnings 24 months before (in euro)	1620.2	1149.6	2108.8	1599.6	0.03	0.23
Single	0.63	0.63	0.65	0.62		
Single parent	0.17	0.18	0.16	0.18		
Cohabit	0.19	0.19	0.20	0.20		
Class I	0.26	0.25	0.26	0.26		
Class II	0.32	0.32	0.32	0.32		
Class III	0.39	0.39	0.39	0.38		
Class IV	0.04	0.04	0.03	0.04		
Joint test (p-value)					0.29	0.47
Observations	565	188	189	188		

Note: Column (1)–(4) report means. Column (5)–(6) show p-values from regressing each background characteristic on treatment dummies, controlling for randomization strata. The second last row reports p-values from a joint hypothesis test. Household composition and claimant classification were used for stratification and are therefore excluded from any test. See Table 2.A.2 in Appendix 2.A for a description of variables. T1: exemption; T2: counseling.

Additionally, Table 2.1 provides information on background characteristics for each experimental group separately. To test for differences between the three groups, we regress each background characteristic on respective treatment indicators. The counseling treatment appears well balanced with the control group. The exemption group seems to have a slightly more favorable labor market position on average. This is expressed in higher educational attainment and larger earnings in the two years preceding the study. Controlling for the background characteristics listed in Table 2.1, our baseline model to estimate treatment effects accounts for these differences in observables. As a sensitivity check, we also estimate unadjusted treatment effects, which leads to qualitatively comparable results. The retrospective nature of our data set allows us to compare groups not only in terms of background characteristics but also in terms of pre-treatment outcomes. As a placebo test, we estimate group differences before the start of treatment for the same period as the treatments lasted (19 months). As we will show in Section 2.7, we find no evidence of systematic differences in outcomes ex ante. The same holds comparing survey outcomes at baseline across groups (see Table 2.B.5 in Appendix 2.B).

Participation in our experiment was voluntary and subject to prior consent, which gives room to sample selection bias. Table 2.B.6 in Appendix 2.B provides information on background characteristics for our full sample and our target population, i.e., all social assistance claimants who were eligible to participate in the first place. Testing for differences between our sample and the target population, we find that, on average, our sample is slightly older, higher educated, more likely to have a Dutch background, and more likely to receive benefits as a single-person household. Moreover, our subjects have shorter welfare histories and are more often assigned to classes I and II. These findings suggest that individuals in our sample have a more favorable labor market position compared to individuals in the target population on average. Our estimates are thus valid for this particular sample of participants, but they may not be valid for the target population of all eligible claimants. In Section 2.7, we assess to what extent our estimates are externally valid if we assume selection in our sample based on observables. To do so, we transform our sample into a pseudo-population using inverse probability weights based on the background characteristics listed in Table 2.1.

2.5 Implementation

2.5.1 Implementation Protocol

The welfare office carried out the implementation of the two treatments. To prevent special treatment by caseworkers, the group assignment of control subjects was not visibly registered in the client management system. Thus, caseworkers could not distinguish control subjects from recipients that were not participating in the experiment.

The experimental protocol required that subjects in the exemption group were not actively approached by the welfare office and not monitored and sanctioned regarding compliance requirements. To prevent accidental deviation from protocol, a pop-up message in the client management system warned and reminded caseworkers about the special status of claimants in that specific treatment group. As mentioned previously, the welfare office contacted subjects at the beginning to inform them about the possibility of opting out of employment services. At this point, 93 percent of subjects decided to do so. There were no opt out or opt in decisions at a later stage.

To carry out the intensive counseling scheme, the welfare office detached six dedicated caseworkers. During the experiment, this group exclusively supervised claimants assigned to the counseling treatment and not, e.g., control subjects or non-participants. Subjects were randomly allocated to a dedicated caseworker. The internal organization and the workflows of the welfare office did not allow for a random selection or rotation of dedicated caseworkers. Instead, caseworkers were recruited among a group of internal applicants. Importantly, this setup poses the limitation that treatment effects are entangled with caseworker effects. Dedicated caseworkers were trained in advance and received peer-to-peer coaching throughout the entire experiment to minimize caseworker effects.

2.5.2 Monitoring and Implementation Checks

We closely monitored the implementation process by meeting with officials from the welfare office every other week and with the group of dedicated caseworkers every three months. In addition to that, we regularly screened data on contact frequency, reclassification, and employment services, looking for signs that indicated deviation from the experimental protocol. Table 2.2 reports a summary of that information for the entire trial period. In what follows, we briefly discuss our findings concerning implementation. In addition to to providing evidence for successful implementation, this exercise serves to present a more detailed picture of the changes in guidance and supervision introduced by the two treatments.

First, consider Panel A of Table 2.2, which reports monthly contacts (letters, e-mails, calls, and meetings) initiated by the welfare office. In line with our experimental protocol, monthly contacts decreased under exemption and increased under counseling. On average, exempted subjects were contacted nearly half as much as control subjects.[27] In contrast, under the counseling treatment, monthly contacts almost doubled compared to the control group. We must keep in mind that the welfare office may contact claimants to provide assistance as well as to monitor their behavior. Differences in monthly contracts may thus reflect changes in both dimensions. Figure 2.C.2 in Appendix 2.C provides a monthly display of the contact data. Notably, the contact intensity in the counseling group diminishes over time.

Next, we look at reclassification, i.e., reassigning claimants to a higher or lower class. We view reclassification as a proxy for the intensity of supervision, assuming that caseworkers are more likely to detect a mismatch in class after getting to know a claimant better. We find that—in line with expectations—counseled subjects were more likely to be reclassified. As shown in Panel B of Table 2.2, a quarter of counseled subjects was reassigned to a different class in the first eight months, compared to 6 percent in the control group. In all three experimental groups, reclassification is almost equally distributed in terms of direction (up and down).

Lastly, we consider employment services, which we measure as the share of claimants that received a certain type of service at least once during the first 16 months. We distinguish five types of services: formal job search assistance, skills training, work experience programs, placements with temporary employment agencies, and private sector instruments.[28] As shown in Panel C of Table 2.2, exempted subjects were on average less likely (9 percent), and counseled subjects more likely (33 percent) to receive services than control subjects (20 percent). In addition, the data reveals quite some variation in the type of services provided to different ex-

[27] Note that the contact registry also includes contact about administrative matters, which explains why contacts initiated by the welfare office did not drop to zero in the exemption group.

[28] Formal job search assistance includes programs or meetings with a focus on enhancing job search efficiency. Skills training encompasses classroom training and coaching programs. Work experience programs are temporary placements, sometimes combined with on-the-job training, aimed at gaining practical experience. Private sector instruments concern wage subsidies to employers, financial incentives offered to workers, and placements for a trial period while receiving benefits.

Table 2.2 Implementation Checks.

	Group			*p*-value		N
	Control	T1	T2	T1 vs. control	T2 vs. control	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Contacts						
Monthly contacts with the welfare office	0.69	0.37	1.31	0.00	0.00	536
Panel B: Proportion of claimants reclassified						
Up and down	0.06	0.03	0.26	0.29	0.00	529
Up	0.03	0.02	0.14	0.78	0.00	529
Down	0.03	0.01	0.12	0.23	0.00	529
Panel C: Proportion of claimants with services						
Any service	0.20	0.09	0.33	0.00	0.01	512
Job search assistance	0.14	0.03	0.13	0.00	0.87	512
Skills training	0.04	0.02	0.04	0.23	0.93	512
Work experience program	0.03	0.01	0.08	0.27	0.04	512
Placement temporary agency	0.04	0.02	0.05	0.19	0.68	512
Private sector instrument	0.07	0.05	0.16	0.37	0.01	512

Note: Column (1)–(3) report means. Column (4)–(5) show *p*-values from regressing each variable on treatment dummies, controlling for randomization strata. Panel A reports monthly contacts (letters, e-mails, calls, and meetings) initiated by the welfare office in the first 16 months. We average contact data over months with an ongoing spell for which data was available. Panel B reports the proportion of claimants reclassified within the first eight months. Panel C shows the proportion of claimants who received a certain type of service at least once in the first 16 months. Data on services is confined to subjects that had not withdrawn by month 16. All data come from the welfare office. T1: exemption; T2: counseling.

perimental groups. Differences between the control and the exemption group are large for formal job search assistance. While 14 percent of control subjects received job search assistance at least once, that is true for only 3 percent of exempted subjects. For counseling, the differences with the control group are the largest for work experience programs and private sector instruments. For these services, counseled subjects were more than twice as likely to receive them at least once compared to subjects in the control group.

In conclusion, the evidence at hand suggests that the welfare office implemented the two treatments in line with protocol, which led to considerable differences in how claimants were guided and supervised. While exempted subjects received little to no guidance and supervision, assignment to the counseling group resulted in a more intensive program on average. The differences in guidance and supervision may also be informative for identifying potential mechanisms. We will therefore come back to the findings of this subsection when discussing potential mechanisms in Section 2.8.

2.6 Empirical Strategy

Assuming linear effects, we use the following baseline specification to estimate the effects of the two treatments on administrative outcomes:

$$Y_{it} = \alpha + \delta_t^1 Z_i^1 + \delta_t^2 Z_i^2 + X_i'\Theta + \gamma + \epsilon_{it} \tag{2.6.1}$$

Y_{it} denotes the outcome of interest for individual i, measured in month t after the start of treatment. We estimate Eq.(2.6.1) separately for each month t to reveal the time dynamics of treatment effects. We consider time dynamics important given that counseling concerns a treatment of immediate exposure, while exemption leaves claimants at their own discretion. This difference may result in different effect patterns over time. Z_i^1 and Z_i^2 are treatment dummies, taking the value 1 if subject i was assigned to treatment 1 and 2, respectively. Thus, our reference category is thus subjects assigned to control. The vector X_i contains background characteristics—dummies for gender, the highest level of education (intermediate, high, and unknown), and migration background (Western, non-Western), as well as age, the duration of the current spell, and cumulative earnings in the 24 months before the invitation.[29] All control variables were measured three months before the start of treatment and are time-invariant. We include X_i to increase the precision of our estimates. γ denotes randomization strata fixed effects (Bruhn and McKenzie, 2009).[30] ϵ_{it} is the error term.

To estimate the effects on survey outcomes, we extend our model in three ways and use the following specification:

$$Y_{it} = \alpha + \delta_t^1 Z_i^1 + \delta_t^2 Z_i^2 + X_i'\Theta + \Psi Y_{iB} + \Phi M_{iB} + S_i'\Omega + \gamma + \epsilon_{it} \qquad (2.6.2)$$

First, to further increase precision, we condition on the baseline level of the outcome of interest, denoted by Y_{iB} (McKenzie, 2012). Second, to avoid dropping observations with missing baseline outcomes, we code missing values at baseline as 0 and include a dummy variable, denoted by M_{iB}, indicating missing values. Third, we include vector S_i, which contains dummies for survey mode (CAWI versus CAPI) and survey language (Dutch versus Non-Dutch). Remember that we administered two follow-up surveys (midline and endline), again to reveal time dynamics of effects. All other features of the model are the same as in Eq.(2.6.1).

Our parameters of interest are δ_t^1 and δ_t^2, which describe the effect of the respective treatment compared to care-as-usual at t time periods after the start of the treatment. As we are facing one-sided noncompliance, Eq.(2.6.1) and (2.6.2) estimate intent-to-treat (ITT) effects. ITT parameters capture the effect of *implementing* the two treatments, which is informative given that both treatments are designed as nonobligatory schemes. To identify the effects of actually receiving treatment, we also estimate local average treatment effects (LATE). However, this is only possible for administrative data, as only in that case do we observe outcomes for both registered subjects and subjects that have withdrawn from the study. We describe our LATE estimation strategy in Appendix 2.D.

To assess the sensitivity of our results, we apply two adaptions to our baseline models. First, we exclude all control variables to estimate unadjusted treatment

[29] Our results do not change if we include age squared or use the log of the duration of the current spell. Results are available upon request.

[30] Remember that we used *living situation* and *class* and as stratifying variables. Therefore, we exclude those variables from the vector X_i.

effects. Second, we estimate our parameters of interest using logistic regression instead of OLS for dichotomous outcome variables. Third, we also test whether our results are sensitive to other ways of operationalizing outcomes. In light of recent advocacy for randomization statistical inference (Gerber and Green, 2012; Young, 2019), we report two types of p-values: standard p-values, which allow us to make inferences about the null hypothesis of no average treatment effect, and randomization inference p-values, which allow us to test the sharp null hypothesis of no treatment effect whatsoever.[31] Furthermore, to account for the fact that we are testing multiple hypotheses (two treatment groups and several outcomes), we also report p-values corrected for multiple comparison using the free step-down methodology of Westfall and Young (1993).[32]

Further, as mentioned earlier in Section 2.4, we report weighted and bounded treatment effects for survey outcomes to address concerns of attrition-related bias. Finally, to address concerns of generalizability, we report results extrapolated to a pseudo-population.

2.7 Results

2.7.1 Summary

This section presents estimated treatment effects for both administrative outcomes and survey outcomes. While this section reports results, we interpret our findings in the light of potential mechanisms in the following section.

Our main findings for administrative outcomes can be summarized as follows: we find that—compared to the control group—exemption leads to higher chances of employment and self-sufficiency. Moreover, exempted subjects are more likely to find employment under a permanent contract, suggesting improvements in reemployment quality. Examining effect heterogeneity, we find that the positive effects of exemption are largely driven by female, lower educated, and younger (below 50) claimants. We find positive but more moderate and largely statistically insignificant effects on employment and self-sufficiency for the counseling treatment. Contrary to exemption, there is no evidence of improvements in reemployment quality.

With regards to survey outcomes, we find evidence that exemption especially leads to gains in experienced autonomy. Unexpectedly, we find no evidence of changes in job search behavior for both treatments. Lastly, there is no evidence that any of the two treatments had a lasting impact on social participation, health, and well-being. We perform several additional analyses and robustness checks to qualify our results.

[31]Following Young (2019), we base our randomization inference p-values on 2,000 iterations. We make use of the user-written Stata package *ritest* (Heß, 2017).

[32]Following the recommendation of Westfall and Young (1993), we base our adjusted p-values on 10,000 bootstrap draws. We calculate adjusted p-values using the user-written Stata package *wyoung* (Jones et al., 2019).

2.7.2 Employment

The primary goal of supervising benefit claimants is their re-entry into the labor market. Figure 2.2 shows the estimated effects on the probability of being employed, i.e., having positive earnings in a certain month.[33] When defining employment, earnings from self-employment are not included due to lags in reporting (see Section 2.4). The results shown in the following figures are all based on estimating the model specified in Eq.(2.6.1) separately for each of the t months before and after randomization. For convenience, we add our point estimates to the control group average, shown in percentages, together with 90 percent confidence intervals (colored areas).[34] The results for the pre-treatment period serve as an additional test for symmetry between the control and treatment groups. We find that differences in employment probabilities are not significant in that period. As will become apparent in later figures, this finding also largely holds for other outcome variables. Table 2.4 further below reports point estimates and p-values, choosing the last treatment month and the last month of observation as focal points.

We continue by discussing treatment effects for the exemption group, as shown in Panel A of Figure 2.2. Throughout the first year of treatment, we do not find evidence of effects on employment. During that time, employment probabilities in both groups, exemption, and control, increase in a similar pattern. Throughout the second year of treatment, employment probabilities in the control group remain largely stagnant at a level of roughly 17 percent. This can, in part, be explained by control subjects entering and exiting the labor force at the same rate. For comparison, among the group of roughly 7,000 claimants invited but not participating in the experiment, employment probabilities stagnate at around 15 percent during the same period.

In contrast to the control group, employment probabilities in the exemption group continue to increase. In the last month of treatment (month 19), they are an estimated 6 percentage points higher in the exemption group on average ($p = 0.092$). This corresponds with a positive effect of 36 percent relative to the control group mean. Looking at the post-treatment period, we find that employment effects are persistent and slightly increasing. One year after the end of treatment (month 31), the effect is 60 percent relative to the control group mean ($p = 0.011$). Interestingly, employment probabilities in the two groups develop more synchronous after the stop of treatment, which may reflect treatment subjects' re-entry into the case-as-usual regime. Furthermore, it is notable that only the control group shows a dip in employment probabilities around the start of the COVID-19 pandemic and the first lockdown in the Netherlands (month 22). This could indicate that exempted subjects found jobs less affected by the economic impact of the first wave of the pandemic.

Next, consider treatment effects for the counseling group, as shown in Panel

[33] For a detailed description of all outcome variables used, see Table 2.A.3 in Appendix 2.A.

[34] Our decision for 90 percent bands follows a suggestion by Romer (2020) to report somewhat narrower intervals than the usual band of two standard errors to avoid an overstated sense of uncertainty.

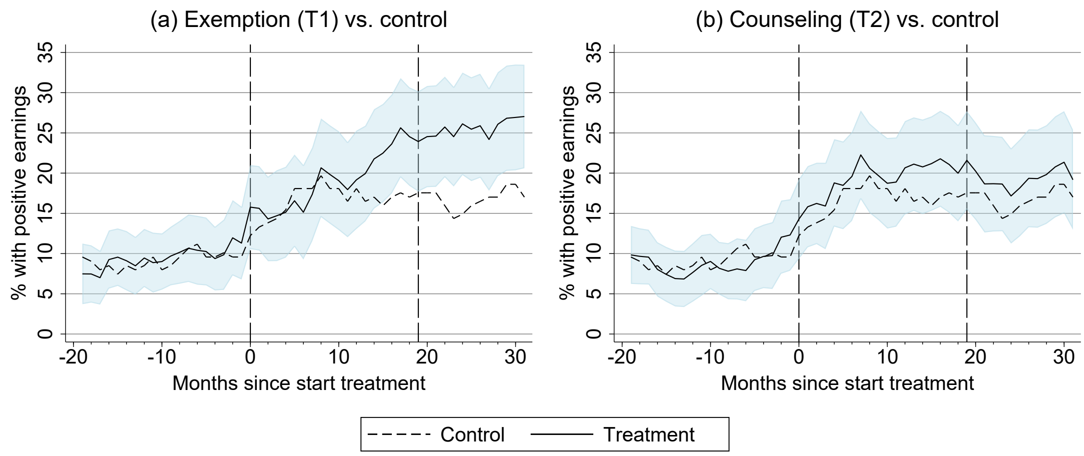


Figure 2.2 Share of Subjects Employed.

Note: Colored areas are 90 percent confidence intervals. Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(2.6.1). Dashed lines indicate the treatment period.

B. In that group, employment probabilities largely follow the same pattern as in the control group, though peaking and stagnating at a slightly higher level before converging back to the level of the control group. Post-treatment, employment probabilities in both groups show the same pattern. Effects are not statistically significant in any of the months. Compared to exemption, employment probabilities in the counseling group are lower in the last months of observation. In the last month, the difference between the two groups is an estimated 8 percentage points ($p = 0.051$).

2.7.3 Self-sufficiency

It is possible to collect partial social assistance benefits while working, e.g., when working part-time for a low wage. This means that having a job (as discussed in the previous subsection) does not necessarily lead to self-sufficiency. We now focus on the impact of the two treatments on becoming self-sufficient due to paid work. We determine self-sufficiency by comparing subjects' monthly earnings to the statutory benefit level, which lies at 70 percent of the monthly minimum wage. Earnings above that level should, in principle, lead to ineligibility for benefits.[35] Therefore, we denote subjects as self-sufficient once their earnings surpass the 70 percent threshold.

As before, Panel A of Figure 2.3 shows the results for the exemption treatment. As we would expect, the probabilities for self-sufficiency lie below those for employment. This reflects the just mentioned fact that not all employment allows for a life

[35] This threshold applies to single-person households. To simplify matters, we do not distinguish the different thresholds applying to different types of households. Instead, we use the 70 percent threshold throughout.

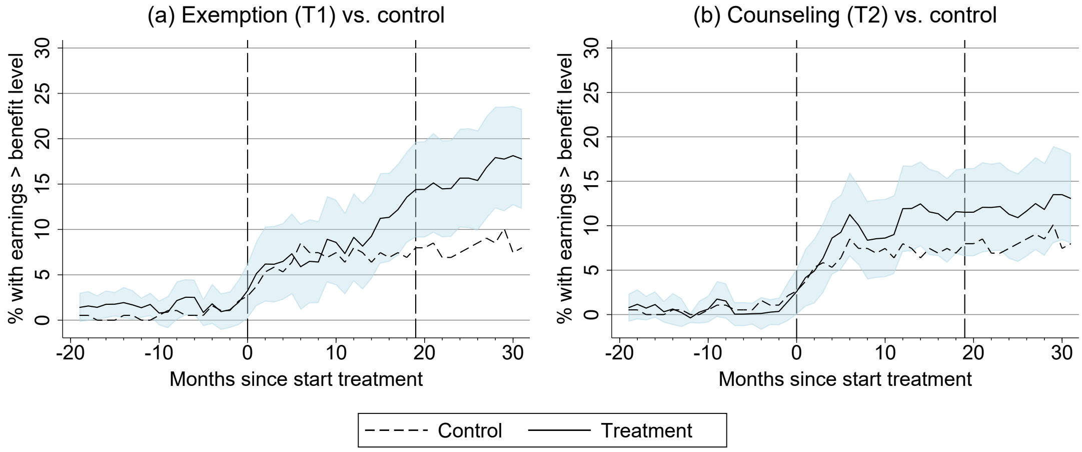


Figure 2.3 Share of Subjects Self-Sufficient.

Note: Colored areas are 90 percent confidence intervals. Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(2.6.1). Dashed lines indicate the treatment period.

independent of benefits. Overall, the patterns we observe are similar to those we saw previously for employment. While the fraction of self-sufficient claimants in the control group stagnates at around 8 percent in the longer term, it increases continuously among those exempted. In the last month of treatment (month 19), exempted subjects are an estimated 6 percentage points more likely to be self-sufficient on average ($p = 0.045$). This corresponds with a positive effect of 75 percent relative to the control group mean. Again, effects are persistent and increase in the period post-treatment. One year after the treatment stopped (month 31), the effect is roughly 120 percent relative to the control group mean ($p = 0.004$).

Next, we look at the counseling group, again shown in Panel B. Like in the control group, self-sufficiency rates stagnate in the longer term, though at a 4–5 percentage point higher level. These differences are largely statistically insignificant, though. Compared to exemption, effects for self-sufficiency appear earlier in the counseling group. In month six, e.g., the difference in effects is 5 percentage points ($p = 0.071$). Toward the end of the treatment period, the two treatment effects are statistically indistinguishable.

2.7.4 Quality of Reemployment

The two treatments might affect not only employment probabilities but also the quality of reemployment. We use the type of contract as a proxy for the quality of jobs people find, distinguishing between temporary and permanent (i.e., indefinite) employment contracts. Figure 2.4 displays effects for the former and Figure 2.5 for the latter. Note that constructing these two outcome variables, we coded the respective opposite options as zero; this means *no contract* or *permanent contract* in the case of a temporary contract and *no contract* or *temporary contract* in the

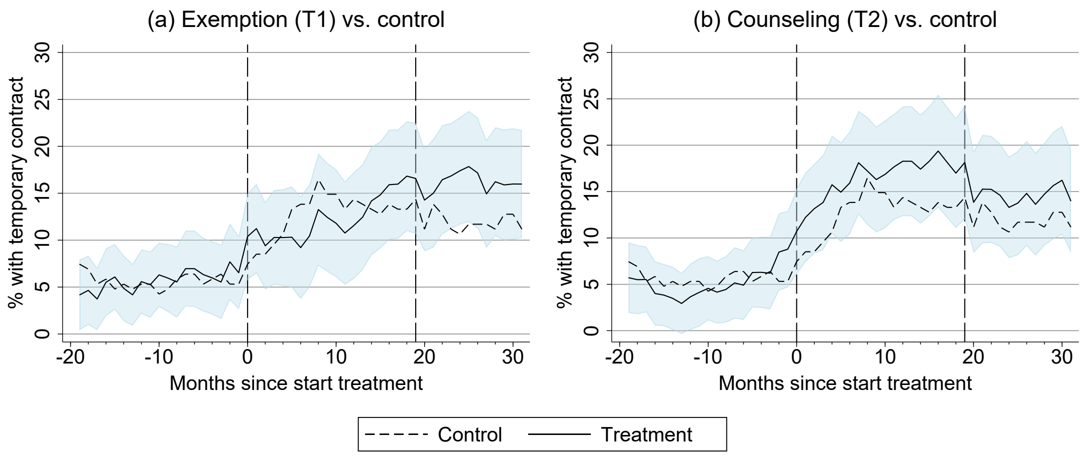


Figure 2.4 Share of Subjects With a Temporary Contract.

Note: Colored areas are 90 percent confidence intervals. Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(2.6.1). Dashed lines indicate the treatment period.

case of a permanent contract. Hence, the effects displayed in Figure 2.4 and 2.5 decompose the employment effects shown in Figure 2.2.[36]

First, consider the fraction of subjects with a temporary and a permanent contract in the control group. Comparing Figure 2.4 and 2.5 in that regard, it becomes clear that subjects in the control group almost exclusively enter temporary jobs. This seems plausible given that, e.g., vacancy referrals and job placements usually concern temporary work.

Next, consider the results for exempted subjects in both figures. Here, a different pattern emerges. With the start of treatment, the fraction of exempted subjects with a temporary contract stagnates and remains almost unchanged throughout the first year of treatment (see Panel A of Figure 2.4). Only in the second year of treatment does the share of subjects with a temporary contract increase. Yet, our estimates suggest a growth in permanent contracts (see Panel A of Figure 2.5). In the last month of treatment (month 19), the estimate on the treatment dummy is nearly 4 percentage points ($p = 0.132$). When comparing these outcomes to the employment effects presented in Figure 2.2, we find that more than 50 percent of the positive employment effects of exemption in the that month can be attributed to permanent employment. In conclusion, it appears that—in contrast to the control group—job finding among exempted subjects also took place in the domain of permanent contracts.

Contrary to exemption, we find compelling evidence that none of the employment patterns for the counseling group can be attributed to permanent contracts. Similar to the control group, changes in employment mainly concern temporary

[36]The point estimates do not add up completely due to a small fraction of contracts that is neither registered as permanent nor as temporary.

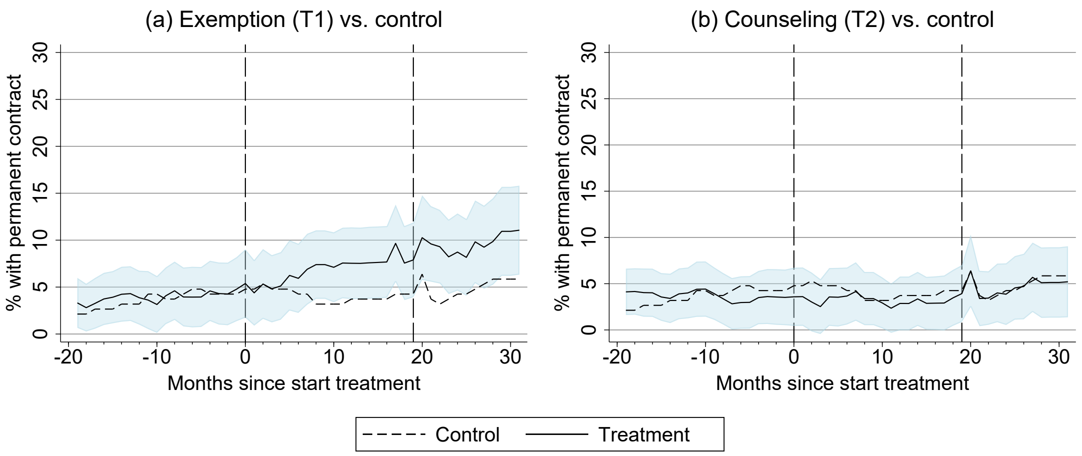


Figure 2.5 Share of Subjects With a Permanent Contract.

Note: Colored areas are 90 percent confidence intervals. Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(2.6.1). Dashed lines indicate the treatment period.

work (see Panel B of Figure 2.4). The fraction of subjects with a permanent contract remains stable throughout the entire observation window (see Panel B of Figure 2.5).

2.7.5 Survey Outcomes

Table 2.3 reports estimates for effects on survey outcomes at midline (Panel A) and endline (Panel B) using unweighted data.[37] First, consider experienced autonomy, which we measure using a survey instrument ($\alpha = 0.82$) consisting of three questions, which could be answered on a 5-point Likert scale (1-5). The questions were as follows: (i) "I have sufficient freedom to make my own choices in my search for work," (ii) "I decide for myself what I do and how I do it," and (iii) "I decide for myself when I do things." Experienced autonomy scores in the control group are relatively stable over time. At both follow-ups, control group scores average roughly 3.3 points. We find a positive effect of exemption on experienced autonomy at both midline and endline. The effects are of small-to-medium size, corresponding to 0.27 ($p = 0.042$) and 0.31 ($p = 0.011$) standard deviations, respectively. For counseling, we find no evidence of effects at midline and a positive effect corresponding to 0.22 standard deviations at endline. The effect is only significant at the 10 percent level, however ($p = 0.092$). These findings suggest that the autonomy component of the treatments translated into subjects' experiences, even though we can be less confident for counseling based on the evidence at hand.

We used two variables to measure job search behavior. First, we asked subjects to indicate if they had done anything in the past four weeks to find paid work

[37]For comparison, we report weighted estimates in Table 2.B.10 in Appendix 2.B. In general, weighted estimates are very similar to unweighted estimates.

Table 2.3 Effects on Survey Outcomes at Midline and Endline.

	Control	T1				T2				N
	Mean (SD) (1)	Coeff. (SE) (2)	*p*-value Std (3)	RI (4)	WY (5)	Coeff. (SE) (6)	*p*-value Std (7)	RI (8)	WY (9)	(10)
Panel A: Midline										
Autonomy score	3.300	0.253	0.042	0.045	0.358	0.077	0.520	0.345	0.971	327
	(0.926)	(0.124)				(0.120)				
Search past 4w	0.258	0.035	0.459	0.439	0.971	0.054	0.222	0.225	0.880	447
	(0.439)	(0.047)				(0.044)				
Hrs search/w	1.800	0.523	0.492	0.510	0.971	0.988	0.136	0.254	0.738	447
	(7.184)	(0.760)				(0.661)				
Volunteer work	0.338	0.040	0.371	0.380	0.967	-0.014	0.767	0.815	0.987	445
	(0.474)	(0.045)				(0.047)				
Bad health	0.639	-0.041	0.363	0.372	0.967	-0.106	0.015	0.177	0.154	447
	(0.482)	(0.045)				(0.043)				
Well-being	6.129	-0.016	0.932	0.941	0.994	-0.005	0.978	0.961	0.994	447
	(2.097)	(0.190)				(0.200)				
Panel B: Endline										
Autonomy score	3.360	0.308	0.011	0.014	0.123	0.215	0.092	0.011	0.617	316
	(1.011)	(0.121)				(0.128)				
Search past 4w	0.279	0.007	0.877	0.885	0.999	0.019	0.674	0.505	0.999	428
	(0.450)	(0.048)				(0.046)				
Hrs search/w	2.612	-0.513	0.521	0.490	0.994	-0.515	0.486	0.253	0.994	428
	(8.310)	(0.798)				(0.739)				
Volunteer work	0.340	-0.011	0.814	0.818	0.999	-0.011	0.832	0.658	0.999	428
	(0.475)	(0.048)				(0.052)				
Bad health	0.610	-0.006	0.906	0.901	0.999	-0.058	0.228	0.445	0.891	425
	(0.490)	(0.048)				(0.048)				
Well-being	6.120	0.075	0.715	0.680	0.999	0.146	0.493	0.333	0.994	425
	(2.083)	(0.206)				(0.212)				

Note: OLS estimates of ITT effects on midline and endline survey outcomes, following Eq.(2.6.2). The model includes strata fixed effects and controls for survey mode and language, the respective baseline value, and the background characteristics listed in Table 2.1. Outcome variables are listed on the left and described in detail in Table 2.A.3 in Appendix 2.A. Column (1) reports control group means with standard deviations in parentheses. Column (2) and (6) report the coefficients of the respective treatment dummies with robust standard errors in parentheses. Column (3) and (7) report the corresponding standard *p*-values. Column (4) and (8) report randomization inference *p*-values based on 2,000 replications. Column (5) and (9) report adjusted *p*-values using the Westfall and Young (1993) methodology and 10,000 bootstrap draws. T1: exemption; T2: counseling.

(yes/no). Second, subjects were asked to report the average number of hours per week spent on job search.[38] In both follow-up surveys, a quarter to a third of control group respondents indicate job searching in the past four weeks. Treatment effects are positive but statistically insignificant at midline, and close to zero, and insignificant at endline. Regarding hours spent on job search, the control group average is 1.8 and 2.6 hours per week at midline and endline, respectively. Point estimates are positive at midline and negative at endline, but again not statistically significant. Hence, we find no evidence for (large) adjustments in terms of job search effort. With a view to the exemption treatment, this finding stands out for two reasons. First, it contrasts with what standard economic theory would predict, namely lower levels of search effort. Second, in light of positive labor market effects of exemption (see Section 2.7.2–2.7.4), our finding may indicate that adjustments

[38]We imputed zero hours of job search for subjects that indicated not having done anything in the past four weeks to find paid work.

took place in dimensions of job search other than effort (e.g., quality).

Our last three survey outcomes concern social participation, health, and well-being. To measure social participation, we asked subjects if they were currently doing any volunteer work (yes/no); this is the case for about a third of the control group respondents at both follow-ups. We find no evidence of effects on volunteering. With a view to exemption, this finding suggests that claimants did not substitute paid for unpaid work once finding paid work was no longer enforced.

As a measure of subjective health, our surveys included the first item of the SF-36 health survey (Ware and Sherbourne, 1992). We transformed the original 5-point scale into a binary variable, indicating a *moderate* or *poor* health status.[39] At both follow-ups, around two-thirds of control respondents report a bad health status. We do not find evidence of health effects for the exemption treatment. For the counseling treatment, we find a significant negative effect at midline, suggesting an 11 percentage point lower chance of bad health ($p = 0.015$). At endline, the effect is a negative 6 percentage points but no longer statistically significant.

As a measure of subjective well-being, we took the simple average of two items, asking subjects about their satisfaction with life and to what extent they find their life meaningful on a scale from 0 to 10.[40] Control group respondents on average report a 6.1 in both follow-up surveys. We do not find evidence for effects on subjective well-being. Hence, it appears that—in contrast to what one may expect—neither exemption nor counseling led to gains in health and well-being. Also, improvements in labor market outcomes, as particularly observed for exemption, do not seem to have translated into higher levels of health and well-being.

2.7.6 Heterogeneous Treatment Effects

Examining the heterogeneity of effects, we choose self-sufficiency as the outcome of interest. We focus on this outcome, as becoming self-sufficient due to finding paid work is the most ambitious and the most relevant and salient variable from a policy perspective. We consider heterogeneous effects alongside the following dimensions of claimant background characteristics: gender (male versus female), education level (lower versus intermediate or higher), and age (younger versus older than the median age of 50). We focus on these characteristics because they have previously shown to matter for the effectiveness of activation policies (see, e.g., Card et al., 2010). Also, our limited sample size does not allow for more finely grained but smaller subgroups, e.g., claimant class.

Figure 2.6–2.8 present treatment effects for the six subgroups. The estimates for each subgroup are based on our baseline model, specified in Eq.(2.6.1), after including an interaction term between treatment and the respective subgroup. In contrast to previous figures, Figure 2.6–2.8 only display point estimates and confidence in-

[39]The item asks respondents to describe their general health and provides the answering options: *excellent*, *very good*, *good*, *moderate*, and *poor*.

[40]Both questions come from the EU Statistics on Income and Living Conditions (EU-SILC 2013) ad-hoc module 'well-being.' The life satisfaction question is comparable to the one used in the World Value Survey.

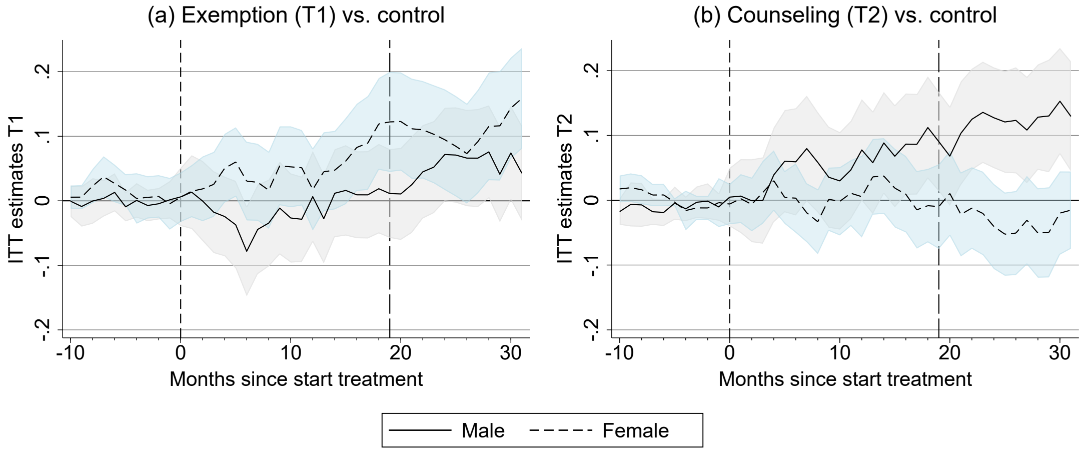


Figure 2.6 ITT Effects on the Probability of Self-Sufficiency (1/0) by Gender.

Note: Gray and colored areas are 90 percent confidence intervals. The graphs show ITT effects per gender and are estimated using separate regressions for each month, following Eq.(2.6.1), yet, including an interaction term between treatment and gender. Dashed lines indicate the treatment period.

tervals. Thus, each graph shows the estimated differences between treatment and control for a respective subgroup.

Figure 2.6 shows how treatment effects vary with gender. From Panel A, it appears that the positive effects of exemption on self-sufficiency (see Figure 2.3) are almost exclusively driven by female subjects during the treatment period. In the last treatment month (month 19), the effect is 12 percentage points for female subjects, compared to close to zero for male recipients. The difference in effect between female and male subjects in that month is marginally significant at the 5 percent level ($p = 0.051$). We can only speculate about potential explanations for this finding. Possibly, not being directed in reemployment activities increases women's chances of finding jobs that are compatible with other responsibilities, such as domestic and care work. Aside from that, compliance requirements may put a considerable strain on the group forced to combine reemployment with other responsibilities. Removing requirements may then bring about substantial motivational effects among that group. Consistent with this conjecture, we observe a reverse pattern in the period post-treatment, i.e., after re-entry into the case-as-usual regime.

Panel B of Figure 2.6 shows that, contrary to exemption, the counseling treatment appears to have only impacted male subjects. This finding is contrary to previous evidence, which indicates larger effects of activation programs for females (Card et al., 2010). A potential reason could be the type of employment services provided. Perhaps the services that made the counseling treatment successful among male recipients were largely incompatible with, e.g., care work, and therefore less suitable for female claimants. On the other hand, if indeed the guidance and assistance provided felt more supportive to male claimants, motivation crowding-in among that group may have been stronger as well.

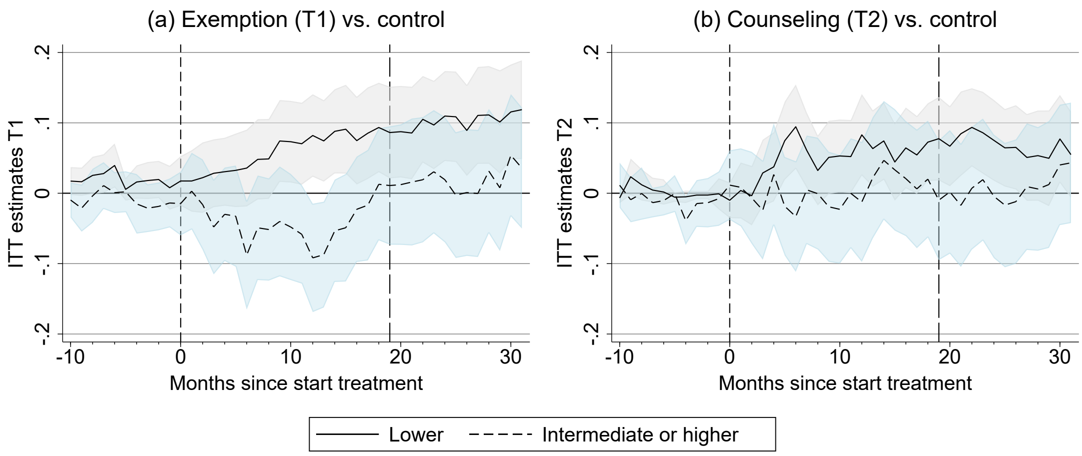


Figure 2.7 ITT Effects on the Probability of Self-Sufficiency (1/0) by Education Level.

Note: Gray and colored areas are 90 percent confidence intervals. The graphs show ITT effects per education level and are estimated using separate regressions for each month, following Eq.(2.6.1), yet, including an interaction term between treatment and education level. Dashed lines indicate the treatment period.

Figure 2.7 displays treatment effects by education level. The results shown in Panel A suggest opposite directions for the effects of exemption. While effects are positive for lower educated subjects, they are negative throughout almost the entire treatment period for intermediate or higher educated subjects. The gap is the largest after one year, amounting to 17 percentage points ($p = 0.003$). A potential explanation for this finding could be that intermediate and higher educated claimants set their reservation wages higher once compliance requirements are lifted. Similarly, the counseling treatment seems to have mostly impacted lower educated subjects. However, the effects for intermediate and higher educated subjects are close to zero instead of negative (see Panel B). Potentially, services that made the counseling treatment successful for lower-educated recipients were less attractive or suitable for their higher-educated counterparts.

Figure 2.8 shows how treatment effects vary for two age groups. For the exemption group (see Panel A), effects move parallel throughout the first year of treatment, after which a strong positive trend emerges for subjects younger than 50 years.[41] In the last month of treatment, the effect amounts to 15 percentage points for the younger group, compared to close to zero for older subjects. This difference in effect between the two age groups is statistically significant at the 1 percent level ($p = 0.002$). The fact that exemption only benefits younger recipients may suggest that older recipients face barriers to outflow that are harder to tackle by benefit recipients alone. Age discrimination might be such a problem. Additionally, the fact that counseling (see Panel B) also does not appear to work in the long

[41]Remember that subjects younger than 27 were excluded beforehand from participating in the experiment.

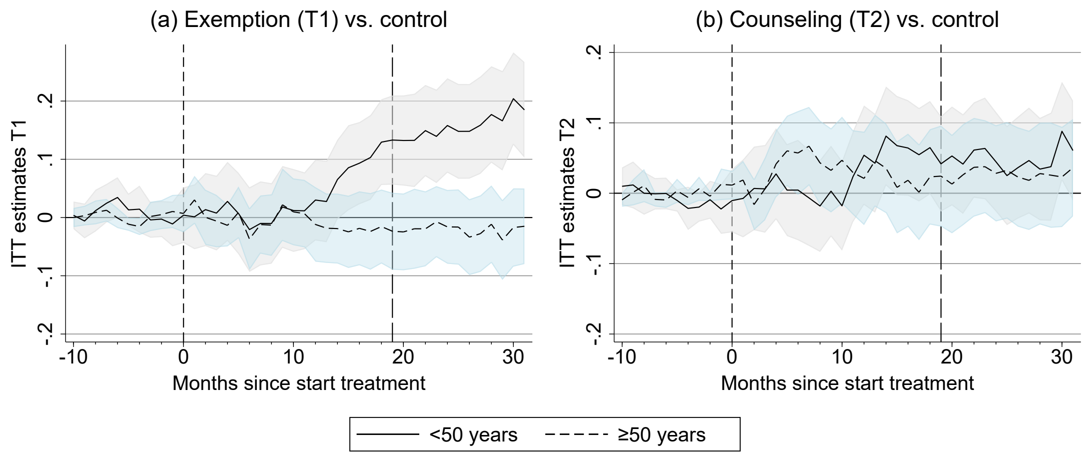


Figure 2.8 ITT Effects on the Probability of Self-Sufficiency (1/0) by Age Group.

Note: Gray and colored areas are 90 percent confidence intervals. The graphs show ITT effects per age group and are estimated using separate regressions for each month, following Eq.(2.6.1), yet, including an interaction term between treatment and age group. Dashed lines indicate the treatment period.

run signals that such problems may indeed be relevant.

In sum, our findings reveal a substantial degree of effect heterogeneity, especially in the results for the exemption treatment. It shows that correlations between the three chosen characteristics do not drive differences in effects. We test this by estimating effects jointly, i.e., using a model which includes interactions of treatment dummies and all the characteristics of interest (gender, education level, and age). We find that the patterns shown in Figure 2.6–2.8 are largely confirmed (see Table 2.B.12 in Appendix 2.B for detailed results). In conclusion, it appears that female, lower educated, and younger (below 50) recipients have particularly benefited from exemption, while counseling seems to have positively affected male and lower educated recipients. Unfortunately, our limited sample size does not allow us to provide results on a more detailed subgroup level. It is advisable for future research to explore the extent to which exemption works for different groups of claimants.

2.7.7 Additional Analyses

We perform several additional analyses to assess whether our findings hold when estimating other types of effects or using different specifications, estimation techniques, or ways of operationalizing our outcome variables.

First, we consider two robustness checks that concern inference. Table 2.4 summarizes the estimation results for administrative outcomes in the last month of treatment (month 19) and one year post-treatment (month 31). In addition to standard p-values that match the confidence bands shown in the figures above, the table also reports randomization inference p-values, which are very similar

Table 2.4 Effects on Administrative Outcomes in Month 19 and 31.

	Control	T1				T2				N
	Mean	Coeff.		p-value		Coeff.		p-value		
	(SD)	(SE)	Std.	RI	WY	(SE)	Std.	RI	WY	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Month 19 (last treatment month)										
Employed	0.176	0.064	0.092	0.109	0.395	0.040	0.277	0.287	0.643	564
	(0.381)	(0.038)				(0.037)				
Self-sufficient	0.080	0.064	0.045	0.048	0.252	0.035	0.240	0.232	0.643	564
	(0.272)	(0.032)				(0.030)				
Temp. contract	0.144	0.022	0.537	0.549	0.481	0.038	0.309	0.321	0.855	564
	(0.352)	(0.036)				(0.037)				
Perm. contract	0.043	0.036	0.132	0.141	0.779	-0.003	0.863	0.873	0.644	564
	(0.202)	(0.024)				(0.019)				
Panel B: Month 31 (one year post-treatment)										
Employed	0.170	0.100	0.011	0.010	0.057	0.022	0.568	0.590	0.812	564
	(0.377)	(0.039)				(0.038)				
Self-sufficient	0.080	0.098	0.004	0.003	0.024	0.051	0.099	0.103	0.331	564
	(0.272)	(0.033)				(0.031)				
Temp. contract	0.112	0.048	0.169	0.164	0.301	0.028	0.408	0.411	0.812	564
	(0.316)	(0.035)				(0.034)				
Perm. contract	0.059	0.052	0.070	0.065	0.451	-0.006	0.783	0.792	0.697	564
	(0.235)	(0.029)				(0.023)				

Note: OLS estimates of ITT effects on administrative outcomes in month 19 and 31, following Eq.(2.6.1). The model includes strata fixed effects and controls for the background characteristics listed in Table 2.1. Outcome variables are listed on the left and described in detail in Table 2.A.3 in Apppendix 2.A. Column (1) reports control group means with standard deviations in parentheses. Column (2) and (6) report the coefficients of the respective treatment dummies with robust standard errors in parentheses. Column (3) and (7) report the corresponding standard p-values. Column (4) and (8) report randomization inference p-values based on 2,000 replications. Column (5) and (9) report adjusted p-values using the Westfall and Young (1993) methodology and 10,000 bootstrap draws. T1: exemption; T2: counseling.

throughout. Additionally, the table reports family-wise adjusted p-values to correct for multiple hypothesis testing.[42] Unsurprisingly, given our limited sample size, only our findings for later months survive this correction. The effects of exemption on employment and self-sufficiency in month 31 are now significant at the 10 and 5 percent level, respectively, with p-values increasing from 0.011 and 0.004 to 0.057 and 0.024. The effect of exemption on permanent contracts is no longer significant, with a p-value of 0.451 instead of 0.070. For survey outcomes, as reported in Table 2.3, none of our findings survive the correction. Only the effect on experienced autonomy at endline remains close to significant at the 10 percent level, with a p-value of 0.123 instead of 0.011. Hence, we cannot reject false positives due to performing multiple comparisons very confidently. Notwithstanding, we find it relevant to report our findings, given that they suggest effects in part inconsistent with standard theoretical predictions. Naturally, we put follow-up research with larger samples high on the research agenda.

Second, as our baseline model includes control variables, we also estimate unadjusted treatment effects as a sensitivity check. We find that point estimates become

[42]Our corrections account for the number of outcomes per outcome group $\times$ 2 hypotheses. That is eight hypotheses in the case of administrative outcomes and 12 hypotheses for both midline and endline survey outcomes.

somewhat larger but do not change materially (see Table 2.B.7 in Appendix 2.B). This finding is reassuring, as it suggests that results are not largely driven by the imbalances observed at baseline. Furthermore, remember that not all subjects assigned to treatment actually received treatment due to withdrawal during the experiment. To identify the effect of actually receiving treatment, we also estimate local average treatment effects (LATE), which leads to qualitatively similar results (see Table 2.B.7 in Appendix 2.B). Finally, we estimate effects using logistic regression instead of OLS. Our results remain the same (see Table 2.B.8 in Appendix 2.B).

Third, we assess the sensitivity of our results to operationalizing some of our outcome variables differently. Figure 2.C.3 in Appendix 2.C shows estimated treatment effects using positive working hours instead of positive earnings to measure employment. We also use an alternative outcome variable for self-sufficiency based on hours worked instead of earnings. Technically, working more than 27 hours per week for a minimum wage should lead to earnings above the benefit level and full exit from social assistance. The estimated treatment effects for that outcome variable are shown in Figure 2.C.4 in Appendix 2.C. In both cases, the results do not change qualitatively. Lastly, we do not obtain qualitatively different results when using an earnings threshold of 60 percent or 80 percent instead of 70 percent to determine self-sufficiency. Figure 2.C.5 and Figure 2.C.6 in Appendix 2.C show the respective results.

Fourth, remember that our sample has different characteristics than the target population due to voluntary participation in the experiment. To address concerns about generalizability, we use inverse probability weights to account for selectivity. Admittedly, this exercise is based on the strong assumption that selection into our sample is driven by observables. Table 2.B.9 in Appendix 2.B reports unweighted and weighted estimates for comparison. For employment and self-sufficiency, weighted estimates point in the same direction and are somewhat larger than their unweighted counterparts. For the type of contract, we find stronger effects for temporary contracts and weaker effects for permanent contracts when using weights. These findings hint at an under-representation of subgroups in our sample, which profit relatively more from the treatments in terms of employment, and an over-representation of subgroups, which profit relatively more from the treatments in terms of reemployment quality.

As a final step, we perform some robustness checks for our survey outcomes. First, we use logistic regression instead of OLS to estimate effects for dichotomous survey outcome variables, which leads to qualitatively similar results (see Table 2.B.8 in Appendix 2.B). Second, as our samples of respondents differ from our full sample, we apply inverse probability weighting to account for these differences. We find that weighted results are very similar to unweighted results (see Table 2.B.10 in Appendix 2.B). Third, we bound treatment effects to preclude that our survey results are subject to attrition-related bias (see Table 2.B.11 in Appendix 2.B). Our first way of bounding effects, which is calculating Horowitz-Manski bounds, leads to an effect range far too wide to be meaningful. As can be expected, our Lee bounds, which we calculate as a second approach, are much tighter but still quite wide for the counseling treatment, where attrition rates are higher. Nonetheless, our Lee bounds suggest treatment effects that point in the same direction as unbounded effects. This leaves us confident that the survey results we interpret are unbiased regarding direction and, if anything, conservative in terms of magnitude.

2.7.8 Limitations

Naturally, our findings come with some limitations. First and foremost, our limited sample size does not allow us to estimate treatment effects more precisely. This frustrates the detection of small effects and—as mentioned before—confident inference in the light of multiple hypotheses. As a consequence, rather small or considerably large effects are often within our estimated confidence bands. Nonetheless, in some cases, we can provide quite strong evidence against *negative* effects, as is the case for the labor market effects of exemption. This is relevant, given that standard economic theory and previous evidence from comparable studies suggest such negative effects.

Moreover, our data does not include information on self-employed individuals at the moment. This might leave us with underestimated or overestimated effects on employment and self-sufficiency. This limitation seems particularly relevant because the Netherlands has one of the highest shares of self-employed workers in the European Union.[43] Data on self-employment will become available at a later point in time and therefore lends itself to follow-up research. For survey outcomes, social-desirability bias poses a potential problem. Yet, all regular precautions to mitigate this risk have been implemented, and we do not think that socially desirable responding has seriously biased our results.

Furthermore, the delivery of the treatments had to stop after 19 months due to the instructions in the government waiver. Hence, we cannot study the effects of longer exposure to treatment, and bearing these limitations in mind, we interpret our findings as "initial" effects. Also, we cannot sort out potential confounding caused by anticipating the end of treatment and following debriefings. Moreover, it was not possible to randomly choose or rotate the dedicated caseworkers tasked with delivering the counseling treatment. Consequently, we must take into consideration that the effects of counseling entail caseworker effects.

Lastly, it is important to note that our findings are valid for a convenience sample of voluntary participants, of which many received benefits for quite some time before the experiment started. As shown in the previous subsection, using inverse probability weighting to account for sample selectivity, we find confirmation of our results. Nonetheless, our results may only have limited validity for the total population of social assistance claimants in Utrecht (e.g., including claimants younger than 27) or populations at other places in the Netherlands (e.g., less urban areas). Moreover, effects may differ when being exposed to the interventions at the beginning of the welfare spell. Future research will therefore be needed to confirm the results in flow samples. Finally, it is important to note that our study took

[43] In 2017, 12.3 percent of the active labor force were own-account workers, according to Statistics Netherlands. This makes the Netherlands rank seventh within the EU-28.

place in a time of sustained economic growth with a relatively tight labor market. Different effects may emerge in an economic downturn.

2.8 Discussing Potential Mechanisms

The treatments implemented by the welfare office concern rather broad regime interventions and included changes on several dimensions. Unfortunately, that makes it impossible to identify the exact mechanisms at play. Yet, some of our findings may be interpreted as circumstantial evidence in favor or against certain mechanisms. In what follows, we will discuss potential explanations for the effects and effect patterns observed.

Regarding the exemption treatment, it appears unlikely that effects are driven by subjects cherry-picking certain services or forms of assistance. After all, our implementation checks in Section 2.5.2 revealed minimal interaction between the welfare office and exempted claimants. Therefore, we turn to the absence of requirements and supervision in search of possible explanations. On the one hand, not directing claimants in their reemployment activities may have led them to wait for better job matches and delay their exit from welfare. Negative treatment effects for intermediate and higher educated subjects provide evidence pointing in that direction. On the other hand, exemption may have reduced motivation crowding-out, possibly over-compensating for the negative effects of reduced regulation. Consistent with that reasoning is our finding that exempted claimants experienced more autonomy and that there is no evidence for large reductions in job search effort. The fact that employment chances increased regardless under exemption may indicate that claimants changed the direction of their search or adapted their search method (van den Berg and van der Klaauw, 2006). It is also thinkable that intrinsic motivation led to higher-quality job search (Koen et al., 2016). All of the above may also explain why exempted subjects were more likely to find permanent employment.

Compared to the exemption group, effects on employment and self-sufficiency appear to have occurred earlier in the counseling group. Still, they leveled off in the longer term and generally remained more moderate. This pattern may be explained by a dilution of the rather work-intensive treatment over a longer period. Consistent with that reasoning, Figure 2.C.2 in Appendix 2.C shows that the contact intensity in the counseling group diminished over time. But stagnating effects in the longer term could also be related to entering less stable employment. Notably, job finding in the counseling group was confined to temporary contracts, a pattern similar to the control group. Furthermore, Section 2.5.2 showed that employment services provided to the counseling group were primarily directed at programs promoting temporary engagements, like placements on trial or work experience programs. Potentially, vacancy referrals also largely concerned temporary work, leading to higher chances of return to unemployment in the longer term.

2.9 Conclusion

We studied a policy experiment in the Netherlands that granted welfare claimants more agency over their return to work. Claimants in the first treatment group were informed that the welfare office would exempt them from all regular compliance requirements tied to benefits receipt, such as reporting job search activities, meeting with a caseworker, or participating in reemployment programs. In contrast to what one would expect from standard economic theory, we find that exemption had positive rather than negative labor market effects. Roughly 1.5 years after the start of treatment, exempted subjects were, on average, 75 percent more likely to be self-sufficient due to finding paid work compared to subjects in the control group. We also find improvements in reemployment quality. In contrast to control subjects, which almost exclusively enter temporary employment contracts, exempted subjects also found permanent employment. We find that effects persist post-treatment. Interestingly, outcomes in the two groups develop almost synchronously in that period, which may reflect return to the care-as-usual regime. While labor market effects are positive on average, we find that they are subject to a substantial degree of heterogeneity. Exemption seems to have worked particularly well for female, lower educated, and younger claimants, while we find negative effects for those with intermediate and higher education.

For the second treatment, while formal requirements remained in place, subjects were assigned to a more intensive counseling program, shaped according to their needs and wishes. We find generally positive but less pronounced and largely statistically insignificant effects on employment and self-sufficiency for this treatment. In contrast to the exemption group, effects appear to have occurred earlier but largely level off toward the end of the observation period. Similar to the control group, job finding in this group was limited to temporary contracts. It appears that those who benefited most from the counseling treatment were male and lower educated claimants.

In addition to labor market outcomes, using survey data, we also studied the effects on experienced autonomy, job search behavior, social participation, health, and well-being. In line with expectations, we find positive effects on experienced autonomy, mostly for the exemption treatment. We find no evidence for lasting treatment effects on the remaining outcomes. Regarding job search behavior, this implies that supervision did not matter much for the effort that claimants put into finding paid work. This finding is particularly relevant for the exemption treatment, for which theory would predict lower search effort. What is more, exempted claimants seem to have achieved better labor market outcomes with the same search effort, which may hint at changes in search behavior (e.g., methods or quality) that unfortunately remain unobserved in our surveys. We deem it possible that these changes are rooted in positive motivational effects of less control. Regarding social participation, health, and well-being, our results may reflect that our interventions as such were not directly targeted at these outcomes. In particular for health and well-being, a variety of impact factors left untouched by the treatments and a short time horizon may explain why we find no evidence of effects.

In sum, our findings yield important insights for the design of income support. Our results suggest that an exemption scheme, which allows welfare claimants a measure of autonomy over their strategies to return to the labor market, outperformed both the current regulatory set-up of conditional benefits and an intensive counseling scheme set up in a cooperative fashion. This finding appears especially relevant given the regulatory simplicity and low cost of exemption compared to other policy options. It is also relevant with a view to recent debates in many welfare states on how the government should approach citizens dependent on income support.

If our findings are not sufficient to justify direct policy changes, at least they warrant larger and more elaborate studies in the same direction in the near future. Important directions for future research include further investigation of effect heterogeneity, as suggested by our findings. The use of administrative data for job search effort and more granular survey instruments on search behavior could shed more light on the channels through which exemption leads to more success in the labor market. In the same spirit, future studies should include more elaborate measures on claimants' motivation. Finally, it will be relevant to study effects on self-employment.

2.A Additional Background Information

Table 2.A.1 List of Exclusion Criteria.

Criterion	Reason
Younger than 27 years.	Government waiver not applicable.
Reaching retirement age during the experiment.	Preventing attrition due to retirement.
Receiving unemployment insurance benefits from the Dutch Employee Insurance Agency (UWV).	Government waiver not applicable.
Subject to the Natural Persons Debt Rescheduling Act (*WSNP traject*).	Government waiver not applicable.
Part-time entrepreneurs and claimants receiving benefits for the self-employed.	Government waiver not applicable.
Claimants admitted to a healthcare institution.	Treatments not applicable.
Asylum status holders with integration obligations.	Government waiver not applicable.

Table 2.A.2 List of Covariates With Description.

Variable	Description	Source
Age	Age in years at the start of treatment (June 1st, 2018).	Civil registry, Statistics Netherlands
Female	1 if subject is female and 0 if subject is male.	Civil registry, Statistics Netherlands
Education level	4 indicator variables (lower, intermediate, higher, and unknown) denoting highest education level attained based on ISCED 2011 classification (record date: October 2017). Lower: less than primary, primary, lower secondary; Intermediate: upper secondary, post-secondary, non-tertiary; Higher: short cycle tertiary, bachelor, master, doctoral.	Civil registry, Statistics Netherlands
Migration background	3 indicator variables (Dutch, Western, and non-Western) based on migration background definition of Statistics Netherlands. Dutch: person in question and parents born in the Netherlands; Western: person in question or at least one of the parents born in Europe (excl. Turkey), North America, Oceania, Indonesia, or Japan; Non-western: person in question or at least one of the parents born in none of the above mentioned regions/countries.	Civil registry, Statistics Netherlands
Current spell	Duration of the ongoing benefit spell at the start of treatment (June 1st, 2018) in months.	Benefit registry, Statistics Netherlands
Earnings 24 months before	Cumulative labor market earnings in the period March 2016 to February 2018 in euro.	Social security records, Statistics Netherlands
Living situation	3 indicator variables (single, single parent, and cohabit) denoting whether in March 2018 social assistance benefits were received as a single-person, single parent, or multi-person household.	Benefit registry, Statistics Netherlands
Class	4 indicator variables (class I, class II, class III, and class IV) denoting claimant classification as registered by the welfare office at the time of sample selection.	Claimant registry, Utrecht welfare office

Table 2.A.3 List of Outcome Variables With Description.

Variable	Description	Source
Employed	1 if subject's labor market earnings in month t are larger than zero and 0 otherwise.	Social security records, Statistics Netherlands
Self-sufficient	1 if subject's labor market earnings in month t are larger than 70 percent of the statutory full-time minimum wage of the same month and 0 otherwise.	Social security records, Statistics Netherlands
Temporary contract	1 if subject's employment contract in month t is for a definite period and 0 otherwise.	Social security records, Statistics Netherlands
Permanent contract	1 if subject's employment contract in month t is for an indefinite period and 0 otherwise.	Social security records, Statistics Netherlands
Experienced autonomy score	Survey instrument consisting of three questions to be answered on a 5-point Likert scale: (1) "I have sufficient freedom to make my own choices in my search for work," (2) "I decide for myself what I do and how I do it," (3) "I decide for myself when I do things." Cronbach's $\alpha = 0.82$.	Surveys
Job search in the past 4 weeks	1 if subject answered *yes* to the question "In the last 4 weeks, did you make any efforts to find paid work? This includes reading job advertisements," and 0 if subject answered *no*.	Surveys
Hours job search/week	Subject's answer to the question "If you consider the last 4 weeks, how many hours did you spend looking for work per week, on average."	Surveys
Volunteer work	1 if subject answered *yes* to the question "Are you involved in doing voluntary work," and 0 if subject answered *no*.	Surveys
Bad health	1 if subject answered the question "How would you describe your health" by choosing the option *moderate* or *bad*, and 0 if subject chose the option *excellent*, *very good*, or *good*. The question is based on the first item of the SF-36 health survey (Ware and Sherbourne, 1992).	Surveys
Well-being	Simple average of two items asking respondents to indicate their satisfaction with life and to what extent they find their life meaningful on a scale from 0 to 10. Both items come from the EU Statistics on Income and Living Conditions (EU-SILC-2013) ad-hoc module 'well-being.'	Surveys

2.B Additional Tables

Table 2.B.1 Survey Response Rates Across Experimental Groups and Survey Waves.

	Group			p-value		N
	Control	T1	T2	T1 vs. control	T2 vs. control	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Survey response relative to the sample						
Baseline	0.92	0.93	0.90	0.84	0.59	565
Midline	0.82	0.80	0.75	0.53	0.08	565
Endline	0.78	0.78	0.71	0.98	0.10	565
Panel B: Survey response relative to subjects approached						
Midline	0.86	0.84	0.83	0.56	0.49	529
Endline	0.83	0.87	0.82	0.30	0.73	510

Note: Column (1)–(3) report the proportion of subjects filling out a survey relative to the full sample (Panel A) or the number of subjects approached per wave (Panel B). At baseline, the full sample was approached. Column (4) and (5) show p-values from regressing a dummy for survey response on treatment dummies. T1: exemption; T2: counseling.

Table 2.B.2 Comparing Midline and Endline Respondents to the Full Sample.

	Full	Respondents		p-value	
	sample	Midline	Endline	Midline	Endline
	(1)	(2)	(3)	(4)	(5)
Age (in years)	47.0	47.2	47.2	0.51	0.40
Female	0.50	0.49	0.49	0.11	0.16
Lower education	0.49	0.44	0.45	0.00	0.00
Intermediate education	0.25	0.27	0.26	0.01	0.24
Higher education	0.23	0.25	0.25	0.01	0.02
Dutch background	0.37	0.40	0.41	0.01	0.00
Western background	0.11	0.11	0.11	0.54	0.35
Non-western background	0.52	0.49	0.47	0.00	0.00
Current spell (in months)	74.5	70.8	68.9	0.03	0.00
Earnings 24 months before (in euro)	1620.2	1845.7	1781.9	0.00	0.07
Single	0.63	0.66	0.67	0.01	0.00
Single parent	0.17	0.16	0.16	0.22	0.11
Cohabit	0.19	0.18	0.18	0.09	0.05
Class I	0.26	0.28	0.28	0.01	0.05
Class II	0.32	0.32	0.33	0.51	0.65
Class III	0.39	0.37	0.36	0.08	0.01
Class IV	0.04	0.04	0.04	0.46	0.05
Observations	565	447	428		

Note: Column (1)–(3) report means. Column (4) and (5) show p-values from a regression comparing responding subjects to attrits at each wave, i.e., regressing each background characteristic on a dummy indicating survey attrition at midline or endline, respectively. See Table 2.A.2 in Appendix 2.A for a description of variables.

Table 2.B.3 Comparing Attrits at Midline Across Experimental Groups.

	Control	T1		T2		N
	Mean (SD)	Coeff.	p-value	Coeff.	p-value	
	(1)	(2)	(3)	(4)	(5)	(6)
Age (in years)	44.3	5.12	0.02	1.40	0.52	118
	(9.2)	(2.15)		(2.19)		
Female	0.55	-0.02	0.87	0.07	0.53	118
	(0.51)	(0.12)		(0.11)		
Lower education	0.82	-0.21	0.05	-0.20	0.04	118
	(0.39)	(0.11)		(0.10)		
Intermediate education	0.09	0.15	0.09	0.08	0.29	118
	(0.29)	(0.09)		(0.08)		
Higher education	0.09	0.07	0.40	0.10	0.19	118
	(0.29)	(0.08)		(0.08)		
Dutch background	0.30	-0.07	0.54	-0.03	0.80	118
	(0.47)	(0.11)		(0.10)		
Western background	0.03	0.15	0.03	0.03	0.48	118
	(0.17)	(0.07)		(0.05)		
Non-western background	0.67	-0.09	0.45	-0.01	0.95	118
	(0.48)	(0.12)		(0.11)		
Current spell (in months)	77.4	26.03	0.16	7.11	0.68	118
	(74.9)	(18.44)		(17.48)		
Earnings 24 months (in euro)	953.7	-478.93	0.32	-83.56	0.88	118
	(2243.9)	(483.29)		(541.00)		
Single	0.42	0.23	0.05	0.09	0.45	118
	(0.50)	(0.12)		(0.11)		
Single parent	0.21	-0.05	0.56	0.04	0.66	118
	(0.44)	(0.09)		(0.10)		
Cohabit	0.36	-0.18	0.09	-0.13	0.22	118
	(0.48)	(0.11)		(0.11)		
Class I	0.15	-0.07	0.35	0.10	0.25	118
	(0.36)	(0.08)		(0.09)		
Class II	0.45	-0.09	0.47	-0.20	0.07	118
	(0.51)	(0.12)		(0.11)		
Class III	0.36	0.16	0.17	0.10	0.35	118
	(0.49)	(0.12)		(0.11)		
Class IV	0.03	-0.00	0.92	-0.01	0.81	118
	(0.17)	(0.04)		(0.04)		

Note: Differences in terms of background characteristics between the control group and each treatment group conditional on attrition at midline, estimated by regressing each background characteristic on treatment dummies restricting the sample to midline attrits. Column (1) reports control group means with standard deviations in parentheses. Column (2) and (4) report the estimated differences with robust standard errors in parentheses. Column (3) and (5) show the corresponding p-values. See Table 2.A.2 in Appendix 2.A for a description of variables. T1: exemption; T2: counseling.

Table 2.B.4 Comparing Attrits at Endline Across Experimental Groups.

	Control	T1		T2		N
	Mean (SD)	Coeff.	*p*-value	Coeff.	*p*-value	
	(1)	(2)	(3)	(4)	(5)	(6)
Age (in years)	43.2	5.73	0.01	3.65	0.07	137
	(9.1)	(2.17)		(1.98)		
Female	0.56	-0.02	0.83	0.00	0.98	137
	(0.50)	(0.11)		(0.10)		
Lower education	0.59	0.10	0.36	0.01	0.89	137
	(0.50)	(0.11)		(0.10)		
Intermediate education	0.24	-0.10	0.27	-0.01	0.93	137
	(0.43)	(0.09)		(0.09)		
Higher education	0.17	-0.02	0.77	-0.01	0.93	137
	(0.38)	(0.08)		(0.08)		
Dutch background	0.20	0.05	0.60	0.11	0.20	137
	(0.40)	(0.09)		(0.09)		
Western background	0.02	0.15	0.02	0.05	0.26	137
	(0.16)	(0.06)		(0.04)		
Non-western background	0.78	-0.20	0.06	-0.16	0.08	137
	(0.42)	(0.10)		(0.09)		
Current spell (in months)	78.0	42.15	0.03	2.89	0.85	137
	(77.4)	(19.02)		(15.70)		
Earnings 24 months (in euro)	1489.5	-1284.66	0.01	24.92	0.97	137
	(3148.0)	(506.52)		(750.50)		
Single	0.39	0.24	0.03	0.16	0.13	137
	(0.49)	(0.11)		(0.10)		
Single parent	0.22	-0.07	0.40	0.05	0.55	137
	(0.43)	(0.09)		(0.09)		
Cohabit	0.39	-0.17	0.09	-0.21	0.03	137
	(0.49)	(0.10)		(0.09)		
Class I	0.29	-0.27	0.00	-0.04	0.68	137
	(0.46)	(0.08)		(0.09)		
Class II	0.34	-0.00	1.00	-0.09	0.36	137
	(0.48)	(0.11)		(0.10)		
Class III	0.34	0.29	0.01	0.13	0.20	137
	(0.48)	(0.11)		(0.10)		
Class IV	0.02	-0.02	0.32	-0.01	0.84	137
	(0.16)	(0.02)		(0.03)		

Note: Differences in terms of background characteristics between the control group and each treatment group conditional on attrition at endline, estimated by regressing each background characteristic on treatment dummies restricting the sample to endline attrits. Column (1) reports control group means with standard deviations in parentheses. Column (2) and (4) report the estimated differences with robust standard errors in parentheses. Column (3) and (5) show the corresponding *p*-values. See Table 2.A.2 in Appendix 2.A for a description of variables. T1: exemption; T2: counseling.

Table 2.B.5 Baseline Balance of Survey Outcomes.

	Full	Group			p-value		N
	sample	Control	T1	T2	T1 vs. control	T2 vs. control	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Autonomy score	3.31	3.33	3.27	3.33	0.68	0.91	426
Search past 4w	0.36	0.33	0.41	0.33	0.20	0.88	516
Hrs search/w	3.07	3.00	3.40	2.79	0.92	0.72	516
Volunteer work	0.36	0.34	0.34	0.40	0.80	0.21	516
Bad health	0.63	0.61	0.63	0.66	0.51	0.23	516
Well-being	5.97	5.81	5.95	6.17	0.43	0.09	516
Baseline response	516	173	174	169			

Note: Column (1)–(4) report means. Column (5)–(7) show p-values from regressing each outcome on treatment dummies, controlling for randomization strata, survey language, and survey mode. See Table 2.A.3 in Appendix 2.A for a description of variables. T1: exemption; T2: counseling.

Table 2.B.6 Comparing the Target Population to the Full Sample.

	Target population	Full sample	p-value
	(1)	(2)	(3)
Age (in years)	45.8	47.0	0.00
Female	0.51	0.50	0.71
Lower education	0.62	0.49	0.00
Intermediate education	0.23	0.25	0.43
Higher education	0.11	0.23	0.00
Dutch background	0.30	0.37	0.00
Western background	0.08	0.11	0.05
Non-western background	0.61	0.52	0.00
Current spell (in months)	86.3	74.5	0.00
Earnings 24 months before (in euro)	1379.4	1620.2	0.19
Single	0.57	0.63	0.00
Single parent	0.14	0.17	0.07
Cohabit	0.28	0.19	0.00
Class I	0.17	0.26	0.00
Class II	0.24	0.32	0.00
Class III	0.57	0.39	0.00
Class IV	0.02	0.04	0.05
Observations	8,338	565	

Note: Column (1)–(2) report means. Column (3) shows p-values from a regression comparing subjects to non-subjects, i.e., regressing each background characteristic on a dummy that indicates being part of the sample. See Table 2.A.2 in Appendix 2.A for a description of variables.

Table 2.B.7 Effects on Administrative Outcomes in Month 19: ITT and LATE.

	Control	T1		T2		N
	Mean (SD) (1)	Coeff. (SE) (2)	p-value Standard (3)	Coeff. (SE) (4)	p-value Standard (5)	(6)
Panel A: Baseline model						
Employed	0.176	0.064	0.092	0.040	0.277	564
	(0.381)	(0.038)		(0.037)		
Self-sufficient	0.080	0.064	0.045	0.035	0.240	564
	(0.272)	(0.032)		(0.030)		
Temp. contract	0.144	0.022	0.537	0.038	0.309	564
	(0.352)	(0.036)		(0.037)		
Perm. contract	0.043	0.036	0.132	-0.003	0.863	564
	(0.202)	(0.024)		(0.019)		
Panel B: Model without controls						
Employed	0.176	0.084	0.049	0.054	0.191	564
	(0.381)	(0.042)		(0.042)		
Self-sufficient	0.080	0.074	0.026	0.043	0.166	564
	(0.272)	(0.033)		(0.031)		
Temp. contract	0.144	0.026	0.493	0.044	0.257	564
	(0.352)	(0.037)		(0.038)		
Perm. contract	0.043	0.053	0.043	0.006	0.796	564
	(0.202)	(0.026)		(0.022)		
Panel C: LATE model						
Employed	0.176	0.066	0.086	0.043	0.267	564
	(0.381)	(0.039)		(0.039)		
Self-sufficient	0.080	0.067	0.041	0.038	0.229	564
	(0.272)	(0.033)		(0.031)		
Temp. contract	0.144	0.023	0.529	0.040	0.300	564
	(0.352)	(0.037)		(0.039)		
Perm. contract	0.043	0.038	0.125	-0.003	0.861	564
	(0.202)	(0.025)		(0.020)		

Note: Estimates of treatment effects on administrative outcomes in month 19. Panel A reports ITT effects estimated following Eq.(2.6.1). Panel B reports ITT effects based on the same model but without controls. Panel C reports local average treatment effects estimated following Eq.(2.D.1)–(2.D.3) in Appendix 2.D. Outcome variables are listed on the left and described in detail in Table 2.A.3 in Appendix 2.A. Column (1) reports control group means with standard deviations in parentheses. Column (2) and (4) report the coefficients of the respective treatment dummies with robust standard errors in parentheses. Column (3) and (5) report the corresponding standard p-values. T1: exemption; T2: counseling.

Table 2.B.8 Effects on Administrative and Survey Outcomes Using Logistic Regression.

	Control	T1		T2		N
	Mean (SD)	Coeff. (SE)	p-value Standard	Coeff. (SE)	p-value Standard	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Administrative outcomes in month 19						
Employed	0.176	1.639	0.088	1.398	0.251	561
	(0.381)	(0.290)		(0.292)		
Self-sufficient	0.080	2.009	0.057	1.608	0.212	559
	(0.272)	(0.366)		(0.380)		
Temp. contract	0.144	1.171	0.599	1.377	0.297	561
	(0.352)	(0.300)		(0.307)		
Perm. contract	0.043	2.210	0.106	1.095	0.856	509
	(0.202)	(0.491)		(0.502)		
Panel B: Survey outcomes at midline						
Search past 4w	0.258	1.373	0.307	1.560	0.149	446
	(0.439)	(0.310)		(0.308)		
Volunteer work	0.338	1.339	0.337	0.904	0.754	442
	(0.474)	(0.304)		(0.321)		
Bad health	0.639	0.725	0.353	0.453	0.012	446
	(0.482)	(0.346)		(0.313)		
Panel C: Survey outcomes at endline						
Search past 4w	0.279	1.102	0.765	1.175	0.627	395
	(0.450)	(0.324)		(0.331)		
Volunteer work	0.340	0.966	0.910	0.921	0.804	424
	(0.475)	(0.304)		(0.332)		
Bad health	0.610	0.959	0.900	0.683	0.237	421
	(0.490)	(0.330)		(0.322)		

Note: Logistic regression estimates of ITT effects for dichotomous administrative and survey outcomes. Estimation results are based on Eq.(2.6.1) for administrative outcomes and Eq.(2.6.2) for survey outcomes. Outcome variables are listed on the left and described in detail in Table 2.A.3 in Appendix 2.A. Column (1) reports control group means with standard deviations in parentheses. Column (2) and (4) report the coefficients of the respective treatment dummies in Odds Ratios with robust standard errors in parentheses. Column (3) and (5) report the corresponding standard p-values. T1: exemption; T2: counseling.

Table 2.B.9 Effects on Administrative Outcomes in Month 19: Unweighted and Weighted Data.

	Control	T1		T2		N
	Mean (SD) (1)	Coeff. (SE) (2)	p-value Standard (3)	Coeff. (SE) (4)	p-value Standard (5)	(6)
Panel A: Baseline model						
Employed	0.176	0.064	0.092	0.040	0.277	564
	(0.381)	(0.038)		(0.037)		
Self-sufficient	0.080	0.064	0.045	0.035	0.240	564
	(0.272)	(0.032)		(0.030)		
Temp. contract	0.144	0.022	0.537	0.038	0.309	564
	(0.352)	(0.036)		(0.037)		
Perm. contract	0.043	0.036	0.132	-0.003	0.863	564
	(0.202)	(0.024)		(0.019)		
Panel B: Model with weights						
Employed	0.176	0.099	0.006	0.068	0.065	564
	(0.381)	(0.036)		(0.037)		
Self-sufficient	0.080	0.087	0.002	0.064	0.035	564
	(0.272)	(0.028)		(0.030)		
Temp. contract	0.144	0.062	0.102	0.052	0.173	564
	(0.352)	(0.038)		(0.038)		
Perm. contract	0.043	0.030	0.114	0.004	0.771	564
	(0.202)	(0.019)		(0.014)		

Note: OLS estimates of ITT effects on administrative outcomes in month 19. The results in Panel A are based on Eq.(2.6.1), using unweighted data. Panel B reports results using the same model but applying inverse probability weights based on the background characteristics listed in Table 2.1. Outcome variables are listed on the left and described in detail in Table 2.A.3 in Appendix 2.A. Column (1) reports unweighted control group means with standard deviations in parentheses. Column (2) and (4) report the coefficients of the respective treatment dummies with robust standard errors in parentheses. Column (3) and (5) report the corresponding standard p-values. T1: exemption; T2: counseling.

Table 2.B.10 Effects on Survey Outcomes at Midline and Endline: Weighted Data.

	Control	T1		T2		N
	Mean (SD) (1)	Coeff. (SE) (2)	*p*-value Standard (3)	Coeff. (SE) (4)	*p*-value Standard (5)	(6)
Panel A: Midline						
Autonomy score	3.300	0.235	0.060	0.069	0.573	327
	(0.926)	(0.125)		(0.121)		
Search past 4w	0.258	0.045	0.328	0.056	0.179	447
	(0.439)	(0.046)		(0.042)		
Hrs search/w	1.800	0.557	0.458	0.975	0.122	447
	(7.184)	(0.749)		(0.629)		
Volunteer work	0.338	0.035	0.430	-0.021	0.650	445
	(0.474)	(0.045)		(0.046)		
Bad health	0.639	-0.040	0.359	-0.105	0.014	447
	(0.482)	(0.044)		(0.042)		
Well-being	6.129	-0.027	0.887	-0.076	0.706	447
	(2.097)	(0.192)		(0.200)		
Panel B: Endline						
Autonomy score	3.360	0.281	0.021	0.196	0.128	316
	(1.011)	(0.121)		(0.128)		
Search past 4w	0.279	0.013	0.788	0.025	0.576	428
	(0.450)	(0.048)		(0.044)		
Hrs search/w	2.612	-0.541	0.481	-0.451	0.521	428
	(8.310)	(0.767)		(0.703)		
Volunteer work	0.340	-0.018	0.701	-0.010	0.840	428
	(0.475)	(0.047)		(0.051)		
Bad health	0.610	-0.001	0.985	-0.056	0.241	425
	(0.490)	(0.047)		(0.048)		
Well-being	6.120	0.097	0.646	0.138	0.521	425
	(2.083)	(0.210)		(0.215)		

Note: OLS estimates of ITT effects on midline and endline survey outcomes. Estimation results are based on Eq.(2.6.2). Observations are weighted by inverse probability weights based on the background characteristics listed in Table 2.1 to account for selective survey attrition. Outcome variables are listed on the left and described in detail in Table 2.A.3 in Appendix 2.A. Column (1) reports unweighted control group means with standard deviations in parentheses. Column (2) and (4) report the coefficients of the respective treatment dummies with robust standard errors in parentheses. Column (3) and (5) report the corresponding standard *p*-values. T1: exemption; T2: counseling.

Table 2.B.11 Treatment Effects Bounds for Survey Outcomes at Midline and Endline.

	Horowitz-Manski bounds				Lee bounds			
	T1		T2		T1		T2	
	Lower (1)	Upper (2)	Lower (3)	Upper (4)	Lower (5)	Upper (6)	Lower (7)	Upper (8)
Panel A: Midline								
Autonomy score	-1.582	1.814	-1.531	1.637	0.150	0.277	0.019	0.138
	(0.134)	(0.124)	(0.130)	(0.126)	(0.213)	(0.206)	(0.191)	(0.220)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.479]	[0.179]	[0.923]	[0.532]
Search past 4w	-0.142	0.234	-0.151	0.256	0.098	0.130	0.030	0.123
	(0.044)	(0.045)	(0.043)	(0.042)	(0.055)	(0.066)	(0.057)	(0.070)
	[0.001]	[0.000]	[0.000]	[0.000]	[0.076]	[0.048]	[0.600]	[0.079]
Hrs search/w	-14.224	15.603	-14.780	17.709	0.626	1.585	0.429	1.877
	(2.030)	(2.041)	(2.105)	(2.098)	(0.843)	(0.773)	(0.871)	(0.658)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.458]	[0.041]	[0.623]	[0.005]
Volunteer work	-0.163	0.216	-0.219	0.200	0.031	0.057	-0.005	0.089
	(0.044)	(0.043)	(0.045)	(0.044)	(0.058)	(0.065)	(0.062)	(0.070)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.592]	[0.386]	[0.933]	[0.207]
Bad health	-0.235	0.138	-0.280	0.127	-0.083	-0.051	-0.131	-0.038
	(0.044)	(0.043)	(0.042)	(0.044)	(0.065)	(0.060)	(0.070)	(0.063)
	[0.000]	[0.001]	[0.000]	[0.004]	[0.206]	[0.393]	[0.063]	[0.552]
Well-being	-1.859	1.875	-2.088	1.989	-0.059	0.238	-0.212	0.500
	(0.265)	(0.269)	(0.277)	(0.284)	(0.329)	(0.272)	(0.304)	(0.285)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.859]	[0.383]	[0.487]	[0.080]
Panel B: Endline								
Autonomy score	-1.511	1.804	-1.689	1.840	0.168	0.324	-0.240	0.454
	(0.134)	(0.119)	(0.136)	(0.125)	(0.222)	(0.195)	(0.192)	(0.183)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.449]	[0.098]	[0.212]	[0.014]
Search past 4w	-0.200	0.242	-0.239	0.258	0.071	0.073	-0.028	0.071
	(0.044)	(0.045)	(0.043)	(0.043)	(0.065)	(0.057)	(0.059)	(0.074)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.269]	[0.204]	[0.632]	[0.338]
Hrs search/w	-18.173	17.148	-20.332	19.449	-0.149	-0.102	-0.894	1.352
	(2.027)	(2.140)	(2.190)	(2.180)	(1.798)	(0.824)	(0.916)	(0.706)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.934]	[0.902]	[0.329]	[0.056]
Volunteer work	-0.220	0.221	-0.253	0.248	0.010	0.012	0.002	0.101
	(0.043)	(0.044)	(0.046)	(0.046)	(0.066)	(0.059)	(0.064)	(0.074)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.876]	[0.842]	[0.975]	[0.174]
Bad health	-0.265	0.186	-0.307	0.200	-0.046	-0.044	-0.118	-0.017
	(0.046)	(0.044)	(0.045)	(0.045)	(0.062)	(0.066)	(0.075)	(0.067)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.465]	[0.501]	[0.115]	[0.796]
Well-being	-2.077	2.458	-2.451	2.661	0.259	0.274	-0.057	0.671
	(0.279)	(0.281)	(0.290)	(0.291)	(0.314)	(0.428)	(0.329)	(0.302)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.409]	[0.522]	[0.863]	[0.027]

Note: Bounds for ITT effects. Outcome variables are listed on the left and described in detail in Table 2.A.3 in Appendix 2.A. Column (1)–(4) provide lower and upper Horowitz-Manski (HM) bounds for each of the two treatments with robust standard errors in parentheses and standard p-values in brackets. HM bounds are estimated following Eq.(2.6.2). Column (5)–(8) provide lower and upper Lee bounds, comparing unconditional means. T1: exemption; T2: counseling.

Table 2.B.12 Treatment Effects on Self-Sufficiency (1/0) With Interactions.

	Month 19		Month 31	
	T1	T2	T1	T2
	(1)	(2)	(3)	(4)
Assigned to treatment	0.109	0.171	0.149	0.178
	(0.067)	(0.066)	(0.076)	(0.077)
	[0.101]	[0.010]	[0.049]	[0.022]
Interaction terms with treatment dummy				
Female	0.092	-0.117	0.105	-0.158
	(0.064)	(0.065)	(0.067)	(0.067)
	[0.156]	[0.072]	[0.118]	[0.019]
Intermediate or higher education	-0.079	-0.094	-0.089	-0.022
	(0.065)	(0.063)	(0.068)	(0.064)
	[0.221]	[0.136]	[0.191]	[0.729]
50 years or older	-0.154	-0.063	-0.186	-0.075
	(0.061)	(0.063)	(0.064)	(0.064)
	[0.012]	[0.322]	[0.004]	[0.239]

Note: OLS estimates of ITT effects in months 19 and 31 on the probability of being self-sufficient (1/0). The model includes interactions of the treatment dummies with dummies for subjects' gender, education level, and age group. Additionally, the model controls for strata, the migration background, the duration of the current spell, and cumulative earnings in the 24 months before the invitation. The columns report coefficients for the respective treatment with robust standard errors in parentheses and standard p-values in brackets. Column (1)–(2) concern estimates for month 19, column (3)–(4) estimates for month 31. T1: exemption; T2: counseling. N=544.

2.C Additional Figures

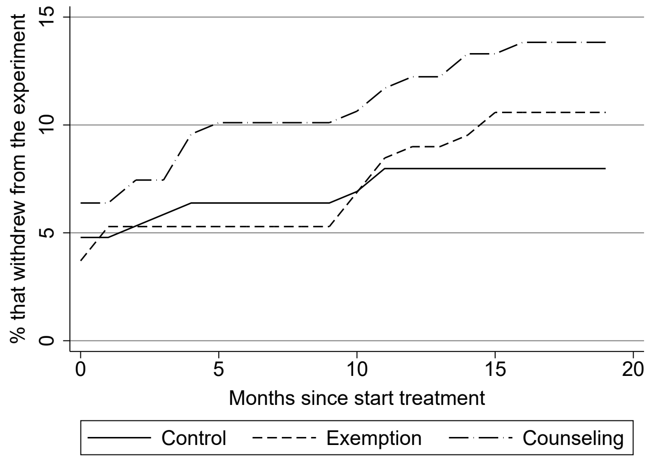


Figure 2.C.1 Cumulative Withdrawal Rates per Experimental Group.

Note: The larger than zero withdrawal rate in month 0 is related to subjects that withdrew in the four weeks between randomization and the start of treatment. The steep increases following month 8 can most likely be attributed to the midline survey, which took place at that time.

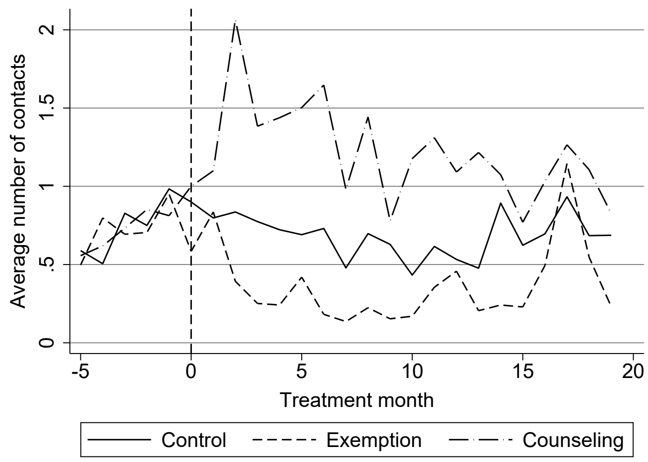


Figure 2.C.2 Average Number of Contacts per Experimental Group and Treatment Month.

Note: Contacts (letters, e-mails, calls, and meetings) initiated by the welfare office. Randomization took place in month 0. The spike in month 17 can be explained by debriefing subjects still receiving benefits. The spike in month 1 for the exemption group can be explained by contacting subjects about the possibility of opting out of employment services.

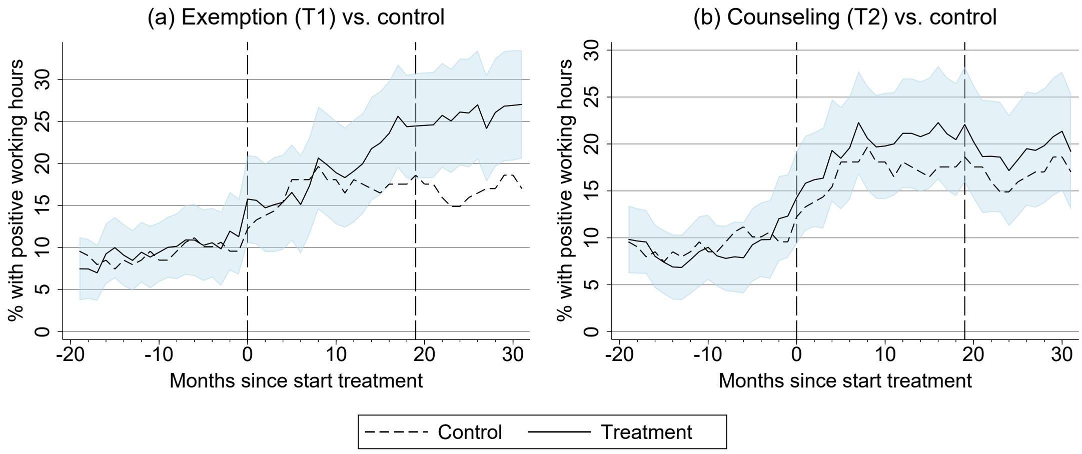


Figure 2.C.3 Share of Subjects With Positive Working Hours.

Note: Colored areas are 90 percent confidence intervals. Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(2.6.1). Dashed lines indicate the treatment period.

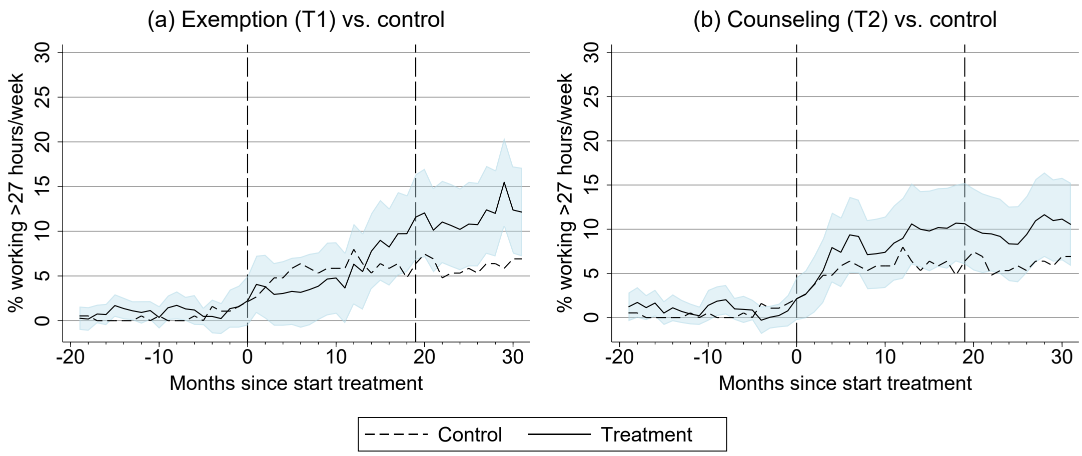


Figure 2.C.4 Share of Subjects Working More Than 27 Hours per Week.

Note: Colored areas are 90 percent confidence intervals. Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(2.6.1). Dashed lines indicate the treatment period.

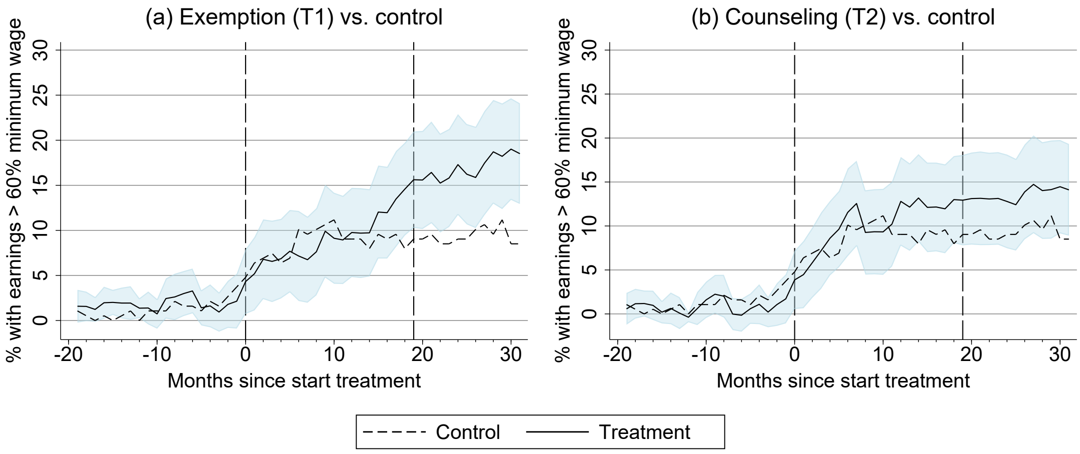


Figure 2.C.5 Share of Subjects Earning More Than 60% of the Minimum Wage.

Note: Colored areas are 90 percent confidence intervals. Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(2.6.1). Dashed lines indicate the treatment period.

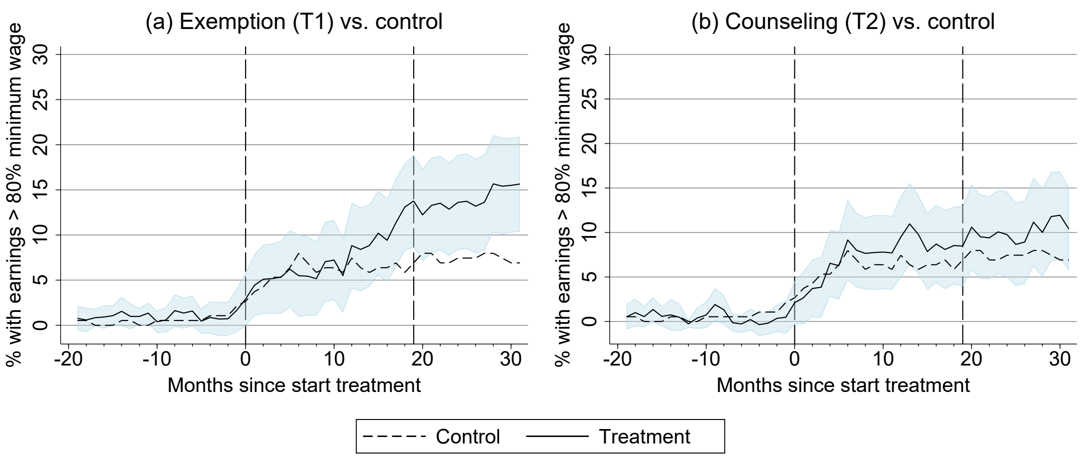


Figure 2.C.6 Share of Subjects Earning More Than 80% of the Minimum Wage.

Note: Colored areas are 90 percent confidence intervals. Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(2.6.1). Dashed lines indicate the treatment period.

2.D Estimation Strategy for Local Average Treatment Effects

We estimate local average treatment effects (LATE) by using the (randomly) assigned treatment status as instrumental variables for the actual treatment status in a two-stage least squares (2SLS) framework (Angrist et al., 1996). To obtain conservative estimates we count partially treated subjects (i.e., subjects that withdrew during the study) as fully treated (Gerber and Green, 2012). The first stage equations estimate actual treatment status, $\hat{T}_i^x$, and are given by:

$$\hat{T}_i^1 = \alpha + \mu Z_i^1 + \pi Z_i^2 + X_i' \Theta + \gamma + \nu_i \tag{2.D.1}$$

$$\hat{T}_i^2 = \alpha + \psi Z_i^1 + \phi Z_i^2 + X_i' \Theta + \gamma + \rho_i \tag{2.D.2}$$

The second stage equation estimates treatment effects and is given by:

$$Y_{it} = \alpha + \delta_t^1 \hat{T}_i^1 + \delta_t^2 \hat{T}_i^2 + X_i' \Theta + \gamma + \epsilon_{it} \tag{2.D.3}$$

All other variables remain the same as in Eq.(2.6.1). In the 2SLS framework, the parameters δ_t^1 and δ_t^2 describe the effects of treatment among *compliers*. Compliers are subjects who receive the treatment if and only if they are assigned to the treatment. As access to the treatment is generally controlled by the welfare office, there are no always treated subjects, irrespective of treatment assignment (also referred to as *always-takers*). Consequently, our LATE estimates equally reflect the average treatment effect on the treated (ATT), i.e., the average effect of treatment among all treated subjects.

2.E Discrepancies With Pre-Analysis Plan

Table 2.E.1 List of Discrepancies With the Pre-Analysis Plan.

Pre-analysis plan	Discrepancy	Location
Treatments ending in September 2019.	Treatments ran until December 2019 after the Municipality of Utrecht had decided on a three-month extension.	Section 2.3.2
Using data on contact frequency and benefit sanctions to monitor implementation.	We complemented the data on contact frequency with data on reclassifications and employment services in our implementation checks, intending to also monitor changes in treatment intensity.	Section 2.5.2
Treatment of partially treated subjects in LATE analyses.	Instead of using and reporting three different approaches to handle partial treatment (conservative estimates, upper and lower bounds, partial credit), we only estimate and report conservative estimates (treating partially treated subjects as fully treated) for space reasons. Other results available upon request.	Section 2.6
Excluded covariates (administrative).	Nationality: Collapsed into migration background variable based on migration background definition of Statistics Netherlands; Marital status: Reflected in household composition (see Table 2.A.2 in Appendix 2.A for more information on covariates).	Section 2.6
Excluded covariates (survey).	We decided to exclude covariates based on survey data due to missingness.	Section 2.6
Additional covariate: earnings 24 months before.	Added as covariate to capture employment history (claimants may have worked in addition to benefits or full-time pre-treatment) in response to suggestion from scientific supervisory committee.	Section 2.6
False Discovery Rate.	We calculated Westfall-Young step-down adjusted p-values instead of a False Discovery Rate because the former procedure allows for dependence among p-values.	Section 2.6
Difference-in-differences model.	Omitted in present version as analysis did not add new insights, but available upon request.	Section 2.6
Labor market outcomes (admin.): outflow from welfare and outflow from welfare to work.	The administrative data registering outflow from welfare proved incomplete. As an alternative, we employed data from social security records and measured employment probabilities and probabilities of being self-sufficient instead.	Section 2.7
Labor market outcomes (admin.): income from work and total income (work + benefits).	Omitted in present version as analysis did not add new insights, but available upon request.	Section 2.7
Health outcomes (admin.): medicine use, hospital visits, and GP visits.	We did not request access to the administrative data sources containing information on these three outcome variables due to a lag in reporting of up to two years.	Section 2.7
Survey outcomes and summary indices.	We decided to give less weight to survey outcomes, among other things, due to attrition at follow-up. Therefore, we refrained from collapsing variables into summary indices and instead reported effects on a subset of outcomes per domain.	Section 2.7
Additional outcome: autonomy score.	We decided to report results for autonomy scores after feedback received at a seminar.	Section 2.7
Heterogeneous effects along household composition, class, benefit duration, and capacity of work.	Omitted in present version, but analyses are available upon request, where group size is sufficient.	Section 2.7.6

CHAPTER 3

Do Earnings Exemptions Stimulate Paid Work Among Welfare Claimants?

3.1 Introduction

Benefit schemes for the unemployed pose the challenge of finding an optimal balance between the insurance component and the potentially adverse incentive effects of income support. Common strategies to restore incentives include adjusting the generosity, duration, and payment profile of benefits. Other approaches rely on activation policies, e.g., skills training, job search assistance, and monitoring. Yet another type of policy tries to make work pay, either through employment-conditional tax credits or by allowing claimants to keep a share of their additional earnings on top of benefits.[1]

The second type of policy, referred to as earnings exemptions or disregards, is a common feature of social benefit systems in many advanced welfare states (OECD, 2018). Incentivizing work in addition to benefits may increase labor force attachment of benefit recipients and facilitate transition toward full-time employment and self-sufficiency—particularly among groups of recipients with low labor market prospects, such as the long term unemployed (see, e.g., Caliendo et al., 2016; Mosthaf et al., 2021). Besides, allowing claimants to support themselves could reduce both welfare expenditures and poverty. In some countries, e.g., in the Netherlands and Germany, this reasoning has led to calls for expanding existing earnings exemption schemes—both in terms of earnings disregarded and time limits applied. However, whether more generous earnings exemptions lead to the desired results is an open question. On the downside, allowing for longer spells of work in addition to benefits could strengthen disincentives for job search and

[1] Examples of employment-conditional tax credits are the Earned Income Tax Credit (EITC) in the United States and the Working Tax Credit (WTC) in the United Kingdom. Comparable programs are in use in the majority of OECD countries (Immervoll and Pearson, 2009).

human capital investments. Moreover, higher disregards may put upward pressure on reservation wages. In this chapter, we provide novel empirical evidence on the effects of increasing the generosity of earnings exemption policies. We present the results of a policy experiment in the Netherlands that offered higher and longer exemptions to claimants of social assistance benefits.

In 2018 the Dutch Ministry for Social Affairs and Employment temporarily allowed several Dutch municipalities, which are responsible for executing the social assistance scheme, to conduct experiments with a new earnings exemption policy. The new policy deviated from the status-quo regulations in two important ways. First, it reduced the implicit tax rate on additional earnings from 75 to 50 percent, allowing working claimants to keep twice as much income (up to a certain earnings ceiling). Second, it removed a six-month time limit for using the exemption. We partnered with one of the municipalities involved, Utrecht, to study the labor market and the budgetary effects of this new policy.

We aim to answer the following research questions. First, what are the employment effects of the new exemption policy? Second, to what extent are employment effects driven by part-time work, and, correspondingly, does the policy promote or harm inflow to full-time work? Third, does the policy lead to improvements in claimants' income situation? Fourth, what is the effect on benefit expenditures?

We use data from a field experiment with 375 current claimants of social assistance benefits sampled in Utrecht to evaluate the new policy.[2] For 19 months, claimants in the treatment group were subject to the new earnings exemption policy, while the current regulatory regime applied to claimants in the control group. We collected administrative data from social security and benefit records at Statistics Netherlands to measure labor market and budgetary outcomes across the 19 treatment months and 12 months post-treatment.

Our results suggest that the more generous exemption policy stimulated employment. Toward the end of the treatment period, employment rates in the treatment group were roughly 40 percent higher than in the control group. These differences are almost entirely driven by part-time work, i.e., work that does not lead to benefit termination. Our findings also suggest temporary improvements in claimants' income situation. Full-time work does not seem to be affected by the policy change, neither during the treatment period nor thereafter. Lastly, our results suggest that, on average, the new policy did not entail additional expenses or savings for the welfare office.

[2]Note that the current study was part of a larger impact evaluation project in Utrecht, which included three treatment groups. This chapter studies one of these three treatments, namely an alternative earnings exemption policy. The other two treatments included changes in the compliance requirements claimants face and how they are supervised and counseled. Specifically, one treatment group was exempted from all requirements, monitoring, and sanctions attached to benefits receipt as well as from obligatory employment services. Subjects in the other treatment group were subject to a more intensive counseling scheme shaped according to their needs and wishes. We evaluate these other two schemes in Chapter 2 and find that exemption improves labor market outcomes while the effects of counseling are small and largely statistically insignificant. When describing the experimental design and procedures, we will limit our attention to the treatment group of interest in this chapter.

Our study contributes to understanding which design features make for an effective earnings exemption policy. In doing so, we contribute to a large body of literature examining financial incentives to work for social benefit recipients. While much of the existing work has focused on incentives conditional on employment (see, e.g., Brewer et al., 2006; Meyer and Rosenbaum, 2001), earnings exemption policies have received less attention. Some studies exploited (in part experimental) changes in incentive structures of different U.S. welfare programs in the 1980s and 1990s (see Blank et al., 1999; Greenberg et al., 1995, for an overview). These studies generally find that earnings exemptions can stimulate employment and raise incomes. In some cases, the effects on part-time and full-time employment are investigated separately, with the finding that employment effects are largely driven by part-time work (Bloom and Michalopoulos, 2001). Yet another group of studies focused on financial incentives to leave welfare altogether and work full time, finding positive results (Card and Robins, 1996; Michalopoulos et al., 2005).

We contribute to this literature in four ways. First, while previous work describes the effects of introducing or enhancing earnings exemptions, time limits have received little attention. The effect of longer exemption periods is theoretically ambiguous. We study a policy that includes a generous extension of the time limit (from six to 19 months), allowing us to extend the literature in this direction. Second, in contrast to several earlier studies, our treatment is not confounded by testing a combination of financial work incentives and other program components, such as changes in case management or an expansion of job search assistance (Blank et al., 1999). This feature allows us to identify the pure effect of changed work incentives. Third, most existing evidence is limited to specific target groups, mostly parents and single mothers receiving welfare benefits (Knoef and van Ours, 2016; Matsudaira and Blank, 2014). The benefit subject to our experiment is a nearly universal safety net program with few eligibility restrictions. This feature allows us to study impacts among a broader group of recipients. Lastly, most existing evidence stems from non-experimental work. Our evaluation builds on random assignment methods.

The chapter is structured as follows. In the following section, we briefly introduce the policy context and describe both the status-quo and the treatment policy. In Section 3.3, we formulate hypotheses which we derive from a basic labor supply framework. Section 3.4 introduces the experimental design and procedure while Section 3.5 discusses data collection and descriptive statistics. In Section 3.6, we present our empirical strategy, and in Section 3.7 our results. Section 3.8 concludes.

3.2 Background and Treatment Policy

3.2.1 Social Assistance in the Netherlands

In the Netherlands, social assistance (or social welfare) is a safety net benefit that provides a minimum income to those with insufficient means for maintaining themselves. Usually, social assistance claimants have either exhausted other types of benefits (e.g., unemployment insurance benefits) or have never been eligible for other benefits due to short work histories. Accordingly, a significant fraction of claimants has relatively low labor market prospects. Social assistance targets people who are—in principle—fit for work. There are other benefit schemes to support those who have a permanent disability. The scheme is non-contributory and foresees a monthly household-based benefit payment linked to the statutory minimum wage. Single-person households receive 70 percent of the monthly net minimum wage, whereas two-person households receive 100 percent. At the start of our study (June 2018), the benefit level was €992 ($1,279 PPP) per month for a single-person household and €1,417 ($1,827 PPP) per month for a two-person household.[3] In addition to social assistance benefits, claimants may be eligible for healthcare, housing, or child allowances.

Social assistance receipt is unlimited in duration but subject to several obligations. To begin with, claimants have to cooperate with the welfare office and disclose all information needed to determine their eligibility and benefit level. Furthermore, claimants are obliged to do everything possible to find paid work. Noncompliance can lead to temporary benefit cuts or termination of benefits altogether. Social assistance is a municipal responsibility, which means that municipal welfare offices administer claims, pay out benefits, and provide employment services. At the time of our study, 5.9 percent of Dutch households received social assistance benefits, compared to 5.5 percent in Utrecht, our study site.[4]

3.2.2 Status-Quo Exemption Policy

Social assistance benefits are means-tested, so benefit payments stop once the household income exceeds the benefit level.[5] In principle, the benefit reduction rate (BRR) lies at 100 percent. This means that benefit payments are reduced one-on-one in the case of earnings in addition to benefits. However, municipalities can introduce an earnings exemption to increase work incentives and stimulate transition from welfare to work. This is the policy at the center of our study. The earnings exemption is regulated by national legislation and thus uniform in design. Implementing the policy is subject to discretion at the municipal level.

Earnings exemptions have formed a part of the Dutch social assistance system since its introduction in the 1960s and have since then been revised and adjusted numerous times, particularly over the past 25 years.[6] The changes that led to the

[3]The benefit level increases with household size but is capped at 190 percent of the net minimum wage for households of five or more adults. Benefit levels had increased by 4 percent during our study. In the Netherlands, benefit payments are net payments. Taxes and other contributions are deducted beforehand by the respective welfare office. At the start of our study, the gross statutory minimum wage was €1,578 ($2,034 PPP) per month for employees of 22 years and older working full-time, that is 36, 38, or 40 hours per week depending on the sector. Note that we convert euros into U.S. dollars using the 2018 OECD purchasing power parity (PPP) exchange rate (OECD, 2021d).

[4]According to Statistics Netherlands and the Municipality of Utrecht.

[5]In addition to that, claimants' assets may not exceed a certain level, and claimants have to be legal residents.

[6]For a review of changes in the period 1992-2002, see Hoff and Jehoel-Gijsbers (2003).

current configuration stem from a parliamentary debate on a new social assistance act in 2003/2004, which discussed the re-introduction of a general earnings disregard. With different design options on the table, today's configuration was expected to strike the best balance between incentives to transition to full-time work (independent of benefits) and incentives for working part-time in addition to benefits.[7]

The current design allows claimants to keep 25 percent of their net earnings up to a maximum of €202 ($260 PPP) per month for a maximum period of six non-consecutive months. The maximum monthly disregard is set at roughly 15 percent of the monthly net minimum wage. Once claimants reach the time limit of six months, the BRR goes back to 100 percent. Thus, claimants can increase their income by a maximum of 20 percent (€202/€992) for a maximum period of six months when working in addition to benefits. In two-person households, both household partners can claim the earnings disregard. The latest available data show that in 2015 roughly two out of three Dutch municipalities had an earnings disregard in place. On average, approximately 8 percent of social assistance claimants work part-time in addition to benefits (Divosa, 2015).

3.2.3 Treatment Policy

A government waiver issued in 2018 allowed municipalities to test an alternative, more generous exemption policy. This alternative policy constitutes the treatment under investigation in this study. The new policy aimed to increase incentives to take up part-time work and facilitate transition toward full-time work. Specifically, the waiver allowed for two adaptations of the status-quo regulations: lowering the benefit reduction rate from 75 to 50 percent and removing the six-month time limit for applying the disregard. Notably, the maximum monthly disregard of €202 ($260 PPP) was maintained. Hence, claimants were allowed to keep more of their earnings under the alternative scheme but reached the maximum disregard earlier. Even though claimants were still bound to a maximum income increase of 20 percent, they could maintain their higher income until the end of the policy experiment, i.e., for a maximum of 19 months instead of six months. In the following section, we will study the work incentives provided by both the status-quo and the treatment policy.

3.3 Theory and Hypotheses

3.3.1 Labor Supply Framework and Budget Constraints

In what follows, we use a basic static labor supply framework to formulate hypotheses about treatment effects. As is common in such a model, we assume that subjects can choose the number of hours they want to work. Furthermore, we assume that

[7]For more information on the different design options discussed, see Note 1 of Amendement Bruls no. 44 from August 27, 2003 (identifier: h-tk-20022003-4983-4983).

wages are fixed and equal to the statutory minimum wage. For simplification, we also assume that subjects are single-person households.

Figure 3.1 shows the hours-income space. On the horizontal axis, we denote hours worked, scaled as weekly working hours. We denote total income (welfare benefits + earnings) on the vertical axis. To facilitate the interpretation, we consider net payments and display total income as a percentage of the monthly net minimum wage for a full-time (40 hours/week) job. Accordingly, point C in Figure 3.1 marks the point at which individuals work full-time and earn the full monthly net minimum wage. The figure includes three stylized budget constraints that are created by three different policies: (i) no exemption policy, represented by line segment ABC, (ii) the standard exemption policy, represented by line segment AFGBC, (iii) and the treatment policy, represented by line segment AEGBC. We will now discuss these budget constraints.

First, consider the budget constraint in the case of no earnings exemption policy (ABC). The claimant receives social assistance benefits amounting to 70 percent of the net minimum wage, and any earnings on top of benefits reduce transfers by the same amount. Benefits are terminated at the break-even point B, which is reached when working roughly 28 hours per week at the hourly minimum wage of roughly €8 ($10 PPP).

Introducing the regular earnings exemption shifts the claimant's budget constraint to AFGBC for a maximum period of six months. During that time, the benefit reduction rate (BRR) reduces from 100 to 75 percent until a maximum disregard of 15 percent of the net minimum wage is reached. Thus, labor market earnings in addition to benefits increase total income up to point F. Between point F and G, additional earnings reduce the benefit payment one-on-one, which leaves total income unchanged. Once claimants surpasses point G, they can no longer claim benefits (or the disregard), which puts them on the no-welfare budget constraint BC. This policy feature introduces a notch at point G.

Assignment to the treatment group shifts the budget constraint to AEGBC until the end of the policy experiment. Reducing the BRR from 75 to 50 percent while leaving the maximum disregard unchanged essentially creates a wedge above segment AF. This wedge introduces ambiguous work incentives, which we will now discuss in greater detail, considering different labor market outcomes. For simplicity, we limit our analysis to incentives within the social welfare system, disregarding, e.g., taxes and other types of benefits and allowances. We also ignore potential administrative hurdles when working in addition to benefits. However, it is important to note that claimants could not be worse off under the treatment policy than the status-quo regime in terms of income (any earnings and any benefits).

3.3.2 Labor Market Effects

First, consider the decision to work or not, for which we take subjects located at point A. Subjects at that point work zero hours and receive the full benefit payment. The incentives at point A are most interesting in our context, as many claimants do not work in addition to benefits. Without an earnings disregard, subjects at point

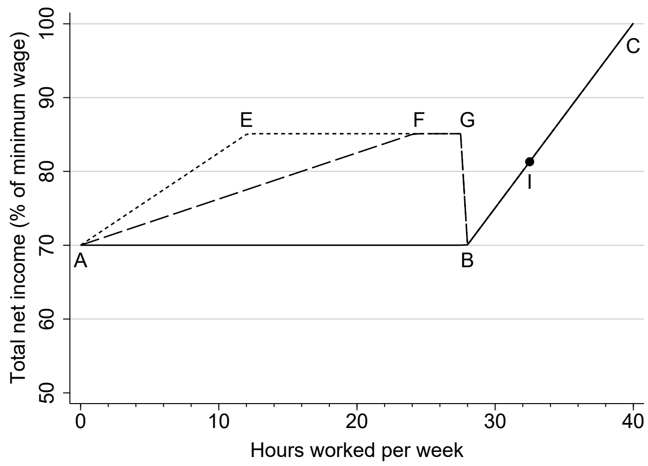


Figure 3.1 Budget Constraints.

Note: Shown is the budget constraint under no exemption policy (ABC), the standard policy (AFGBC), and the treatment policy (AEGBC).

A have no incentive to enter the labor force for some hours per week, as earnings would reduce their transfer by the same amount.

Introducing an earnings disregard could lead to two outcomes, depending on the rate at which subjects are willing to give up leisure for income while maintaining the same level of satisfaction. A subject with strong preferences for leisure might decide to remain working zero hours. A subject willing to trade off leisure and income more easily might decide to start working and move to a point right of A. In doing so, the subject increases total income until point F. It is easy to see how the treatment policy provides stronger incentives to start working for subjects at point A. Thanks to the lower BRR, entering the labor force now becomes attractive for subjects that—given their preferences—would have previously decided to remain non-working.

We thus expect to find higher employment rates among those treated. Moreover, given the extended time limit under the treatment policy (19 months instead of six months), the incentive to work remains in place longer. For this reason, we expect to find longer employment spells among treated subjects and to see effects on employment probabilities increase over time.

Next, consider part-time and full-time employment.[8] Both exemption policies only incentivize part-time work in addition to benefits. They do not provide a direct incentive to enter a full-time job. Nonetheless, we may expect treatment effects on full-time employment. On the one hand, allowing claimants to work part-time in addition to benefits for longer may increase their chances of transitioning to a full-time job, e.g., if it takes more than six months for stepping stone effects to

[8]For the context of our study, we define full-time employment as work that leads to independence of benefits (segment BC) and part-time employment as work in addition to benefits (segment AB).

materialize. On the other hand, the treatment may hamper inflow in full-time employment, e.g., if longer part-time spells lead to lower search effort and higher reservation wages. Also, treated subjects reach their maximum income at lower weekly working hours (point E instead of F). This feature may make it harder for them to switch to full-time employment by increasing their working hours. In conclusion, we expect that employment effects are driven by part-time work. Given that mechanisms are pulling in opposite directions, the effects on full-time employment will essentially be an empirical question.

Finally, consider earnings, which may also be affected in different directions. While higher employment rates should increase average earnings, treated subjects reach maximum utility at lower working hours (point E instead of F). For this reason, treated subjects entering employment may do so for fewer hours, while treated subjects working more than 12 hours a week may reduce their working hours. Treated subjects working less than 12 hours a week may decide to lower or increase working hours depending on whether substitution effects dominate income effects or vice versa. In addition, effects on full-time employment may lead to differences in earnings. In conclusion, the net effect on earnings will mainly depend on the magnitude of employment effects and subjects' location on the budget constraint.

Note that we did not discuss incentives for subjects working full-time to reduce their working hours and start collecting benefits. This concerns subjects at point I, which would be better off by working part-time and entering social assistance. We assume that such entry or eligibility effects (Ashenfelter, 1983) are of little actual relevance in our study context. First, the treatment policy does not alter incentives for subjects at point I. Second, including only current social assistance recipients, the setup of the experiment does not allow us to document entry effects.

3.3.3 Budgetary Effects

Earnings exemptions constitute costs for the welfare office due to the foregone benefit reduction. Hence, introducing the wedge above segment AF and extending the eligibility period should increase benefit expenditures. Employment leads to lower benefit expenditures, either because subjects exit the scheme altogether (full-time work) or because their labor market earnings reduce benefit payments according to the benefit reduction rate in place (part-time work). Thus, to the extent that the more generous exemption stimulates employment, the welfare office may realize savings. Similar to the effects on earnings, it will be an empirical question of whether benefit expenditures are positively or negatively affected.

3.4 Experimental Design

3.4.1 Procedure and Timing

We evaluate the more generous exemption policy through a field experiment with claimants currently receiving social assistance in the city of Utrecht. We distinguish two experimental groups: a control group subject to the status-quo policy and a treatment group subject to the alternative policy configuration. We decided to randomize a stock sample of claimants as we deemed the validity of the government waiver too short to recruit a sufficiently sized sample based on new inflow. Importantly, the government waiver required that claimants give consent before subjected to the alternative policy. Following that requirement, we sampled and randomized subjects in two steps. First, we recruited a sample of social assistance claimants consenting to be included in the study. Second, we randomly allocated that sample to our experimental groups. We acknowledge that this procedure is likely to produce a selective sample of subjects and pay attention to this particular issue in Section 3.7.4.

Various groups of claimants were not eligible for participation in the study.[9] For example, claimants younger than 27 are excluded by law from any earnings exemptions, a regulation the government waiver did not change. Eventually, we counted 8,338 claimants eligible for participation (around two-thirds of the total population of social assistance claimants at that time), and we recruited subjects among that group of claimants. Our recruiting campaign was scheduled for two months in early 2018 and included personal invitation letters to all eligible claimants, posters, advertisements, and information desks at the welfare office. On the day of randomization, 752 claimants (roughly 9 percent of those eligible and invited) had consented to participate in the study and were included in the randomization.

We allocated subjects to experimental groups using a stratified randomization design. Stratified randomization is a method to increase the precision of treatment effect estimates. For this procedure, subjects are first grouped into different strata based on background information and then allocated to experimental groups within each stratum. We used two background variables to form our strata: *living situation*, indicating whether a claimant receives benefits as a single-person household or not, and *class*, which is a four-level classification applied by the welfare office indicating claimants' labor market prospects.[10] We allocated subjects to experimental groups within each stratum with equal probabilities.[11] 188 subjects were allocated to the control group and 187 to the treatment group, which leaves us with a sample of 375 subjects. The remaining 377 subjects were allocated to the two treatment groups that are part of the same impact evaluation project but subject to Chapter 2.

The new earnings exemption policy came into effect in June 2018 and lasted for 19 months, until December 2019. The end date was laid down in the government waiver. Hence, the treatment stopped before the COVID-19 pandemic unfolded. We informed treatment and control subjects about their group assignment two weeks before the start of treatment. The information letters sent to treatment

[9]See Table 3.A.1 in Appendix 3.A for a full list of exclusion criteria.

[10]At the time of randomization, we only had access to limited data on claimant background characteristics which we obtained from the welfare office. Among the available information, we chose the two variables for stratification that we deemed most predictive for employment chances.

[11]We used the user-written Stata package *randtreat* for randomization (Carril, 2017). The randomization syntax is available upon request.

subjects again laid out the new policy regulations.[12] In addition to that, the welfare office organized two information events for treatment subjects where caseworkers explained the new regulations and subjects could ask questions. Claimants' group assignment was upheld when exiting and re-entering social assistance. From month 17 on, the welfare office contacted subjects in the treatment group still receiving benefits to inform them about the upcoming termination of the exemption period and potential consequences given their situation. Figure 3.2 provides a complete timeline of the study.

Lastly, note that both experimental groups were surveyed three times as part of the larger evaluation project. The surveys took place before randomization, after eight months, and after 16 months. We will not make use of survey data for this study. Chapter 2 provides more information on the surveys.

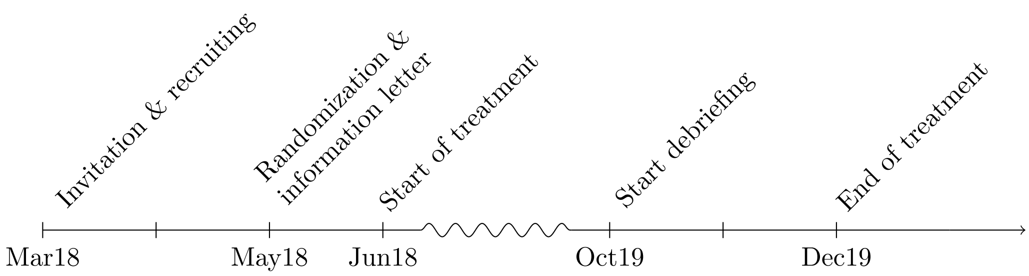


Figure 3.2 Study Timeline.

3.4.2 Implementation

The welfare office of the Municipality of Utrecht was responsible for implementing the treatment. Implementation required changes to the automatic benefit payment administration system. Furthermore, subjects assigned to treatment were labeled in the welfare office's client management system. In this way, caseworkers responsible for benefit payment administration would know which scheme to apply in case of questions from claimants. Claimants in the control group were not visibly labeled. The treatment did not include changes in the way claimants were supervised and counseled. Claimants in both experimental groups continued on their existing trajectories.

We had regular meetings with the project team at the welfare office to monitor implementation. Additionally, we used data obtained from the welfare office to check if subjects in both groups received the same treatment in terms of supervision and counseling. Contact registration data reveals that—on average—the welfare office contacted subjects in both groups with the same intensity. Data on reclassifying claimants (i.e., changing their status regarding labor market prospects in the client management system) suggest equal treatment in terms of supervision.

[12] All communication material, which is only available in Dutch, can be accessed at `https://dspace.library.uu.nl/handle/1874/395951`.

Table 3.1 Implementation Checks.

	Group		p-value	N
	Control (1)	Treatment (2)	Treatment vs. control (3)	(4)
Monthly contacts with the welfare office	0.69	0.80	0.20	351
Proportion of claimants reclassified	0.06	0.08	0.33	351

Note: Column (1)–(2) report means. Column (3) shows p-values from regressing each variable on a treatment dummy, controlling for randomization strata. Monthly contacts refer to contacts (letters, e-mails, calls, and meetings) initiated by the welfare office in the first 16 months. We average contacts over months with an ongoing spell for which data was available. The proportion of claimants reclassified includes upward and downward classification within the first eight months. All data are obtained from the welfare office.

We summarize both findings in Table 3.1. Our findings suggest that assignment to the treatment group only entailed administrative changes and no changes in supervision and counseling. As the content of caseworker meetings are confidential, we cannot investigate if claimants in the treatment group were explicitly steered toward part-time work by their caseworkers. However, we picked up no such signals in discussions with caseworkers and the project team at the welfare office.

3.4.3 Compliance

Participation in the experiment was voluntary. This means that subjects could withdraw from the study at any time, introducing one-sided noncompliance (or failure-to-treat) into the experiment. In case of withdrawal, subjects fell under the status-quo exemption policy again. During the study, 14 percent of subjects in the treatment group used the withdrawal option. In the control group—where withdrawal did not lead to a change in treatment—9 percent of subjects stopped participating. The difference in chance of withdrawal between the two groups is significant at the 10 percent level ($p = 0.072$). Withdrawal does not lead to missing data as the anonymized data sources of Statistics Netherlands allow us to follow subjects regardless of their participation status. When estimating treatment effects, we will account for one-sided noncompliance by estimating intent-to-treat (ITT) effects. We rule out accidental treatment of control subjects (two-sided noncompliance) given that the welfare office controlled access to the treatment.

3.5 Data and Descriptive Statistics

3.5.1 Data Sources and Data Collection

Using unique personal identifiers, we can link our subjects to data available at Statistics Netherlands. We collect data from social security records, that is, monthly information on employment and earnings. Additionally, we obtain data on social assistance benefit payments from the benefit registry. The civil registry at Statistics Netherlands provides us with information on socio-demographics, including gender, age, the highest education level, household composition, and migration background.

From the welfare office, we obtain information on claimants' class assignments.

The retrospective nature of the data at Statistics Netherlands allows us to collect information on subjects' labor market outcomes before treatment. We use this information to construct covariates reflecting subjects' welfare (duration of the current spell) and employment histories (earnings in the 24 months before the invitation). In addition, the data allows us to compare outcomes ex ante, which we regard as an additional test for symmetry between the experimental groups. Eventually, we construct a longitudinal data set with monthly observations, following subjects 19 months pre-treatment, during the 19 treatment months, and 12 months post-treatment (October 2016–December 2020).

As the administrative records at Statistics Netherlands are complete, we do not encounter attrition. An important data limitation concerns labor market information on self-employed individuals, which is not (yet) available at the time of writing this dissertation.[13]

3.5.2 Outcome Variables

We are interested in the labor market and the budgetary effects of the more generous earnings exemption policy. To study employment effects, we construct a set of binary outcome variables indicating different types of employment. The first variable indicates general employment and takes the value 1 if—in a given month—a subject has worked more than zero hours, and 0 otherwise.

The second and third variables indicate full-time and part-time employment, respectively. As a threshold to distinguish full-time and part-time work, we use the kink at point B in the budget constraint shown in Figure 3.1. We thus define full-time work as work that technically leads to independence from benefits and part-time work as work in addition to benefits. Specifically, our variable indicating full-time employment takes the value 1 if—in a given month—a subject has worked at least 28 hours a week, and 0 otherwise.[14] Our variable indicating part-time employment takes the value 1 if—in a given month—a subject has worked between 1 and 27 hours a week, and 0 otherwise. By construction, treatment effects for these two variables decompose the treatment effects on general employment.

In addition to these binary outcomes, we construct three continuous outcome variables.[15] Those are: (i) labor market earnings (incl. zero earnings), (ii) welfare benefits, and (iii) total income (earnings + welfare benefits). These outcomes allow us to assess the effects on benefit expenditures and claimants' income situation. Table 3.A.3 in Appendix 3.A lists and describes all outcomes variables.

[13] With self-employed individuals, we refer to own-account workers.

[14] As we do not observe weekly working hours, we have to determine part-time and full-time work based on hours worked per month. We convert weekly hours into monthly hours by multiplying the former by the factor 4.33 (52/12). In this way, 28 hours per week are converted into 121.24 hours per month, for instance.

[15] To limit extreme values in continuous variables, we replace all data above the 99th percentile by the 99th percentile (winsorization).

3.5.3 Descriptive Statistics and Balance

Column (1) of Table 3.2 provides information on background characteristics for our target group of 8,338 eligible claimants. Column (2)–(4) report the same information, but then for our full sample and separately for each experimental group. The descriptive information in Column (2) shows that our sample is quite heterogeneous. Half of our sample is lower educated, i.e., no more than lower secondary education. The other half splits up almost equally into intermediate and higher educated subjects.[16] One-third of our sample has a Dutch background. Heterogeneity regarding labor market prospects shows in the distribution across claimant classes. To tailor their employment services, the welfare office groups claimants into four broad categories: (i) expected to find a job within three months (class I), (ii) expected to find a job within three to 12 months (class II), (iii) currently unfit for work (class III), and (iv) occupational disabilities (class IV). Among our sample, 25 percent are classified as class I, while a third fall into class II and roughly 40 percent into class III. The average spell duration of roughly 6.5 years shows that many claimants have extensive welfare histories. Lastly, two-thirds of our sample are single-person claimants.

Comparing our sample to the target group, we find that the two groups differ on several dimensions. For instance, our subjects are—on average—higher educated, more likely to have a Dutch or Western background, more likely to live in a single-person household, and classified as closer to the labor market. Statistically significant differences between the target group and our sample are indicated by the p-values in Column (5). It seems plausible that our sample has a more favorable labor market position on average, given that participation in the experiment required consent beforehand. In our analysis, we will assess to what extent the findings for our convenience sample can be generalized to the target group when assuming that selection is based on observables. We report on this exercise in Section 3.7.4.

Looking at the descriptive statistics for our two experimental groups, shown in Column (3)–(4), it appears that the two groups are mostly balanced. Statistically significant differences between the two groups are indicated by the p-values in Column (6). The only significant difference (at the 10 percent level) concerns the education level, with treatment subjects being higher educated on average. Though not statistically significant ($p = 0.106$), we find that subjects assigned to treatment have more extensive work histories, as indicated by higher cumulative earnings in the two years before the invitation. We control for these differences with the set of covariates included in our model and compare adjusted and unadjusted effects in Section 3.7.4.

[16] We classify education levels according to the International Standard Classification of Education (ISCED) 2011.

Table 3.2 Background Characteristics for Target Group, Sample and Experimental Groups.

	Target group	Full sample	Group: Control	Group: Treatment	p-value: Sample vs. target	p-value: Treatment vs. control
	(1)	(2)	(3)	(4)	(5)	(6)
Age (in years)	45.8	46.7	47.0	46.4	0.06	0.56
Female	0.51	0.49	0.49	0.50	0.50	0.91
Lower education	0.62	0.48	0.54	0.42	0.00	0.02
Intermediate education	0.23	0.25	0.23	0.27	0.52	0.38
Higher education	0.11	0.23	0.19	0.26	0.00	0.09
Education unknown	0.04	0.05	0.04	0.05	0.57	0.84
Dutch background	0.30	0.36	0.35	0.37	0.02	0.63
Western background	0.08	0.12	0.11	0.13	0.03	0.51
Non-western background	0.61	0.52	0.54	0.50	0.00	0.36
Current spell (in months)	86.3	78.4	79.1	77.7	0.04	0.84
Earnings 24 months before (in euro)	1379.4	1456.5	1149.6	1765.2	0.70	0.11
Single	0.57	0.65	0.63	0.67	0.00	
Single parent	0.14	0.16	0.18	0.14	0.37	
Cohabit	0.28	0.19	0.19	0.19	0.00	
Class I	0.17	0.25	0.25	0.25	0.00	
Class II	0.24	0.33	0.32	0.33	0.00	
Class III	0.57	0.39	0.39	0.39	0.00	
Class IV	0.02	0.03	0.04	0.03	0.14	
Joint test (p-value)						0.54
Observations	8,338	375	188	187		

Note: Column (1)–(4) report means. Column (5) shows p-values from regressing each background characteristic on a sample dummy (N = 8,338). Column (6) shows p-values from regressing each background characteristic on a treatment dummy, controlling for randomization strata (N = 375). The second last row reports p-values of a joint hypothesis test. Background characteristics reflecting household composition and claimant class were used for stratification and therefore excluded from the balancing test in Column (6). We measured all background characteristics at the time of invitation. See Table 3.A.2 in Appendix 3.A for a description of variables.

3.6 Empirical Strategy

Assuming linear effects, we use the following baseline specification to estimate treatment effects:

$$Y_{it} = \alpha + \delta_t T_i + X_i'\Theta + \gamma + \epsilon_{it} \tag{3.6.1}$$

where i indexes the subject and t the month for which we estimate the effect. To investigate effect dynamics, we estimate Eq.(3.6.1) separately for each month t. We are interested in effect dynamics as certain impacts, e.g., switches to full-time employment may take some time to materialize. Y_{it} denotes an outcome variable of interest. T_i is a dummy variable taking the value 1 if a subject was randomly assigned to the treatment group and 0 if a subject was assigned to the control group, thus, the reference group in our model.

To increase the precision of our estimates, we include a vector of covariates, denoted by X_i. X_i includes variables for age, gender, the education level (4 categories), the migration background (3 categories), cumulative earnings in the 24 months before the invitation, and the duration of the current benefit spell in months. Including these covariates, we also control for differences between the two groups in terms of observable background characteristics. To account for randomization within strata, we include strata fixed effects, denoted by γ (Bruhn and McKenzie, 2009).[17] ϵ_{it} denotes the error term.

Our parameter of interest is δ_t, which describes the effect of the more generous earnings exemption in month t. Since we do not know to what extent treated subjects were aware of the new policy (e.g., if subjects read the information letters) or understood its implications, we interpret our estimates as intent-to-treat (ITT) effects. This interpretation is also appropriate given that subjects could withdraw from the study (see Section 3.4.3). Hence, we interpret our findings as the effects of *implementing* a more generous earnings exemption in the way it was done for the experiment.

In light of recent advocacy for randomization inference, we calculate two types of p-values—standard p-values that correspond with the null hypothesis of no average treatment effect and randomization inference p-values that correspond with the null hypothesis of no treatment effect whatsoever (Gerber and Green, 2012; Young, 2019).[18]

3.7 Results

3.7.1 Employment

Figure 3.3 shows estimated treatment effects on employment probabilities. We obtain estimates by running our baseline model specified in Eq.(3.6.1) separately for each month shown. To facilitate interpretation, we plot employment rates in percentages for the control group and add point estimates with 90 percent confidence intervals (colored areas) to the control group outcomes.[19] All graphs cover the following three periods: 19 months before randomization, 19 months of treatment, and 12 months post-treatment. We include pre-treatment months into our observation period to assess the symmetry of our two groups in terms of outcomes ex ante. We find that, for most outcomes, there are no statistically significant differences pre-treatment.

First, consider Panel A of Figure 3.3, which shows effects on employment probabilities (i.e., having positive working hours). The graphs show that already pre-treatment, some subjects (roughly 10 percent) in both experimental groups were working. This is not surprising given that claimants can collect benefits to supplement income from work. In the control group, employment rates increase in the first treatment months before stagnating at around 17 percent.[20] This development

[17]As we formed randomization strata based on the two variables *living situation* and *class*, we do not include these two variables in the vector X_i.

[18]Following Young (2019), we base our randomization inference p-values on 2,000 iterations. We use the user-written Stata package *ritest* (Heß, 2017).

[19]In our decision to plot 90 percent bands, we follow Romer (2020), who suggests reporting narrower intervals than the usual band of two standard errors to avoid an overstated sense of uncertainty.

[20]We find that employment probabilities in the control group are very similar to those of the

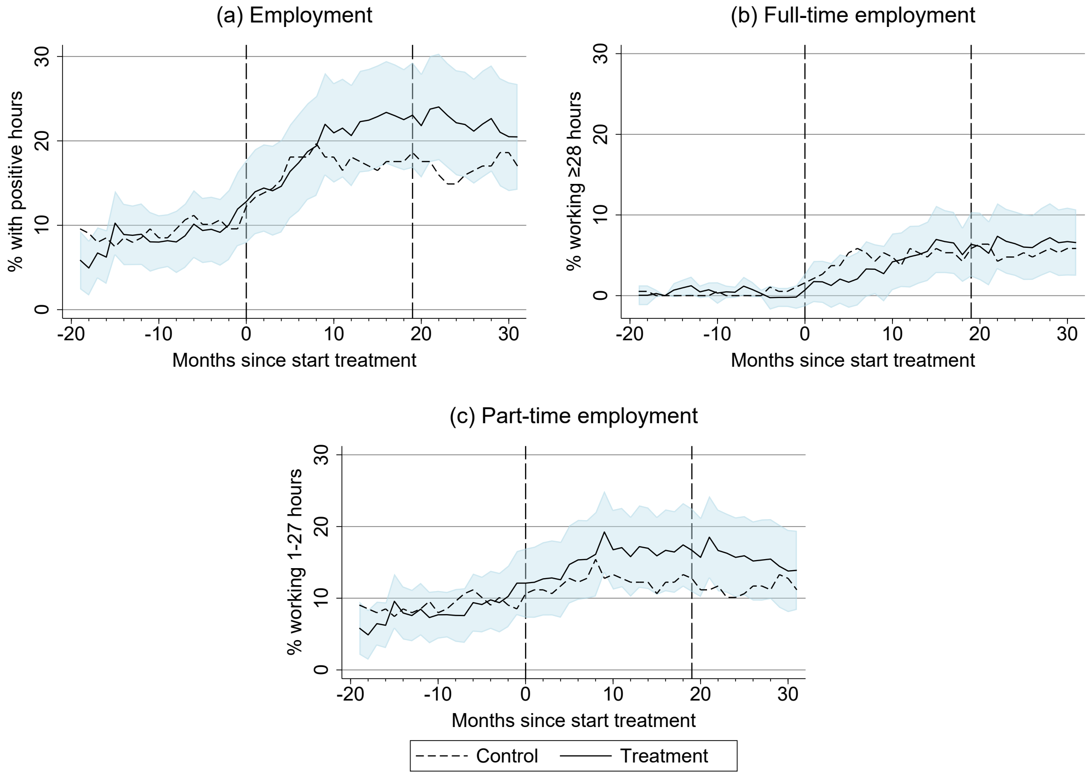


Figure 3.3 Employment Rates in the Treatment and Control Group.

Note: Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(3.6.1). Colored areas are 90 percent confidence intervals. Dashed lines indicate the treatment period.

continues in the post-treatment period, except for a brief dip around month 22. In month 22, the Netherlands experienced the first lockdown related to the outbreak of the COVID-19 pandemic, which may explain the drop in employment.

Turning to treatment effects, we find that (estimated) employment rates in the treatment group follow a mostly continuous upward trend, leading to larger differences between the two groups over time. Effects reach up to 40 percent relative to the control group mean but are only statistically significant (at the 10 percent level) in one month. Toward the end of the post-treatment period, employment probabilities in both groups converge. Potentially, this development reflects treated subjects giving up work due to discontinued earnings exemptions.

As shown in Panel B and C of Figure 3.3, the differences in employment rates between the two groups can almost completely be ascribed to part-time employment. Note that the effects in Panel B and C decompose the effects shown in Panel A. While the chances to work full-time do not differ significantly in most treatment months, results suggest higher chances to work part-time among treated subjects.

circa 7,000 non-participants. Roughly 15 percent of non-participants are employed in the last months of observation.

Again, the effects are only significant (at the 10 percent level) in one treatment month. The convergence of employment rates in post-treatment months also appears to be driven by part-time employment, which is consistent with the idea that discontinued earnings exemptions could drive this development.

Concerning full-time employment (see Panel B), there is some suggestive evidence for detrimental effects in the first months. Roughly six months into treatment, treated subjects are nearly 4 percentage points less likely to work full-time ($p = 0.060$). This corresponds with a negative effect of 40 percent relative to the control group mean. This gap in full-time employment rates may stem from treated subjects redirecting their search toward part-time work in the first treatment months. Notably, the gap closes soon after, and in the longer term, full-time employment rates do not differ significantly between the two groups.

Our empirical findings so far are consistent with the stylized predictions derived from our basic labor supply model. First, the treatment seems to have stimulated employment, with effects increasing over time. Second, it appears that employment effects are driven by part-time work. Third, effects on part-time work seem to disappear after discontinuing the earnings exemption. In addition, we find no strong evidence for effects when looking at full-time employment, neither during nor after the treatment period. Only in the first treatment months may the effects have been negative. Overall, our estimates are surrounded by greater uncertainty than would be desirable due to our limited sample size. Some caution is thus advised when interpreting the results.

3.7.2 Earnings, Welfare Benefits and Total Income

Figure 3.4 presents estimated treatment effects in the same fashion as the previous figure. As shown in Panel A and B, the differences in employment probabilities also translate into differences in total earnings (incl. zero earnings). There is suggestive evidence for negative effects in the first treatment months, which may be attributed to entering part-time employment at the cost of finding a full-time job. In later treatment months, differences turn positive. As before, the differences observed are mostly not significant at the 10 percent level.

Panel B shows treatment effects on monthly welfare benefits. Welfare benefits constitute income for claimants but costs for the welfare office. Over time, welfare benefits go down in both groups, which matches the increasing employment rates. With higher employment rates in the treatment group, average welfare benefits should be lower than in the control group. At the same time, lowering the benefit reduction rate and extending the time limit should lead to higher welfare benefits on average. We find that the differences in welfare benefits between the two groups are small and not statistically significant. This would suggest that the two opposing effects largely cancel each other out and that there are no additional costs or savings involved from the perspective of the welfare office.

Importantly, we find statistically significant differences in some pre-treatment months, with lower average benefits in the treatment group. This could mean that our estimates for the effects on welfare benefits are downward biased and that

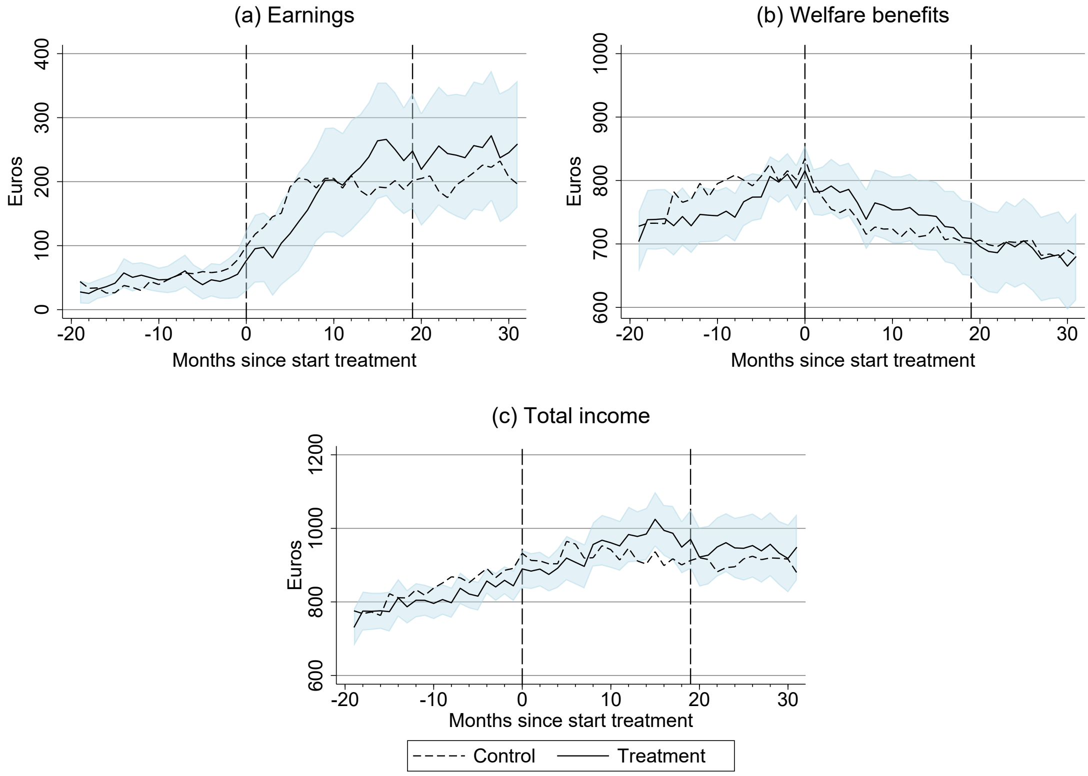


Figure 3.4 Earnings, Welfare Benefits, and Total Income in the Treatment and Control Group.

Note: Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(3.6.1). Colored areas are 90 percent confidence intervals. Dashed lines indicate the treatment period.

introducing a more generous earnings exemption may come at the cost of higher average benefits. With this limitation in mind, a more careful conclusion seems that quite likely, the treatment did not result in lower benefit expenditures.

Panel C shows treatment effects on total income (welfare benefits + earnings). Effects on earnings and welfare benefits were insignificant when looking at these two outcomes individually. Taking both sources of income together, we find significant positive effects toward the end of the treatment period. The effect on total income is the largest in month 16 with an estimate of roughly €100 ($p = 0.023$). This is a positive effect of roughly 10 percent relative to the control group mean. In line with the previous results for welfare benefits, we also find statistically significant pre-treatment differences for total income, with a lower average income in the treatment group. The effects on total income may thus be downward biased as well. In conclusion, we find compelling evidence that welfare claimants were, on average, better off in terms of total income in later treatment months. Consistent with previous results, differences in total income disappear post-treatment.

Table 3.3 Treatment Effects on Cumulative Outcomes.

	Control	Treatment			N
	Mean	Coeff.	*p*-value		
	(SD)	(SE)	Standard	RI	
	(1)	(2)	(3)	(4)	(5)
Panel A: Month 19 (last treatment month)					
Months with work	3.362	0.502	0.379	0.385	375
	(6.535)	(0.570)			
Earnings	3,593	63	0.934	0.939	375
	(8,447)	(759)			
Welfare benefits	14,915	415	0.399	0.402	375
	(5,612)	(491)			
Total income	18,523	460	0.417	0.435	375
	(5,915)	(566)			
Panel B: Month 31 (one year post-treatment)					
Months with work	5.378	1.131	0.215	0.219	375
	(9.929)	(0.910)			
Earnings	6,064	575	0.661	0.655	375
	(13,624)	(1,310)			
Welfare benefits	23,314	351	0.683	0.677	375
	(9,483)	(858)			
Total income	29,405	891	0.379	0.391	375
	(10,013)	(1,012)			

Note: OLS estimates of ITT effects on cumulative outcomes, following Eq.(3.6.1). The model includes strata fixed effects and controls for the background characteristics listed in Table 3.2. Outcome variables are listed on the left and described in detail in Table 3.A.3 in Appendix 3.A. Column (1) reports the control group means with standard deviations in parentheses. Column (2) reports coefficients of the treatment dummy with robust standard errors in parentheses. Column (3) reports the corresponding standard *p*-values, and Column (4) randomization inference *p*-values based on 2,000 replications. Cumulative outcomes cover the period from month 0 to month 19 (Panel A) or month 31 (Panel B), respectively.

3.7.3 Cumulative Outcomes

Table 3.3 reports estimated treatment effects on outcomes measured cumulatively. More specifically, we sum up outcomes from month 0 up to and including month 19 (last treatment month) and month 31 (last month observed), respectively. We include month 0 as subjects learned about their group assignment at the beginning of that month and may have already changed their behavior in expectation of treatment.

For months with work during the treatment period, the point estimate is positive, which fits the findings from previous subsections. The effect corresponds with an increase in months worked of 15 percent relative to the control group mean but is not statistically significant. Relative magnitudes for earnings, welfare benefits, and total income are smaller, corresponding with increases of 2–3 percent relative to the control group mean. These effects are not statistically significant either. Results do not change when including post-treatment months.

Hence, when looking at the treatment (and post-treatment) period as a whole, we find no evidence that treated claimants are significantly better off. This finding is not surprising, though, given that—if anything—labor market effects materialize halfway, are driven by part-time work, and do largely not persist post-treatment. From a cost perspective, we again find no evidence that the more generous exemption resulted in significant extra savings or expenses for the welfare office.

3.7.4 Additional Analyses

In what follows, we conduct some additional analyses to qualify our results. First, we estimate unadjusted treatment effects, omitting covariates and randomization strata fixed effects from our basic model specified in Eq.(3.6.1). Figure 3.C.1 in Appendix 3.C shows the results for employment probabilities, Figure 3.C.2 the results for continuous outcomes. We find that there are some differences in outcomes before treatment. For instance, employment probabilities and earnings appear higher in the treatment group ex ante. This finding is consistent with the imbalances observed in Table 3.2, where we found the treatment group to have higher cumulative earnings in the 24 months before the experiment on average. Comparing unadjusted (in Figure 3.C.1 and 3.C.2) with adjusted effect estimates (in Figure 3.3 and 3.4), it shows that differences ex ante disappear and differences post-randomization become somewhat smaller. This finding leaves us less concerned about the imbalances detected at baseline and leads us to believe that we can address imbalances by controlling for the covariates included in vector X_i (see Section 3.6).

Second, we operationalize employment outcomes differently, i.e., based on earnings instead of hours. Figure 3.C.3 in Appendix 3.C shows the results. In Panel A, we use positive earnings instead of positive working hours as the outcome variable. In Panel B, we count subjects as full-time employed if their earnings surpass the statutory benefit level, which lies at 70 percent of the monthly minimum wage. Claimants should become ineligible for social assistance when their earnings surpass that threshold. In Panel C, we measure part-time employment as having positive earnings up to the 70 percent threshold. In sum, we find that results do not change when using different outcome variables.

Lastly, we use inverse probability weighting based on the variables listed in Table 3.2 to generalize our results to the target population and address concerns about external validity. Admittedly, this exercise requires the strong assumption that selection into the sample is based on observables. Table 3.B.1 in Appendix 3.B reports the results. For comparison, Panel A shows the results of our unweighted model. We find that point estimates are smaller and change direction in the case of cumulative benefits and total income but are still statistically insignificant throughout.

3.7.5 Limitations

Our study is subject to a few limitations. First, we cannot estimate effects very precisely due to a limited sample. Consequently, even substantial effects are surrounded by greater uncertainty than would be desirable. Higher powered studies will be needed to confirm our findings and confidently detect more modest effects. Second, our results are based on a convenience sample of current social assistance claimants in Utrecht. Different effects may materialize in other places (e.g., less urban areas), other welfare systems, or with a sample of new inflow. Future studies will have to confirm if our findings hold for different populations.

Third, our study setup did not allow us to explore potential entry effects, i.e., to what extent working individuals outside the scheme enter welfare in response to changed incentives. This may be unproblematic given that the incentives for entry are not much higher under the policy tested than under the status-quo regulation. Nonetheless, entry effects may be non-negligible (see, e.g., Moffitt, 1992, 1996) and require attention when studying comparable incentive structures in the future.

Fourth, the policy studied combined a larger disregard with a change in time limits, which does not allow for disentangling individual effects. Lastly, it should be noted that the Netherlands experienced a time of sustained economic growth with a relatively tight labor market at the time of the experiment. Effects may be smaller under less favorable labor market conditions, for instance.

3.8 Conclusion

Programs that aim to make work pay among benefit recipients are a common component of social benefit schemes in many countries. Recently, such policies have received renewed attention in political debates about financial work incentives, e.g., in the Netherlands and Germany. Common are demands for more generous earnings exemptions, i.e., allowing benefit claimants to keep more earnings on top of benefits and for longer periods. Such modifications—the reasoning goes—encourage part-time work in addition to benefits and facilitate the transition toward full-time employment and self-sufficiency.

In this chapter, we reported on a direct (experimental) test of such a policy measure. We evaluated a more generous earnings exemption policy for Dutch social assistance benefits claimants. The policy under investigation introduced two changes. First, it lowered the implicit tax rate on earnings in addition to benefits from 75 to 50 percent (up to a maximum of €202 ($260 PPP) per month). Second, it removed the six-month time limit that typically applies to earnings exemptions. Hence, claimants could keep more of their earnings for a longer period. We studied the labor market and the budgetary effects of this policy change in a field experiment with 375 current benefit claimants sampled in Utrecht.

In line with expectations, we find higher employment rates in the treatment group, subject to the new policy. Roughly 1.5 years after the start of the experiment, employment probabilities are nearly 40 percent higher in that group than in the control group. Our results suggest that differences between the two groups are almost entirely driven by part-time work, which we define as work for less than 28 hours a week (as work for 28 hours a week or more would terminate benefits). We do not find much evidence for effects on full-time employment. Results indicate that exit to full-time work fell behind in the first months. However, full-time employment rates remained unaffected in the longer run and post-treatment. This finding suggests that—on average—there is no displacement of full-time work. At the same time, adaptations to the policy, such as removing the time limit, were not successful in promoting transition to full-time employment either.

We also find evidence that the policy translates into higher total income (welfare benefits + earnings) toward the end of the experiment. However, this effect is short-lived, and measuring total income cumulatively across the entire treatment period shows no evidence of effects. Lastly, we find no evidence for budgetary effects when looking at benefit expenditures. Potentially, lower benefit payments due to (part-time) employment compensate for higher expenses (due to a lower benefit reduction rate and an extended eligibility period).

On a cautionary note, our limited sample size does not allow for very precise effect estimates. Although effect dynamics are suggestive of positive labor market effects, it is only in some treatment months that estimates are also statistically significant (at the 10 percent level). In light of this limitation, we draw careful conclusions. Overall, we interpret the results as speaking in favor of a more generous earnings exemption. While possibly stimulating part time-work among benefit recipients, the policy does not seem to discourage exit to full-time employment on a large scale or lead to sizeable additional welfare expenses. Results also imply that it probably needs other incentives or instruments to stimulate full-time exits.

Further research will be needed to study the interaction of financial incentives targeting part-time and full-time work. In the same line, future work should explore whether stronger financial incentives result in larger effects. Moreover, financial work incentives may interact in complex ways with taxes and other types of benefits and allowances. Also, the administrative hassle involved in working in addition to benefits may have a large discouraging effect. Complementary research is needed to identify and target such difficulties and test solutions for making work in addition to benefits less cumbersome.

3.A Additional Background Information

Table 3.A.1 List of Exclusion Criteria.

Criterion	Reason
Younger than 27 years.	Government waiver not applicable.
Reaching retirement age during the experiment.	Preventing attrition due to retirement.
Receiving unemployment insurance benefits from the Dutch Employee Insurance Agency (UWV).	Government waiver not applicable.
Subject to the Natural Persons Debt Rescheduling Act (*WSNP traject*).	Government waiver not applicable.
Part-time entrepreneurs and claimants receiving benefits for the self-employed.	Government waiver not applicable.
Claimants admitted to a healthcare institution.	Treatments not applicable.
Asylum status holders with integration obligations.	Government waiver not applicable.

Table 3.A.2 List of Covariates With Description.

Variable	Description	Source
Age	Age in years at the start of treatment (June 1st, 2018).	Civil registry, Statistics Netherlands
Female	1 if subject is female and 0 if subject is male.	Civil registry, Statistics Netherlands
Education level	4 indicator variables (lower, intermediate, higher, unknown) denoting highest education level attained based on ISCED 2011 classification (record date: October 2017). Lower: less than primary, primary, lower secondary; Intermediate: upper secondary, post-secondary non-tertiary; Higher: short cycle tertiary, bachelor, master, doctoral.	Civil registry, Statistics Netherlands
Migration background	3 indicator variables (Dutch, Western, non-Western) based on migration background definition of Statistics Netherlands. Dutch: person in question and parents born in the Netherlands; Western: person in question or at least one of the parents born in Europe (excl. Turkey), North America, Oceania, Indonesia, or Japan; Non-western: person in question or at least one of the parents born in none of the above mentioned regions/countries.	Civil registry, Statistics Netherlands
Current spell	Duration of the ongoing benefit spell at the start of treatment (June 1st, 2018) in months.	Benefit registry, Statistics Netherlands
Earnings 24 months before	Cumulative labor market earnings in the period March 2016 to February 2018 in euro.	Social security records, Statistics Netherlands
Living situation	3 indicator variables (single, single parent, cohabit) denoting whether in March 2018 social assistance benefits were received as a single-person, single parent, or multi-person household.	Benefit registry, Statistics Netherlands
Class	4 indicator variables (class I, class II, class III, class IV) denoting claimant classification as registered by the welfare department at the time of sample selection.	Claimant registry, Utrecht welfare office

Table 3.A.3 List of Outcome Variables With Description.

Variable	Description	Source
Employed	1 if subject's working hours in month t are greater than zero and 0 otherwise.	Social security records, Statistics Netherlands
Part-time work	1 if subject's working hours in month t are greater than zero but below 28 and 0 otherwise.	Social security records, Statistics Netherlands
Full-time work	1 if subject's working hours in month t are greater than or equal to 28 and 0 otherwise.	Social security records, Statistics Netherlands
Months with work	Number of months in the reference period with positive working hours.	Social security records, Statistics Netherlands
Earnings	Labor market earnings (excl. income from self-employment).	Social security records, Statistics Netherlands
Welfare benefits	Actual payments of social assistance benefits, i.e., after accounting for earnings and other settlements.	Benefit registry, Statistics Netherlands
Total income	Sum of welfare benefits and earnings.	Social security records and benefit registry, Statistics Netherlands

3.B Additional Tables

Table 3.B.1 Treatment Effects on Cumulative Outcomes in Month 19: Unweighted and Weighted Data.

	Control	Treatment		N
	Mean (SD)	Coeff. (SE)	p-value	
	(1)	(2)	(3)	(4)
Panel A: Baseline model				
Months with work	3.362	0.502	0.379	375
	(6.535)	(0.570)		
Earnings	3,593	63	0.934	375
	(8,447)	(759)		
Welfare benefits	14,915	415	0.399	375
	(5,612)	(491)		
Total income	18,523	460	0.417	375
	(5,915)	(566)		
Panel B: Model with weights				
Months with work	3.362	0.395	0.384	375
	(6.535)	(0.453)		
Earnings	3,593	32	0.952	375
	(8,447)	(530)		
Welfare benefits	14,915	-237	0.622	375
	(5,612)	(479)		
Total income	18,523	-213	0.689	375
	(5,915)	(530)		

Note: OLS estimates of ITT effects on cumulative outcomes (month 0 to 19). The results in Panel A are based on Eq.(3.6.1), using unweighted data. Panel B reports results using the same model but applying inverse probability weights based on the background characteristics listed in Table 3.2. Outcome variables are listed on the left and described in detail in Table 3.A.3 in Appendix 3.A. Column (1) reports the control group means with standard deviations in parentheses. Column (2) reports coefficients of the treatment dummy with robust standard errors in parentheses. Column (3) reports the corresponding p-values.

3.C Additional Figures

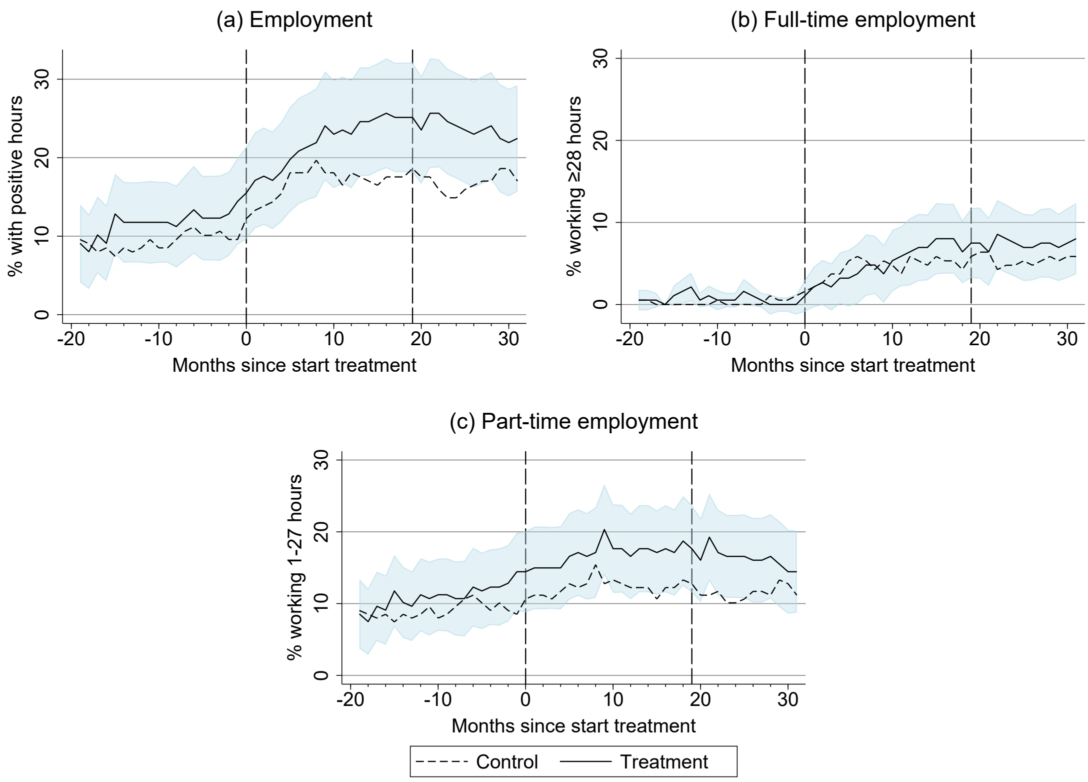


Figure 3.C.1 Employment Rates in the Treatment and Control Group (Unadjusted Effects).

Note: Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(3.6.1) but excluding covariates and randomization strata fixed effects. Colored areas are 90 percent confidence intervals. Dashed lines indicate the treatment period.

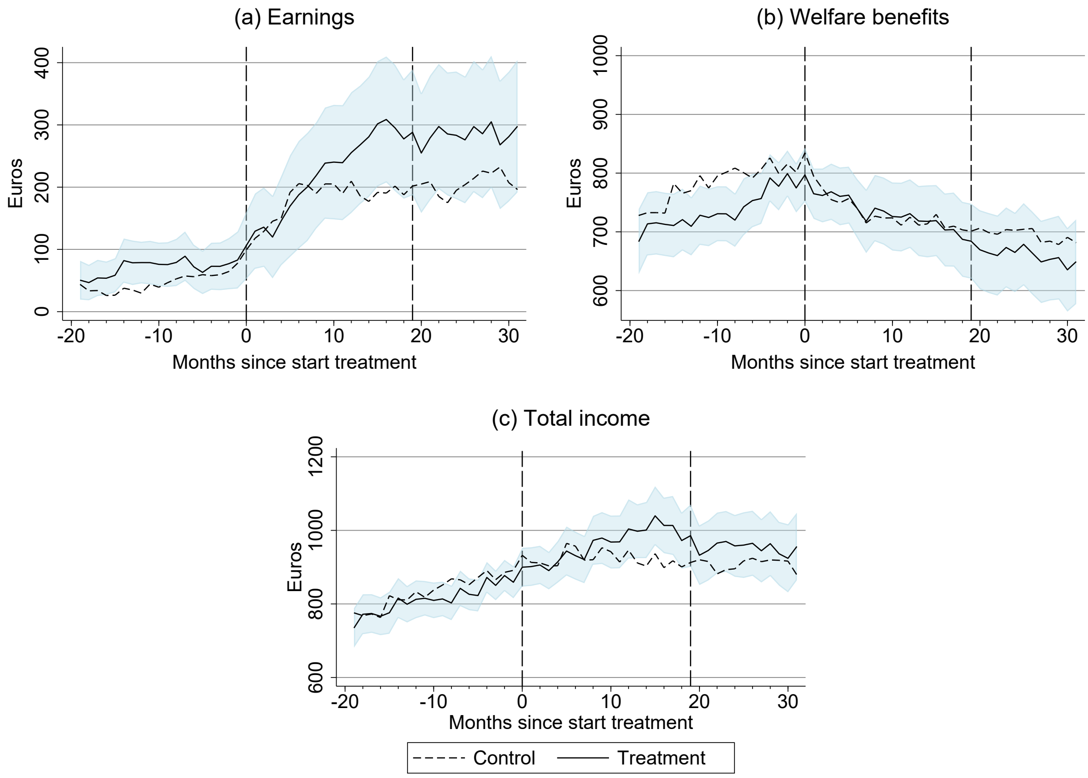


Figure 3.C.2 Earnings, Welfare Benefits, and Total Income in the Treatment and Control Group (Unadjusted Effects).

Note: Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(3.6.1) but excluding covariates and randomization strata fixed effects. Colored areas are 90 percent confidence intervals. Dashed lines indicate the treatment period.

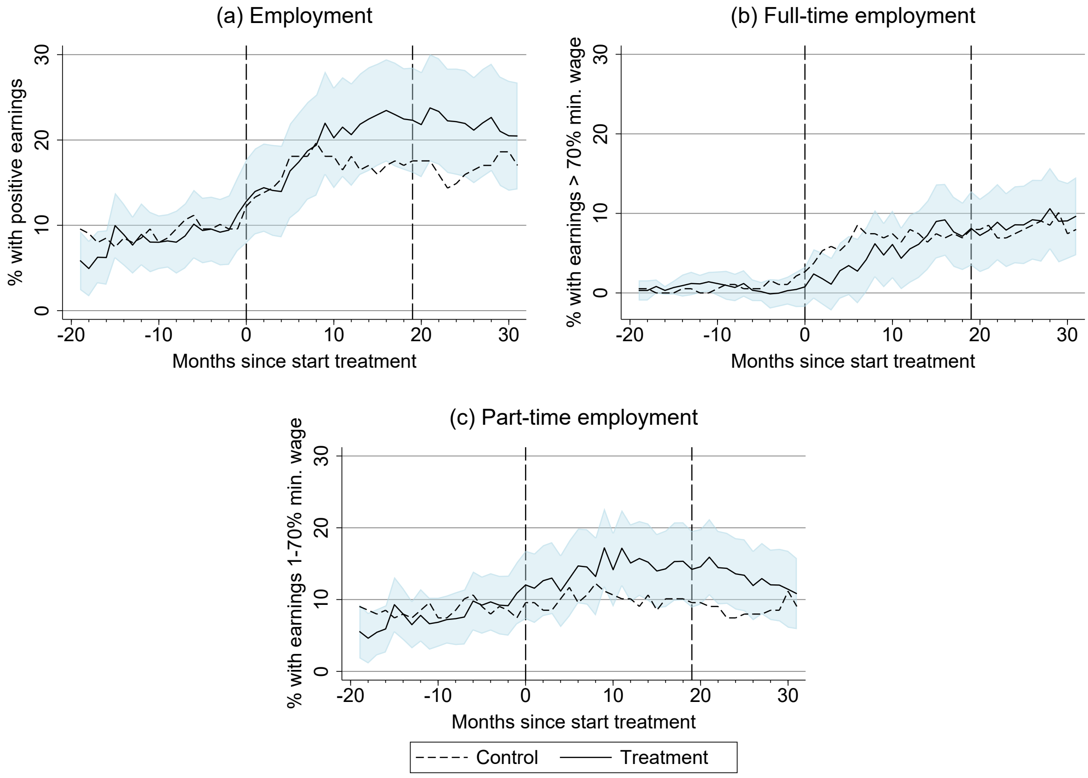


Figure 3.C.3 Employment Rates in the Treatment and Control Group (Based on Earnings).

Note: Treatment graphs show ITT effects and are estimated using separate regressions for each month, following Eq.(3.6.1). Colored areas are 90 percent confidence intervals. Dashed lines indicate the treatment period.

CHAPTER 4

The Effects of Sanctions and Reprimands in Unemployment Insurance*

4.1 Introduction

In many countries, job seekers claiming benefits are required to show that they actively search for a new job. Benefit authorities monitor claimants' behavior and may impose a benefit sanction, i.e., a partial withdrawal of benefits, if claimants fail to provide the required search effort. By introducing a cost for unwanted behavior, monitoring and sanctioning are thought to restore incentives to stay active and search for a job, thus countering moral hazard in unemployment insurance (UI).

Monitoring and sanctioning are thought of as attractive policy options for various reasons. First, the bureaucratic means needed to process and verify job search efforts and impose sanctions often falls short of what is required to implement comprehensive training programs, counseling schemes, public employment initiatives, or other forms of activation. Second, there are concerns about the mixed effectiveness of activation policies (Martin, 2015). Third, contrary to other forms of activation, monitoring and sanctioning comes without potential lock-in effects, i.e., they do not keep job seekers away from job search. On the downside, benefit sanctions may impose (often substantial) costs on individuals already financially constrained. In addition, sanctions could push job seekers into lower-quality jobs, resulting in potentially long persisting disadvantages. Sanctions could also drive people out of the labor force altogether and have other unintended side effects, such as increases in crime.[1]

It is important to mention that both the presence of a sanctioning threat (*ex*

*This research was pre-registered at the registry of the Open Science Foundation (Verlaat et al., 2020b). We report discrepancies with the pre-registration in Table 4.D.1 in Appendix 4.D.

[1]Machin and Marie (2006) report evidence from the United Kingdom that suggests a positive relationship between benefit sanctions and crime.

ante effects) and the actual imposition of a sanction (*ex post* effects) may affect claimant behavior. In this study, we focus on the latter. Our primary research question reads: What are the causal effects of imposing benefit sanctions on claimant behavior and economic outcomes? In addition to sanctions, there exist other, less severe means to discipline claimants. Softer disciplinary measures may impact outcomes differently. Therefore, we also study the causal effects of one such measure, namely, issuing a reprimand.

We study sanctions and reprimands in the Dutch UI system. In the Netherlands, UI benefit claimants are required to report four job search activities every four weeks. Noncompliance with this effort requirement may result in a reprimand or a benefit cut of 25 percent for four months. Decisions on disciplinary measures lie with caseworkers at the Public Employment Services (PES). For our study, we collect extensive register data available at the PES. This data includes information on sanction and reprimand receipt and a broad range of outcomes. Identifying the causal effects of sanctions and reprimands comes with the empirical problem that caseworker decisions are selective. Hence, sanctioned claimants differ from unsanctioned claimants in observed and unobserved characteristics. Most likely, these characteristics determine both the likelihood of receiving a sanction and the success in, e.g., finding a job and leaving unemployment. Estimation techniques that fail to account for selectivity in sanction receipt are likely to produce biased results.

We address the problem of selection effects by employing an instrumental variable (IV) approach. In the Dutch UI system, infringement cases are quasi-randomly assigned to caseworkers, which, due to discretionary power, systematically vary in their stringency to impose disciplinary measures. This feature causes exogenous variation in sanctioning probabilities. We exploit this setting and identify causal effects using caseworker stringency as an instrument for the actual treatment. To calculate stringency, we take a caseworker's mean treatment rate across all cases, except the respective individual case at hand (leave-out mean). Using the stringency of decision makers as an instrument for individual treatments is a strategy originating from work on the causal effects of incarceration (Kling, 2006).[2] Recently, similar IV designs have found frequent application in many research areas.[3] Our research design recovers the causal effects of sanctions and reprimands for those claimants, whose treatment depends on whether they are assigned a stricter or a more lenient caseworker (local average treatment effect).

Our first-stage results indicate a strong relationship between our caseworker stringency instruments and the probability of treatment. Assignment to a caseworker who is 10 percentage points more likely to impose a sanction increases the likelihood of receiving a sanction by 5 percentage points. The probability of being

[2] See also Aizer and Doyle (2015) and Bhuller et al. (2020).

[3] Prominent examples concern Maestas et al. (2013), French and Song (2014), and Autor et al. (2019), who exploit the stringency of judges deciding on disability insurance appeals to study the causal effects of disability benefit receipt on labor supply and various other outcomes. Another example comes from Sampat and Williams (2019), who use the leniency of randomly assigned patent examiners to identify the causal effects of patenting.

reprimanded increases by 9 percentage points if the caseworker is 10 percentage points stricter. In contrast to treatment probabilities, both instruments are only very weekly related to observable claimant characteristics, suggesting (conditional) independence. Moreover, our instruments are unrelated to other treatments often imposed by PES caseworkers (workshops, coaching meetings, and exemptions from search requirements), which provides evidence supporting the exclusion restriction. Finally, the first-stage relationship holds in different subsamples, also when we use cases from outside the respective subsample to calculate our instruments. This last finding supports the monotonicity assumption.

Our study contributes to understanding the effectiveness of different disciplinary measures in unemployment insurance. It thereby extends a broad literature on benefit sanctions for job seekers, which covers different countries, especially the Netherlands, Denmark, Germany, and Switzerland, and various benefit schemes, such as unemployment insurance, universal social assistance, or youth unemployment programs.[4] Similar to our study, most of the existing work examines sanctions received for violating effort requirements and focuses on the effects of imposing a sanction (*ex post* effects) rather than the effects of a sanctioning threat (*ex ante* effects).[5] This literature consistently shows that imposing a sanction increases probabilities of job entry and exit from unemployment. Most studies document large effects; van den Berg et al. (2004), e.g., find that receiving a sanction more than doubles the transition rate from welfare to work, while Abbring et al. (2005), Svarer (2011), and Hillmann and Hohenleitner (2015) document increases in job-finding rates of 60–100 percent after a sanction.

Only a few studies look beyond exit from benefits and job finding and examine the effects of sanctions on post-unemployment outcomes. These studies consistently document detrimental effects. Arni et al. (2013), e.g., report negative effects of a sanction on post-unemployment earnings and job stability. Similarly, van den Berg and Vikström (2014) find that receiving a sanction leads to lower wages, higher chances of working part-time rather than full-time, and job-finding at a lower occupational level. These negative effects are persistent in both studies, lasting for up to two and up to four years post-unemployment, respectively. Similarly, only a few studies examine transitions to non-employment, documenting increased exits from the labor force after a sanction (Arni et al., 2013; Busk, 2016). In sum, the existing empirical evidence suggests that benefit sanctions successfully promote exit from unemployment and job finding. However, leaving unemployment more quickly may come at the cost of re-employment quality and involve higher chances of withdrawal from the labor force.

We contribute to this literature in three ways. First, we combine rich register

[4]See McVicar (2020), for a review.

[5]Using laboratory experiments, Boone et al. (2009) find that *ex ante* effects are substantial and often larger than *ex post* effects. To the best of our knowledge, the only studies investigating *ex ante* effects in a field setting come from Lalive et al. (2005) and Arni et al. (2013). Both studies provide evidence in favor of *ex ante* effects, showing that unemployment duration is shorter (Lalive et al., 2005) and job entry rates are higher (Arni et al., 2013) in Swiss regions with a stronger sanctioning threat.

data from different sources to include a wide range of outcomes. This approach allows us to provide a more comprehensive view of the effects of benefit sanctions and complement the scarce evidence on how outcomes other than exit from benefits and job finding are affected. In addition to examining effects on benefit dependency and employment, we study the preventive effect of sanctions, i.e., their effect on recidivism rates. We also examine how sanctions affect job search behavior. Moreover, we investigate if sanctions promote inflow in sickness benefits. Inflow in sickness benefits may reflect evasion behavior or potential health effects. Lastly, similar to some previous studies, we study the effects on re-employment quality, which we proxy by earnings and chances of finding a temporary versus a permanent job.

Second, although the effects of imposing sanctions have been documented rather extensively, little is known about the effectiveness of less severe measures, such as reprimands. In contrast to sanctions, reprimands do not involve financial consequences. Previous studies show that it does not need a financial penalty to change benefit claimants' behavior—signaling a threat can already suffice.[6] To the best of our knowledge, ours is only the second study to examine the impacts of softer disciplinary measures in UI. The other study—also from the Netherlands—focuses on a very particular group of job seekers, that is, unemployed workers from the primary education sector (van der Klaauw et al., 2008). The authors find that reprimands have comparable positive effects on re-employment chances as sanctions. Our study includes all non-compliant UI benefit recipients in the Netherlands. This allows us to investigate whether previous findings hold for a sample more representative of the general population of UI benefit claimants.

Our third contribution concerns the methodological approach. Almost all existing studies rely on timing-of-events models to identify the causal effects of imposing sanctions. This method deals with selectiveness in sanctioning decisions by simultaneously modeling the chances of receiving a sanction and re-employment or exit from benefits. Although this approach is commonly accepted and widely applied, it also relies on quite strong assumptions, e.g., that claimants do not anticipate the imposition and exact timing of a sanction (Abbring and van den Berg, 2003). In contrast to earlier work, our analysis uses an IV approach to identify causal effects. We are aware of only two other studies investigating UI policies that apply a similar IV design. Both examine the causal effects of search requirements on effort provision and labor market outcomes.[7] To the best of our knowledge, we are

[6]Some studies look at the effects of sending warning letters, which announce that an infraction took place and that a sanction may follow (Arni et al., 2013; Lalive et al., 2005). These studies find that warning letters and sanctions have similar effects on benefit exit and job entry rates. Moreover, warnings can have the same negative effects on post-unemployment earnings and job stability as actual sanctions (Arni et al., 2013). Other studies show that already the announcement of mandatory re-employment services can increase exit rates from unemployment (see, e.g., Black et al., 2003; Geerdsen, 2006; Rosholm and Svarer, 2008). In contrast to these examples, which signal an imminent threat, reprimands issue a warning that refers to future infractions.

[7]The first study uses data from Switzerland, where caseworkers have discretionary power in determining the number of applications to be reported each month (Arni and Schiprowski, 2019). Using caseworker stringency in setting search targets as an instrument for individual requirements, the authors find that higher effort targets reduce unemployment duration. The second

the first to use an IV approach to identify the causal effects of imposing benefit sanctions.

Our main findings can be summarized in four parts. First, contrary to previous studies, we find no evidence that imposing a benefit sanction affects employment probabilities and benefit dependency. Second, we find tentative evidence for reduced earnings in the longer term. This finding is consistent with evidence from previous studies. It suggests that sanctions may impose a double cost on sanctioned claimants—the cost of foregone benefits and foregone earnings. Third, we find evidence that imposing a sanction promotes future compliance. Compared to the sample average, sanctioned claimants are roughly 40 percent less likely to commit another infraction in the 12 months post-sanction. We also find tentative evidence for reporting more job search activities in the month after the infringement. Fourth, while reprimands also seem to promote reporting of job search activities, they leave other outcomes unaffected. In sum, our findings suggest an important trade-off. Although imposing a sanction promotes compliance effectively, adverse effects on long-term economic outcomes may follow. Reprimands may constitute a less harmful alternative.

The remainder of this chapter is organized as follows. In Section 4.2 we discuss theoretical considerations. Section 4.3 introduces the policy context and describes the process of sanctioning in the Dutch UI system. In Section 4.4 we present our data and sample, while Section 4.5 introduces our empirical strategy. In Section 4.6 we report results. Section 4.7 concludes.

4.2 Theoretical Considerations

Economic theory is ambiguous about the effects of benefit sanctions on economic outcomes. On the one hand, theoretical models commonly predict that imposing a sanction increases job finding and benefit exit rates. On the other hand, sanctioned claimants may accept lower-quality jobs or drop out of the labor force altogether. Furthermore, the effectiveness of sanctions may also depend on whether sanctions induce job seekers to substitute informal with formal (and thus verifiable) job search activities and the effectiveness of formal relative to informal search channels (van der Klaauw and van Ours, 2013). In what follows, we will briefly review the standard approaches to modeling incentives and policy choices in UI and the predictions one can derive from these models regarding the effectiveness of sanctioning policies. Given that these models have been formalized elaborately in previous work, we leave it at providing some stylized considerations.

Most theoretical contributions build on a classical job search model. Such a model assumes that benefit claimants derive utility from unemployment benefits and the expected change in welfare when entering employment. Individuals can

study investigates the causal effects of imposing broader job search requirements on labor market outcomes and UI benefits receipt (Vethaak and van der Klaauw, 2021). Using caseworkers' tendency to assign broader search requirements as an instrument for individual requirements, the authors find that claimants required to broaden their search stay unemployed longer.

influence their chances of finding work in two (complementary) ways: the intensity of looking for a job and their reservation wage. Claimants try to maximize the value of unemployment. In doing so, they trade-off the immediate benefit of receiving income support and the expected benefit of future income against the cost of searching for a job and settling for a lower wage. Claimants will increase their search effort until the marginal cost equals the expected marginal gain of searching. Some implications for the optimal design of UI can already be inferred from this basic model. Increasing the benefit level, e.g., will lead to lower search efforts, as higher benefits reduce the marginal gain of searching for a job. The same result can be derived for expanding the duration of UI benefits.

We can extend this basic model of UI to include a sanctioning policy. In that case, another cost component enters the value function of the benefit claimant—the expected cost of incurring a benefit cut. The probability of receiving a sanction typically depends on three factors: individual search effort, the rate at which the claimant is monitored, and a term reflecting imperfect monitoring efforts. That the sanctioning probability depends on search effort changes optimal behavior. Choosing their effort level, claimants now have to consider that higher effort reduces the likelihood of incurring a sanction and vice versa. In other words: The claimant's value function now includes an additional reward for incurring search cost. Accordingly, in optimum, claimants will search harder than in a basic model without sanctioning policy. This result forms one of the theoretical foundations for introducing benefit sanctions; the threat of a sanction alone may lead claimants to search more intensively, resulting in higher chances of finding a job and exiting benefits.

However, sanctions do not merely operate through the channel of deterrence but should also affect the search effort and reservation wages of those *actually* sanctioned. Specifically, a benefit cut (temporarily) reduces income from benefits, and claimants may want to compensate their loss in income by realizing labor market earnings. Job seekers can increase their chances of finding a job (and generating earnings) by—again—searching more intensively, lowering their reservation wage, or both. To the extent that job seekers search harder after receiving a sanction, they also run a lower risk of being sanctioned again. While the term *ex ante* effect describes the change in behavior resulting from a sanctioning threat, the *ex post* effect includes any action produced by actually receiving a sanction.

Hence, basic models of UI posit that imposing a sanction should promote the transition from unemployment to employment. At the same time, claimants may respond to a sanction by lowering their reservation wage, leading them to accept lower-paying jobs. Lower-paying jobs may also score lower on other quality dimensions, such as the permanency of a position. Thus, although sanctions have the potential to speed up job finding and exit from benefits, they may harm post-unemployment outcomes. Also, sanctions could reduce the value of being unemployed so that other states, such as non-employment, become attractive outside options.

Another important caveat concerns the type of job search that job seekers perform. Monitoring and sanctioning schemes typically require verifiable job search efforts, such as applications sent, CVs uploaded, or job interviews scheduled. In response to a sanctioning threat or an actual sanction, jobs seekers may substitute informal ways of search, such as networking or talking to friends and former colleagues, for verifiable search activities. To the extent that formal, more verifiable activities are less effective than informal channels, sanctions may harm chances of re-employment. As we will describe in the next section, the Dutch PES explicitly allows claimants to report informal search activities. Therefore, we do not expect sanctions to operate through changes in job search channels in our study context.

4.3 Policy Background

4.3.1 Unemployment Insurance Benefits in the Netherlands

Unemployment insurance in the Netherlands is a compulsory, contribution-based social protection scheme.[8] Employees are entitled to UI benefits when having worked for 26 weeks in the 36 weeks immediately preceding unemployment.[9] The basic UI benefit lasts three months. Workers with longer employment histories can claim benefits for longer; usually, one year of employment prolongs entitlement by one month. The maximum duration was 38 months until 2016 and is 24 months since then. Among OECD countries, the Netherlands ranks in the upper third in terms of benefit generosity (OECD, 2021b); UI benefits pay for 75 percent of the last earned gross wage in the first two months of entitlement and 70 percent after that. As of January 1st, 2014, the maximum benefit was roughly €3,000 ($3,708 PPP) per month.[10] Claimants are allowed to combine UI benefit receipt and work, in which case they may keep 30 percent of their labor income. The execution of the scheme lies with an autonomous administrative authority, the Employee Insurance Agency (UWV). The agency operates 37 Public Employment Service (PES) offices throughout the country. These offices are responsible for registering claimants, monitoring compliance, and providing employment services.

4.3.2 Job Search Requirements and Benefit Sanctions

Like other OECD countries, claimants of UI benefits in the Netherlands have to meet several obligations, e.g., being available for work, accepting suitable job offers, and actively searching for a job. Moreover, claimants must cooperate with the PES office and its caseworkers and disclose all relevant documents and information. Given the focus of this study, we now describe a particular obligation—the job

[8]In this subsection, we describe the setup and parameters of the scheme as of January 1st, 2014, the starting point for our data series. We convert euros into 2014 purchasing power-adjusted U.S. dollars using the OECD purchasing power parity (PPP) exchange rate (OECD, 2021d).

[9]Claimants have to meet additional eligibility criteria, of which the most important are: residence in the Netherlands, involuntary unemployment, being capable and available for work, and age below the legal retirement age. Ineligible workers may claim universal social assistance.

[10]For comparison, the statutory monthly minimum wage at the same time was €1,486 ($1,837 PPP) for a full-time job of 36, 38, or 40 hours depending on the sector.

search requirement—in more detail.[11]

Claimants are obliged to report four job search activities every four weeks through an online portal (*Werkmap*) that the PES offices use to communicate with claimants. When reporting their activities, claimants can choose from a list of valid job search activities. Valid activities include job applications, job interviews, registering with a temporary employment agency, contacting a potential employer, taking part in an assessment procedure, following an employment-related workshop, talking to people in one's network, and uploading one's CV on a job platform. The PES office informs claimants about the job search requirement and the reporting procedure at the start of their benefit spell. In principle, the requirement applies to all claimants, although there may be temporary exemptions in exceptional situations (e.g., a death or a severe illness in the family).[12] Working claimants who use the UI benefit as a supplement are usually required to continue looking for work, i.e., for a job that allows them to exit benefits altogether.

Noncompliance with the search requirement may result in a benefit sanction, i.e., a temporary cut in one's benefit. Specifically, a benefit sanction may follow in case claimants reported too few job search activities or in case they reported activities too late. The severity of the sanction depends on the type of infringement. The default sanction for reporting *too few* activities is a cut in benefits of 25 percent for at least four months. However, the sanction may be as low as 15 percent or reach 100 percent. Reporting activities *too late* usually results in a cut of 5 percent for at least one month, but sanctions as low as 2 percent or reaching 20 percent are also possible. In 2017, which is the latest year included in our sample, the PES imposed 60,800 sanctions in total, among which 12,400 sanctions for violating the job search requirement (UWV, 2017). For comparison, the PES counted roughly 330,000 UI benefit spells in the same year (UWV, 2017). Sanctioned claimants remain registered as unemployed. Instead of imposing a benefit sanction, the PES office may also issue a reprimand, which does not entail financial consequences for the claimant.

4.3.3 Monitoring and Infringement Procedure

The PES office monitors compliance with the job search requirement through its online portal. If a claimant registers too few search activities within a reference period, the portal automatically starts an infringement procedure. The PES office at which the claimant is registered is responsible for handling that procedure. The infringement procedure consists of five steps. The flowchart in Figure 4.1 illustrates the process. The figure also reports percentages for first-time infringement cases processed by the PES between 2016 and 2017, our main sampling period. The five steps are:

[11] Compared to other OECD and EU member countries, the Netherlands ranks in the upper third regarding the strictness of job search monitoring (Immervoll and Knotz, 2018).

[12] Claimants are also exempted if they are currently in training, starting their own business, or reaching the legal pension age within one year.

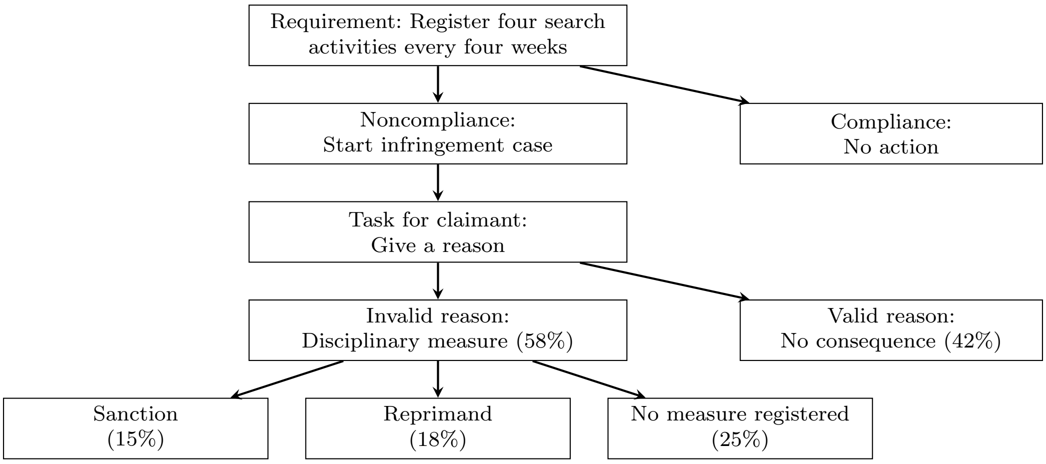


Figure 4.1 Processing of Infringement Cases.

Note: Percentages refer to first-time infringement cases processed by the PES between 2016 and 2017 (N = 51,848).

1. The PES office sends an automated digital message asking the claimant to give a reason for violating the search requirement. The claimant is given one week to respond to that message and, if applicable, register activities retrospectively. For an example message, see Message 1 in Appendix 4.C. At this stage, the claimant does not know which caseworker will process the case. Therefore, we have no reason to assume strategic behavior of the claimant in response to the observed or assumed strictness of the caseworker, as has been observed, for example, in courtroom settings (Gaudet et al., 1933).

2. The PES office assigns the infringement case to one of its caseworkers for further processing. Usually, claimants are not assigned a permanent caseworker. For that reason, PES offices allocate infringement cases to caseworkers through automatically generated and randomly ordered lists. On a daily basis, these lists are used to evenly distribute cases among caseworkers tasked with reviewing cases. It is not common for caseworkers to switch cases with each other. Also, claimants cannot request their cases to be handled by a different caseworker. At some offices, cases may be allocated within a team of caseworkers, e.g., teams for certain industry sectors or occupations.

3. The designated caseworker reviews the infringement case and the clarification provided by the claimant. In case of unclarities, the caseworker can contact the claimant. Based on the information at hand, the caseworker decides whether the claimant provided a valid reason for violating the search requirement. According to the review protocol, valid reasons include personal circumstances, lack of knowledge or skills, force majeure, incorrect information, or other exceptional circumstances. Between 2016 and 2017, claimants were deemed to have a valid reason in 42 percent of first-time infringement cases.

4. If the reason is invalid, the caseworker has to decide on the disciplinary measure, i.e., between a reprimand and a sanction. In the case of a sanction, the caseworker also determines the severity and duration of the sanction. Reprimands and sanctions cannot be stacked. If multiple violations are present, the most severe violation determines the outcome of the process. In some cases, no disciplinary measure is registered despite an invalid reason. This may happen if claimants find work or terminate their spell for other reasons during the review. Another reason may be registration errors. Between 2016 and 2017, 25 percent of all infringement cases resulted in a sanction, 18 percent resulted in a reprimand, and 15 percent had no measure registered.
5. The PES office digitally informs the claimant about the outcome of the review. If the caseworker judged the reason for noncompliance to be valid, a message informs the claimant accordingly (see Message 2 in Appendix 4.C, for an example). If the caseworker judged the reason invalid, the message announces a letter detailing all consequences (see Message 3 in Appendix 4.C, for an example).[13]

It is important to mention that the default employment services at the PES are automated online employment services and automated digital contact. Only in some instances do caseworkers at the PES provide individual employment services. As mentioned before, claimants are not assigned a permanent caseworker.

4.4 Data and Sample

4.4.1 Data Collection

We use individual-level data from different data sources administered by the PES. All data can be merged using unique identifiers. Information on cases of noncompliance with job search requirements comes from the infringement registry at the PES. This data details the start and end date of the infringement review, the caseworker's decision regarding a valid or invalid reason for noncompliance, the caseworker's decision on the disciplinary measure, the PES office responsible for the review, and unique identifiers for case, claimant, and caseworker. From the client management system at the PES, we obtain data on claimants' background characteristics. We collect information on gender, age, the highest education level, and the sector of the previous job. From other PES databases, we obtain information on the monthly wage in the previous job, employment and UI outcomes in the month before the infringement, and the months of UI entitlement left at the time of the infringement.

We only have limited access to data on caseworkers. Specifically, we collect information on other treatments caseworkers assign (e.g., sending claimants to workshops or exempting them from job search requirements) and their work experience.

[13]Unfortunately, we are not authorized to provide examples of letters announcing a sanction or a reprimand.

We proxy work experience by counting the months between the first and last case handled within our sampling period.

We merge this data with information on outcomes, which we obtain from various sources. We collect data on labor market outcomes from social security records available at the PES. These records contain monthly information on earnings, hours worked, and the type of employment contract. In addition to that, we collect monthly information on UI and sickness benefit payments from the payment management system at the PES. Furthermore, we obtain data on the number of job search activities registered through the PES online portal in the month after the infringement. Lastly, we collect information on subsequent violations of the search requirement from the PES infringement registry.

4.4.2 Sample Selection

To form our baseline sample, we select all cases of first-time infringements of job search requirements started between January 2014, the moment from which this data is available, and December 2017. We exclude more recent years, as a policy change in January 2018 has largely eliminated discretionary room for caseworkers to impose sanctions for first-time infringements.[14] This gives us roughly 150,000 unique cases handled by around 2,900 caseworkers at 37 PES offices. Selecting first-time infringements only, every claimant appears once in our dataset.

We exclude five types of cases: (i) cases for which it is not clear which PES office handled them (0.4 percent), (ii) cases without information on start and end dates of the benefit spell (6.0 percent), (iii) cases which took more than 31 days to decide (1.8 percent), (iv) cases of claimants younger than 18 or older than 64 (4.0 percent), and (v) cases of claimants previously employed in the sector Public Administration and Defense (1.3 percent). We exclude observations from that sector as former public sector employees are subject to a specific set of rules and regulations in UI.

Moreover, we drop observations whose caseworkers have handled less than 20 cases during the baseline sampling period (6.0 percent of cases; the median caseworker handled 23 cases). Thus, we use at least 20 cases to construct our caseworker stringency instruments. This measure is in line with previous work (see, e.g., French and Song, 2014; Kling, 2006) and aims at reducing noise in our instruments. Our results are unaffected by modifying this cutoff point and introducing an upper limit of handled cases. We report the results of this sensitivity check in Section 4.6.5. Our sample restrictions leave us with a baseline sample of 124,913 cases handled by 1,488 caseworkers. We use this sample to construct our caseworker stringency instruments.

As information from social security records is only available from January 2016, we have to restrict our sample for estimation purposes. Specifically, our main estimation sample includes cases started between February 2016 and December 2017. Starting the sample in February 2016 allows us to control for outcomes in the month before infringement. Our main estimation sample includes 51,848 cases handled by

[14] The new policy instructed caseworkers to always impose a reprimand in case of a first-time infringement and a sanction in case of a recurring infringement.

1,303 caseworkers. Still, the cases come from all 37 PES offices in the Netherlands. For outcomes available before January 2016, we will also report results using a larger estimation sample reaching back to February 2014. Table 4.B.1 in Appendix 4.B shows how different restrictions affect the number of cases and caseworkers in our samples.

4.4.3 Summary Statistics

Table 4.1 reports summary statistics for our main estimation sample of 51,848 first-time infringement cases. Panel A summarizes claimant characteristics. The claimants included in our sample are on average 41 years old ($SD = 11.5$); roughly half of our sample are women. The largest share of claimants (38 percent) has completed an intermediate level of education, while a quarter of our sample has a lower level of education, and 14 percent of our sample is highly educated. For 24 percent of claimants, the education level is unknown.[15] Previous employment spreads across various sectors, though 80 percent of claimants had a job in either Service, Retail, Health, Manufacturing, Temporary Employment, or Transport (in descending order) before registering for UI benefits.[16] The average claimant in our estimation sample earned roughly €2,100 ($2,596 PPP) per month in the last job before entering benefits and had roughly ten months of UI entitlement left at the time of the infringement. Roughly half of our sample had work in the month before the infringement, which is not surprising given that UI benefits can also be collected as a supplementary income.[17] Most claimants (87 percent) received UI benefit payments in the month before infringement. Claimants may collect zero benefits due to, e.g., settlements with other incomes.

Panel B of Table 4.1 reports summary statistics for the outcomes of infringement cases. Caseworkers deemed the reasons for noncompliance invalid in 58 percent of all cases. Decomposing this share into the three possible subsequent outcomes shows that 25 percent of all cases resulted in a sanction, 18 percent resulted in a reprimand, and 15 percent resulted in no measure registered. Thus, conditional on giving an invalid reason, a sanction followed in 43 percent, a reprimand in 31 percent, and no penalty in 26 percent of the cases.[18] Panel C shows the fraction

[15] The PES classifies education levels as follows: lower education includes primary until upper secondary education; intermediate education comprises post-secondary non-tertiary education; higher education is tertiary education.

[16] The sector classification of the PES does not follow the NACE system but the system of the Dutch Tax Authority. That explains why a category for workers at temporary employment agencies exists, for example.

[17] Remember that job search requirements also apply to working claimants.

[18] As mentioned before, the treatment category 'no penalty despite invalid reason' may result from administrative errors or claimants finding employment or terminating their spell for other reasons during the review process. To shed more light on potential explanations, we regress a dummy variable indicating this treatment on the claimant characteristics listed in Panel A of Table 4.1. Table 4.B.3 in Appendix 4.B reports the results. We find few significant correlations, all of which are economically small. This finding leads us to believe that the third treatment category largely comprises random administrative errors. In Section 4.6.5, we exclude cases in the third treatment category from our estimation sample. This does not change our results qualitatively.

Table 4.1 Summary Statistics Main Estimation Sample.

	Mean (1)	SD (2)	Min. (3)	Max. (4)	Observations (5)
Panel A: Claimant characteristics					
Female	0.53	0.50	0	1	51,848
Age	40.91	11.47	19	64	51,848
Lower education	0.24	0.43	0	1	51,848
Intermediate education	0.38	0.49	0	1	51,848
Higher education	0.14	0.35	0	1	51,848
Education unknown	0.24	0.43	0	1	51,848
Manufacturing	0.13	0.33	0	1	51,848
Health	0.14	0.34	0	1	51,848
Retail	0.16	0.37	0	1	51,848
Services	0.18	0.39	0	1	51,848
Transport	0.08	0.27	0	1	51,848
Temporary employment agency	0.12	0.33	0	1	51,848
Other sector	0.20	0.40	0	1	51,848
Monthly wage (euro)	2,099	1,057	0	4,718	51,848
UI months left	10.14	8.7	0	283	51,848
Employed (t-1)	0.48	0.50	0	1	51,768
UI benefits (t-1)	0.87	0.34	0	1	51,763
Panel B: Case outcomes					
Invalid reason	0.58	0.49	0	1	51,848
Sanction	0.25	0.43	0	1	51,848
Reprimand	0.18	0.38	0	1	51,848
None	0.15	0.36	0	1	51,848
Panel C: Case year					
2016	0.37	0.48	0	1	51,848
2017	0.63	0.48	0	1	51,848

Note: The PES registries may include faulty entries, which explains the unusually high maximum for UI months left.

of cases per year. Roughly two-thirds of cases were handled in 2017, and one-third in 2016. The increase in case numbers is related to the PES expanding automatic infringement cases.

Table 4.2 reports summary statistics for the 1,303 caseworkers that handled the cases in our estimation sample. As data on caseworkers is not accessible, we have to use case data to construct caseworker variables. We use the number of months between the first and the last case handled within our baseline sampling period (January 2014 to December 2017) as a proxy for caseworkers' work experience. The average work experience is 38 months (SD = 12.0), which shows that most caseworkers handled cases during the entire period. We can also calculate the average number of cases handled; 88 cases (Median = 58) during the whole sampling period and 2.5 cases per active month. Figure 4.A.1 in Appendix 4.A shows the distribution of cases per caseworker. The distribution is right-skewed and reveals that most caseworkers (80 percent) handled less than 100 cases in total and less than 4 cases per month.

Table 4.2 Summary Statistics Sample of Caseworkers.

	Mean (1)	SD (2)	Min. (3)	Max. (4)	Observations (5)
Months worked within sampling period	38.07	11.97	2	48	1,303
Total infringement cases	88.42	113.82	20	1,639	1,303
Infringement cases / month	2.54	2.93	0	36	1,303

Note: Months worked within sampling period are the months between the first and the last infringement case handled within the baseline sampling period (January 2014 to December 2017).

4.5 Empirical Strategy

4.5.1 IV Model

We aim to identify the causal effects of sanctions and reprimands in UI on outcomes related to recidivism, job search behavior, employment, and benefit dependency. We must consider that caseworkers' decisions are endogenous due to selection based on job seekers' (un)observed characteristics. For instance, caseworkers may be more inclined to sanction less motivated job seekers or job seekers with lower employment prospects. In that case, estimating effects with an ordinary least squares (OLS) regression of outcomes on a variable indicating the respective treatment leads to biased results. Sanctions may show to negatively affect employment chances, although these correlations are driven by (un)observables instead of exposure to a sanction.

We follow an IV approach to circumvent endogeneity problems, using caseworker stringency as an instrument for the actual treatment. We exploit the fact that—due to discretionary power—-caseworkers systematically vary in their tendency to impose a certain disciplinary measure and that within each PES office, infringement cases are quasi-randomly assigned to caseworkers. This setup causes exogenous variation in the probability of receiving a particular treatment, and we use this variation to identify the causal effects of receiving a sanction or a reprimand. Our approach builds on various recent studies from different contexts that use decision makers' stringency as an instrument for the actual treatment. Illustrative examples come from French and Song (2014), Arni and Schiprowski (2019), and Bhuller et al. (2020). We could measure caseworker stringency as a simple average treatment rate among all cases handled, an approach equivalent to using a full set of caseworker dummy variables as instruments for the actual treatment. Instead, we follow the previous literature and construct our instruments as leave-out mean treatment rate, i.e., the average sanctioning or reprimanding rate for all cases handled by a caseworker but excluding the case considered. This approach should remove any mechanical relationship between an individual's case outcome and the instruments.

As mentioned in Section 4.3, caseworker decisions in infringement cases are multi-leveled and multidimensional. In a first step, the caseworker decides whether the claimant gave a valid reason for violating search requirements. If the reason is deemed invalid, the caseworker determines the disciplinary measure, which can take the form of a sanction or a reprimand. Remember that no measure is registered in some cases despite giving an invalid reason. We specify a model with multiple endogenous predictors to account for the multidimensionality of caseworker decisions. We estimate this model using a two-stage least squares (2SLS) procedure.[19] The second stage equation of our baseline model takes the following form:

$$Y_{it} = \alpha + \beta_t^S D_i^S + \beta_t^R D_i^R + \beta_t^N D_i^N + X_i'\Theta + \eta + \epsilon_{it} \tag{4.5.1}$$

The endogenous variable D_i^S is an indicator variable taking the value 1 in case claimant i received a benefit sanction and 0 otherwise. Likewise, the endogenous variables D_i^R and D_i^N indicate receiving a reprimand and not being disciplined despite giving an invalid reason, respectively. Hence, the omitted reference category in our model comprises claimants that violated job search requirements but were not sanctioned or reprimanded because they gave a reason deemed valid. Note that caseworkers cannot impose multiple disciplinary measures, i.e., all treatments are mutually exclusive.

The vector X_i contains background characteristics of claimant i. We include the variables listed in Panel A of Table 4.1.[20] η denotes interacted PES office $\times$ calendar month fixed effects, which we include to account for the possibility that regulatory regimes may vary across PES offices and time. This approach also controls for any differences in the claimant base across PES offices. Our results remain unaffected when using PES office $\times$ quarter fixed effects instead. Also, we obtain qualitatively similar results when excluding X_i from the model. Furthermore, our results do not change when controlling for caseworkers' experience or their tendency to assign other treatments such as workshops or exemptions (measured as corresponding leave-out means). We report the results of all these sensitivity checks in Section 4.6.5. Lastly, Y_{it} denotes the dependent variable of interest, measured in month t after the month in which claimant i's infringement case started. We thus normalize all procedures to start in month zero. Due to data availability, we can only measure outcomes up to and including month 12. ϵ_{it} is the error term.

As a result of including three endogenous predictors, we have three first stage equations, specified in Eq.(4.5.2)—(4.5.4):

$$D_i^S = \alpha + \gamma Z_{j,-i}^S + \mu Z_{j,-i}^R + \sigma Z_{j,-i}^N + X_i'\Theta + \eta + \nu_i \tag{4.5.2}$$

$$D_i^R = \alpha + \delta Z_{j,-i}^S + \pi Z_{j,-i}^R + \rho Z_{j,-i}^N + X_i'\Theta + \eta + \omega_i \tag{4.5.3}$$

$$D_i^N = \alpha + \psi Z_{j,-i}^S + \phi Z_{j,-i}^R + \tau Z_{j,-i}^N + X_i'\Theta + \eta + \upsilon_i \tag{4.5.4}$$

In each equation, we regress a respective endogenous treatment indicator on individual covariates, PES office $\times$ month fixed effects, and measures of caseworker

[19] Note that the 2SLS estimator using the leave-out mean treatment rate as an instrument for the actual treatment is numerically equivalent to a jackknife instrumental variables estimator (JIVE) with a full set of caseworker dummy variables as instruments.

[20] We account for missing values in the variables Employment (t-1) and UI benefits (t-1) by coding missing values as 0 and including a dummy variable indicating missing values.

stringency. These measures, denoted by $Z^S_{j,-i}$, $Z^R_{j,-i}$, and $Z^N_{j,-i}$, are our instrumental variables. We calculate the treatment-specific stringency of caseworker j assigned to claimant i's case as their respective average treatment rate (e.g., sanctioning rate) across all cases of j excluding the case of i (leave-out mean). In Section 4.6.5 we show that our results are robust to using a split-sample approach when calculating caseworker stringency. Using a split-sample approach, we can ensure no relationship between a claimant's case outcome and our stringency measures. We use this approach as a sensitivity check and not our main strategy due to the smaller sample size and reduced statistical power.

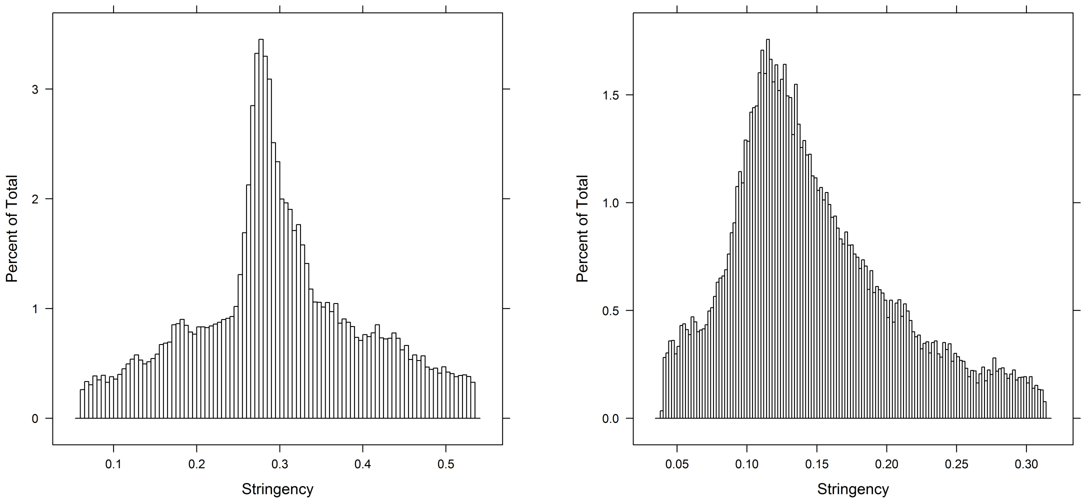
Our parameters of interest are the coefficients β^S_t and β^R_t in Eq.(4.5.1). These parameters estimate the average causal effect of receiving a sanction and a reprimand, respectively, on outcome Y_{it}. The parameters estimate local average treatment effects (LATE), i.e., effects for the subgroup of claimants who could have received a different treatment had a different caseworker handled their case.[21] When discussing results, we focus on the effects of sanctions and reprimands and disregard the third treatment category (no penalty despite invalid reason), as it is largely unaccounted for what that treatment exactly entails. As such, we also have no theoretically informed expectations regarding treatment effects. For completeness, we report estimates of β^N_t in Table 4.B.8 in Appendix 4.B.

4.5.2 Assessing the Instruments

Relevance. Figure 4.2 shows the identifying variation in the instruments for the sanctioning and reprimanding decision, respectively. The variation for sanctioning stringency is shown on the left-hand side, and the variation for reprimanding stringency is on the right-hand side. The histograms plot the distribution of each instrument after controlling for interacted PES office × calendar month fixed effects and individual covariates. Our instruments have a mean of 0.30 (sanction) and 0.15 (reprimand), and a standard deviation of 0.14 (sanction) and 0.09 (reprimand). The two histograms show that (conditional on the factors mentioned above) there is quite some variation left in caseworkers' stringency. Caseworkers' tendency to sanction is more widespread than their tendency to reprimand. For example, a caseworker at the 10th percentile imposes a sanction in 13 percent of cases and a reprimand in 7 percent of cases. A caseworker at the 90th percentile sanctions 47 percent and reprimands 26 percent of cases.

Table 4.3 reports estimates for the first-stage Eq.(4.5.2)–(4.5.3).[22] The models in the uneven columns only include office × month fixed effects, while the models in the even columns also include individual covariates. The point estimate of 0.53 for sanctioning stringency in Column (1) indicates that the likelihood of receiving a sanction increases by roughly 5 percentage points when a 10 percentage point stricter caseworker handles the case. The point estimate of 0.91 for reprimanding stringency in Column (3) can be interpreted similarly: assignment to a 10 percentage point stricter caseworker increases the chances of receiving a reprimand by

[21] This interpretation holds under the assumptions of instrument exogeneity and monotonicity.
[22] We report estimates for the first-stage Eq.(4.5.4) in Table 4.B.2 in Appendix B.



(a) Decision to Sanction. **(b)** Decision to Reprimand.

Figure 4.2 Identifying Variation in Stringency Instruments.

Note: The values plotted in the histograms are mean-standardized residuals from a regression of the respective instrument on interacted PES office × calendar month fixed effects and individual covariates (as listed in Panel A of Table 4.1). The histograms show the distribution of the mean-standardized residuals excluding the top and bottom 5 percent. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. The figure is inspired by Bhuller et al. (2020).

roughly 9 percentage points.

Both point estimates are highly significant. The F-statistics of joint significance tests of all three instruments in Eq.(4.5.2) and (4.5.3) are 105 and 654, respectively. This result provides additional evidence for the strength of the two instruments. As we would expect, the different treatments show the strongest association with their corresponding instrument, whereas instruments for other treatment decisions explain little variation. Note that our estimates do not change substantially when including individual covariates. This finding provides a first indication that our instruments are (conditionally) independent. Remember that we restricted our sample to caseworkers with at least 20 cases. Our first-stage results are robust to choosing higher cutoff points for cases handled. They are also robust to using a split-sample instrument, i.e., when using a random half of the sample to construct instruments and using these instruments for estimation in the other half. We report on both sensitivity checks in Section 4.6.5.

Conditional Independence. For our stringency measures to be valid instruments, they should be (conditionally) independent from individual covariates related to our outcomes of interest. We can test this assumption for the covariates

Table 4.3 First-Stage Estimates.

	Pr(Sanction)		Pr(Reprimand)	
	(1)	(2)	(3)	(4)
Stringency (Sanction)	0.525***	0.517***	0.026*	0.026*
	(0.021)	(0.021)	(0.012)	(0.012)
Stringency (Reprimand)	−0.119***	−0.115***	0.905***	0.890***
	(0.026)	(0.026)	(0.031)	(0.030)
Stringency (None)	0.027	0.028	0.087***	0.081***
	(0.034)	(0.033)	(0.022)	(0.022)
Office/month FE	Yes	Yes	Yes	Yes
Individual covariates	No	Yes	No	Yes
Dependent mean	0.249	0.249	0.181	0.181
F-statistic	105	107	654	630
R2 (adj.)	0.328	0.336	0.230	0.240
Observations	51,848	51,848	51,848	51,848

Note: Estimates from OLS regressions. Dependent variables are dummies indicating actual treatment (sanction, reprimand). Explanatory variables are caseworker stringency instruments. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. Standard errors are clustered at the caseworker level (N=1,303) and reported in parentheses. *p < 0.05; **p < 0.01, ***p < 0.001.

we observe in our data (i.e., the variables listed in Panel A of Table 4.1). Including individual covariates in our first-stage regressions left our estimates pretty much unchanged, which already provided evidence in favor of the conditional independence assumption. As a second test, we examine whether our instruments are systematically related to claimant background characteristics. In doing so, we focus on the instruments for the sanctioning and the reprimanding decision.[23]

We begin by regressing a dummy indicating sanction and reprimand receipt, respectively, on individual covariates and office $\times$ month fixed effects. Column (1) and (3) of Table 4.4 report the corresponding estimates. We find highly significant correlations between both dependent variables and several covariates. Female claimants, e.g., are more likely to receive a sanction, while the sanctioning risk decreases with higher levels of education. The opposite is true for the risk of being reprimanded. These results provide evidence in favor of our initial assumption that treatment decisions in UI are endogenous due to selection based on (un)observable characteristics.

Column (2) and (4) of the same table report estimates of regressing our stringency instruments on the same set of covariates and office $\times$ month fixed effects. Overall, we find weaker and fewer significant correlations. Still, some correlations are statistically significant. Female claimants, e.g., have caseworkers that are on average 0.3 percentage points stricter with regards to sanctions. For instance, claimants with previous jobs in the sectors Retail, Service, and Transport have on average 0.4 percentage points stricter caseworkers regarding reprimands. We also have to reject the null hypothesis of a join significance test for both instruments.

The correlations between the probability of treatment and our instruments may reflect a certain degree of caseworker specialization within offices. As mentioned in Section 4.3, caseworkers at some PES offices may, e.g., specialize in claimants from

[23]Table 4.B.3 in Appendix 4.B reports results for the third treatment category, i.e., 'no penalty despite invalid reason'.

Table 4.4 Relation of Actual Treatment and Stringency Instruments With Claimant Background Characteristics.

	Pr(Sanction)	Stringency (Sanction)	Pr(Reprimand)	Stringency (Reprimand)
	(1)	(2)	(3)	(4)
Female	0.0434***	0.0034*	−0.0245***	−0.0024*
	(0.0040)	(0.0015)	(0.0036)	(0.0009)
Age	−0.0011***	−0.0000	−0.0009***	−0.0002**
	(0.0002)	(0.0001)	(0.0002)	(0.0001)
Intermed. education	−0.0148***	−0.0000	0.0048	0.0006
	(0.0044)	(0.0019)	(0.0039)	(0.0010)
Higher education	−0.0556***	−0.0097***	0.0280***	0.0036*
	(0.0058)	(0.0026)	(0.0055)	(0.0015)
Education unknown	−0.0205***	−0.0020	−0.0012	0.0033*
	(0.0049)	(0.0023)	(0.0044)	(0.0016)
Health	−0.0023	0.0006	0.0069	0.0026
	(0.0067)	(0.0027)	(0.0062)	(0.0018)
Retail	−0.0055	0.0025	0.0118*	0.0038*
	(0.0065)	(0.0025)	(0.0060)	(0.0017)
Services	−0.0089	−0.0019	0.0109	0.0038*
	(0.0061)	(0.0025)	(0.0060)	(0.0017)
Transport	−0.0017	0.0026	0.0083	0.0040*
	(0.0073)	(0.0029)	(0.0069)	(0.0018)
Temp. employment	−0.0021	−0.0032	0.0031	0.0005
	(0.0069)	(0.0027)	(0.0056)	(0.0016)
Other sector	−0.0041	0.0040	−0.0049	0.0005
	(0.0061)	(0.0024)	(0.0056)	(0.0016)
Monthly wage (x1.000)	−0.0141***	−0.0019*	0.0118***	0.0028***
	(0.0020)	(0.0008)	(0.0020)	(0.0006)
UI months left (x10)	−0.0126***	−0.0018	0.0090***	0.0018*
	(0.0028)	(0.0012)	(0.0027)	(0.0008)
Employed (t-1)	0.0062	−0.0022	−0.0116**	0.0028*
	(0.0036)	(0.0014)	(0.0038)	(0.0011)
UI benefits (t-1)	0.0390***	0.0011	0.1004***	0.0078***
	(0.0052)	(0.0021)	(0.0067)	(0.0014)
Dependent mean	0.2493	0.2970	0.1809	0.1515
R2 (adj.)	0.3042	0.6176	0.1953	0.7264
F-statistic join test	21.74	3.13	24.39	4.76
p-value	[.000]	[.000]	[.000]	[.000]
Observations	51,848	51,848	51,848	51,848

Note: Estimates from OLS regressions. Dependent variables are dummies indicating actual treatment (sanction, reprimand) or caseworker stringency instruments. Explanatory variables are individual covariates. All covariates except age, UI months left, and monthly wage are binary variables. The reference category for education level is *lower education*; the reference category for sector is *manufacturing*. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. Standard errors are clustered at the caseworker level (N=1,303) and reported in parentheses. All regressions include interacted PES office × calendar month fixed effects. *p < 0.05; **p < 0.01, ***p < 0.001.

specific industry sectors. Nevertheless, we think of these correlations as unproblematic, given that they are economically small. In Column (2), the largest estimate (-0.0097 for higher education) is equivalent to 0.07 standard deviations of the dependent variable, for example.[24] In Column (4), the largest estimate (0.0078 for UI benefits in t-1) corresponds with 0.09 standard deviations of the dependent variable. Moreover, individual covariates only explain an additional 0.5–0.7 percentage points of the variation in our stringency instruments when adding them to a model only including office × month fixed effects.

Note that PES offices do not have access to other administrative datasets than those obtained for this study. Therefore, we deem it unlikely that caseworker assignment is related to unobserved characteristics. In conclusion, we assume that by controlling for individual characteristics in our main estimation models, we can address the small amount of non-randomness encountered. As we will show in Section 4.6.5, most of our results remain unaffected when excluding individual covariates from our models.

Exclusion Restriction. Conditional independence of our instruments allows us to interpret the reduced form estimates of our 2SLS model as the causal effects of being assigned a stricter caseworker. We need an exclusion restriction to interpret our 2SLS estimates as the causal effects of receiving a sanction or a reprimand. For the exclusion restriction to hold, caseworker stringency should affect outcomes only through the sanctioning and reprimanding channel and not in any other way. As is well described in previous studies using stringency as an instrument (see, e.g., Arni and Schiprowski, 2019; Bhuller et al., 2020), the potential relation between stringency and other treatments or caseworker characteristics poses a challenge in that regard. Stricter caseworkers may be more inclined to assign claimants to active labor market programs, e.g., or have more experience, both of which may directly affect outcomes of interest. To assess whether the exclusion restriction holds, we follow the steps suggested by the earlier empirical work mentioned above, in particular, Arni and Schiprowski (2019).

First, we examine whether our instruments are correlated with the probability of receiving other treatments often assigned by PES caseworkers. We construct three binary variables indicating (i) assignment to training programs and workshops, (ii) assignment to coaching meetings with a caseworker, and (iii) exemptions from job search requirements. The three variables take the value 1 in case claimant i received the respective treatment at least once six months post-procedure. Again, we limit our analysis to the instruments for the sanctioning and the reprimanding decision.

Column (1) and (2) of Table 4.5 report summary statistics for the three variables measuring other treatment assignments. Column (3) and (5) report estimates from regressing a dummy indicating sanction and reprimand receipt, respectively, on the three variables, controlling for office × month fixed effects and individual covariates. As expected, we find that both sanctions and reprimands are related to other

[24] Remember that the standard deviation is 0.14 for the sanctioning instrument and 0.09 for the reprimanding instrument (conditional on office × month fixed effects).

Table 4.5 Relation of Sanctioning and Reprimanding Probability and Respective Stringency Instruments With Other Treatments Assigned by Caseworkers.

	Explanatory variables		Dependent variables			
	Mean	SD	Pr(Sanction)	Stringency (Sanction)	Pr(Reprimand)	Stringency (Reprimand)
	(1)	(2)	(3)	(4)	(5)	(6)
Pr(Workshop)	0.049	0.216	0.005	0.005	0.024**	0.002
			(0.008)	(0.003)	(0.008)	(0.002)
Pr(Coaching)	0.078	0.269	0.000	0.000	0.005	−0.002
			(0.007)	(0.003)	(0.007)	(0.002)
Pr(Exemption)	0.182	0.386	−0.021***	−0.000	0.017***	0.000
			(0.004)	(0.002)	(0.005)	(0.001)
Dependent mean			0.249	0.297	0.181	0.152
Observations			51,848	51,848	51,848	51,848

Note: Estimates from OLS regressions. Dependent variables are dummies indicating actual treatment (sanction, reprimand) or caseworker stringency instruments. Explanatory variables are dummies indicating other treatments assigned by PES caseworkers in the six months post-procedure. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level (N=1,303) and reported in parentheses. *p < 0.05; **p < 0.01, ***p < 0.001.

treatment assignments in the six months post-procedure. Sanctioned claimants are less likely, and reprimanded claimants are more likely to receive an exemption from job search requirements. Those reprimanded are also more likely to participate in training programs or workshops. Column (4) and (6) report estimates using our instruments as dependent variables. In contrast to the previous results, the correlations between our instruments and the probabilities of receiving other treatments are small and statistically insignificant. Caseworker stringency thus appears to be unrelated to other treatments usually assigned by PES caseworkers. Additionally, we calculate leave-out means for each of the other three treatment assignments (in the same way as we did for our instruments) and include these leave-out means as control variables in our main 2SLS model. This exercise leaves our results unaffected, providing additional evidence favoring the exclusion restriction. We report detailed results in Section 4.6.5.

Second, we examine the relationship between our stringency instruments and our proxy for caseworker experience. Remember that, in the absence of data on caseworkers, we measure caseworker experience as the number of months between the first and last case that a caseworker handled within the baseline sampling period. Table 4.6 shows the results of regressing our stringency instruments on that measure and office × month fixed effects, using our sample of caseworkers. We find significant associations of both stringency instruments with caseworker experience. Caseworkers with ten months more experience are 4 percentage points stricter in sanctioning and 1 percentage point more lenient in reprimanding. As we will show in Section 4.6.5, including caseworker experience as a control variable in our main 2SLS model leaves results unchanged, however. This result alleviates concerns that our stringency instruments reflect caseworker experience.

Table 4.6 Relation of Stringency Instruments With Caseworker Experience.

	Stringency (Sanction) (1)	Stringency (Reprimand) (2)
Months worked within baseline sampling period	0.004***	−0.001***
	(0.001)	(0.000)
Dependent mean	0.396	0.082
Observations	1,303	1,303

Note: Estimates from OLS regressions. Dependent variables are caseworker stringency instruments. The explanatory variable is the months between the first and last case handled by a caseworker within the baseline sampling period (January 2014 to December 2017). All regressions include interacted PES office × calendar month fixed effects. The sample concerns caseworkers, which explains the smaller sample size. Robust standard errors are reported in parentheses. *p < 0.05; **p < 0.01, ***p < 0.001.

Monotonicity. It seems plausible that the causal effects of sanctions and reprimands are not the same across all types of claimants. Assuming that effects are heterogeneous, our instruments must satisfy the monotonicity assumption to interpret our estimates as LATE. In our setting, monotonicity requires that claimants sanctioned or reprimanded by a lenient caseworker would have received a sanction or reprimand if their case had been assigned to a stricter caseworker and vice versa.

As suggested by Bhuller et al. (2020), the monotonicity assumption in stringency instrument designs comes with two testable implications. According to the first implication, the first-stage estimates should be non-negative in any given subsample. Negative estimates in certain subsamples would indicate that there are groups of individuals among which caseworker stringency is negatively correlated with the risk of treatment. Following Bhuller et al. (2020), we split our sample into 14 subsamples along socio-demographic dimensions and sectors of previous employment. We then estimate the first-stage relationships for each of these subsamples using our baseline instruments. We report the results in Column (1) of Table 4.B.4 in Appendix 4.B. For brevity, the table only reports the results for the sanctioning instrument. The results for the reprimanding instrument are available upon request. For comparison, the first panel of the table shows the first-stage estimates for the full sample. Panel A splits the sample into two subsamples according to gender, Panel B into two subsamples according to claimants' age, Panel C into three subsamples according to the education level, and Panel D into six subsamples according to the sector of previous employment. Consistent with the first implication, first-stage estimates are non-negative in all subsamples and do not differ substantially in terms of magnitude; point estimates lie between 0.49 and 0.55 compared to 0.52 for the full sample.

According to the second implication, the first-stage estimates should also be non-negative in a given subsample when using every case outside the respective subsample to construct the instrument instead of the full sample. For example, using cases of intermediate and higher educated claimants when constructing the instrument for the subsample of lower educated claimants. Following Bhuller et al. (2020), we construct such reverse-sample instruments for each of the previous 14 subsamples. We then use these 14 reverse-sample instruments to re-estimate the first-stage relationship for each subsample. Column (2) of Table 4.B.4 reports the results of this exercise. As before, all point estimates are non-negative and indicate a strong first-stage relationship. In conclusion, both tests provide evidence in favor of the monotonicity assumption.

4.6 Results

4.6.1 Recidivism

An important objective of monitoring and sanctioning policies is to discourage non-compliant behavior in the future. As sanctions reduce the marginal cost of job search, they should, in theory, result in higher search effort and fewer incidences of noncompliance with search requirements. To study the effectiveness of sanctions and reprimands in this regard, we look at recidivism. Our recidivism measure is a dummy indicating the occurrence of a new infringement case (due to noncompliance with search requirements) at least once in the 12 months after the current procedure.[25]

Using this outcome measure, we have to consider two important caveats. First, exit from UI benefits affects recidivism probabilities. After all, compliance requirements are conditional on receiving benefits; there is no risk of noncompliance when exiting the scheme. Consequently, besides capturing the preventive success of sanctions and reprimands, the effects on recidivism may in part be attributed to accelerated or delayed exit from benefits. In Section 4.6.4, we find no effects on benefit dependency, which gives us reason to assume that the entanglement of effects is negligible in our case.

Second, aside from actual job search effort, recidivism probabilities also depend on claimants learning to avoid detection. Remember that UI claimants must register their job search activities through the online portal of the PES and that infringement cases are started based on the information registered. Claimants can avoid infringement procedures by actually providing and registering the required effort. However, they may also register a sufficient number of activities without having performed (all of) them or register activities that took place outside the respective four-week time window. As we can only observe incidences of new cases and not claimants' actual effort, the effects on recidivism probabilities may capture both actual compliance and formal compliance with search requirements.

Table 4.7 reports estimates for the effects of sanctions and reprimands on our recidivism measure. For comparison, the table also includes estimates from an OLS regression, controlling for office $\times$ month fixed effects and individual covariates. The OLS results in Column (1) show a small negative correlation between sanctions and reprimands and our recidivism measure. Both treatments are associated with a 1–2 percentage point lower chance of being involved in a new case at least once by month 12. These results stand in stark contrast to the causal effects reported in the other two columns.

[25] To avoid including potential double-registrations of the current case, we only regard cases processed two months after the current case or later.

Table 4.7 Effects of Sanctions and Reprimands on Recidivism.

	Pr(New case)		
	OLS	RF	2SLS
	(1)	(2)	(3)
Sanction	−0.017**		−0.110***
	(0.006)		(0.031)
Reprimand	−0.012*		0.006
	(0.006)		(0.030)
Stringency (Sanction)		−0.057***	
		(0.017)	
Stringency (Reprimand)		0.017	
		(0.025)	
Office/month FE	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes
Dependent mean	0.283	0.283	0.283
Observations	51,848	51,848	51,848

Note: Shown are results from the main estimation sample (cases started between February 2016 and December 2017). The dependent variable is a dummy indicating the occurrence of a new infringement case between two and 12 months after the current case. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are reported in parentheses. OLS standard errors are clustered at the claimant level, while RF and 2SLS standard errors are clustered at the caseworker level (N=1,303). *p < 0.05; **p < 0.01, ***p < 0.001.

First, consider Column (2), which presents reduced-from (RF) effects obtained by regressing our recidivism measure on caseworker stringency instruments. The results indicate that assignment to a caseworker, who is 10 percentage points stricter regarding sanctions, lowers the chance of being involved in a new case by 0.6 percentage points. Stringency with regards to reprimands is not significantly related to recidivism. The 2SLS results in Column (3) show how these effects scale up to the causal effects of sanctions and reprimands. We find that a sanction lowers the chance of recidivism by 11 percentage points, which corresponds with a negative effect of nearly 40 percent relative to the sample mean. The point estimate for the effect of a reprimand is close to zero and statistically insignificant. Hence, our IV estimates suggest much stronger effects for sanctions and no effects for reprimands compared to OLS. The divergence between estimates may hint at caseworkers sanctioning claimants who are less likely to re-offend, supporting our initial assumption that caseworkers' treatment decisions are endogenous. In Table 4.B.5 in Appendix 4.B we decompose the effects on recidivism rates. We use dummies indicating the outcome of a new infringement case distinguishing between a sanction, a reprimand, and giving a valid reason.[26] It shows that both a sanction and a reprimand increase the chance of receiving a sanction in the case of a new infringement. This finding is consistent with PES policy, which prescribes imposing a sanction if job search requirements are violated again.

As data on recidivism dates back to February 2014, we can also use our more extensive baseline sample for estimation. Table 4.B.6 in Appendix 4.B presents the results. The findings for sanctions are identical; relative to a lower sample mean

[26] In the case of several follow-up infringements, our dummies indicate the outcome of the first follow-up infringement case.

recidivism rate, the effect of a sanction still corresponds with a roughly 40 percent decrease in recidivism probability. The results for reprimands are different, however. Instead of a null effect, we now find a strong positive effect of 19 percentage points, an increase of roughly 80 percent relative to the sample mean. What could explain the divergence in effects between the two samples? One potential explanation concerns a change in treatment. While the sanctioning treatment is stable over time and tangible, the PES office may have changed how it communicates reprimands to claimants. For instance, including a more severe threat or making the consequences of a sanction more salient may have eliminated previous boomerang effects. It is also thinkable that in earlier sample years, reprimands involved a more thorough check of compliance activities in subsequent months. In that case, the strong positive effect may also enclose higher chances of being monitored. The expansion of automated monitoring, which started in 2017, may explain why effects disappear in the later sample.

In conclusion, it appears that sanctions lead to significantly lower chances of recidivism. However, it remains an open question to what extent this effect is driven by increased job search effort (i.e., actual compliance) or changes in registration behavior that lower the chances of being detected (i.e., formal compliance). For reprimands, the findings remain ambiguous, though we can quite confidently preclude effects in the direction of lower recidivism rates.

4.6.2 Job Search

Examining the effects of sanctions and reprimands on job search may help explain our findings for recidivism. In our data, we observe the total number of job search activities that claimants registered through the online portal of the PES in the month after the month of infringement. As mentioned in Section 4.3.2, claimants may register various activities, including job applications, job interviews, or contacts with a potential employer or people in one's network. Table 4.8 shows the effects on total job search activities in the same way as the previous table.[27] On average, claimants in our sample registered 2.7 activities. This number is lower than the requirement of four activities every four weeks. The divergence may be explained by different time windows (four weeks versus the following month), claimants with lower effort targets, and claimants violating the requirement.

Column (3) shows the 2SLS estimates for the effects of both treatments on registered activities. It shows that issuing a reprimand leads claimants to register significantly more activities—one more activity in absolute terms and 40 percent more activities relative to the sample mean. A sanction also results in significantly more activities registered, though the effect size is much smaller—12 percent relative to the sample mean. When using our larger baseline sample for estimation, we find confirmation of the former effect, while the effect of a sanction is smaller and not statistically significant (see Table 4.B.7 in Appendix 4.B for results).

[27]We eliminate extreme values by replacing the top 0.1 percent with the value at the 99.9th percentile (winsorization).

Table 4.8 Effects of Sanctions and Reprimands on Job Search.

	OLS	Reduced form	2SLS
	(1)	(2)	(3)
Sanction	−0.106***		0.325*
	(0.030)		(0.137)
Reprimand	0.543***		1.079***
	(0.033)		(0.138)
Stringency (Sanction)		0.191**	
		(0.074)	
Stringency (Reprimand)		0.966***	
		(0.111)	
Office/month FE	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes
Dependent mean	2.683	2.683	2.683
Observations	42,237	42,237	42,237

Note: Shown are results from the main estimation sample (cases started between February 2016 and December 2017). The dependent variable is the number of registered job search activities with the PES portal in the month after the month of infringement. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are reported in parentheses. OLS standard errors are clustered at the claimant level, while RF and 2SLS standard errors are clustered at the caseworker level (N=1,303). *p < 0.05; **p < 0.01, ***p < 0.001.

In conclusion, the results for job search suggest that claimants show a stronger response in registration behavior following a reprimand. Noticeably, the previous section showed that chances of recidivism are unaffected by reprimands and decline after a sanction. What could explain these diverging results? Potentially, reprimanded claimants report more activities directly after infringement but revert to the same behavior as non-reprimanded claimants in later months, both in terms of effort and timing of registration. Sanctions, on the other hand, may motivate claimants to follow the requirement of four activities in four weeks more thoroughly and persistently, however, without inducing higher (registered) effort.

4.6.3 Labor Market Outcomes

We continue by studying the effects of sanctions and reprimands on different labor market outcomes. We examine both effect dynamics and cumulative effects. Figure 4.3 displays IV estimates for the effects on four outcomes over time. Those outcomes are employment probabilities, earnings, and chances of working under a temporary and a permanent contract, respectively. We obtain estimates by running our baseline 2SLS model separately for each month t. In addition to point estimates, the graphs also plot 95 percent confidence intervals (gray and colored areas).

Panel A shows the effects on employment probabilities. We operationalize employment as having positive labor market earnings in a given month. Note that we only observe the earnings of employed workers while income from self-employment remains unobserved. We find no evidence of effects in the 12 months post-procedure for reprimands. Point estimates remain largely constant and close to zero. For sanctions, point estimates turn negative (up to minus 6 percentage points) in later months, but effects remain statistically insignificant at conventional levels. Still,

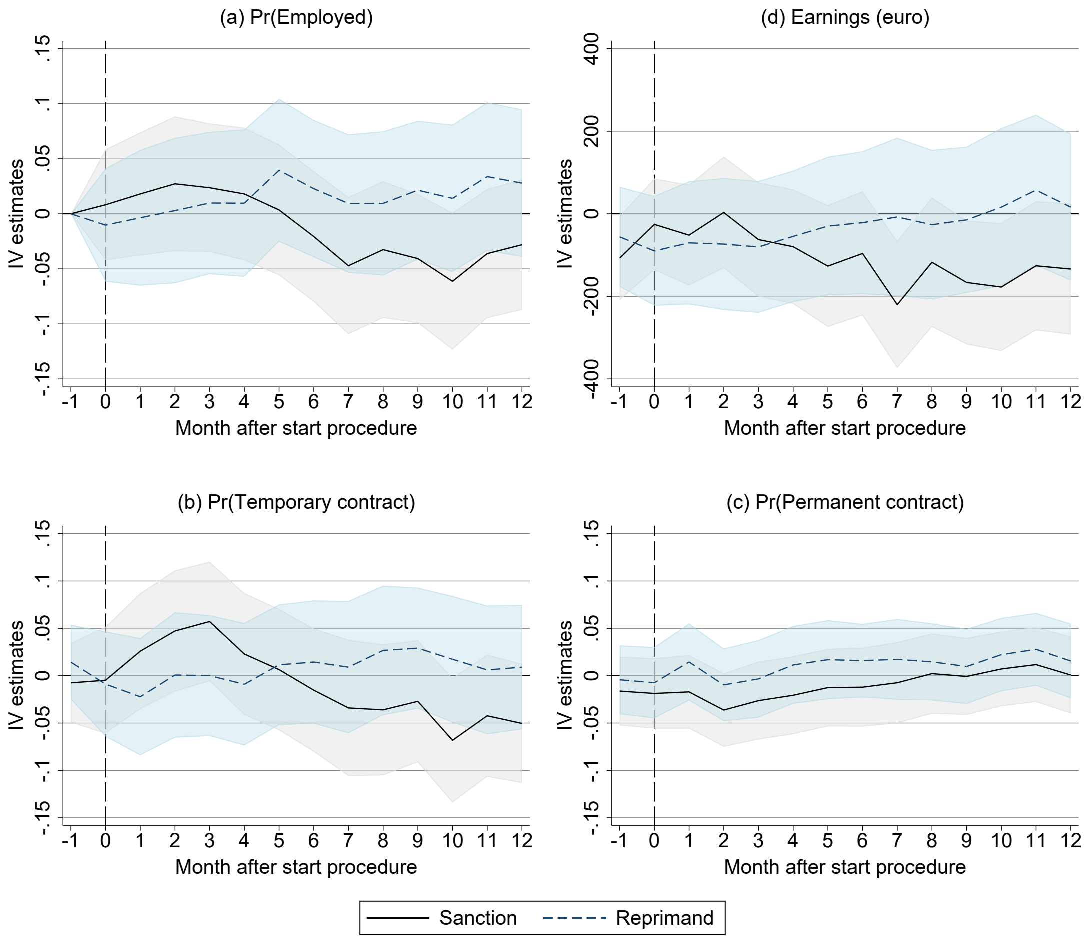


Figure 4.3 Effects of Sanctions and Reprimands on Labor Market Outcomes Over Time.

Note: Shown are results from the main estimation sample, i.e., cases started between February 2016 and December 2017 (N=51,848). All regressions include interacted PES office $\times$ calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level (N=1,303). Gray and colored areas show 95 percent confidence intervals.

Table 4.9 Effects of Sanctions and Reprimands on Cumulative Earnings.

	Earnings in months 1-12			Earnings in months 6-12		
	OLS	RF	2SLS	OLS	RF	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Sanction	-587^{***}		$-1,339$	-349^{***}		-932^{*}
	(148)		(743)	(86)		(424)
Reprimand	$-1,220^{***}$		-281	-412^{***}		43
	(172)		(913)	(100)		(518)
Stringency (Sanction)		-705			-484^{*}	
		(383)			(216)	
Stringency (Reprimand)		139			303	
		(743)			(420)	
Office/month FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes	Yes	Yes	Yes
Dependent mean	13,836	13,836	13,836	7,678	7,678	7,678
Observations	51,848	51,848	51,848	51,848	51,848	51,848

Note: Shown are results from the main estimation sample (cases started between February 2016 and December 2017). The dependent variables measure earnings in months 1-12 and 6-12 after the current case, respectively. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office $\times$ calendar month fixed effects and individual covariates. Standard errors are reported in parentheses. OLS standard errors are clustered at the claimant level, while RF and 2SLS standard errors are clustered at the caseworker level (N=1,303). *p < 0.05; **p < 0.01, ***p < 0.001.

with a view to confidence intervals, the results provide quite compelling evidence against (modest to large) *positive* employment effects of sanctions in the longer term.

Decomposing employment effects, Panel B and C show effects on the probabilities of working under a temporary contract and a permanent contract, respectively.[28] We view permanent contracts as a proxy for durable employment. Chances of permanent employment seem unaffected by both sanctions and reprimands—point estimates are close to zero throughout. Hence, receiving a sanction does not appear to harm re-employment quality in terms of durability. In line with the results for permanent employment, the insignificant employment effects shown in Panel A can largely be ascribed to temporary work.

Lastly, Panel D shows the effects on monthly earnings (incl. zero earnings).[29] We find significant negative effects in later months for sanctions. The average loss in earnings amounts to roughly €200 ($247 PPP) per month, a drop of roughly 16 percent relative to the sample mean. The timing of the negative effects on earnings matches with the negative point estimates for employment effects shown in Panel A. Relatively speaking, though, the effects on earnings are larger, suggesting (additional) adjustments in hours worked or wages. For reprimands, there is no evidence of effects on earnings.

Table 4.9 presents estimates from OLS, reduced form, and 2SLS regressions when measuring earnings cumulatively. The results in Column (1)–(3) concern the whole 12 month post-procedure period and the results in Column (4)–(6) the second

[28] For subjects with multiple employment contracts in a given month, we only count the highest-ranking contract, whereby permanent contracts rank higher than temporary contracts.

[29] Again, we eliminate extreme values by replacing the top 0.1 percent with the value at the 99.9th percentile.

six months. We find that a sanction reduces cumulative earnings in the second six months by €900 ($1,112 PPP) on average, which is a negative effect of 12 percent relative to the sample mean. The effect of a sanction on total cumulative earnings is slightly smaller in relative terms (–10 percent) and not statistically significant at conventional levels. In line with monthly results, we find no evidence of effects of reprimands on cumulative earnings.

4.6.4 Benefit Dependency

We now turn to the effects of sanctions and reprimands on benefit payments. We study benefit payments as they reflect both full independence from benefits (exit from benefits) and partial independence from benefits (working in addition to benefits). Figure 4.4 shows IV estimates for two types of benefit payments over time, UI benefits and sickness benefits. Panel A shows the effects on UI benefit payments. The results suggest that reprimands do not affect UI benefit payments—point estimates are close to zero and relatively precise. We find significant negative effects for sanctions in the month of infringement and the three subsequent months. In these four months, sanctions lead to roughly €140 ($173 PPP) lower payments per month on average, a decrease of roughly 16 percent relative to the sample mean. The effect pattern and magnitude correspond with the cut in benefits induced by a sanction: 25 percent for four months.[30] The timing of effects indicates that benefit cuts are processed and come into effect immediately. In later months, there are no effects of receiving a sanction. Taken together, these findings suggest that neither reprimands nor sanctions affect dependency on UI benefits. Data on benefit payments date back to February 2014, which is why we can re-estimate effects using our larger baseline sample. As Panel A of Figure 4.A.2 in Appendix 4.A shows, the results remain unchanged. Table 4.10 presents results for cumulative UI benefits in the same fashion as Table 4.9. The cut in benefits results in an average cumulative loss of €500 ($618 PPP) after six months, roughly 15 percent relative to the sample mean. After 12 months, the cumulative loss is 7 percent relative to the sample average and not statistically significant.

In addition to affecting job finding and exit from UI benefits, sanctions may also promote the use of other benefits. Trying to avoid future sanctions, individuals may switch to schemes that are subject to fewer or no requirements at all. Sanctions may also lead claimants to drop out of the labor force altogether (non-participation). Both effects have been documented by previous empirical work.[31] The data obtained from the PES allows us to examine the effects of sanctions and reprimands on sickness benefits receipt. These benefits can be claimed during sickness spells and involve fewer compliance requirements. For instance, claimants of

[30] That we find more moderate effects hints at some subjects receiving less severe cuts.

[31] The results of Manning (2009) and Petrongolo (2009), e.g., suggest that UI benefit claimants in the United Kingdom responded to a tightening of job search requirements by moving into non-participation and incapacity benefits, respectively. Arni et al. (2013) and Busk (2016) find that benefit sanctions encourage switches to non-participation. The findings of van den Berg et al. (2019) suggest that the use of sickness benefits increases after job vacancy referrals.

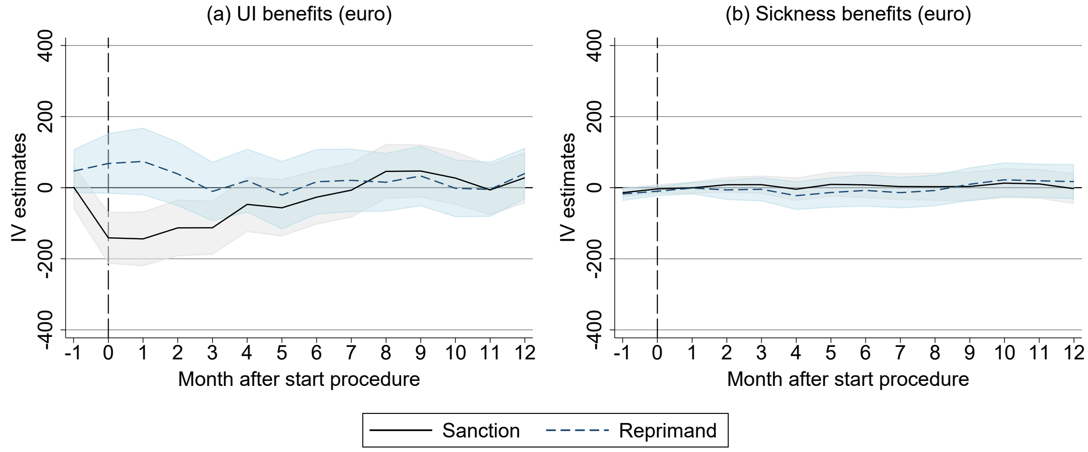


Figure 4.4 Effects of Sanctions and Reprimands on Benefit Payments Over Time.

Note: Shown are results from the main estimation sample, i.e., cases started between February 2016 and December 2017 (N=51,848). All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level (N=1,303). Gray and colored areas show 95 percent confidence intervals.

sickness benefits are largely exempted from search requirements. As the PES also administers sickness benefits, switching involves relatively little cost compared to other benefits.[32] Panel B of Figure 4.4 shows IV estimates for the effects of sanctions and reprimands on sickness benefit payments. Point estimates are close to zero and quite precise. This result suggests that neither sanctions nor reprimands promote a transition to sickness benefits. We obtain the same result when estimating effects using our larger baseline sample (see Panel B of Figure 4.A.2 in Appendix 4.A).

Lastly, Figure 4.5 shows effects on the probability of non-participation, which we operationalize as receiving neither labor income nor UI or sickness benefits. Again, we find no evidence of effects for both treatments. For reprimands, point estimates in later months are negative (up to minus 5 percentage points) but not statistically significant at conventional levels.

4.6.5 Sensitivity Analyses

We conduct several additional analyses to assess the sensitivity of our results. First, we test whether our results are robust to choosing other cutoff points regarding caseworker caseloads and different ways of constructing our instruments. Second, we check whether our results change when controlling for other treatments often assigned by PES caseworkers or caseworker experience. Lastly, we perform sensitivity checks across some further dimensions, i.e., excluding individual covariates,

[32]It should be noted that next to strategic behavior, claiming sickness benefits may also reflect actual illness and thus document the health effects of sanctions. Lacking data on health outcomes, we cannot distinguish between these two mechanisms.

Table 4.10 Effects of Sanctions and Reprimands on Cumulative UI Benefits.

	UI benefits in months 1-12			UI benefits in months 1-6		
	OLS	RF	2SLS	OLS	RF	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Sanction	−362***		−357	−331***		−497*
	(73)		(347)	(40)		(194)
Reprimand	951***		225	632***		123
	(86)		(416)	(47)		(234)
Stringency (Sanction)		−181			−255*	
		(180)			(101)	
Stringency (Reprimand)		353			226	
		(298)			(168)	
Office/month FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes	Yes	Yes	Yes
Dependent mean	5,476	5,476	5,476	3,428	3,428	3,428
Observations	51,848	51,848	51,848	51,848	51,848	51,848

Note: Shown are results from the main estimation sample (cases started between February 2016 and December 2017). The dependent variables measure benefit payments in months 1-12 and 1-6 after the current case, respectively. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are reported in parentheses. OLS standard errors are clustered at the claimant level, while RF and 2SLS standard errors are clustered at the caseworker level (N=1,303). *p < 0.05; **p < 0.01, ***p < 0.001.

excluding cases with no penalty registered, including PES office × quarter fixed effects instead of month fixed effects, and clustering standard errors at the level of the PES office instead of the caseworker level.

Cutoff Points. Selecting our baseline sample, we excluded cases assigned to caseworkers who had handled less than 20 infringement cases during our sampling period. We apply other sample selections for our sensitivity analysis, moving the lower limit to 30 and 40 cases, respectively, and introducing an upper limit of 500 cases. For convenience, Panel A of Table 4.B.9 in Appendix 4.B reproduces a selection of our baseline result and reports estimates for the first-stage relationship and IV effects on four outcomes: (i) recidivism, (ii) employment in month 12, (iii) cumulative earnings in months 6-12, and (iv) cumulative UI benefits in months 1-6. In Panel B, we restrict the sample to caseworkers with at least 30 cases. This restriction excludes 16 percent of caseworkers and 5 percent of observations relative to our baseline sample. Our first-stage and 2SLS estimates remain largely unaffected. In Panel C, we increase the lower limit to 40 cases (excluding 31 percent of caseworkers and 12 percent of observations relative to the baseline sample). Again, the estimated effects do not change materially. However, the effect of a sanction on earnings loses its statistical significance. In Panel D, we introduce an upper limit of 500 cases (excluding 1 percent of caseworkers and 8 percent of observations relative to the baseline sample). Once more, we obtain qualitatively similar results. These findings give us confidence that our results are not affected by caseworker caseloads.

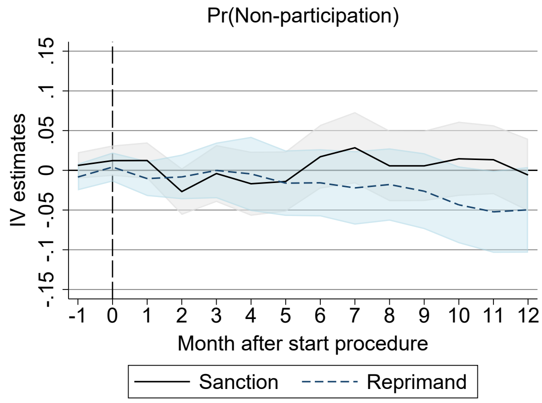


Figure 4.5 Effects of Sanctions and Reprimands on the Probability of Non-Participation Over Time.

Note: Shown are results from the main estimation sample, i.e., cases started between February 2016 and December 2017 (N=51,848). All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level (N=1,303). Gray and colored areas show 95 percent confidence intervals.

Sample Selection and Construction of Instruments. In Table 4.B.10 in Appendix 4.B, we assess whether our results are robust to choosing a different sample and constructing our instruments differently. In Panel B, we use our estimation sample (N=51,848) instead of our baseline sample (N=124,913) to construct our instruments. We obtain qualitatively similar results. In Panel C, we use a split-sample approach to construct our instruments. This sensitivity check follows Bhuller et al. (2020). Following this approach, we randomly split our sample into two equal parts and use one half to construct leave-out mean stringency instruments, following the same procedure as before (see Section 4.5.1). We then use these instruments for estimation in the other half. In terms of absolute size, most of our coefficients remain unchanged. Only for the effect of a sanction on earnings do we observe a smaller effect that is no longer statistically significant. The effect on UI benefits also loses its statistical significance. This change, however, can be ascribed to reduced statistical power, splitting the sample in half.

Other Treatment Assignments and Caseworker Experience. As mentioned in Section 4.5.2, our identification strategy relies on the assumption that caseworker stringency is not related to other treatment choices made by caseworkers or certain caseworker characteristics. Following Arni and Schiprowski (2019), we now provide additional checks of this assumption.

First, we examine whether our results are robust to controlling for other treatments often assigned by PES caseworkers. In Table 4.5 we already showed that our caseworker stringency instruments are not correlated with the probability of following a workshop, receiving a coaching meeting, or obtaining an exemption from job search requirements, respectively. Here, we calculate leave-out means for each of these three treatment choices and include these leave-out means as control variables in our 2SLS specifications. Panel B of Table 4.B.11 in Appendix 4.B presents the results. We find that most results remain the same. We find some larger changes in coefficient size for earnings and UI benefits. Compared to the baseline estimates, coefficients are larger for UI benefits and smaller and no longer statistically significant for earnings.

Second, we assess whether our results change when adding our measure of caseworker experience as a control variable. Remember that we measured caseworker experience as the number of months between the first and last case handled within our baseline sampling period. Panel C of Table 4.B.11 shows that coefficients remain largely the same, except for earnings, where the estimate is smaller and no longer significant. This result alleviates concerns that our stringency instruments reflect caseworker experience.

Additional Sensitivity Checks. Table 4.B.12 in Appendix 4.B reports the result of some additional sensitivity checks. Again, Panel A reproduces our baseline results for comparison. In Panel B, we exclude individual covariates from our 2SLS specifications. We already showed in Table 4.3 that the first-stage estimates hardly change after excluding individual covariates. For 2SLS estimates, this holds for the effects on recidivism, job search, and employment probabilities. For the effects on earnings, coefficients considerably increase in size. For the effects on UI benefits, coefficients also become larger.

Remember that 15 percent of the cases in our estimation sample concerned cases with an invalid reason for noncompliance but no disciplinary measure registered. In Panel C, we exclude these cases from our sample and estimate effects removing the respective dummy variable and instrument from our main 2SLS model. The results remain qualitatively unchanged. Only for the effect of a sanction on earnings do we find a smaller effect that is not statistically significant anymore.

In Panel D, we include interacted PES office $\times$ calendar quarter fixed effects instead of month fixed effects, which leads to similar results throughout. In Panel E, we cluster standard errors at the level of the PES office ($N = 37$) instead of the caseworker level. Most effects remain statistically significant, though at a higher α level.

4.6.6 Limitations

Our study is subject to a few limitations. First, even though our main estimation sample includes more than 50,000 observations, we lack the power to precisely estimate the effects of sanctions and reprimands on some of our outcome variables. As a result, we cannot distinguish potential negative employment effects of receiving a sanction from null effects with enough confidence, for instance. In other cases, we find statistically significant effects, but rather large confidence bands still surround our estimates. This is the case for the effects of imposing a sanction on earnings, for example. Hence, using an IV approach in combination with sample size restrictions limits our ability to estimate effects very precisely and detect small effects. Nonetheless, we can provide compelling evidence against the large effects reported in previous empirical work. For example, van den Berg et al. (2004) find that receiving a sanction increases the transition rate from welfare to work by 140 percent. Similarly, Svarer (2011) reports that exit rates from unemployment increase by more than 100 percent after imposing a first sanction.

Second, we had to limit our observation window to 12 months post-infringement due to sample restrictions and data availability. As a result, we could not study the persistence of effects over a longer time horizon. Notably, some of the previous empirical work finds long-persisting effects of sanctions. For instance, van den Berg and Vikström (2014) show that receiving a sanction decreases wages for up to four years (as far as their data go). Accordingly, we put a longer follow-up period high on the agenda for follow-up research.

Third, even though the default sanction foresees a 25 percent withdrawal for four months, some variation exists regarding sanction height and duration. Naturally, this fact raises the question of whether more severe sanctions have larger effects, as suggested by some earlier studies (Busk, 2016; Svarer, 2011). Unfortunately, data on sanction severity was not accessible, which leaves it up to follow-up research to study the importance of sanction severity.

Fourth, it is important to note that we could only study the effects of actually imposing a sanction or a reprimand (*ex post* effects). Our policy context did not allow for studying *ex ante* effects, i.e., the effects of having in place a sanctioning threat. Moreover, we did not have access to data on self-employment, which may leave us with overestimated or underestimated results for the effects on labor market outcomes. Similarly, we lack access to data on receipt of benefits not administered by the PES, e.g., universal social assistance or disability benefits. Therefore, potential effects on inflow into other benefit schemes may remain unnoticed.

Lastly, we could only observe registered job search activities, which could differ from actual search efforts. Future research may have access to more comprehensive data on search behavior, e.g., online search through the portal of the PES, or surveys, allowing for a more detailed examination of claimants' responses in terms of job search.

4.7 Conclusion

Monitoring and sanctions are common elements of UI and other social protection schemes around the world. For the optimal design of these schemes, it is essential to know how benefit sanctions translate into claimant behavior and economic outcomes. Previous empirical work consistently shows that imposing a sanction increases rates of job finding and exit from unemployment. However, there is limited evidence on how sanctions affect other outcomes such as recidivism, job search, inflow into other benefits, or re-employment quality. Moreover, little is known about the effects of imposing softer disciplinary measures, such as reprimands. Our study aimed to offer contributions in both directions.

Studying the causal effects of sanctions and reprimands comes with the empirical challenge of potential selectivity in the imposition of these measures. In contrast to existing empirical work, which relies almost entirely on a timing-of-events approach to solve potential endogeneity problems, we identified causal effects using an instrumental variable approach. Our study design exploits the quasi-random assignment of infringement cases in the Dutch UI system to caseworkers who systematically differ in their stringency to impose different disciplinary measures (reprimands or sanctions). With our design, we recover the impacts for those claimants at the margin of receiving a certain treatment, i.e., claimants whose treatment depends on whether they are assigned a stricter or more lenient caseworker.

In contrast to earlier empirical work, we find no evidence of effects on employment rates and outcomes measuring benefit dependency. We also find no evidence for switching to other benefits (sickness benefits) or moving out of the labor force altogether. If anything, our results tentatively suggest adverse effects on labor income, indicating that receiving a sanction could lead to roughly 16 percent lower earnings on average six months later. Lower earnings may result from accepting lower-quality jobs, a finding documented by earlier studies (see, e.g., Arni et al., 2013). In light of large confidence bands, our results for earnings should be taken with caution, however. Considering that sanctioned claimants neither find a job more quickly nor more often, the potentially negative effects on earnings remain uncompensated. In conclusion, our findings suggest that sanctions may impose substantial costs on UI benefit claimants, even more so when considering the forgone benefit payments resulting from the sanction itself.

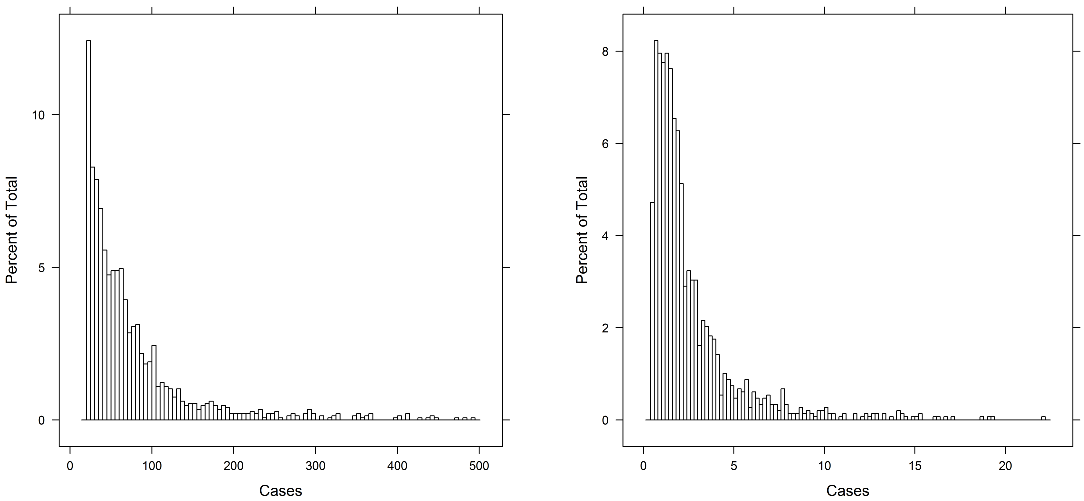
On the other hand, sanctions seem an effective tool to encourage compliant behavior in the future. Imposing a sanction lowers the probability of re-offense in the subsequent 12 months by 40 percent on average. We also find tentative evidence for reporting more job search activities one month after the infringement. Still, better compliance does not seem to result in higher chances of finding employment or exiting benefits. Potentially, imposing a sanction has a limited effect on actual effort or encourages job search activities that are less relevant for success in the labor market.

In the Dutch UI system, caseworkers can decide between imposing a reprimand or a sanction, which allows us to contrast the impacts of a rather severe disciplinary measure with a softer one. We find that imposing a reprimand leaves outcomes largely unchanged. Only for job search activities do we observe a rather large increase in the month after the infringement. Again, reporting more activities does not seem to go in hand with higher job finding. Hence, reprimands seem to change behavior successfully while neither improving nor harming labor market outcomes.

In conclusion, our results suggest that inflicting benefit sanctions comes with a trade-off. Although sanctions constitute an effective means to discipline benefit claimants, they may be detrimental to post-unemployment outcomes and impose considerable costs on those sanctioned. In that respect, the effects of benefit sanctions on health outcomes and personal debt remain understudied and provide important directions for future research. Reprimands appear to constitute a less harmful alternative. Importantly, our conclusions concern the *ex post* effects of both treatments. We cannot speak to the question of whether the existence of a sanctioning threat as such is effective (*ex ante* effects). Although there is initial evidence on the importance of threat effects (Arni et al., 2013; Lalive et al., 2005), more research will be needed to study the interplay and relative importance of *ex post* and *ex ante* effects.

Finally, higher compliance does not lead to higher rates of job finding or exit from unemployment, which raises questions about the effectiveness of a purely quantitative job search criterion (in our context, four job search activities in four weeks). It may be worth considering criteria that incorporate other dimensions of job search (e.g., quality or type of activity) or prioritizing instruments that target self-regulation or other essential skills. Future research will be needed to study the effectiveness of alternative criteria and instruments. Generally, using elaborate measures of job search may allow for a better understanding of how sanctions do or fail to translate into search behavior and successful re-employment.

4.A Additional Figures



(a) Total Cases Handled.

(b) Cases Handled per Month.

Figure 4.A.1 Distribution of Cases per Caseworker.

Note: The two histograms show the distribution of cases handled per caseworker. For the histogram on the right-hand side, we divide the total number of cases by the months worked in the baseline sampling period (January 2014 to December 2017). We crop the distribution at 500 cases in total, which excludes 14 caseworkers (1 percent).

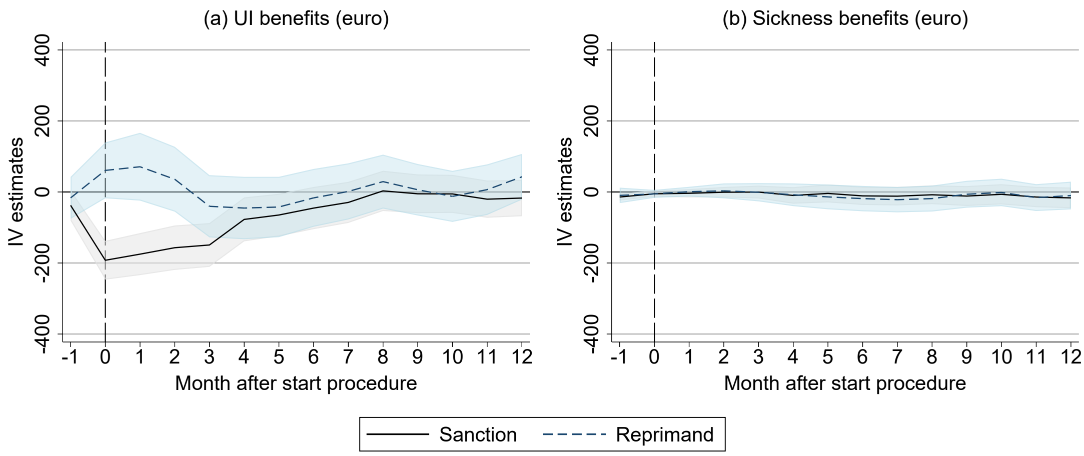


Figure 4.A.2 Effects of Sanctions and Reprimands on Benefit Payments Over Time (Baseline Sample).

Note: Shown are results from the baseline sample, i.e., cases started between February 2014 and December 2017 (N=124,913). All regressions include interacted PES office $\times$ calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level (N=1,488). Gray and colored areas show 95 percent confidence intervals.

4.B Additional Tables

Table 4.B.1 Sample Restrictions and Sample Size.

Restriction	Sample sizes (after restriction)		
	No. of cases/individuals (1)	No. of caseworkers (2)	No. of PES offices (3)
First-time infringement cases started between January 2014 and December 2017.	151,310	2,885	37
Exclude cases without information on PES office.	150,673	2,868	37
Exclude cases without information on start and end dates of benefit spell.	141,583	2,680	37
Exclude cases which took more than 31 days to decide.	138,911	2,678	37
Exclude cases of claimants younger than 18 or older than 64.	134,190	2,414	37
Exclude cases of claimants previously employed in the sector Public Administration and Defense.	132,886	2,394	37
Exclude cases of caseworkers who have handled less than 20 cases.	124,913	1,488	37
Baseline sample.	**124,913**	**1,488**	**37**
Exclude cases started before February 2016.	51,848	1,303	37
Main estimation sample.	**51,848**	**1,303**	**37**

Note: Table inspired by Bhuller et al. (2020).

Table 4.B.2 First-Stage Estimates for Eq.(4.5.4).

	Pr(None)	
	(1)	(2)
Stringency (Sanction)	−0.001	−0.002
	(0.022)	(0.022)
Stringency (Reprimand)	0.113***	0.115***
	(0.022)	(0.022)
Stringency (None)	0.463***	0.464***
	(0.053)	(0.053)
Office/month FE	Yes	Yes
Individual covariates	No	Yes
Dependent mean	0.152	0.152
F-statistic	225	231
R2 (adj.)	0.076	0.078
Observations	51,848	51,848

Note: Estimates from OLS regressions. Dependent variable is a dummy indicating actual treatment (no penalty despite invalid reason). Explanatory variables are caseworker stringency instruments. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. Standard errors are clustered at the caseworker level (N=1,303) and reported in parentheses. *p $<$ 0.05; **p $<$ 0.01, ***p $<$ 0.001.

Table 4.B.3 Relation of Actual Treatment and Stringency Instrument With Claimant Background Characteristics for the Third Treatment Category.

	Pr(None)	Stringency (None)
	(1)	(2)
Female	0.0117**	0.0009
	(0.0036)	(0.0008)
Age	−0.0006**	0.0000
	(0.0002)	(0.0001)
Intermed. education	−0.0066	0.0007
	(0.0042)	(0.0010)
Higher education	0.0020	−0.0020
	(0.0053)	(0.0014)
Education unknown	0.0078	−0.0004
	(0.0045)	(0.0014)
Health	0.0012	0.0014
	(0.0063)	(0.0017)
Retail	−0.0063	0.0001
	(0.0059)	(0.0016)
Services	0.0005	0.0020
	(0.0058)	(0.0016)
Transport	0.0116	0.0012
	(0.0071)	(0.0017)
Temp. employment	0.0061	−0.0006
	(0.0064)	(0.0015)
Other sector	0.0033	0.0006
	(0.0057)	(0.0015)
Monthly wage last job (x1.000)	−0.0049**	0.0005
	(0.0017)	(0.0005)
UI months left (x10)	−0.0009	−0.0005
	(0.0026)	(0.0008)
Employed (t-1)	−0.0080*	−0.0015
	(0.0036)	(0.0009)
UI benefits (t-1)	−0.0113*	0.0026*
	(0.0057)	(0.0013)
Dependent mean	0.1524	0.1524
R-squared	0.0653	0.4514
Observations	51,848	51,848

Note: Estimates from OLS regressions. Dependent variables are a dummy indicating actual treatment (no penalty despite invalid reason) or caseworker stringency instrument. Explanatory variables are individual covariates. All covariates except age, UI months left, and monthly wage are binary variables. The reference category for education level is *lower education*; the reference category for Sector is *manufacturing.* Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. Standard errors are clustered at the caseworker level (N=1,303) and reported in parentheses. All regressions include interacted PES office × calendar month fixed effects. *p < 0.05; **p < 0.01, ***p < 0.001.

Table 4.B.4 Testing the Monotonicity Assumption.

	Baseline instrument	Reverse-sample instrument
	Pr(Sanction)	Pr(Sanction)
	(1)	(2)
Full sample		
Stringency (Sanction)	0.517***	
	(0.021)	
Dependent mean	0.249	
Observations	51,848	
Panel A: Gender		
Subsample: Male claimants		
Stringency (Sanction)	0.499***	0.598***
	(0.027)	(0.025)
Dependent mean	0.219	0.219
Observations	24,618	24,613
Subsample: Female claimants		
Stringency (Sanction)	0.526***	0.637***
	(0.023)	(0.022)
Dependent mean	0.277	0.277
Observations	27,230	27,228
Panel B: Age		
Subsample: Age ≤ 40 years		
Stringency (Sanction)	0.532***	0.645***
	(0.024)	(0.023)
Dependent mean	0.261	0.261
Observations	26,799	26,799
Subsample: Age >40 years		
Stringency (Sanction)	0.485***	0.586***
	(0.027)	(0.025)
Dependent mean	0.237	0.237
Observations	25,049	25,042
Panel C: Education level		
Subsample: Lower education		
Stringency (Sanction)	0.523***	0.641***
	(0.031)	(0.030)
Dependent mean	0.305	0.305
Observations	12,596	12,589
Subsample: Intermediate education		
Stringency (Sanction)	0.545***	0.651***
	(0.026)	(0.025)
Dependent mean	0.281	0.281
Observations	19,670	19,670
Subsample: Higher education		
Stringency (Sanction)	0.484***	0.578***
	(0.043)	(0.042)
Dependent mean	0.195	0.195
Observations	7,285	7,285
Panel D: Sector		
Subsample: Manufacturing		
Stringency (Sanction)	0.544***	0.659***
	(0.042)	(0.041)
Dependent mean	0.270	0.270
Observations	6,607	6,606

Continued on next page

Table 4.B.4 – continued from previous page

	Baseline instrument	Reverse-sample instrument
	Pr(Sanction)	Pr(Sanction)
	(1)	(2)
Subsample: Health		
Stringency (Sanction)	0.506***	0.603***
	(0.041)	(0.040)
Dependent mean	0.230	0.230
Observations	7,096	7,094
Subsample: Retail		
Stringency (Sanction)	0.534***	0.634***
	(0.037)	(0.037)
Dependent mean	0.243	0.243
Observations	8,290	8,290
Subsample: Services		
Stringency (Sanction)	0.458***	0.563***
	(0.036)	(0.037)
Dependent mean	0.214	0.214
Observations	9,427	9,425
Subsample: Transport		
Stringency (Sanction)	0.503***	0.619***
	(0.058)	(0.057)
Dependent mean	0.266	0.266
Observations	4,068	4,068
Subsample: Temporary employment		
Stringency (Sanction)	0.541***	0.651***
	(0.043)	(0.042)
Dependent mean	0.291	0.291
Observations	6,234	6,233

Note: Each column provides the results of a test for the monotonicity assumption as suggested by Bhuller et al. (2020). Column (1) reports first-stage estimates for the full sample and 14 subsamples using the baseline caseworker stringency instrument. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. Column (2) reports first-stage estimates for the same 14 subsamples using a reverse-sample instrument. Reverse-sample instruments are calculated as a caseworker's leave-out mean treatment rate across all cases outside the given subsample. All regressions estimate Eq.(4.5.2) and control for interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level (N=1,303). *p < 0.05; **p < 0.01, ***p < 0.001.

Table 4.B.5 Effects of Sanctions and Reprimands on Recidivism Decomposed.

	2SLS			
	Pr(New case)	Pr(Sanction)	Pr(Reprimand)	Pr(Valid reason)
	(1)	(2)	(3)	(4)
Sanction	-0.110^{***}	0.076^{***}	0.031^{**}	-0.158^{***}
	(0.031)	(0.016)	(0.011)	(0.025)
Reprimand	0.006	0.074^{***}	-0.001	-0.066^{*}
	(0.030)	(0.015)	(0.014)	(0.026)
Office/month FE	Yes	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes	Yes
Dependent mean	0.283	0.058	0.033	0.159
Observations	51,848	51,848	51,848	51,848

Note: Shown are results from the main estimation sample (cases started between February 2016 and December 2017). The dependent variable in Column (1) is a dummy indicating occurrence of a new infringement case between two and 12 months after the current case. The dependent variables in Column (2)–(4) are dummies indicating the outcome of the new case. They indicate the outcome of the first follow-up case in the case of several infringements. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level (N=1,303) and reported in parentheses. *p $<$ 0.05; **p $<$ 0.01, ***p $<$ 0.001.

Table 4.B.6 Effects of Sanctions and Reprimands on Recidivism (Baseline Sample).

	Pr(New case)		
	OLS	RF	2SLS
	(1)	(2)	(3)
Sanction	0.021^{***}		-0.095^{***}
	(0.004)		(0.022)
Reprimand	0.044^{***}		0.191^{***}
	(0.005)		(0.031)
Stringency (Sanction)		-0.055^{***}	
		(0.012)	
Stringency (Reprimand)		0.147^{***}	
		(0.021)	
Office/month FE	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes
Dependent mean	0.247	0.247	0.247
Observations	124,913	124,913	124,913

Note: Shown are results from the baseline sample (cases started between February 2014 and December 2017). The dependent variable is a dummy indicating occurrence of a new infringement case between two and 12 months after the current case. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are reported in parentheses. OLS standard errors are clustered at the claimant level, while RF and 2SLS standard errors are clustered at the caseworker level (N=1,488). *p $<$ 0.05; **p $<$ 0.01, ***p $<$ 0.001.

Table 4.B.7 Effects of Sanctions and Reprimands on Job Search (Baseline Sample).

	OLS	Reduced form	2SLS
	(1)	(2)	(3)
Sanction	-0.239^{***}		0.181
	(0.023)		(0.117)
Reprimand	0.360^{***}		1.007^{***}
	(0.027)		(0.155)
Stringency (Sanction)		0.091	
		(0.065)	
Stringency (Reprimand)		0.737^{***}	
		(0.108)	
Office/month FE	Yes	Yes	Yes
Individual covariates	Yes	Yes	Yes
Dependent mean	2.918	2.918	2.918
Observations	101, 581	101, 581	101, 581

Note: Shown are results from the baseline sample (cases started between February 2014 and December 2017). The dependent variable is the number of registered job search activities with the PES portal in the month after the month of infringement. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are reported in parentheses. OLS standard errors are clustered at the claimant level, while RF and 2SLS standard errors are clustered at the caseworker level (N=1,488). *p < 0.05; **p < 0.01, ***p < 0.001.

Table 4.B.8 Effects of Sanctions, Reprimands, and No Penalty Despite Invalid Reason.

	First-stage			2SLS				
	Pr(Sanction)	Pr(Reprimand)	Pr(None)	Pr(New case)	Search activities	Pr(Employed in month 12)	Earnings months 6-12	UI benefits months 1-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Stringency (Sanction)	0.517***	0.026*	−0.002					
	(0.021)	(0.012)	(0.022)					
Stringency (Reprimand)	−0.115***	0.890***	0.115***					
	(0.026)	(0.030)	(0.022)					
Stringency (None)	0.028	0.081***	0.464***					
	(0.033)	(0.022)	(0.053)					
Sanction				−0.110***	0.325*	−0.028	−932*	−497*
				(0.031)	(0.137)	(0.030)	(424)	(194)
Reprimand				0.006	1.079***	0.028	43	123
				(0.030)	(0.138)	(0.034)	(518)	(234)
None				−0.012	0.140	0.128*	1,371	513
				(0.059)	(0.270)	(0.061)	(830)	(376)
Dependent mean	0.249	0.181	0.152	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	51,848	42,237	51,833	51,848	51,848

Note: Shown are results from the estimation sample (cases started between February 2016 and December 2017). Column (1)–(3) report first-stage estimates, Column (4)–(8) 2SLS estimates on different dependent variables. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office $\times$ calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level (N=1,303) and reported in parentheses. *$p < 0.05$; **$p < 0.01$, ***$p < 0.001$.

Table 4.B.9 Sensitivity Analysis: Cutoff Points.

	First-stage		2SLS				
	Pr(San.)	Pr(Rep.)	Pr(New case)	Search activities	Pr(Employed in month 12)	Earnings months 6-12	UI benefits months 1-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline							
Stringency (San.)	0.517^{***}	0.026^{*}					
	(0.021)	(0.012)					
Stringency (Rep.)	-0.115^{***}	0.890^{***}					
	(0.026)	(0.030)					
Sanction			-0.110^{***}	0.325^{*}	-0.028	-932^{*}	-497^{*}
			(0.031)	(0.137)	(0.030)	(424)	(194)
Reprimand			0.006	1.079^{***}	0.028	43	123
			(0.030)	(0.138)	(0.034)	(518)	(234)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51,848	51,848
Panel B: ≥30 cases							
Stringency (San.)	0.531^{***}	0.032^{*}					
	(0.023)	(0.013)					
Stringency (Rep.)	-0.110^{***}	0.902^{***}					
	(0.027)	(0.033)					
Sanction			-0.131^{***}	0.357^{*}	-0.025	-911^{*}	-479^{*}
			(0.034)	(0.143)	(0.031)	(437)	(202)
Reprimand			-0.017	1.123^{***}	0.018	-222	193
			(0.032)	(0.145)	(0.035)	(528)	(240)
Dependent mean	0.247	0.186	0.282	2.690	0.668	7,691	3,424
Observations	49,013	49,013	49,013	39,923	48,998	49,013	49,013
Panel C: ≥40 cases							
Stringency (San.)	0.545^{***}	0.040^{**}					
	(0.025)	(0.015)					
Stringency (Rep.)	-0.111^{***}	0.910^{***}					
	(0.029)	(0.035)					
Sanction			-0.141^{***}	0.302^{*}	-0.032	-874	-461^{*}
			(0.038)	(0.147)	(0.034)	(466)	(221)
Reprimand			-0.007	1.114^{***}	0.025	-202	226
			(0.034)	(0.146)	(0.036)	(543)	(248)
Dependent mean	0.242	0.194	0.282	2.711	0.668	7,714	3,444
Observations	45,527	45,527	45,527	37,130	45,513	45,527	45,527
Panel D: <500 cases							
Stringency (San.)	0.502^{***}	0.033^{**}					
	(0.021)	(0.011)					
Stringency (Rep.)	-0.111^{***}	0.871^{***}					
	(0.027)	(0.028)					
Sanction			-0.100^{**}	0.294^{*}	-0.035	$-1,117^{*}$	-466^{*}
			(0.032)	(0.141)	(0.032)	(451)	(204)
Reprimand			0.004	1.116^{***}	0.022	-14	133
			(0.032)	(0.144)	(0.037)	(560)	(248)
Dependent mean	0.264	0.167	0.282	2.654	0.667	7,662	3,398
Observations	47,933	47,933	47,933	39,091	47,920	47,933	47,933

Note: Shown are results from the estimation sample (cases started between February 2016 and December 2017). Column (1)–(2) report first-stage estimates, Column (4)–(7) 2SLS estimates on different dependent variables. Panel A reports baseline results for comparison. Panel B reports results excluding cases of caseworkers with less than 30 cases in total. Panel C shows results excluding cases of caseworkers with less than 40 cases in total, and Panel D only including cases of caseworkers with less than 500 cases in total. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level and reported in parentheses. *p < 0.05; **p < 0.01, ***p < 0.001.

Table 4.B.10 Sensitivity Analysis: Sample Selection and Split-Sample Instrument.

	First-stage		2SLS				
	Pr(San.)	Pr(Rep.)	Pr(New case)	Search activities	Pr(Employed in month 12)	Earnings months 6-12	UI benefits months 1-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline							
Stringency (San.)	0.517***	0.026*					
	(0.021)	(0.012)					
Stringency (Rep.)	−0.115***	0.890***					
	(0.026)	(0.030)					
Sanction			−0.110***	0.325*	−0.028	−932*	−497*
			(0.031)	(0.137)	(0.030)	(424)	(194)
Reprimand			0.006	1.079***	0.028	43	123
			(0.030)	(0.138)	(0.034)	(518)	(234)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51,848	51,848
Panel B: Estimation sample							
Stringency (San.)	0.578***	−0.012					
	(0.018)	(0.009)					
Stringency (Rep.)	−0.103***	0.780***					
	(0.017)	(0.017)					
Sanction			−0.152***	0.295*	−0.038	−870*	−483**
			(0.030)	(0.129)	(0.029)	(417)	(185)
Reprimand			−0.039	0.907***	0.006	−548	522*
			(0.027)	(0.128)	(0.029)	(438)	(213)
Dependent mean	0.230	0.193	0.288	2.692	0.669	7,758	3,419
Observations	46,973	46,973	46,973	38,283	46,959	46,973	46,973
Panel C: Split-sample instrument							
Stringency (San.)	0.406***	−0.008					
	(0.018)	(0.011)					
Stringency (Rep.)	−0.092***	0.625***					
	(0.020)	(0.026)					
Sanction			−0.109**	0.530*	−0.054	−608	−496
			(0.040)	(0.206)	(0.042)	(611)	(292)
Reprimand			−0.050	0.990***	0.031	−417	530
			(0.040)	(0.187)	(0.045)	(704)	(312)
Dependent mean	0.249	0.174	0.281	2.660	0.665	7,591	3,442
Observations	27,116	27,116	27,116	22,053	27,110	27,116	27,116

Note: Shown are results from the estimation sample (cases started between February 2016 and December 2017). Column (1)–(2) report first-stage estimates, Column (4)–(7) 2SLS estimates on different dependent variables. Panel A reports baseline results for comparison. Panel B reports results using the estimation sample for calculating stringency instruments. Panel C presents results using a split-sample instrument, i.e., using stringency instruments constructed in one random half of the sample for estimation in the other half of the sample. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level and reported in parentheses. *p < 0.05; **p < 0.01, ***p < 0.001.

Table 4.B.11 Sensitivity Analysis: Other Treatment Choices and Caseworker Experience.

	First-stage		2SLS				
	Pr(San.)	Pr(Rep.)	Pr(New case)	Search activities	Pr(Employed in month 12)	Earnings months 6-12	UI benefits months 1-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline							
Stringency (San.)	0.517^{***}	0.026^{*}					
	(0.021)	(0.012)					
Stringency (Rep.)	-0.115^{***}	0.890^{***}					
	(0.026)	(0.030)					
Sanction			-0.110^{***}	0.325^{*}	-0.028	-932^{*}	-497^{*}
			(0.031)	(0.137)	(0.030)	(424)	(194)
Reprimand			0.006	1.079^{***}	0.028	43	123
			(0.030)	(0.138)	(0.034)	(518)	(234)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51,848	51,848
Panel B: Controlling for other treatment choices							
Stringency (San.)	0.549^{***}	0.016					
	(0.021)	(0.014)					
Stringency (Rep.)	-0.120^{***}	0.899^{***}					
	(0.025)	(0.031)					
Sanction			-0.103^{**}	0.236	-0.008	-555	-604^{**}
			(0.034)	(0.147)	(0.032)	(429)	(208)
Reprimand			0.013	1.021^{***}	0.028	227	75
			(0.031)	(0.140)	(0.035)	(503)	(226)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51,848	51,848
Panel C: Controlling for caseworker experience							
Stringency (San.)	0.544^{***}	0.010					
	(0.021)	(0.013)					
Stringency (Rep.)	-0.133^{***}	0.900^{***}					
	(0.027)	(0.030)					
Sanction			-0.113^{***}	0.287^{*}	-0.032	-735	-595^{**}
			(0.031)	(0.137)	(0.030)	(419)	(192)
Reprimand			0.008	1.098^{***}	0.030	-64	176
			(0.030)	(0.138)	(0.034)	(519)	(233)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51848	51,848

Note: Shown are results from the estimation sample (cases started between February 2016 and December 2017). Column (1)–(2) report first-stage estimates, Column (4)–(7) 2SLS estimates on different dependent variables. Panel A reports baseline results for comparison. Panel B reports results including caseworkers' tendencies to assign other treatments (workshops, meetings, exemptions from job search requirements) as control variables (all three variables operationalized as respective leave-out mean treatment rate). Panel C shows results including caseworker experience as a control variable. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level and reported in parentheses. *p < 0.05; **p < 0.01, ***p < 0.001.

Table 4.B.12 Additional Sensitivity Analyses.

	First-stage		2SLS				
	Pr(San.)	Pr(Rep.)	Pr(New case)	Search activities	Pr(Employed in month 12)	Earnings months 6-12	UI benefits months 1-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline							
Stringency (San.)	0.517^{***}	0.026^{*}					
	(0.021)	(0.012)					
Stringency (Rep.)	-0.115^{***}	0.890^{***}					
	(0.026)	(0.030)					
Sanction			-0.110^{***}	0.325^{*}	-0.028	-932^{*}	-497^{*}
			(0.031)	(0.137)	(0.030)	(424)	(194)
Reprimand			0.006	1.079^{***}	0.028	43	123
			(0.030)	(0.138)	(0.034)	(518)	(234)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51,848	51,848
Panel B: No individual covariates							
Stringency (San.)	0.525^{***}	0.026^{*}					
	(0.021)	(0.012)					
Stringency (Rep.)	-0.119^{***}	0.905^{***}					
	(0.026)	(0.031)					
Sanction			-0.112^{***}	0.200	-0.032	$-1,758^{**}$	-951^{**}
			(0.032)	(0.153)	(0.033)	(540)	(347)
Reprimand			0.025	1.144^{***}	0.054	990	686^{*}
			(0.031)	(0.156)	(0.038)	(674)	(343)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51,848	51,848
Panel C: Exclude cases with no measure registered							
Stringency (San.)	0.475^{***}	0.035^{***}					
	(0.017)	(0.010)					
Stringency (Rep.)	-0.081^{**}	0.899^{***}					
	(0.025)	(0.029)					
Sanction			-0.089^{***}	0.247^{*}	-0.003	-578	-345^{*}
			(0.027)	(0.120)	(0.026)	(367)	(171)
Reprimand			-0.004	0.792^{***}	0.047	261	155
			(0.024)	(0.107)	(0.027)	(410)	(173)
Dependent mean	0.291	0.216	0.286	2.689	0.668	7,738	3,460
Observations	42,955	42,955	42,955	34,863	42,943	42,955	42,955
Panel D: PES office × calendar quarter fixed effects							
Stringency (San.)	0.529^{***}	0.029^{*}					
	(0.021)	(0.012)					
Stringency (Rep.)	-0.106^{***}	0.893^{***}					
	(0.025)	(0.031)					
Sanction			-0.121^{***}	0.350^{**}	-0.030	$-1,008^{*}$	-464^{*}
			(0.030)	(0.132)	(0.029)	(409)	(188)
Reprimand			0.002	1.101^{***}	0.023	-32	163
			(0.030)	(0.136)	(0.034)	(519)	(234)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51,848	51,848
Panel E: Standard errors clustered at level of PES office (*N*=37)							
Stringency (San.)	0.517^{***}	0.026					
	(0.048)	(0.022)					
Stringency (Rep.)	-0.115^{*}	0.890^{***}					
	(0.044)	(0.063)					
Sanction			-0.110^{**}	0.325	-0.028	-932	-497^{*}
			(0.040)	(0.163)	(0.037)	(474)	(217)
Reprimand			0.006	1.079^{***}	0.028	43	123
			(0.039)	(0.138)	(0.046)	(505)	(261)
Dependent mean	0.249	0.181	0.283	2.683	0.668	7,678	3,428
Observations	51,848	51,848	51,848	42,237	51,833	51848	51,848

Note: Shown are results from the estimation sample (cases started between February 2016 and December 2017). Column (1)–(2) report first-stage estimates, Column (4)–(7) 2SLS estimates on different dependent variables. Panel A reports baseline results for comparison. Panel B reports results excluding individual covariates. Panel C shows results excluding cases with an invalid reason but no measure registered from the sample. Panel D present results controlling for PES office × calendar quarter fixed effects instead of office × month fixed effects. Panel E shows results clustering standard errors at the level of the PES office (N=37) instead of the level of the caseworker. Stringency is measured as a caseworker's respective leave-out mean treatment rate across all cases except the case of claimant i. All regressions include interacted PES office × calendar month fixed effects and individual covariates. Standard errors are clustered at the caseworker level and reported in parentheses. *p < 0.05; **p < 0.01, ***p < 0.001.

4.C Messages Sent by the PES Office

Message 1: Asking for clarification (own translation)

Subject: Give a reason for not providing enough job search activities

You have not reported enough job search activities for the period [Start] to [End]. We would like to know the reason for this. Did you do enough activities but not report them? If so, please indicate the reason and complete your activities in 5 steps:

1. Click on '+ Add application activities.' This will take you to another page.
2. There, click on 'Submit new application activity' and fill in what you have done.
3. Click 'Back to task' to go back to this page.
4. At the bottom of this task, enter your reason in the explanation.
5. Click on 'Submit now.'

Did you not do enough activities? Then let us know the reason in 2 steps:

1. At the bottom of this task, enter your reason in the explanation.
2. Click on 'Submit now.'

With your reaction you can prevent a temporary reduction of your benefits. Did you not give a reason? Then we must assume that you do not have a valid reason. Do we consider your reason not to be valid? In that case, your benefit payments may be temporarily reduced.

Message 2: Update if reason is valid (own translation)

Subject: Valid reason for not performing job search task

Dear Sir/Madam,
We asked you to let us know why you did not report enough job search activities or why you reported them too late. You have provided us with a valid reason in your response. Therefore, there are no consequences for your benefit.
Do you fail to meet an obligation attached to your benefit in the future? In that case, it is important that you contact us as soon as possible. A caseworker will then examine with you what is possible in your situation.
Kind regards,

Message 3: Update if reason is invalid (own translation)

Subject: Follow-up on unfulfilled obligation

Dear Sir/Madam,
You have not fulfilled an obligation that is attached to your benefit. For example, you did not do a task in your *Werkmap* (on time), or you did not come to an appointment. We asked you to tell us what the reason for this was.
We will now assess whether failure to do so has consequences for your benefit. If you have given us a reason, we will use this information in our assessment. You will receive our decision within four weeks by letter.

Kind regards,

4.D Discrepancies With Pre-Registration

Table 4.D.1 List of Discrepancies With the Pre-Registration.

Pre-registration	Discrepancy	Location
Sample cases between 2014–2019.	We had to exclude cases after January 2018 due to a policy change that largely eliminated discretionary room for caseworkers to impose sanctions for first-time infringements.	Section 4.4.2
Exclusion criteria.	We also excluded cases without information on start and end dates of benefit spells and cases which took more than 31 days to decide.	Section 4.4.2
Second sample of observations with digital contact only.	Eventually, the data available did not allow for that sample selection.	Section 4.4.2
Analyze instrument quality.	We included a test for monotonicity.	Section 4.5.2
Outcome variables.	Eventually, we gained access to monthly data instead of data averaged over several months. For this reason, we decided to show outcomes cumulated over different periods instead of monthly averages (if possible) and to also report effect dynamics.	Section 4.6
Outcome: hours worked and wages.	Omitted in present version as analysis did not add new insights but available upon request.	Section 4.6
Additional outcome: temporary contract.	We included this variable to show a complete decomposition of employment effects (permanent contract versus temporary contract).	Section 4.6
Additional outcome: sickness benefits.	We included this variable after a suggestion received at a seminar.	Section 4.6
Additional outcome: non-participation.	We included this variable after a suggestion received at a seminar.	Section 4.6
Effect heterogeneity.	Not included in the study but available upon request.	

CHAPTER 5

The Labor Supply Effects of Generous and Unconditional Cash Support

5.1 Introduction

Governments worldwide rely on financial transfers to eradicate poverty and fight social exclusion. A common concern is the negative work incentives that such programs may entail. Therefore, antipoverty programs traditionally include activity-related criteria for monetary support, such as investments in training or education, or participation in labor market insertion programs. A relatively recent development in antipoverty efforts is unconditional forms of support. Particularly in low- and middle-income countries, an increasing number of programs favoring unconditional transfers has emerged over the past years. Frequently, arguments such as lower program cost and the psychological benefits of self-determined spending go hand in hand with the implementation of such programs. Evaluations have shown these programs to improve health (Pega et al., 2017), education (Baird et al., 2014), and psychological well-being outcomes (Haushofer and Shapiro, 2016), while labor supply effects appear largely absent (Banerjee et al., 2017; Bastagli et al., 2016).

Although the impacts of unconditional transfer programs in low- and middle-income countries are well documented, less is known about the effectiveness of similar schemes in higher-income countries. Several factors make the setting of a higher-income country distinct. Transfers may complement existing, potentially extensive safety nets. Moreover, labor markets and other economic institutions may be structured differently. With this study, we aim to bring forward the literature on unconditional transfer programs by describing their labor supply effects when implemented in the setting of a European welfare state. Our analysis uses data from a field experiment in Barcelona (Spain) trialing a generous and unconditional municipal cash transfer program.

The negative labor supply effects of income support are well described (Moffitt, 2002). Beneficiaries of a transfer program may decide to work less due to the pure income effect. Additionally, to the extent that a program is means-tested, substitution effects may play a role. In that case, potentially high implicit tax rates and fear of losing access to the program may prevent beneficiaries from realizing earnings. In the case of unconditional support, there are no conditions that may counter these effects, such as work requirements, possibly resulting in stronger labor supply responses.

However, some effects may operate in the opposite direction. First, fear of scarring effects or human capital deterioration could prevent beneficiaries from working less or leaving the workforce altogether. Second, a cash transfer without strings attached may allow for human capital or other investments that result in higher wages and thus increase work incentives. Lastly, research indicates that financial stress can impede cognitive capacity and promote poverty-reinforcing behaviors (Mani et al., 2013; Shah et al., 2012). Unconditional cash support may counter such effects and free up cognitive resources that can be invested in income-earning activities.

Our study aims to answer the following main research question: What is the effect of generous and unconditional cash support on adult labor force participation? Additionally, we seek to answer several sub-questions. First, we are interested in the effects on choices between salaried work versus self-employment, full-time versus part-time work, and temporary versus permanent contracts. An unconditional transfer may affect these choices by alleviating credit constraints (self-employment), funding leisure (part-time work), or providing liquidity while searching for a better job (permanent positions). Our second sub-question is: What is the effect on engaging in other activities, such as job search, human capital formation, and social participation. It is interesting to study these outcomes, as an unconditional transfer may allow for different time allocations between work-related and other activities. Our third and last sub-question is: Do different program modalities help maintain work incentives? We focus on two modalities: a social activation component and a generous transfer withdrawal rate.

We use data from a two-year randomized controlled trial (RCT) that tested a new municipal antipoverty program. The program targeted economically vulnerable households in disadvantaged neighborhoods of Barcelona. It consisted of two components: (i) a household-based cash transfer, depending on household income, size, and composition, and (ii) different social activation policies. The cash transfer amounted to roughly €492 ($779 PPP) on average per month, which is nearly half the monthly statutory minimum wage, and was paid to a designated adult household member (henceforth: *main recipient*).[1] Activation policies targeted social entrepreneurship and community involvement.

We apply three comparisons. First, we compare households randomly chosen to become beneficiaries of the program with households assigned to a control group. Second, we compare recipient households randomly assigned to participate in a

[1]We convert euros into purchasing power-adjusted U.S. dollars using the OECD purchasing power parity (PPP) exchange rate (OECD, 2021d).

social activation plan with households that received the cash transfer without further obligations. Third, we compare recipient households assigned to a 100 percent transfer withdrawal rate with households assigned to a 25–35 percent withdrawal rate. Our main data sources are two waves of surveys; conducted at baseline and endline. We complement this data with employment information from administrative data sources.

Our main findings can be summarized in three parts. First, we find that the program had negative labor supply effects overall. Roughly two years after the start of the program, main recipients in households assigned to receive transfers were 20 percent less likely to work than their counterparts in control households. We find confirmation of negative employment effects when pooling outcomes at the household level. Probabilities of job search, social participation, and participation in education activities appear unaffected. Notably, six months after the end of transfers, the negative employment effects persist.

Second, we find tentative evidence that assigning beneficiaries to a social activation plan amplifies negative labor supply effects but may lead to higher rates of social participation and higher chances of following education. Implementing a more generous benefit reduction rate appears to alleviate but not eliminate negative labor supply effects.

Third, studying effect heterogeneity we find indications that negative labor supply effects might be limited to households with care responsibilities. Although labor supply effects are largely absent among households without children, they are sizeable and negative among households with children living at home. This result suggests that reductions in labor supply may be explained by a substitution of labor for care tasks.

Our findings complement the existing literature on the labor supply effects of income-enhancing programs in advanced economies. This literature includes an extensive number of studies—many of them using data from the US—examining the effects of different welfare programs on work effort (see, e.g., Bargain and Doorley, 2011; Moffitt, 1992; Schoeni and Blank, 2000). Closely related is a strand of literature studying the effects of changes in program features such as welfare generosity (Lemieux and Milligan, 2008) or time limits (Grogger, 2001). The effects of providing unconditional cash assistance, however, remain understudied. So far, only a few studies have investigated unconditional programs.

Among these studies are evaluations of dividend programs. Dividend programs differ from common income support schemes by providing cash assistance irrespective of household income, i.e., without a means test. Examples of such programs are the Alaska Permanent Fund Dividend and the Eastern Band of Cherokee Indians Casino Dividend. The former program distributes oil-production revenues in the form of a yearly dividend to every Alaska resident that has lived within the state for a full calendar year. In 2019, the transfer was $1,600. Using a synthetic control group method, Jones and Marinescu (2018) study the labor market effects of the dividend. The authors find no impact on labor force participation but increases in part-time work. The latter program distributes biannual casino dividends among members of the Eastern Band of the Cherokee Nation, on average $4,000–$6,000 a year. Using a differences-in-difference method, Akee et al. (2010) find no effects on adult employment outcomes.

Further work on unconditional programs includes the negative income tax (NIT) experiments conducted at five study sites in the United States and Canada in the 1970s. The experiments would randomly assign households to a monthly guaranteed income without any work requirements but subject to a withdrawal rate, each site testing different combinations of parameters (see Burtless, 1986, for a detailed description of the experiments). The experiments showed moderate declines in employment and hours worked, which are probably exaggerated due to misreporting and selective attrition (Ashenfelter and Plant, 1990; Burtless, 1986; Robins, 1985).

In sum, there is little evidence on the labor supply effects of poverty alleviation efforts that refrain from any activity-related criteria. To the best of our knowledge, the program we study is the first cash transfer in a high-income country that provides subsistence level assistance without strings attached. In addition, exploiting a randomized design and collecting social security data next to self-reported employment information from surveys allows us to circumvent some of the internal validity concerns encountered by earlier studies.

The remainder of this chapter is organized as follows. Section 5.2 describes the program studied. In Section 5.3, we discuss the policy and local context. In Section 5.4, we set out the experimental design and method, while Section 5.5 covers data collection and outcome variables. Section 5.6 discusses experiment integrity and Section 5.7 our empirical strategy. In Section 5.8, we present our results, while Section 5.9 concludes.

5.2 Treatment Program

The treatment subject to our study is a municipal antipoverty program introduced by the City Council of Barcelona. The program, named B-MINCOME, aimed to improve the socio-economic situation of participating households as a means to combat poverty and social exclusion in deprived areas of the city. The program was tested in a randomized controlled trial of nearly two years. We exploit this trial for our analysis (see Section 5.4 for a description of the trial). The B-MINCOME program consisted of two main components: (i) an income support component, called *Municipal Inclusion Support Benefit* (henceforth: *SMI benefit*), and (ii) an activation component. We now describe both components in more detail.

SMI benefit. The SMI benefit involved monthly payments and had to top up household income to the household's imputed subsistence level. Accordingly, the benefit level depended on household income, size, and composition. Appendix 5.A provides a detailed account of how the city council determined the benefit level. Transfers could vary between €100 ($154 PPP) and €1,676 ($2,586 PPP) per month. The maximum level corresponds with twice the 2016 at-risk-of-poverty threshold for single-person households in Catalonia. For comparison, the national monthly minimum wage was €826 ($1,309 PPP) when implementing the program. Although the program targeted households, payments were made to one designated household member (henceforth: *main recipient*) selected by the household. Hence, a maximum of one person per household could register for the program. Other (potential) household members were treated as joint beneficiaries. The benefit level responded to changes in household size, composition, and income. For some households, additional income would reduce the transfer one-on-one (reduction rate of 100 percent). Other households faced a reduction rate of 23–35 percent. We will describe these modalities in more detail in Section 5.4.3.

Activation policies. The B-MINCOME program included two social activation plans, directed at community involvement and social entrepreneurship, respectively.[2] The first plan encouraged participation in the social and community life of a respective neighborhood. Under the second plan, participants were trained to become social entrepreneurs or gained work experience in a social entrepreneurship initiative. A maximum of one person per household could take part in the activities. Participants in the activities could also be household members other than the main recipient.

Participation in the B-MINCOME program was voluntary and subject to application. To be eligible for the program, main recipients and their households had to meet six criteria. First, household members had to be registered as Barcelona residents for at least two years and live in the target area of the trial. Second, households needed to have an open file at the municipal social service office for legal reasons.[3] Third, at least one household member had to be aged between 25 and 60. Fourth, the household members had to share (not divide) household expenses. Fifth, household income at the start of the program had to lie below an eligibility threshold, such that the household would receive a monthly transfer (see Appendix 5.A for more information on that threshold). Lastly, excluding the household's primary residence, household assets could not exceed four times the maximum annual SMI benefit.

5.3 Background

5.3.1 Geographic and Socio-Economic Context

This subsection introduces the local setting in which the B-MINCOME trial took place. We will focus on the socio-economic situation in 2017 (when the trial started) and 2016 (the year before). As mentioned before, the program was implemented in Barcelona. Barcelona, which is the capital of the autonomous community of

[2]B-MINCOME included two additional plans, promoting housing renovations for room rental and offering vocational training, respectively. Due to implementation problems, we exclude households assigned to these two plans from our analysis. Excluding households is unproblematic due to random assignment.

[3]The municipal social service office provides information, assistance, and financial aid to vulnerable citizens. The services offered are diverse and may include, e.g., financial emergency aid, access to soup kitchens, temporary housing, child allowances, or counseling.

Catalonia, is the second-most populated city in Spain and the fifth-most populated urban area in the European Union. In 2017, roughly 1.6m people lived within Barcelona city limits, while the Barcelona urban area counted 5.0m inhabitants (Eurostat, 2021). In economic terms, Catalonia is one of the strongest regions of the country, accounting for one-fifth of total Spanish GDP (National Statistics Institute, 2021); 30 percent of which is generated in Barcelona (Statistical Institute of Catalonia, 2021).

Nonetheless, significant socio-economic disparities exist within the city of Barcelona. Among the most disadvantaged districts are ten neighborhoods located at the North-Eastern city limits known as the *Eix Besòs* (Besòs Axis) area.[4] Figure 5.1 displays a map of Barcelona highlighting these districts. *Eix Besòs*, which comprises roughly 7 percent of Barcelona's total population (114.000 inhabitants), was chosen by the city council as a target area for trialing the B-MINCOME program. The area's structural vulnerability shows in several indicators.[5]

First, *Eix Besòs* has some of the highest unemployment rates in Barcelona. Roughly 7 percent of Barcelona's total working-age population was registered unemployed at the start of the trial. In some of the target neighborhoods, unemployment rates were almost twice as high. Further, all target neighborhoods turn up in the lowest quartile in terms of household disposable income. In 2016, the mean household income per capita in the target area was roughly 50 percent of an average Barcelona household. For illustration, Panel A of Figure 5.2 shows a map on the neighborhood-level with the share of households earning less than €5,000 ($7,925 PPP) per year, circling the ten target neighborhoods in black. Panel B shows a map with the mean annual household income per capita. The maps extend into neighboring communities in the North-East, revealing that the target area also stands out compared to close urban areas outside Barcelona city limits. The vulnerability of inhabitants also shows in education indicators. In most target neighborhoods, roughly 40 percent of the adult population have either no degree or completed no more than primary education—a rate almost twice the city's average. Lastly, the area is characterized by household crowding. While the size of the housing stock in the area is small compared to the rest of the city (up to 65 percent is less than 60m2, compared to 28 percent in all of Barcelona), the fraction of large households is comparatively high (up to 30 percent are four-person households or larger, compared to 14 percent in all of Barcelona).

5.3.2 Existing Income Support Schemes

This subsection will relate the B-MINCOME program to existing income support schemes available in the target area. We describe the situation at the start of the program, end-2017. Due to Spain's decentralized political structure, both the

[4]The ten neighborhoods are Ciutat Meridiana, Vallbona, Torre Baró, Roquetes, Trinitat Nova, Trinitat Vella, Baró de Viver, Bon Pastor, Verneda-La Pau, and Besòs-Maresme.

[5]Data on all indicators come from the Statistical Office of the Municipality of Barcelona. The data can be accessed at `https://ajuntament.barcelona.cat/estadistica/angles/index.htm`.

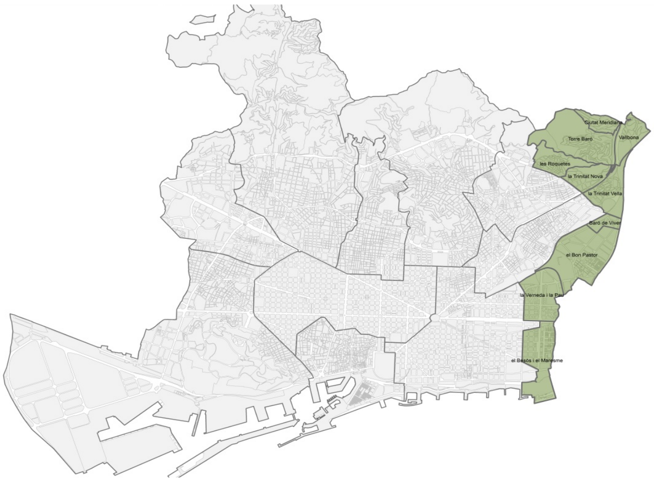


Figure 5.1 Map of Barcelona Highlighting the Target Area.

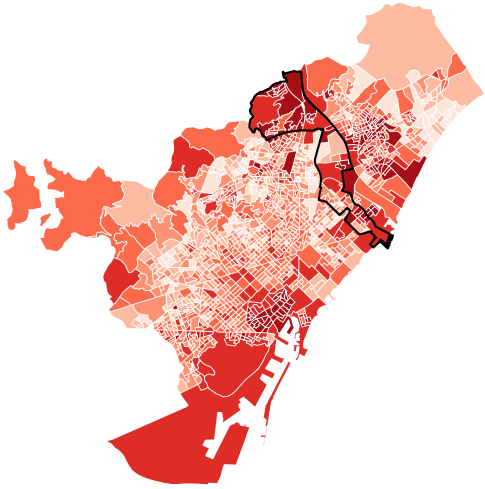
Source: Municipality of Barcelona.

central government and the governments of Spain's autonomous communities have the power to draw up and execute social protection programs. This setup makes for a scattered and complex system with many regional features. While programs targeting poverty and social exclusion are largely region-specific, social insurance schemes, like unemployment insurance (UI) benefits, remain centralized.

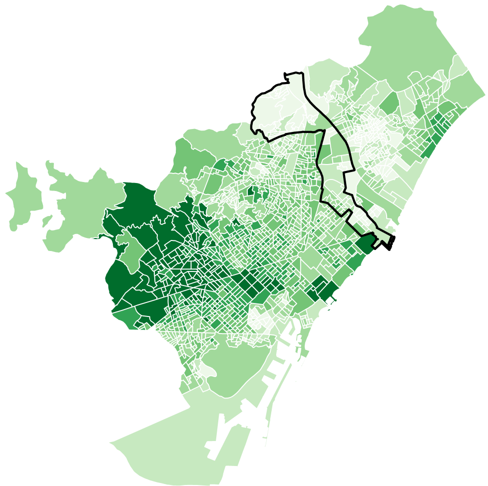
People living in the target area of the trial may have access to four types of monetary transfers. First, they may be claimants of national UI benefits. These benefits could be contribution-based (*prestación contributiva de desempleo*), non-contributory (*prestaciones por desempleo de nivel asistencial*), or specifically target the long-term unemployed. All UI benefits are means-tested, time-limited (between 6–24 months), and either pay for 50–70 percent of reference earnings (contributory benefits), or roughly 60 percent of the national monthly minimum wage (non-contributory and targeted benefits).[6] Second, households may receive the national family allowance (*prestaciones familiares*), which is a non-contributory and means-tested transfer paying €24 ($38 PPP) per month per dependent child below 18 years of age. Third, households may be eligible for Catalonia's guaranteed citizenship income (*renta garantizada de ciudadanía*, or RGC), a region-specific, household-based social assistance benefit. The RGC is unlimited in time, means-tested, and conditional on household members not working.[7] The benefit starts with €564 ($894 PPP) per month for single-person households and may reach €1,062 ($1,683 PPP) per month for households with five or more members. These amounts correspond with roughly 70 and 130 percent of the national monthly minimum wage, respectively. To receive the benefit, claimants have to remain registered with the Public Employment Services of Catalonia and accept suitable job offers. A tranche of €150

[6]In 2017, the national monthly minimum wage was €826 ($1,309 PPP), taking 12 payments per year into account. For comparison, the average monthly wage of a full-time private-sector employee was €2,234 ($3,541 PPP), according to OECD estimates (OECD, 2021a).

[7]Only single parents working part-time are eligible despite having a job.



(a) Share of Households with Annual Earnings below €5,000 in 2016.



(b) Mean Annual Household Income per Capita in 2016.

Figure 5.2 Maps of Barcelona Showing Household Income.

Note: Both maps display neighborhoods of Barcelona and neighboring communities to the North-East (Badalona, Sant Adrià del Besòs, and Santa Coloma de Gramanet). The target area of the trial is circled in black. Breaks of intervals are the 10th, 25th, 50th, 75th and 90th percentile of the distribution of the respective variable.

Source: Own calculations based on data from the National Statistics Institute's experimental statistics (*INE Estadística Experimental*). The data can be accessed at `https://www.ine.es/experimental/experimental.htm`.

($238 PPP) is conditional on complying with a social inclusion or job finding plan. Fourth, families may claim municipal support, including financial emergency aid, municipal child allowance, or municipal housing allowance.

The B-MINCOME program complements the existing social safety net by—in large part—covering families ineligible for most of the national or regional programs described above. Relying on low-wage jobs and small or unstable employment, these families often fail to meet the eligibility requirements of existing schemes, such as extensive formal employment records or complete withdrawal from the labor force. Consequently, many families in the target area remain trapped between state support and self-sufficiency.

For instance, only 10 percent of the households participating in the trial received the regional social assistance benefit (RGC) at the start of the recruitment. Roughly 80 percent received municipal financial support in the 12 months before the trial started. However, existing municipal transfers are often too small to cover basic needs (households received €200 ($317 PPP) per month on average). These two examples illustrate that the B-MINCOME program reaches households previously excluded from income support or at least from programs ensuring a minimum subsistence level.

5.4 Design and Methods

5.4.1 Sampling

Participants for the B-MINCOME trial were recruited among households living in ten target neighborhoods belonging to the *Eix Besòs* area (see Section 5.3.1 for more information). Recruitment for the trial took two months (September–October 2017) and included three steps. First, the Municipality of Barcelona identified 4,305 households that were expected to meet the eligibility criteria of the B-MINCOME program based on information collected from municipal social services records. That is roughly 10 percent of all households living in the target area. Second, the municipality sent letters to the selected households informing them about the program and inviting them to apply. Households were also invited to join one of 400 information events that took place in the target area. Applying households signed an informed consent sheet that gave approval for being followed through surveys and administrative records during and after the trial. Third, the municipality screened the received applications and selected all households that truly fulfilled the eligibility criteria. Of the 4,305 households invited, 2,339 households (54 percent) had applied for the program, of which 1,518 (35 percent) met all criteria.[8] All 1,518 eligible households were enrolled into the trial and approached to fill in a baseline survey. Figure 5.3 shows a study timeline including the different recruitment steps.

5.4.2 Randomization

Randomization took place after the baseline survey. Households were assigned to different experimental conditions through a public lottery. Participants were informed about their assignment via SMS. The lottery followed a stratified randomization design. Two variables were used to form randomization strata: (i) the expected size of the SMI benefit that a household would receive (three categories: high, medium, low), and (ii) a dummy variable indicating the employability of at least one household member (yes, no).[9] In Appendix 5.B we describe the randomization mechanism in detail, while Table 5.E.1 in Appendix 5.E reports the number of households per stratum.

5.4.3 Treatment Arms

Of the 1,518 households included in the lottery, 378 households were assigned to the control group and 822 households to the treatment group. The remaining 318

[8]The high share of eventually ineligible households has to do with the quality and up-to-dateness of information in the municipal social services records.

[9]Low: up to 50 percent of the maximum SMI benefit; Medium: between 50 and 75 percent of the maximum SMI benefit; High: more than 75 percent of the maximum SMI benefit. Employability was included as a stratum due to a treatment arm offering vocational training. As mentioned before, we exclude this arm from our analysis due to implementation issues. A seventh stratum comprised households eligible for household renovations. These households are also excluded from the study and only mentioned for completeness.

Table 5.1 Number of Households per Treatment Arm.

		Activation		No activation	Total
		Social entrepreneurship	Community involvement		
Withdrawal	Full	–	138	197	335
	Partial	100	138	249	487
Total		100	276	446	822

households were assigned to groups outside the scope of this study.[10] Control households did not receive any intervention and were only approached to fill in surveys. Treatment households were randomly allocated to different treatment arms testing program modalities. The program modalities concerned activation policies and benefit withdrawal rates.[11] Table 5.1 cross-tabulates the number of households assigned to each treatment arm. As is shown, treatment arms were cross-randomized, except for the social entrepreneurship arm. In Section 5.8, we will show that excluding this arm from our analyses as part of a robustness check leaves our main results unchanged. Figure 5.3 shows the allocation of households across all experimental conditions. The treatment arms were set up as follows:

Activation versus no activation. All treatment households were randomly assigned to one of four social activation plans or no activation plan. As mentioned before, we only include households assigned to the plans targeting social entrepreneurship and community involvement, next to households assigned no plan.

Full versus partial withdrawal. Remember that increases in income reduced the SMI benefit. All treatment households were randomly assigned to two different withdrawal rates—full or partial withdrawal. Households subject to the full withdrawal rate saw their transfer decrease one-on-one with any additional income (100 percent withdrawal rate). Households assigned to the partial withdrawal arm faced a 25–35 percent withdrawal rate, depending on their extra income. Each additional euro up to €250 ($396 PPP) per month would reduce the benefit by 25 cents, and each euro above €250 per month would reduce the benefit by 35 cents.

5.4.4 Program Implementation

In what follows, we will describe the implementation of the two components of the B-MINCOME program, the SMI benefit, and two activation policies in more detail.

[10]The 318 households comprise 24 households assigned to an activation plan targeting household renovations, 150 households assigned to an activation plan offering vocational training, and 144 households forming a reserve pool. All three groups are excluded from our analysis.

[11]A third modality concerned the obligation to participate in an activation plan (obligatory versus optional). Due to signals that the municipality did not enforce obligatory participation, we disregard these two treatment arms and treat all activation plans as optional.

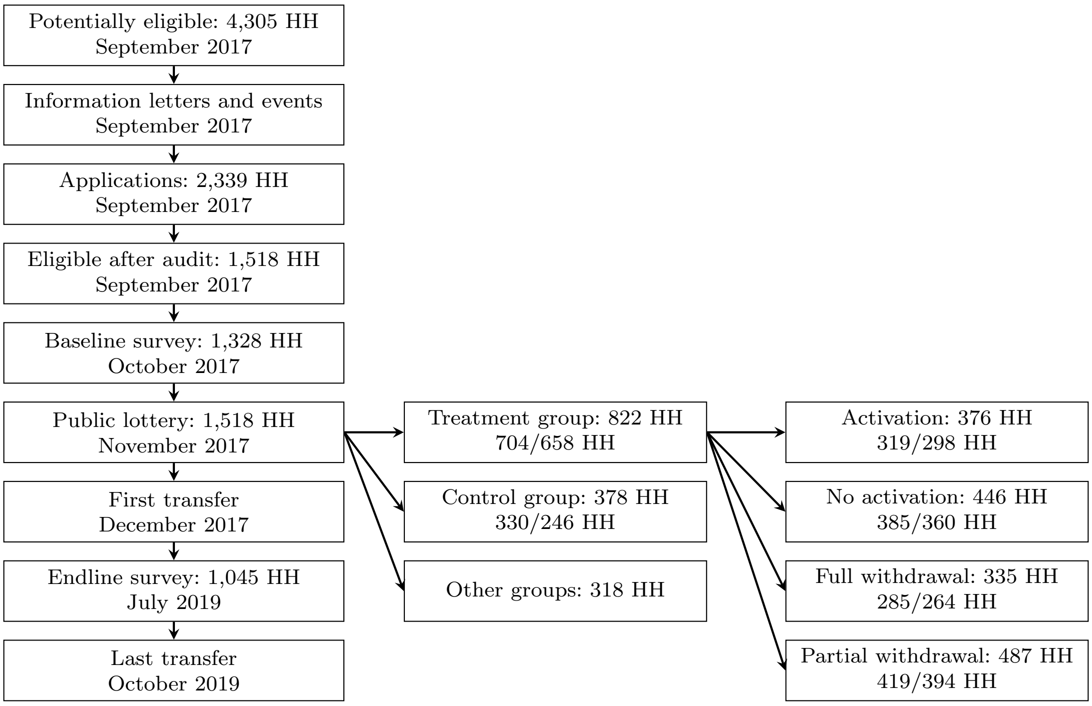


Figure 5.3 Study Timeline and Treatment Arms.

Note: Numbers separated by a slash indicate survey response at baseline and endline, respectively.

SMI Benefit. Treatment households participated in the B-MINCOME program for 23 months in total. The first transfer of the SMI benefit took place in December 2017; the last transfer occurred in October 2019. All payments were transferred to a private bank account of the main recipient. Treatment households were obliged to report changes in household income, size, and composition every quarter to recalculate the benefit level. If applicable, benefit adjustments came into effect with the next payment. Potential overpayment or underpayment in the preceding quarter was settled with payments in the upcoming quarter in equal parts.

On average, the 882 households assigned to treatment received a monthly transfer of €422 ($668 PPP) during the 23-month treatment period. Roughly 14 percent of households received no payments at all, which can be explained by non-take-up (we discuss non-take-up in Section 5.6.3). Conditional on positive transfers, the average monthly transfer was €492 ($779 PPP) (Median = €463; *SD* = €286). Transfers did not exceed €1,500 ($2,376 PPP). Per capita, households received €166 ($263 PPP) per month on average (conditional on positive transfers). In the second year of the trial (from September 2018 onward), 25 percent of the monthly SMI benefit was paid out in a local digital currency, called REC (Real Economy Currency). Participants could use this currency for payment in designated shops and organizations within Barcelona. The REC was at parity with the euro and could be used with a mobile app or a payment card. Figure 5.F.1 in Appendix 5.F plots mean monthly transfers and mean monthly transfers per capita across the 23-

month treatment period. Figure 5.F.2 shows the distribution of average monthly transfers.

Social Activation Plans. The activation plan targeting community involvement consisted of a series of workshops organized by two non-government organizations in each of the ten target neighborhoods. The workshops aimed to promote and facilitate micro-projects of participants that would benefit their neighborhood's community. For example, participants in the workshops were working on developing a neighborhood campaign, collecting community stories, organizing photo and video exhibitions, or developing a neighborhood tour.

The activation plan directed at social entrepreneurship consisted of three phases, an intake phase and two training phases. During the intake phase, households assigned to the plan were invited to interview sessions with program implementers at local social facilities. The goal of the interviews was to provide information about the plan, assess the capabilities of different household members, and select a household member that would take part in the following phases. At the end of the intake phase, groups were formed according to individual profiles and interests. During the first training phase, participants followed two courses of one month, covering basic entrepreneurial skills, such as financial planning. Classes took place three times a week. During the second training phase, participants could choose between two training tracks. In the first track, participants developed a business plan, supported by coaching (200 hours) and further skills training (235 hours). In the second track, participants joined existing local social entrepreneurship initiatives to gain work experience (at least 6 hours per week for 3–6 months).

5.5 Data Collection and Outcomes

5.5.1 Administrative Data Sources

We use administrative data sources to collect information on participant and household characteristics, households' welfare histories, and labor market participation. Data on participant and household characteristics come from the municipal civil registry. We observe main recipients' age and gender, and household size, household composition, and residency (city district) at the time of recruitment. Further, we collect data on household income and municipal transfers received in the 12 months pre-treatment from the municipal benefit registry. Municipal transfers are household-based and include, e.g., schooling, housing, and healthcare allowances, transport subsidies, and child benefits. Lastly, we observe whether households received Catalonia's guaranteed citizenship income (RGC) at the time of recruitment.

Information on labor market participation comes from social security records. These records contain individual-level employment information. We have access to records covering the period June 2019 to April 2020. Hence, this data is only available for the last five treatment months and six months post-treatment. Note that the records only include employed individuals; we do not have access to administrative data on self-employed individuals. The records detail an individual's labor market status (employed versus not in the records) on fixed reference dates separated by windows of usually ten days. Unfortunately, the records do not include any further employment information, such as hours worked, earnings, or the type of contract.

5.5.2 Survey Data and Waves

We complement the information obtained from administrative data sources with data collected through surveys. The surveys included a module on background information with questions on, e.g., socio-demographics and household characteristics. Another module asked questions on time use, including work, job search, social participation, and education and training. These two modules are of interest to this study. Other modules collected information on, e.g., deprivation, health and well-being, and the financial situation. All surveys covered two levels of observation—the household and the individual. Only main recipients filled in surveys. Hence, main recipients provided information on themselves, their household, and other household members. Questions about other household members only concerned factual information, e.g., age or labor market status.

Participants were surveyed three times. The first wave (baseline) took place in the four weeks between enrollment and the public lottery (October 2017). Thus, respondents knew about their participation in the trial but had not yet been assigned to a group. A survey bureau administered the baseline questionnaire through computer-assisted telephone interviewing (CATI). The second wave (midline) took place about one year into the pilot (October 2018). The third and last wave (endline) took place three months before the end of the pilot (July 2019). In contrast to the baseline survey, follow-up rounds included computer-assisted personal interviewing (CAPI) as an alternative survey mode. CAPI was meant to facilitate surveying households with language difficulties and households not answering the phone. We only use data from the baseline and the endline survey in our analysis.

5.5.3 Outcome Variables

We construct fourteen outcome variables based on survey data. Ten of these variables measure outcomes at the level of the main recipient, and four variables pool outcomes at the household level. All variables are based on information reported by the main recipient. For a list and detailed description of all outcome variables, see Table 5.C.2 in Appendix 5.C.

The first group of variables measures labor market outcomes. First, we construct a dummy variable indicating whether the main recipient was *working* (either employed or self-employed) when surveyed. Additionally, we create three sets of dummy variables to decompose treatment effects on the probability of having work. The first two dummy variables indicate whether the main recipient was *employed* or *self-employed*, respectively. The second set of dummies indicates *full-time* or *part-time work*, respectively. As we do not observe hours worked, these two variables will serve as proxies for labor supply decisions at the intensive margin. A third set of dummies indicates work under a *permanent* or a *temporary* contract, respectively. These two dummies serve as proxies for the quality of employment. Two additional variables pool labor market outcomes at the household level. The first variable counts the *number of adult household members* working (employed and self-employed).[12] The second variable is a dummy variable taking the value 1 if *at least one adult household member* is working and 0 otherwise.

The remaining variables measure activities related to job search, social participation, and human capital formation. First, we include a dummy variable indicating whether the main recipient tried to *find paid employment* in the past four weeks. Second, we construct a dummy variable taking the value 1 if the main recipient was active in any *civil society organization or initiative* in the past year.[13] Another dummy variable indicates whether the main recipient took part in any *study or vocational training* in the past year. Two additional variables measure human capital formation at the household level. The first variable counts the *number of adult household members* that took part in any study or vocational training in the past year. The second variable is a dummy indicating whether *at least one adult household member* took part in any study or vocational training.

The self-reported nature of our survey data may raise concerns about data accuracy. Therefore, we construct an additional variable measuring the main recipient's labor market status, this time using administrative data obtained from social security records. We consolidate the 10-day observation intervals into monthly observations. This operation leaves us with eleven dummy variables, one for each month between June 2019 and April 2020. Each dummy takes the value 1 if the main recipient was *employed at least once in the respective month* and 0 otherwise.

5.5.4 Sample Characteristics

Table 5.2 reports descriptive information for the 1,200 households included in our sample. We show information obtained from administrative data sources (Panel A) and the baseline survey (Panel B). As we will explain in more detail in Section 5.6.1, we encounter missing data in the surveys and some administrative records. Therefore, some descriptive statistics do not include the full sample.

As Section 5.3.1 showed, the B-MINCOME program was trialed in disadvantaged urban areas of Barcelona. Our sample descriptive statistics complete this picture, illustrating the economic and social vulnerability of households included in the trial. In sum, we find that participating households are relatively large on average and, for the most part, families with children, while household income and labor market attachment are low. Furthermore, figures on educational attainment suggest low levels of human capital formation on average. We now discuss sample characteristics in more detail.

[12]Adult household members are members between 18 and 65 years of age.

[13]Civil society organizations and initiatives include neighborhood organizations, school organizations or parents' associations, non-profit organizations, religious groups or organizations, political parties, and any voluntary work.

Table 5.2 Sample Descriptive Statistics.

	Mean (1)	SD (2)	Min. (3)	Max. (4)	N (5)
Panel A: Administrative data					
No. of hh members	3.442	1.529	1	11	1,200
No. of hh members 25-65	1.712	0.668	0	6	1,200
No. of children (cond.)	1.753	0.828	1	5	741
Monthly hh income	535.620	416.512	0	1,768	1,200
Monthly transfers	172.948	184.301	0	2,084	1,200
Main recipient female	0.734	0.442	0	1	1,150
Main recipient age	42.967	9.907	9	91	1,192
Panel B: Baseline survey data					
No hh member working	0.397	0.490	0	1	1,034
Single-person hh	0.030	0.171	0	1	1,034
Single-parent hh	0.141	0.348	0	1	1,034
Adults without children	0.139	0.346	0	1	1,034
Adults with children	0.690	0.463	0	1	1,034
Compulsory education or less	0.506	0.500	0	1	1,034
Secondary education	0.395	0.489	0	1	1,034
Tertiary education	0.100	0.300	0	1	1,034
All hh members Spanish	0.472	0.499	0	1	1,034
No hh member Spanish	0.219	0.413	0	1	1,034
Mixed nationalities	0.309	0.463	0	1	1,034
Owner-occupied house	0.249	0.433	0	1	1,029

Note: See Table 5.C.1 in Appendix 5.C for a description of variables. Data on recipient age may be erroneous, which explains the odd minimum and maximum values.

Most households in our sample are adults with children (69 percent). The remaining third comprises single-parent households (13 percent), adults without children (14 percent), and single-person households (3 percent). Households have 3.5 members on average ($SD = 1.5$), which makes them somewhat larger than an average household at risk of poverty in Barcelona (2.5 members).[14] On average, half of the household is between 25 and 65 years old. Households with children have 1.8 children living at home on average (we denote members younger than 16 as children). In terms of housing situation, three quarters of households rents their domicile, while one quarter owns their house (25 percent).

To understand households' economic situation, we look at household income and municipal transfers received. The average monthly net household income in the year before the trial is €536 ($850 PPP; $SD =$ €417).[15] Hence, the average household in our sample lives off an income that equals 30 percent of the 2016 at-risk-of-poverty threshold for households with two adults and two children in Catalonia. Roughly 80 percent of households in our sample claimed municipal transfers at some point in the year before the trial. On average, households received €173 ($274 PPP) in monthly transfers. The most common transfers are food subsidies, safety-net benefits, and family assistance. From the baseline survey, we learn that no member is working in 40 percent of responding households. For comparison, among the

[14]Data on the population at risk of poverty in Barcelona stems from the 2016 EU-SILC survey, which included a proprietary Barcelona sample.

[15]Income data is retrieved from tax income statements covering the period April 2016 until July 2017.

population at risk of poverty in Barcelona, that rate is 17 percent.

The baseline survey also provides information on the highest education level of household members and their nationalities. We pool that information on the household level. In half of all households, no household member has a degree that exceeds compulsory education (primary and lower secondary education). In 40 percent of households, the highest level attained by any member is secondary education (higher secondary education or vocational training). Only in 10 percent of households does at least one member hold a tertiary degree. For comparison, among the population at risk of poverty in Barcelona, 40 percent of households fall into the first category, 26 percent into the second, and 29 percent into the third. Regarding nationalities, in roughly half of all households, all members hold a Spanish nationality. In 22 percent of households, no member is a Spanish citizen. In the remaining fraction, both Spanish and non-Spanish nationalities occur. Lastly, data on main recipients shows that a large majority is female (73 percent). On average, main recipients are 43 years old ($SD = 10$).

5.6 Experiment Integrity

5.6.1 Attrition

Participation in the B-MINCOME trial did not depend on filling in surveys, which introduces a risk of attrition-related bias for outcomes based on survey data. We follow a three-step procedure to diagnose this risk. First, we assess whether survey response is correlated with treatment status. We test for differences in survey response rates between the treatment and the control group using the following specification:

$$response_{ht} = \alpha + \beta_{1t} T_h + \gamma + \epsilon_{ht} \tag{5.6.1}$$

In this equation, the variable $response_{ht}$ is a dummy taking the value 1 if household h was surveyed during survey wave t and 0 otherwise. t may denote the baseline or endline survey. T_h is a treatment dummy indicating the assignment of household h to the treatment group. γ denotes randomization strata fixed effects. ϵ_{ht} is the error term. For comparisons between different treatment arms, we use the following, slightly adapted specification:

$$response_{ht} = \alpha + \beta_{1t} T_h^x + \beta_{2t} C_h + \gamma + \epsilon_{ht} \tag{5.6.2}$$

Here, the dummy variable T_h^x indicates assignment to treatment arm x. x may denote the activation policy arm or the partial withdrawal arm. The dummy variable C_h indicates assignment to the control group. Hence, the reference category is households assigned to the treatment arm without activation policy or full benefit withdrawal, respectively. All other features remain the same as in Eq.(5.6.1). In both specifications, we are interested in the parameters denoted by β_{1t}, which describe differences in response rates between groups or treatment arms of interest at wave t, respectively.

Table 5.3 Attrition: Differences in Survey Response Rates Across Treatment Conditions.

	Control mean (SD) (1)	Treatment group (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Baseline	0.873	-0.006	-0.015	0.010	1,200
	(0.333)	(0.022)	(0.025)	(0.025)	
		[0.789]	[0.551]	[0.704]	
Endline	0.651	0.152	-0.015	0.021	1,200
	(0.477)	(0.029)	(0.028)	(0.029)	
		[0.000]	[0.605]	[0.457]	
Baseline and endline	0.585	0.116	-0.021	0.025	1,200
	(0.493)	(0.031)	(0.032)	(0.033)	
		[0.000]	[0.524]	[0.452]	

Note: Differences in survey response rates between treatment and control groups and treatment arms. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.6.1). Column (3) and (4) report coefficients on the treatment dummy, estimating Eq.(5.6.2). We report robust standard errors in parentheses and p-values in brackets.

Table 5.3 reports differences in response rates at baseline, endline, and endline conditional on baseline. Column (1) shows response rates in the control group. Column (2) reports estimated differences in response rates between the treatment and the control group. Column (3) and (4) compare response rates in the activation versus no activation arm and the partial withdrawal versus full withdrawal arm, respectively.

At baseline, 87 percent of control households filled a survey. The response rate in the treatment group is not significantly different. Neither do baseline response rates differ significantly between treatment arms. Expectedly, response rates are lower at endline. In the control group, 65 percent of households filled the endline survey; 59 percent were surveyed at both baseline and endline. Response rates at follow-up do not differ significantly between treatment arms. However, treatment households have a significantly higher probability of responding at follow-up than control households. The difference in endline response rates between the two groups is an estimated 15 percentage points (23 percent) and slightly lower but still statistically significant when conditioning on baseline response.

Hence, while baseline response shows no relation with treatment status, the response at follow-up correlates with treatment assignment. This finding seems plausible, assuming that receiving transfers increases the attachment to the program. To diagnose attrition in more detail, we perform two additional analyses. The results of these analyses make us confident that, even though response rates at follow-up differ between treatment and control groups, attrition is unlikely to bias our results.

As a first additional analysis, we test for differences in baseline outcomes between attrition and non-attrition households at endline. Regressing our baseline outcome variables on a dummy indicating attrition at follow-up, we find no significant differences between the two groups of households. Table 5.D.1 in Appendix 5.D reports detailed results. Second, we assess whether attrition households in the control group are different from those in the treatment group—again concerning baseline outcomes. We conduct this analysis by regressing our baseline outcome variables on a treatment dummy, restricting our sample to attrition households at endline. Results show no significant differences, except for one outcome variable. Attrition households assigned to the treatment group are 20 percent more likely to have at least one member working at baseline ($p = 0.089$). We report detailed results in Table 5.D.2 in Appendix 5.D. We repeat this analysis comparing households in the activation treatment arms (results shown in Table 5.D.3) and withdrawal treatment arms (results shown in Table 5.D.4). We find no significant differences except for social participation in the activation arms. Attrition households assigned to activation are 70 percent more likely to have their main recipient show civic engagement at baseline ($p = 0.010$).

5.6.2 Baseline Balance

We perform two tests of baseline balance. First, we assess baseline balance in terms of covariates. We include variables measuring household size, household income, and receipt of welfare transfers pre-treatment. Second, we test for differences in survey outcomes at baseline. While the first test builds on administrative data and includes our full sample, the second test restricts the sample to households surveyed at baseline.[16] Both tests aim to assess the integrity of the public lottery mechanism used for randomization.

We compare households assigned to control and treatment groups and households assigned to different treatment arms. For the former comparison, we use the following specification:

$$Y_{hB} = \alpha + \beta_1 T_h + \gamma + \epsilon_h \tag{5.6.3}$$

In that equation, Y_{hB} denotes the variable of interest for household h measured at baseline. T_h is a treatment dummy indicating the assignment of household h to the treatment group. γ denotes randomization strata fixed effects and ϵ_h is the error term. Comparing different treatment arms, we use the following slightly adapted specification:

$$Y_{hB} = \alpha + \beta_1 T_h^x + \beta_2 C_h + \gamma + \epsilon_h \tag{5.6.4}$$

Here, T_h^x is a dummy variable indicating assignment to a treatment arm x. As before, x may denote the activation policy arm or the partial withdrawal arm. C_h is a dummy variable indicating assignment to the control group. Hence, the reference category is households assigned to the treatment arm without activation policy or full benefit withdrawal, respectively. All other terms remain the same as in Eq.(5.6.3). Assessing baseline balance, we are interested in the parameters denoted by β_1, which either describe the differences at baseline between treatment and control households or households assigned to different treatment arms.

Table 5.4 reports the first set of results—baseline balance in terms of covariates. Column (1) shows control group means and standard deviations. Column (2)

[16]Remember that we miss outcome information at baseline for roughly 13 percent of our sample.

Table 5.4 Baseline Balance: Covariates.

	Control mean (SD) (1)	Treatment group (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
No. of hh members	3.463	0.088	-0.004	-0.028	1,200
	(1.596)	(0.094)	(0.101)	(0.101)	
		[0.352]	[0.972]	[0.778]	
No. of children	1.101	0.039	-0.001	-0.005	1,200
	(1.076)	(0.065)	(0.073)	(0.074)	
		[0.551]	[0.992]	[0.948]	
No. of hh members 25-65	1.704	0.042	0.001	-0.030	1,200
	(0.719)	(0.043)	(0.045)	(0.045)	
		[0.332]	[0.990]	[0.502]	
Monthly hh income	424.650	23.598	2.970	15.498	1,200
	(381.618)	(18.041)	(20.480)	(20.616)	
		[0.191]	[0.885]	[0.452]	
Monthly transfers	173.043	4.451	-6.396	22.686	1,200
	(184.509)	(11.433)	(12.584)	(12.822)	
		[0.697]	[0.611]	[0.077]	
RGC recipient	0.056	0.025	0.051	-0.018	1,200
	(0.229)	(0.016)	(0.022)	(0.022)	
		[0.116]	[0.019]	[0.409]	

Note: Differences in covariates between treatment and control groups and treatment arms. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.6.3). Column (3) and (4) report coefficients on the treatment dummy, estimating Eq.(5.6.4). We report robust standard errors in parentheses and p-values in brackets. See Table 5.C.1 in Appendix 5.C for a description of variables.

reports differences between the treatment and the control condition. Column (3) presents differences between households assigned to an activation policy and households that are not; Column (4) shows differences between households assigned to partial withdrawal and full withdrawal.

We find that households assigned to the treatment group do not differ significantly from those assigned to the control group. The same holds for households assigned to different treatment arms, with two exceptions. First, households assigned to an activation policy are 5 percentage points (90 percent) more likely to receive Catalonia's guaranteed citizenship income (RGC) at the time of recruitment ($p = 0.019$). Second, households assigned to the partial withdrawal arm on average received roughly €23 (13 percent) more in monthly transfers pre-treatment. This difference is only significant at the 10 percent level ($p = 0.077$), however. In conclusion, the results of the first balancing test strongly suggest that the public lottery was executed correctly.

Table 5.5 presents estimated differences in survey outcomes at baseline. A few significant differences appear. Main recipients assigned to treatment are 7 percentage points (20 percent) more likely to show civil engagement ($p = 0.027$). They are also more likely to have followed education in the past six months, although this difference is barely significant at the 10 percent level. Main recipients in the activation arm are also more likely of being involved in civil ($p = 0.025$) and educational activities ($p = 0.051$). Moreover, households in the activation arm have more members studying on average ($p = 0.018$) and a higher chance of at least one member following education ($p = 0.014$). For the partial withdrawal arm, the

Table 5.5 Baseline Balance: Survey Outcomes.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Working	0.400	-0.025	-0.024	0.028	1,032
	(0.491)	(0.034)	(0.037)	(0.037)	
		[0.459]	[0.515]	[0.452]	
Job search past 4w	0.506	-0.008	0.056	0.072	1,031
	(0.501)	(0.033)	(0.037)	(0.038)	
		[0.809]	[0.130]	[0.056]	
Social participation	0.345	0.073	0.083	0.019	1,034
	(0.476)	(0.033)	(0.037)	(0.038)	
		[0.027]	[0.025]	[0.616]	
Education in past 12m	0.203	0.047	0.064	0.006	1,034
	(0.403)	(0.028)	(0.033)	(0.033)	
		[0.099]	[0.051]	[0.867]	
No. of members working	0.755	0.017	0.033	0.027	1,034
	(0.770)	(0.054)	(0.060)	(0.060)	
		[0.746]	[0.581]	[0.655]	
At least one member working	0.579	0.005	-0.032	0.012	1,034
	(0.495)	(0.034)	(0.038)	(0.038)	
		[0.894]	[0.400]	[0.752]	
No. of members in education	0.576	-0.012	0.136	-0.036	1,034
	(0.852)	(0.055)	(0.057)	(0.058)	
		[0.821]	[0.018]	[0.535]	
At least one member in education	0.418	-0.009	0.092	-0.019	1,034
	(0.494)	(0.034)	(0.037)	(0.038)	
		[0.797]	[0.014]	[0.606]	

Note: Differences in baseline outcomes between treatment and control groups and treatment arms. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.6.3). Column (3) and (4) report coefficients on the treatment dummy, estimating Eq.(5.6.4). We report robust standard errors in parentheses and p-values in brackets. The sample is restricted to baseline respondents. See Table 5.C.2 in Appendix 5.C for a description of variables.

only significant difference appears for the probability of job search in the past four weeks—main recipients assigned to activation are more likely to have looked for work ($p = 0.056$). In conclusion, the number of imbalances lies a bit higher than to be expected by chance, given the number of hypotheses tested (three comparisons, eight outcomes). Potentially, some selectiveness is introduced by baseline non-response. We will condition on the baseline value of the respective outcome when estimating treatment effects, which should control for any baseline imbalances encountered. As part of a sensitivity analysis, we will also report unadjusted estimates.

5.6.3 Compliance

We determine compliance (or take-up) for the program as a whole and separately for the income support and activation component of the program. Of the 822 households assigned to the treatment group, 717 households (87 percent) actually participated in the program. The remaining 105 households were excluded before the start of the program due to various reasons. Those reasons include refusal, no show, residency outside the target area, and ineligibility due to income or assets. Table 5.E.2 in Appendix 5.E lists the share of households per reason. Table 5.E.3 in Appendix 5.E shows that participation rates are comparable across treatment arms.

For the income support component of the program, we find that all households eligible for receiving the SMI benefit in a respective month actually received the transfer (both in euro and in the local digital currency). At the same time, none of the households assigned to the control group received any payment associated with the B-MINCOME program. In the activation arm, roughly two-thirds of households took up their assigned treatment (conditional on joining the program as a whole). Take-up rates were similar for both activation plans, with 65.5 percent in the community involvement arm and 66.7 percent in the social entrepreneurship arm. None of the households assigned to the control group or the treatment arm without activation participated in an activation plan.

We account for noncompliance by estimating intent-to-treat (ITT) effects. Following this approach, we compare households according to their original group assignment, regardless of actual participation in the program or an activation plan. We elaborate on this strategy in more detail in the following section.

5.7 Empirical Strategy

We are interested in the overall impacts of the B-MINCOME program and the effects of different program modalities, as implemented in the treatment arms. We use different specifications for survey outcomes and administrative outcomes. To assess the overall impact of the program on survey outcomes, we estimate the following specification:

$$Y_{hE} = \alpha + \beta T_h + X_h'\Theta + \Psi Y_{hB} + \Phi M_{hB} + \nu + \gamma + \epsilon_h \tag{5.7.1}$$

In that equation, Y_{hE} describes the outcome of interest for household h at endline. T_h is a dummy indicating the assignment of household h to the treatment group. Hence, our reference category is households assigned to the control group. X_h' is a vector of covariates, which we include to increase the precision of our estimates. The vector contains the variables listed in Table 5.4. As a second means to increase precision, we follow McKenzie (2012), and condition on the baseline value of the respective outcome, denoted by Y_{hB}. As mentioned in Section 5.6.1, we encounter survey non-response at baseline. To avoid losing observations at follow-up due to missing baseline data, we replace missing baseline outcomes with 0 and include a dummy variable, denoted by M_{hB}, indicating missingness at baseline. With this approach, we follow Haushofer and Shapiro (2016). Remember that our survey mode at follow-up included both CATI and CAPI. To control for survey mode, we include ν, a dummy indicating the CAPI method. γ denotes randomization strata fixed effects and ϵ_h is the error term. Our parameter of interest is β, which describes the estimated overall impact of the B-MINCOME program at follow-up.

To estimate the relative effects of different program modalities, we use a slightly modified specification:

$$Y_{hE} = \alpha + \beta_1 T_h^x + \beta_2 C_h + X_h'\Theta + \Psi Y_{hB} + \Phi M_{hB} + \nu + \gamma + \epsilon_h \tag{5.7.2}$$

As in previous specifications, T_h^x is a dummy variable indicating assignment to a treatment arm x, while C_h indicates assignment to the control group. All other features remain the same as in Eq.(5.7.1). As before, x may denote the activation policy arm or the partial withdrawal arm. Our parameter of interest is β_1, which describes the estimated treatment effect of assignment to an activation policy versus no activation, or assignment to a partial versus a full withdrawal rate.

When estimating treatment effects on administrative outcomes, we omit outcomes at baseline and controls for survey mode. Accordingly, Eq.(5.7.1) simplifies to:

$$Y_{ht} = \alpha + \beta_t T_h + X_h' \Theta + \gamma + \epsilon_{ht} \tag{5.7.3}$$

and Eq.(5.7.2) changes to:

$$Y_{ht} = \alpha + \beta_{1t} T_h^x + \beta_{2t} C_h + X_h' \Theta + \gamma + \epsilon_{ht} \tag{5.7.4}$$

Note that our treatment dummies indicate assignment to a particular treatment instead of actual receipt of that treatment in all specifications. This distinction is relevant because of non-participation in the program and the different activation plans. Accordingly, we interpret our effect estimates as intent-to-treat (ITT) effects. More specifically, our results describe the impacts of *implementing* the B-MINCOME program, and not the effects of participation in the program as intended. We consider this limitation unproblematic as non-take up of the program as a whole or certain program features can also be expected under program roll-out. Viewed in this light, the impacts of program implementation are the main parameters of interest.

In addition to average effects, we also study effect heterogeneity. In doing so, we focus on the overall impact of the program and on survey outcomes. We examine effect heterogeneity by interacting the treatment dummy with a dummy indicating a subgroup of interest. Accordingly, Eq.(5.7.1) changes into:

$$\begin{aligned} Y_{hE} = \alpha + & \beta_1 T_h + \beta_2 S_h + \beta_3 T_h S_h + \\ & X_h' \Theta + \Psi Y_{hB} + \Phi M_{hB} + \nu + \gamma + \epsilon_h \end{aligned} \tag{5.7.5}$$

In that equation, S_h denotes a dummy variable indicating a subgroup of interest. All other terms remain the same as in Eq.(5.7.1). The parameter of interest is β_3, which describes the difference in treatment effects between the units inside and outside the respective subgroup.

While our tables in the main section show naive p-values, we report p-values adjusted for investigating multiple outcomes in Appendix 5.E. To adjust p-values, we follow the free step-down methodology of Westfall and Young (1993). Following the recommendation of Westfall and Young (1993), we base our adjustment on 10,000 bootstrap draws.[17]

[17] We adjust our p-values for testing hypotheses on eight outcomes. We exclude six outcome variables, which are meant to decompose effects on labor participation. We calculate adjusted p-values using the user-written Stata package *wyoung* (Jones et al., 2019).

5.8 Results

Remember that we aim to study the overall impact of the B-MINCOME program, in addition to investigating differences between treatment modalities. While Section 5.8.1 addresses the overall impact, Section 5.8.2 and 5.8.3 contrast the effects of activation versus no activation, and partial versus full transfer withdrawal, respectively. We discuss program impacts in four steps. First, we present and decompose labor market effects at the individual level using survey data. Second, we confirm our survey results using administrative data. Third, we discuss labor market effects at the household level. Lastly, we report effects on adjacent outcomes: job search, social participation, and education.

5.8.1 Overall Impact of the Program

Table 5.6 presents estimated treatment effects on survey outcomes at endline, that is, three months before the last transfer. The table only includes outcomes measured at the level of the main recipient. Further below, we discuss results for outcomes pooled at the household level. Column (1) shows control group means and standard deviations. Column (2) reports coefficients on the treatment dummy, estimating Eq.(5.7.1). We report robust standard errors in parentheses and p-values in brackets.

We find that assignment to the program has a significant negative effect on the probability of working at endline. The point estimate is –9.5 percentage points (p = 0.005), which corresponds with a negative effect of 20 percent relative to the control group mean of 47 percent. The effect remains significant at the 10 percent level after correction for multiple inference (see Table 5.E.4 in Appendix 5.E). In Table 5.7 we further decompose this general labor force participation effect. For comparison, the first row again reports the non-decomposed effect.

The results show that negative labor supply effects are confined to paid employment rather than self-employment. Moreover, the results suggest reductions in both full-time and part-time work, though full-time work appears more affected in relative terms. While chances to work full-time are 25 percent lower relative to the control group ($p = 0.045$), the relative effect is 16 percent for part-time work ($p = 0.184$). Lastly, both permanent and temporary work seems to be affected, though the effect is larger for permanent contracts in relative terms. While chances to have a permanent contract are 27 percent lower relative to the control group (p = 0.040), the relative effect is 18 percent for temporary contracts ($p = 0.131$).

To confirm our finding of a negative labor supply effect, we estimate treatment effects on our administrative measure of labor force participation. Panel A of Figure 5.4 plots point estimates and 95 percent confidence intervals for treatment effects in several months. For now, we direct our attention to the estimate in the month of the endline survey (indicated by a black dashed line and labeled accordingly). We find that using administrative data leaves the result unchanged; the point estimate of –9.0 percentage points ($p = 0.004$) is very similar to the coefficient reported in Table 5.6. This result makes us confident that our finding is not distorted by

Table 5.6 Treatment Effects at Endline: Main Recipient.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Working	0.473	-0.095	-0.043	0.031	901
	(0.500)	(0.034)	(0.033)	(0.033)	
		[0.005]	[0.190]	[0.343]	
Job search past 4w	0.024	-0.015	-0.002	-0.002	904
	(0.155)	(0.011)	(0.008)	(0.009)	
		[0.157]	[0.785]	[0.835]	
Social participation	0.378	0.008	0.084	-0.021	904
	(0.486)	(0.035)	(0.037)	(0.037)	
		[0.818]	[0.023]	[0.572]	
Education past 6m	0.212	0.032	0.090	0.030	900
	(0.410)	(0.032)	(0.033)	(0.033)	
		[0.321]	[0.007]	[0.356]	

Note: OLS estimates of treatment effects on survey outcomes at endline (individual level). Outcome variables are listed on the left and described in detail in Table 5.C.2 in Appendix 5.C. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.7.1). Column (3) and (4) present coefficients on the treatment dummy, estimating Eq.(5.7.2). We report robust standard errors in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4.

inaccurate reporting or biased due to survey attrition.

Concordant with the results for main recipients, we find negative labor supply effects when pooling outcomes at the household level. We report these results in Table 5.8. We find that—on average and controlling for household size—treatment households have significantly fewer members working than control households ($p = 0.003$). Likewise, chances of at least one member working are significantly lower among households assigned to treatment ($p = 0.007$). Both effects survive the correction for multiple inference (see Table 5.E.4 in Appendix 5.E).

Lastly, we find no evidence of overall impacts on outcomes measuring other types of activities (see again Table 5.6). It appears that, in general, job search is not a very common activity. Merely 2.4 percent of control respondents report having looked for work in the past four weeks. The point estimate on the treatment dummy is negative and sizable in relative terms (roughly 60 percent lower chances compared to the control group) but not estimated precisely enough. Participation in civil society organizations and following a study or vocational training are more common activities. Among control respondents, 38 percent report civic engagement, while 21 percent indicate having followed education in the past six months. For both outcomes, the point estimate on the treatment effect is positive but relatively small and not statistically significant. When measuring education-related activities at the household level (see again Table 5.8), we find results consistent with findings at the individual level; point estimates are positive but not statistically significant.

5.8.2 Effects of Activation Versus No Activation

We now consider the effects of being assigned to a social activation plan versus receiving the SMI benefit without activation. Column (3) of Table 5.6 reports results for outcomes measured at the level of the main recipient. We find no evidence

Table 5.7 Treatment Effects at Endline: Decomposition of Labor Supply Effects.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Working	0.473	-0.095	-0.043	0.031	901
	(0.500)	(0.034)	(0.033)	(0.033)	
		[0.005]	[0.190]	[0.343]	
Employed	0.457	-0.101	-0.046	0.017	901
	(0.499)	(0.034)	(0.033)	(0.033)	
		[0.003]	[0.159]	[0.615]	
Self-employed	0.016	0.007	0.003	0.015	901
	(0.127)	(0.009)	(0.011)	(0.011)	
		[0.474]	[0.817]	[0.191]	
Working full-time	0.229	-0.058	0.015	0.023	901
	(0.421)	(0.029)	(0.027)	(0.027)	
		[0.043]	[0.590]	[0.390]	
Working part-time	0.245	-0.039	-0.058	0.010	901
	(0.431)	(0.030)	(0.029)	(0.030)	
		[0.197]	[0.045]	[0.743]	
Permanent contract	0.186	-0.050	-0.011	0.023	895
	(0.390)	(0.024)	(0.023)	(0.024)	
		[0.040]	[0.626]	[0.318]	
Temporary contract	0.264	-0.048	-0.031	0.003	895
	(0.442)	(0.032)	(0.030)	(0.030)	
		[0.131]	[0.294]	[0.916]	

Note: OLS estimates of treatment effects on survey outcomes at endline (individual level). Outcome variables are listed on the left and described in detail in Table 5.C.2 in Appendix 5.C. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.7.1). Column (3) and (4) present coefficients on the treatment dummy, estimating Eq.(5.7.2). We report robust standard errors in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4.

that recipients assigned to activation have different chances of working at endline compared to their counterparts receiving nothing but the benefit. The point estimate on the treatment dummy is negative but not statistically significant. For comparison, Panel B of Figure 5.4 shows treatment effects when using our administrative measure of labor force participation. The results are consistent with those obtained using survey data—there are no significant differences in labor supply effects between households assigned to activation versus no activation.

As before, Table 5.7 decomposes the effect on the probability of working. We find a significant negative effect for part-time work ($p = 0.045$). This effect is compensated by a small (and insignificant) increase in the likelihood of working full-time. This finding suggests that activation may indeed harm employment chances more, though negative effects are confined to the domain of part-time work. We find no evidence of effects for other decompositions.

Column (3) of Table 5.8 presents the relative effects of activation for outcomes pooled at the household level. We find that households assigned to the activation arm are less likely to have at least one member working than households receiving the benefit without activation. Finding labor supply effects at the household level rather than at the individual level is consistent with activation policies potentially targeting household members other than the main recipient (see Section 5.2). On a cautionary note, the effect is only statistically significant at the 10 percent level and does not survive the correction for multiple inference. We find no evidence for

Table 5.8 Treatment Effects at Endline: Household.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
No. of members working	0.870	-0.154	-0.066	0.000	904
	(0.823)	(0.051)	(0.052)	(0.052)	
		[0.003]	[0.201]	[0.994]	
At least one member working	0.638	-0.089	-0.060	-0.013	904
	(0.481)	(0.033)	(0.034)	(0.034)	
		[0.007]	[0.075]	[0.709]	
No. of members in education	0.533	0.053	0.075	0.011	904
	(0.806)	(0.054)	(0.054)	(0.054)	
		[0.328]	[0.164]	[0.836]	
At least one member in education	0.394	0.044	0.051	0.004	904
	(0.490)	(0.036)	(0.037)	(0.037)	
		[0.222]	[0.169]	[0.908]	

Note: OLS estimates of treatment effects on survey outcomes at endline (household level). Outcome variables are listed on the left and described in detail in Table 5.C.2 in Appendix 5.C. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.7.1). Column (3) and (4) present coefficients on the treatment dummy, estimating Eq.(5.7.2). We report robust standard errors in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4.

effects on other pooled outcomes.

In contrast to overall treatment effects, we find evidence that recipients assigned to activation are more likely to spend time on social participation and education than their benefit-only counterparts (see again Table 5.6). The effects are statistically significant at the 5 and 1 percent level, respectively. However, only the effect on education is still significant after correction for multiple inference. We interpret these results with caution for two reasons. First, both outcomes already differed significantly at baseline (see again Table 5.5). In Section 5.8.6 we estimate unadjusted effects as part of different sensitivity checks. Results show that coefficients only slightly increase in size when excluding control variables, among which the baseline value of the respective outcome. We believe that this finding provides some reassurance but that the imbalances at baseline warrant caution nonetheless. Second, respondents may have interpreted their participation in an activation plan or components thereof as social participation or education activities. Therefore, results on both outcomes may in part reflect program participation rather than outcomes realized outside the program. Consistent with that reasoning, we find that the point estimate for the effect on education roughly halves when excluding the training-heavy social entrepreneurship arm from the sample (see Table 5.E.5 in Appendix 5.E).

5.8.3 Effects of Partial Withdrawal Versus Full Withdrawal

This subsection presents the effects of being assigned to a partial withdrawal rate versus a full withdrawal rate. Column (4) of Table 5.6 reports results for outcomes measured at the level of the main recipient. We find no evidence for differences in effects between the two treatment arms—neither for employment outcomes nor for outcomes measuring other types of activities. The same holds when pooling

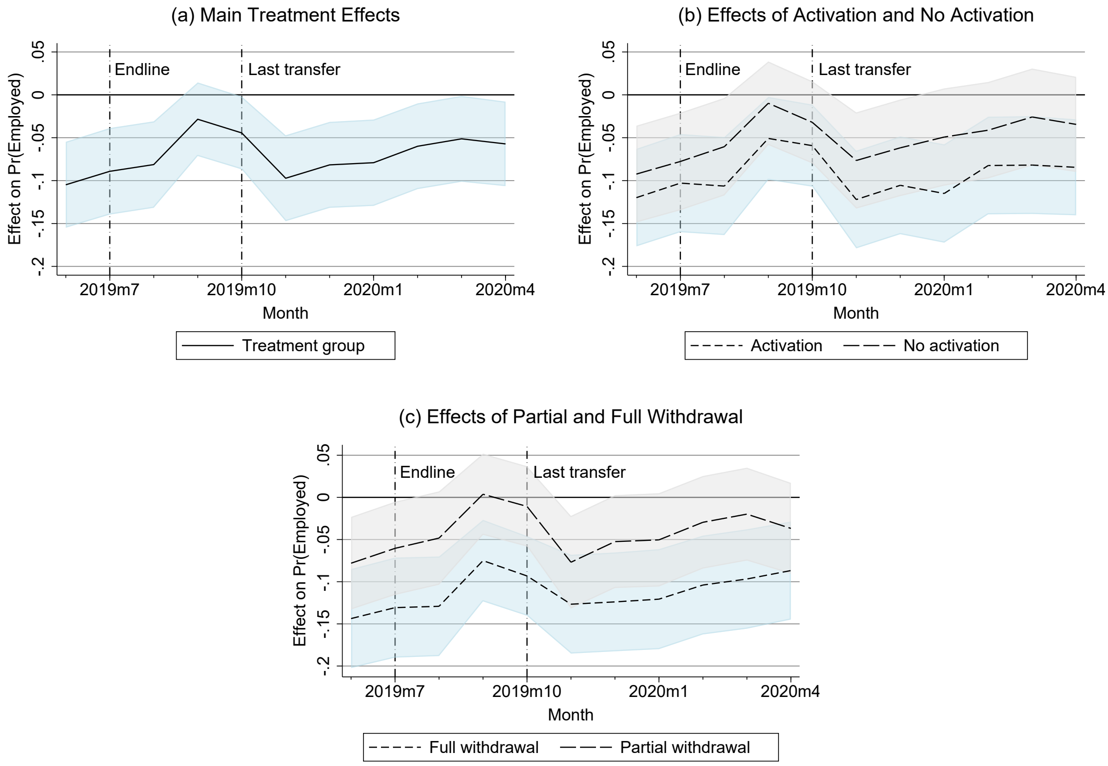


Figure 5.4 Treatment Effects on Employment Probabilities Using Administrative Data.

Note: Grey and colored areas are 95 percent confidence intervals. Graphs show ITT effects, which are estimated using separate regressions for each month. Panel A estimates Eq.(5.7.3), while Panel B and C estimate Eq.(5.7.4).

outcomes at the household level (see Table 5.8). Remember that the social entrepreneurship arm was not cross-randomized and solely faced a partial withdrawal rate. As a robustness check, we exclude this arm from our sample, which leaves results unchanged (see Table 5.E.5 in Appendix 5.E).

We obtain different results when estimating effects using administrative data (see Panel C of Figure 5.4). We find that in the month of the endline survey, recipients under the full withdrawal regime were 6.0 percentage points less likely to be employed than their counterparts under the partial withdrawal regime (p = 0.033). This finding suggests that the full withdrawal regime provided stronger disincentives to work, which fits the predictions provided by standard labor supply models.

It remains an open question why administrative data leads to different results than survey data when studying the effects of withdrawal modalities. Potentially, differential attrition at endline plays a role. As Table 5.D.4 in Appendix 5.D shows, there are no significant differences in baseline outcomes when comparing attrition households in both withdrawal arms. However, point estimates indicate that attrition recipients in the partial withdrawal arm were more likely to work at baseline

($p = 0.300$). Hence, there is a possibility that the results obtained from surveys are downward biased due to higher chances of missing outcomes of working recipients in the partial withdrawal arm. Given that we do not encounter attrition in administrative data sources, the effects estimated on administrative data may prove more reliable when comparing partial and full withdrawal.

5.8.4 Persistence of Effects Post-Treatment

We now assess the persistence of effects post-treatment. On the one hand, we would expect that households try to compensate for their loss in income once the program ends. This behavior may attenuate negative labor supply effects toward the end or after the trial. On the other hand, negative effects may persist, e.g., if being out of work resulted in human capital depreciation or had other scarring effects. As our survey data only reaches as far as July 2019, three months before the end of the trial, we rely on administrative data to examine effect persistence. Using administrative data, we can follow subjects until April 2020, which is six months after the last transfer. Note that Spain imposed a full lockdown due to the unfolding COVID-19 pandemic at the end of March 2020.

Consider again Figure 5.4, which plots monthly effect estimates from June 2019—the month before the endline survey—until April 2020. Panel A shows estimates of the overall treatment effect. We find that negative employment effects briefly diminish toward the end of the trial but quickly revert to previous levels. In the longer term, the effects diminish in size but remain negative throughout. On the one hand, this finding may suggest that the adverse effect of the program on labor supply is persistent. On the other hand, the pattern of returning negative effects after the termination of the program could indicate the presence of compensatory efforts by authorities. For example, social workers may have advertised other support programs among treated households.

Panel B and C of Figure 5.4 plot effect estimates for the different treatment arms. We find that effects for treatment arms follow the same dynamic as overall treatment effects. The difference in effects between the respective arms remains essentially constant over time. This is to say that—also in the longer term—there is no evidence that assignment to activation leads to significantly different effects than receiving no more than the benefit. For the treatment arms testing different withdrawal rates, the difference in effects observed at endline persists until the end of our observation window.

5.8.5 Heterogeneous Treatment Effects

We now study treatment effect heterogeneity to assess whether treatment effects differ for households with and without care responsibilities. For our analysis, we compare effects between households with and without children. Understanding to what extent effects are driven by households with care responsibilities may help to uncover potential treatment mechanisms.

Table 5.9 reports the results of three models. In each model, we interact the

Table 5.9 Heterogeneous Treatment Effects at Endline.

	Main recipient working			At least one member working		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.023 (0.056) [0.677]	-0.012 (0.054) [0.818]	-0.003 (0.052) [0.959]	-0.015 (0.062) [0.808]	-0.032 (0.060) [0.590]	-0.009 (0.057) [0.877]
Interaction terms with treatment dummy						
HH with children (16 years or younger)	-0.108 (0.069) [0.121]			-0.123 (0.074) [0.097]		
HH with children (15 years or younger)		-0.128 (0.068) [0.060]			-0.100 (0.072) [0.164]	
HH with children (14 years or younger)			-0.150 (0.067) [0.025]			-0.143 (0.070) [0.043]
N	901	901	901	904	904	904

Note: OLS estimates of treatment and interaction effects on survey outcomes at endline. Column (1)–(3) report effects on the probability that the main recipient is working. Column (4)–(6) report effects on the probability of any household member working. Outcome variables are described in detail in Table 5.C.2 in Appendix 5.C. The model in Column (1) and (4) includes a term interacting the treatment dummy with a dummy indicating that there are children of 16 years or younger in the household, the model in Column (2) and (5) a term interacting the treatment dummy with a dummy indicating that there are children of 15 years or younger in the household, and Column (3) and (6) a term interacting the treatment dummy with a dummy indicating that there are children of 14 years or younger in the household. We report robust standard errors in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4.

treatment dummy with a dummy indicating children, following Eq.(5.7.5). For the first model, we set the cutoff age for children at 16 years (65 percent of households have children of 16 years or younger). For the second and third model, we lower the cutoff age to 15 and 14 years, respectively (62 and 59 percent of households have children of 15 or 14 years or younger, respectively).[18]

We choose these cutoff points as compulsory secondary education in Catalonia lasts until the age of 16. In contrast to primary schools (for students between 6 and 12 years of age), which offer full-day care, secondary schools usually finish the day at lunchtime. For reasons of brevity, we focus on the overall impact of the program and on two work-related outcome variables: (i) the probability that the main recipient is working and (ii) the probability that any household member is working.

We find tentative evidence that negative labor supply effects are larger among main recipients with care responsibilities. Remember that the estimate for the overall impact of the program on the probability that the main recipient is working was –9.5 percentage points (see Table 5.6). We find much smaller point estimates of –2.3 to –0.3 percentage points for households without children, while the coefficients on the interaction terms are a sizeable –10.8 to –15.0 percentage points. The two

[18] In Table 5.E.6 in Appendix 5.E we further vary the cutoff age to 19, 12, and 5 years or younger, respectively. The results still suggest that households with children decrease their labor supply more, although not all interaction effects are statistically significant.

largest negative coefficients are significant at the 10 and 5 percent level, respectively. This finding suggests that the negative labor supply effects of the program are mainly driven by recipients with children. When pooling work probabilities at the household level, effect heterogeneity shows the same pattern as individual outcomes.

In sum, our findings for effect heterogeneity are consistent with the idea that the negative labor supply effects of the program stem from the substitution of labor for domestic and care work among recipients with children. We have no detailed survey data on time use, which impedes directly investigating this channel. Still, it seems plausible that main recipients with children work less mainly because of the care duties they face. The finding is also consistent with evidence from comparable programs tested in the past. Robins (1985) and Burtless (1986), for instance, report stronger reductions in work effort among single female heads in some of the 1970s U.S. negative income tax experiments.

5.8.6 Sensitivity Analyses

We assess the sensitivity of our results in three steps. First, we estimate effects on survey outcomes excluding most control variables from our main models specified in Eq.(5.7.1) and (5.7.2). The only variables we leave included are randomization strata fixed effects and a dummy variable indicating the survey mode. Table 5.E.7 in Appendix 5.E reports the results. We find that unadjusted effect estimates are somewhat larger but do not differ much from those obtained when including control variables. This results makes us confident that the few imbalances observed at baseline (see Table 5.4 and 5.5) are not concerning.

Second, we estimate effects on survey outcomes, including additional covariates in our models. These covariates measure individual or household background characteristics and are constructed using information from the baseline survey.[19] We include a dummy variable indicating the gender of the main recipient, dummies for the neighborhood in which the household is located (ten neighborhoods), dummies for the type of household (four types), dummies for household composition regarding nationalities (three types), and dummies for the highest education level reached by any household member (three levels). Table 5.C.1 in Appendix 5.C provides a detailed description of all covariates. Table 5.E.8 in Appendix 5.E reports results, which hardly change.

Third, we use logistic regression instead of OLS to estimate effects when the dependent variable is binary. Table 5.E.9 in Appendix 5.E reports the results, which do not change materially.

[19]To account for missingness in these covariates due to baseline non-response, we code missing values as zero and include an additional dummy variable indicating non-response at baseline.

5.9 Discussion and Conclusion

Concerned by potentially negative work incentives, antipoverty programs usually provide monetary support in return for fulfilling certain activity-related criteria, e.g., efforts directed at human capital formation or labor market insertion. This chapter studied the labor supply effects of a poverty alleviation program that does not include any such conditions. The two-year program, drawn up by the City Council of Barcelona, consisted of a monthly cash transfer to households with income below the subsistence level. The benefit level depended on the household income, size, and composition. On average, households received roughly €492 ($779 PPP) per month, which is equivalent to roughly 50 percent of the national monthly minimum wage. Although the benefit was household-based, transfers were made to the account of a designated household member, the main recipient.

We studied the impacts of the program on outcomes related to employment and activities that indicate investment in human capital (following training or education) and the community (social participation). Our analysis uses data from social security records and survey data. For identification, we exploit the fact that the program got trialed in an RCT including roughly 1,500 households recruited in ten target neighborhoods. Our main findings can be summarized in four parts. First, we find strong evidence for negative labor supply effects. After two years, households assigned to the cash transfer were 14 percent less likely to have at least one member working compared to households assigned to the control group; main recipients were 20 percent less likely to work. Second, negative labor supply effects persisted until at least six months after the last payment. Third, we find tentative evidence that negative labor supply effects are mainly driven by households with care responsibilities. Fourth, there is no evidence for effects on social participation and education-related activities.

In addition to studying overall impacts, we contrasted different program modalities implemented in treatment arms. These modalities were: assignment to an activation plan (directed at community involvement or social entrepreneurship) versus pure benefit receipt and a 100 percent transfer withdrawal rate versus a 25–35 percent withdrawal rate. We find suggestive evidence that activation matters. Although some employment-related outcomes worsen under activation, there could be a positive impact on social participation and education-related activities. However, it remains unclear whether this result merely reflects participation in an activation plan. Expectedly, the transfer withdrawal rate shows to matter, too. Labor supply effects are less negative under a more generous withdrawal rate.

In sum, our results suggest that the negative labor supply effects of unconditional transfers should not be underestimated. While the negative effects reported in previous studies are usually neglectable or moderate, our findings suggest sizeable effects on labor force participation. What may explain our results? First, the transfers under the B-MINCOME program were rather generous compared to comparable interventions. Possibly, the income effect was large enough to significantly affect labor supply decisions at the extensive margin. Second, B-MINCOME transfers were subject to a withdrawal scheme, which may amplify substitution effects.

Our findings for treatment arms with different withdrawal rates suggest that such effects indeed played a role. Third, in line with existing evidence, we find stronger labor supply responses among recipients with care responsibilities. In fact, our results suggest that, potentially, effects are almost entirely be driven by this group of participants. If lower labor supply is indeed related to care duties, we may expect improvements in children's outcomes. For instance, children's education outcomes or health could improve. Adolescents may be less likely to commit (minor) crimes. Follow-up research will be needed to examine program effects in such domains and come to conclusions about broader welfare effects. Finally, an important finding concerns the persistence of effects. Employment rates in the treatment group remain on a lower level even six months after the last transfer, which indicates that households' labor supply decisions may be hard to reverse. We can only speculate whether the persistence of effects is related to scarring, the local labor market situation, or other aspects.

Naturally, our study is subject to some limitations, one of which concerns data availability. The social security records that we could access only contain binary information on an individual's employment status. Other important employment-related outcomes such as hours worked, earnings or occupations remain unobserved. Access to such information would allow for a more comprehensive investigation of labor supply effects, including, e.g., decisions at the intensive margin. Equally unobserved remain effects later than six months after the program, which leaves open the question of how long-lasting impacts are. Moreover, we do not dispose of data on household income more broadly. Consequently, we cannot examine to what extent the program affected receipt of other public transfers or total disposable household income. Lastly, lacking more comprehensive data on time use leaves open the question of whether beneficiaries substituted work for other tasks. Another obvious limitation has to do with external validity. The same program may affect households in less disadvantaged areas differently. Likewise, effects in rural areas may differ from effects observed in urban places. Future research will be needed to confirm our findings in such settings.

Still, our results yield some important lessons for policymaking. First, if monetary transfers are generous and unconditional, one must embrace the possibility of reduced labor market participation. Second, activation policies not directed at paid work may reinforce this effect rather than alleviate it. Hence, if one wishes to counteract the negative work incentives of generous and unconditional transfers with activation, such policies may be more successful if targeting self-supporting employment. Supporting evidence comes from Markussen and Røed (2016), who report on the positive employment effects of a Norwegian antipoverty program that pairs generous transfers with tailored rehabilitation, training, and job practice. At the same time, the potential lock-in effects of such programs should not be underestimated either. Third, when designing a transfer scheme, the potentially strong deterring effect of a 100 percent withdrawal rate should be taken seriously. Lower withdrawal rates, which reduce implicit tax rates on earnings in addition to transfers, are policy options worth considering if one wishes to strengthen work incentives. Finally, our finding that households with care responsibilities drive negative labor supply effects hints at an important trade-off. If unconditional transfers allow households to substitute labor for care duties, child outcomes may improve. Reduced labor participation may then occur at the advantage of impeding intergenerational transmission of poverty. Evaluations of other unconditional programs report promising results in that regard (see, e.g., Akee et al., 2010).

Lastly, our findings provide some interesting directions for future research. First of all, the program may have achieved other potential policy objectives. In addition to the aforementioned impact on children, such objectives may include improving health and psychological well-being, alleviating financial hardship, promoting home improvements, or preventing evictions. Local project reports already provide evidence pointing in that direction (Todeschini and Sabes-Figuera, 2019). Evaluating the program in broader terms will allow for a more comprehensive understanding of "positive" and "negative" program effects and potential trade-offs between the different goals. In addition to the above mentioned outcomes, it may be worthwhile studying effects on household composition, marital status, and intra-household bargaining. Second, it will be interesting for future research to examine heterogeneity in effects more comprehensively. For instance, labor supply responses may differ between occupations or baseline income levels. Lastly, more research is needed to understand the community effects of unconditional antipoverty efforts. Cash transfers that can be spent or invested with no strings attached may have distinct effects on the local economy, crime rates, or other neighborhood quality indicators.

5.A Determining the SMI Benefit Level

The SMI benefit level equals the difference between a household's imputed subsistence level and monthly income. We will now describe both items in more detail.

Imputed subsistence level. The sum of a household's imputed living and housing costs. Living costs include costs for energy and water utilities. The fixed values to impute a household's living costs are €402.60 ($638 PPP) per month for the first adult and €148.00 ($235 PPP) for every additional household member. Housing costs comprise rent, mortgage payments, municipal taxes, and property taxes. The fixed values to impute a household's housing costs are €260.00 ($412 PPP) per month for the first adult, €110.00 ($174 PPP) for a second household member, and €40.00 ($63 PPP) for every additional household member. If imputed housings costs exceed actual housing costs, the latter is considered.

Household income. The sum of the incomes of all household members in a given month. This includes income from work, homeownership, financial investments, and economic activities. Household income cannot fall below zero.

For illustration, Table 5.A.1 provides an example calculation for a four-person household consisting of two adults and two children. The example household would receive a monthly transfer of €396.60 ($586 PPP). Table 5.A.1 also shows the eligibility threshold for the household under consideration. To be eligible for the program, the household's income cannot exceed the imputed subsistence level of €1,296.60 ($2,055 PPP).

Table 5.A.1 Example Calculation for SMI Benefit.

Member	Income	Subsistence level		
		Living costs	Housing costs	
		Imputed	Imputed	Actual
Adult 1	€450.00	€402.60	€260.00	
Adult 2	€450.00	€148.00	€110.00	€650.00 (rent) + €50.00 (taxes)
Child 1	–	€148.00	€40.00	
Child 2	–	€148.00	€40.00	
Sum	€900.00	€846.60	€450.00	€700.00
		€846.60 (living costs) + €450.00 (housing costs; lower value) = €1,296.60 (imputed subsistence level)		
Total SMI	€1,296.60 (imputed subsistence level) – €900.00 (household income) = €396.60 (monthly benefit)			

5.B Randomization Mechanism

Households were assigned to experimental conditions per stratum. The randomization mechanism was modeled after a lottery that assigns places in the city's public nurseries. The mechanism works as follows:

1. Each household at random receives a unique administration number between 1 and the total number of households in the stratum.
2. From a bag containing ten balls with the numbers 0 to 9, nine balls are taken with replacement to obtain a nine-digit number.
3. Dividing this number by the number of households in the respective stratum, one obtains a quotient and a remainder.
4. Households are sorted consecutively according to their administration number. The sorted list starts with the household whose administration number is the one next to the remainder. For instance, if the remainder is 6, the first position on the list goes to the household with administration number 7, the second position to the household with number 8, etc.
5. Households are assigned to an experimental condition going through the ordered list from top to bottom, allocating the first x number of households to the first condition, the second x number of households to the second condition, etc. Although conditions are assigned in the same order in each stratum, the number of available places in each condition differs between strata. Consequently, assignment probabilities in the different strata are not the same. Table 5.B.1 lists the assignment probabilities per stratum.

Table 5.B.1 Assignment Probabilities per Stratum.

No.	Strata		No activation		Community involvement		Social entrepreneurship	Control group	Other groups
	Expected SMI	Employable	Full	Partial	Full	Partial	Partial		
1	High	Yes	9%	11%	6%	6%	4%	37%	26%
2	High	No	15%	17%	10%	10%	8%	41%	–
3	Medium	Yes	10%	13%	7%	7%	6%	42%	16%
4	Medium	No	14%	16%	10%	10%	8%	43%	–
5	Low	Yes	18%	22%	12%	12%	8%	23%	4%
6	Low	No	17%	23%	13%	13%	8%	26%	–

Note: Percentages do not add up to 100 percent due to rounding. Other groups comprise an activation plan offering vocational training. This experimental condition is excluded from the study. The table omits stratum no. 7 (see Table 5.E.1 in Appendix 5.E), which is excluded from the study, too.

5.C Lists of Variables

Table 5.C.1 List of Covariates With Description.

Variable	Description	Source
Monthly household income	Average monthly household income in the period April 2016 to July 2017.	Municipal benefit registry
Monthly transfers	Average monthly municipal transfers received in the 12 months before the start of treatment. Municipal transfers may include schooling, housing, and healthcare allowances, transport subsidies, and child benefits.	Municipal benefit registry
RGC recipient	1 if household received Catalonia's guaranteed citizenship income (*renta garantizada de ciudadanía*, or RGC) at the time of recruitment and 0 otherwise.	Municipal benefit registry
Main recipient female	1 if main recipient is female and 0 otherwise.	Municipal civil registry
Main recipient age	Age in years.	Municipal civil registry
Single-person hh	1 if household has one adult member and 0 otherwise. Adult members are members of age 16 or older.	Survey
Single-parent hh	1 if household has one adult member living with a child under age 16 and 0 otherwise.	Survey
Adults without children	1 if household has more than one adult member and 0 otherwise.	Survey
Adults with children	1 if household has more than one adult member living with at least one child under age 16 and 0 otherwise.	Survey
Compulsory education or less	1 if no household member completed compulsory education or at least one household member completed compulsory education and 0 otherwise. Compulsory education comprises primary education and lower secondary education.	Survey
Secondary education	1 if at least one household member completed secondary education and 0 otherwise. Secondary education comprises higher secondary education and vocational education.	Survey
Tertiary education	1 if at least one household member completed tertiary education and 0 otherwise. Tertiary education comprises university education.	Survey
All hh members Spanish	1 if all household members are Spanish citizens and 0 otherwise.	Survey
No hh members Spanish	1 if no household member is a Spanish citizen and 0 otherwise.	Survey
Mixed nationalities	1 if at least one household member is a Spanish citizen and 0 otherwise.	Survey
Owner-occupied house	1 if the household lives in owned property and 0 otherwise.	Survey

Table 5.C.2 List of Outcome Variables With Description.

Variable	Description	Source
Working	1 if main recipient indicated to currently work in paid employment or to be self-employed and 0 otherwise.	Survey
Employed	1 if main recipient indicated to currently work in paid employment and 0 otherwise.	Survey
Self-employed	1 if main recipient indicated to currently be self-employed and 0 otherwise.	Survey
Working full-time	1 if main recipient indicated to work full-time (employed or self-employed) and 0 otherwise.	Survey
Working part-time	1 if main recipient indicated to work part-time (employed or self-employed) and 0 otherwise.	Survey
Permanent contract	1 if main recipient indicated to work under an indefinite contract and 0 otherwise.	Survey
Temporary contract	1 if main recipient indicated to work under a fixed-term contract and 0 otherwise.	Survey
Employed (admin.)	1 if main recipient is listed as employed in social security records at least once in a given month and 0 otherwise.	Social security records
Job search past 4w	1 if main recipient answered *yes* to the question: "In the past four weeks, have you tried to find paid employment (including work of any type and even if it was just for a few hours)?", and 0 if main recipient answered *no*.	Survey
Social participation	1 if main recipient indicated to have taken active part in at least one of the following groups, organizations, or initiatives in the past 12 months and 0 otherwise: neighborhood organization, school organization, parents' association, non-profit organization, religious group, political party, any other organization offering volunteer opportunities.	Survey
Education past 6m	1 if main recipient indicated to have followed a study (vocational or tertiary education) or non-school education (e.g., a private course) in the past six months and 0 otherwise.	Survey
No. of members working	Number of household members aged between 18 and 65 in paid employment or self-employed.	Survey
At least one member working	1 if at least one household member aged between 18 and 65 is in paid employment or self-employed and 0 otherwise.	Survey
No. of members in education	Number of household members aged between 18 and 65 that followed a study (vocational or tertiary education) or non-school education (e.g., a private course) in the past six months.	Survey
At least one member in education	1 if at least one household member aged between 18 and 65 has followed a study (vocational or tertiary education) or non-school education (e.g., a private course) in the past six months and 0 otherwise.	Survey

5.D Attrition Analyses

To test for differences in baseline outcomes between attrition and non-attrition households, we estimate the following specification:

$$Y_{hB} = \alpha + \beta_1 attrition_h + \gamma + \epsilon_h \tag{5.D.1}$$

Here, Y_{hB} describes the outcome of interest for household h at baseline. The variable $attrition_h$ is a dummy taking the value 1 if a household was surveyed at baseline, but not at endline, and 0 otherwise. γ denotes randomization strata fixed effects and ϵ_h is the error term. We report the results of this analysis in Table 5.D.1 below. Column (1) shows the means and standard deviations for the group of non-attrition households. Column (2) reports coefficients on the attrition dummy.

We use the following specification to test for differences in baseline outcomes between attrition households assigned to treatment and control groups:

$$Y_{hB} = \alpha + \beta_1 T_h + \gamma + \epsilon_h \tag{5.D.2}$$

In this equation, all features are the same as in Eq.(5.D.1), except for the dummy variable T_h, which indicates assignment to the treatment group. Estimating Eq.(5.D.2), we restrict the sample to households that filled in the baseline survey but not the endline survey. Table 5.D.2 below reports the results of this second attrition analysis. Column (1) shows control group means and standard deviations. Column (2) presents coefficients on the treatment dummy.

Lastly, we test for differences in baseline outcomes between attrition households assigned to different treatment arms using a slightly adapted specification:

$$Y_{hB} = \alpha + \beta_1 T_h^x + \beta_2 C_h + \gamma + \epsilon_h \tag{5.D.3}$$

Here, T_h^x indicates assignment to a treatment arm x. As previously, x may denote the activation policy arm or the partial withdrawal arm. C_h indicates assignment to the control group. All other terms remain unchanged compared to Eq.(5.D.2). Again, we restrict the sample to households that filled in the baseline survey but not the endline survey. Table 5.D.3 below shows the results of this analysis for the activation arms and Table 5.D.4 for the withdrawal arms.

Table 5.D.1 Attrition: Differences Between Attrition and Non-Attrition Households.

	Non-attrition mean (SD) (1)	Attrition (2)	N (3)
Working	0.387	-0.054	1,032
	(0.487)	(0.035)	
		[0.125]	
Job search past 4w	0.504	-0.034	1,031
	(0.500)	(0.036)	
		[0.347]	
Social participation	0.401	-0.001	1,034
	(0.490)	(0.036)	
		[0.969]	
Education in past 12m	0.249	-0.038	1,034
	(0.433)	(0.030)	
		[0.208]	
No. of members working	0.732	-0.016	1,034
	(0.762)	(0.059)	
		[0.779]	
At least one member working	0.568	-0.047	1,034
	(0.496)	(0.036)	
		[0.195]	
No. of members in education	0.586	-0.090	1,034
	(0.811)	(0.055)	
		[0.105]	
At least one member in education	0.424	-0.051	1,034
	(0.494)	(0.035)	
		[0.148]	

Note: Differences in baseline outcomes between attrition and non-attrition households. Attrition households are households that filled in the baseline survey but not the endline survey. Column (1) reports means and standard deviations for non-attrition households. Column (2) shows the coefficient on the attrition dummy, estimating Eq.(5.D.1). We report robust standard errors in parentheses and p-values in brackets. The sample does not comprise 1,200 observations due to baseline non-response. See Table 5.C.2 in Appendix 5.C for a description of variables.

Table 5.D.2 Attrition: Differences Between Attrition Households in Treatment and Control Groups.

	Control mean (SD) (1)	Treatment (2)	N (3)
Working	0.358 (0.482)	-0.027 (0.064) [0.674]	244
Job search past 4w	0.486 (0.502)	-0.026 (0.065) [0.690]	243
Social participation	0.349 (0.479)	0.066 (0.064) [0.306]	244
Education in past 12m	0.202 (0.403)	-0.000 (0.054) [0.998]	244
No. of members working	0.679 (0.815)	0.159 (0.114) [0.167]	244
At least one member working	0.495 (0.502)	0.110 (0.065) [0.089]	244
No. of members in education	0.495 (0.753)	-0.003 (0.098) [0.978]	244
At least one member in education	0.394 (0.491)	-0.041 (0.065) [0.531]	244

Note: Differences in baseline outcomes between attrition households in the treatment and control groups. Attrition households are households that filled in the baseline survey but not the endline survey. Column (1) reports means and standard deviations for attrition households in the control group. Column (2) shows the coefficient on the treatment dummy, estimating Eq.(5.D.2). We report robust standard errors in parentheses and p-values in brackets. See Table 5.C.2 in Appendix 5.C for a description of variables.

Table 5.D.3 Attrition: Differences Between Attrition Households in the Activation and No Activation Arm.

	No activation mean (SD) (1)	Activation (2)	N (3)
Working	0.333	-0.032	244
	(0.475)	(0.082)	
		[0.701]	
Job search past 4w	0.366	0.140	243
	(0.485)	(0.086)	
		[0.105]	
Social participation	0.306	0.223	244
	(0.464)	(0.085)	
		[0.010]	
Education in past 12m	0.153	0.080	244
	(0.362)	(0.070)	
		[0.251]	
No. of members working	0.819	-0.122	244
	(0.845)	(0.137)	
		[0.372]	
At least one member working	0.597	-0.079	244
	(0.494)	(0.083)	
		[0.339]	
No. of members in education	0.458	0.005	244
	(0.768)	(0.131)	
		[0.971]	
At least one member in education	0.319	0.041	244
	(0.470)	(0.083)	
		[0.616]	

Note: Differences in baseline outcomes between attrition households in the activation and no activation treatment arm. Attrition households are households that filled in the baseline survey but not the endline survey. Column (1) reports means and standard deviations for attrition households in the no activation arm. Column (2) shows the coefficient on the treatment dummy, estimating Eq.(5.D.3). We report robust standard errors in parentheses and p-values in brackets. See Table 5.C.2 in Appendix 5.C for a description of variables.

Table 5.D.4 Attrition: Differences Between Attrition Households in the Partial and Full Withdrawal Arm.

	Full withdrawal mean (SD)	Partial withdrawal	N
	(1)	(2)	(3)
Working	0.276	0.085	244
	(0.451)	(0.082)	
		[0.300]	
Job search past 4w	0.448	-0.007	243
	(0.502)	(0.087)	
		[0.937]	
Social participation	0.362	0.089	244
	(0.485)	(0.086)	
		[0.299]	
Education in past 12m	0.241	-0.064	244
	(0.432)	(0.071)	
		[0.365]	
No. of members working	0.810	-0.062	244
	(0.868)	(0.139)	
		[0.657]	
At least one member working	0.569	-0.004	244
	(0.500)	(0.084)	
		[0.960]	
No. of members in education	0.500	-0.039	244
	(0.731)	(0.129)	
		[0.760]	
At least one member in education	0.379	-0.046	244
	(0.489)	(0.083)	
		[0.585]	

Note: Differences in baseline outcomes between attrition households in the partial and full withdrawal treatment arm. Attrition households are households that filled in the baseline survey but not the endline survey. Column (1) reports means and standard deviations for attrition households in the full withdrawal arm. Column (2) shows the coefficient on the treatment dummy, estimating Eq.(5.D.3). We report robust standard errors in parentheses and p-values in brackets. See Table 5.C.2 in Appendix 5.C for a description of variables.

5.E Additional Tables

Table 5.E.1 Number and Share of Households per Randomization Strata.

No.	Strata		Households	
	Expected SMI	Employable	No.	Percent
1	High	Yes	274	18.0
2	High	No	81	5.3
3	Medium	Yes	379	25.0
4	Medium	No	164	10.8
5	Low	Yes	419	27.6
6	Low	No	165	10.9
7	Other		36	2.4
Total			1,518	100.0

Note: Households in stratum no. 7 are excluded from the study and only listed for completeness. The stratum comprises households eligible for a housing renovation program.

Table 5.E.2 Number and Share of Households Excluded From the Program per Reason.

Reason	No. of Households	Share of households (%)
Not eligible due to income or assets	38	36.2
No show	29	27.6
Refusal	22	21.0
Residency outside target area	16	15.2
Total	105	100.0

Table 5.E.3 Participation Rates per Treatment Arm.

		Activation		No activation	Total
		Social entrepreneurship	Community involvement		
Withdrawal	Full	–	92.0%	85.2%	88.1%
	Partial	90.0%	86.2%	85.5%	86.7%
Total		90.0%	89.1%	85.4%	87.2%

Note: Number of households actually participating in the B-MINCOME program in each treatment arm divided by the number of households assigned to each treatment arm.

Table 5.E.4 Treatment Effects at Endline: Adjusted p-values.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Working	0.473	-0.095	-0.043	0.031	901
	(0.500)	(0.034)	(0.033)	(0.033)	
		[0.076]	[0.534]	[0.928]	
Job search past 4w	0.024	-0.015	-0.002	-0.002	904
	(0.155)	(0.011)	(0.008)	(0.009)	
		[0.504]	[0.754]	[1.000]	
Social participation	0.378	0.008	0.084	-0.021	904
	(0.486)	(0.035)	(0.037)	(0.037)	
		[0.832]	[0.158]	[0.984]	
Education past 6m	0.212	0.032	0.090	0.031	900
	(0.410)	(0.032)	(0.033)	(0.033)	
		[0.664]	[0.050]	[0.928]	
No. of members working	0.870	-0.155	-0.066	-0.000	904
	(0.823)	(0.051)	(0.052)	(0.052)	
		[0.048]	[0.534]	[1.000]	
At least one member working	0.638	-0.089	-0.060	-0.013	904
	(0.481)	(0.033)	(0.034)	(0.034)	
		[0.076]	[0.348]	[0.996]	
No. of members in education	0.533	0.053	0.076	0.008	904
	(0.806)	(0.054)	(0.054)	(0.054)	
		[0.664]	[0.534]	[1.000]	
At least one member in education	0.394	0.044	0.051	0.003	904
	(0.490)	(0.036)	(0.037)	(0.037)	
		[0.544]	[0.534]	[1.000]	

Note: OLS estimates of treatment effects on survey outcomes at endline (individual and household level). Outcome variables are listed on the left and described in detail in Table 5.C.2 in Appendix 5.C. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.7.1). Column (3) and (4) present coefficients on the treatment dummy, estimating Eq.(5.7.2). We report robust standard errors in parentheses and adjusted p-values using the Westfall and Young (1993) methodology and 10,000 bootstrap draws in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4.

Table 5.E.5 Treatment Effects at Endline: Excluding the Social Entrepreneurship Arm.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Working	0.473	-0.097	-0.052	0.032	830
	(0.500)	(0.034)	(0.035)	(0.035)	
		[0.005]	[0.137]	[0.351]	
Job search past 4w	0.024	-0.014	0.001	0.000	833
	(0.155)	(0.011)	(0.009)	(0.009)	
		[0.204]	[0.917]	[0.967]	
Social participation	0.378	0.014	0.114	-0.012	833
	(0.486)	(0.036)	(0.040)	(0.039)	
		[0.689]	[0.005]	[0.760]	
Education past 6m	0.212	0.006	0.044	-0.009	829
	(0.410)	(0.032)	(0.035)	(0.033)	
		[0.857]	[0.208]	[0.789]	
No. of members working	0.870	-0.155	-0.085	-0.005	833
	(0.823)	(0.052)	(0.056)	(0.055)	
		[0.003]	[0.133]	[0.924]	
At least one member working	0.638	-0.086	-0.065	-0.010	833
	(0.481)	(0.034)	(0.037)	(0.036)	
		[0.011]	[0.076]	[0.783]	
No. of members in education	0.533	0.035	0.046	-0.020	833
	(0.806)	(0.055)	(0.059)	(0.056)	
		[0.516]	[0.430]	[0.719]	
At least one member in education	0.394	0.029	0.022	-0.022	833
	(0.490)	(0.037)	(0.040)	(0.038)	
		[0.435]	[0.580]	[0.563]	

Note: OLS estimates of treatment effects on survey outcomes at endline (individual and household level) excluding units assigned to the social entrepreneurship treatment arm. Outcome variables are listed on the left and described in detail in Table 5.C.2 in Appendix C. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.7.1). Column (3) and (4) present coefficients on the treatment dummy, estimating Eq.(5.7.2). We report robust standard errors in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4.

Table 5.E.6 Heterogeneous Treatment Effects at Endline (Varying the Age of Children).

	Main recipient working			At least one member working		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.057	-0.040	-0.074	-0.006	-0.026	-0.092
	(0.063)	(0.049)	(0.037)	(0.069)	(0.053)	(0.038)
	[0.371]	[0.416]	[0.048]	[0.934]	[0.622]	[0.016]
Interaction terms with treatment dummy						
HH with children (19 years or younger)	-0.060			-0.127		
	(0.074)			(0.080)		
	[0.421]			[0.110]		
HH with children (12 years or younger)		-0.097			-0.122	
		(0.066)			(0.068)	
		[0.144]			[0.073]	
HH with children (5 years or younger)			-0.111			-0.022
			(0.087)			(0.086)
			[0.204]			[0.802]
N	895	901	895	898	904	898

Note: OLS estimates of treatment and interaction effects on survey outcomes at endline. Column (1)–(3) report effects on the probability that the main recipient is working. Column (4)–(6) report effects on the probability that any household member is working. Outcome variables are described in detail in Table 5.C.2 in Appendix 5.C. The model in Column (1) and (4) includes a term interacting the treatment dummy with a dummy indicating that there are children of 19 years or younger in the household, the model in Column (2) and (5) a term interacting the treatment dummy with a dummy indicating that there are children of 12 years or younger in the household, and Column (3) and (6) a term interacting the treatment dummy with a dummy indicating that there are children of 5 years or younger in the household. We report robust standard errors in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4.

Table 5.E.7 Unadjusted Treatment Effects at Endline.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Working	0.473	-0.130	-0.066	0.044	901
	(0.500)	(0.038)	(0.037)	(0.038)	
		[0.001]	[0.076]	[0.242]	
Employed	0.457	-0.139	-0.068	0.031	901
	(0.499)	(0.038)	(0.037)	(0.037)	
		[0.000]	[0.064]	[0.413]	
Self-employed	0.016	0.008	0.002	0.014	901
	(0.127)	(0.010)	(0.012)	(0.011)	
		[0.394]	[0.870]	[0.229]	
Working full-time	0.229	-0.069	0.006	0.020	901
	(0.421)	(0.031)	(0.029)	(0.029)	
		[0.027]	[0.838]	[0.495]	
Working part-time	0.245	-0.061	-0.072	0.025	901
	(0.431)	(0.032)	(0.031)	(0.031)	
		[0.059]	[0.019]	[0.433]	
Permanent contract	0.186	-0.067	-0.027	0.011	895
	(0.390)	(0.029)	(0.027)	(0.027)	
		[0.022]	[0.321]	[0.686]	
Temporary contract	0.264	-0.069	-0.044	0.019	895
	(0.442)	(0.033)	(0.031)	(0.031)	
		[0.036]	[0.150]	[0.544]	
Job search past 4w	0.024	-0.015	-0.002	-0.001	904
	(0.155)	(0.011)	(0.008)	(0.008)	
		[0.180]	[0.824]	[0.927]	
Social participation	0.378	0.035	0.100	-0.024	904
	(0.486)	(0.037)	(0.038)	(0.039)	
		[0.340]	[0.009]	[0.538]	
Education past 6m	0.212	0.035	0.101	0.028	900
	(0.410)	(0.032)	(0.034)	(0.034)	
		[0.273]	[0.003]	[0.397]	
No. of members working	0.870	-0.183	-0.051	0.028	904
	(0.823)	(0.060)	(0.061)	(0.060)	
		[0.002]	[0.397]	[0.646]	
At least one member working	0.638	-0.115	-0.081	-0.002	904
	(0.481)	(0.037)	(0.039)	(0.040)	
		[0.002]	[0.038]	[0.960]	
No. of members in education	0.533	0.047	0.133	-0.006	904
	(0.806)	(0.060)	(0.060)	(0.061)	
		[0.432]	[0.027]	[0.926]	
At least one member in education	0.394	0.038	0.078	-0.002	904
	(0.490)	(0.038)	(0.039)	(0.039)	
		[0.324]	[0.043]	[0.968]	

Note: OLS estimates of treatment effects on survey outcomes at endline (individual and household level). Outcome variables are listed on the left and described in detail in Table 5.C.2 in Appendix 5.C. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.7.1). Column (3) and (4) present coefficients on the treatment dummy, estimating Eq.(5.7.2). We report robust standard errors in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode.

Table 5.E.8 Treatment Effects at Endline With Additional Controls.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Working	0.473	-0.092	-0.044	0.035	901
	(0.500)	(0.034)	(0.033)	(0.034)	
		[0.008]	[0.176]	[0.297]	
Employed	0.457	-0.097	-0.050	0.023	901
	(0.499)	(0.034)	(0.033)	(0.034)	
		[0.004]	[0.130]	[0.499]	
Self-employed	0.016	0.007	0.005	0.012	901
	(0.127)	(0.009)	(0.011)	(0.011)	
		[0.445]	[0.649]	[0.257]	
Working full-time	0.229	-0.059	0.003	0.019	901
	(0.421)	(0.029)	(0.027)	(0.027)	
		[0.043]	[0.906]	[0.476]	
Working part-time	0.245	-0.035	-0.050	0.017	901
	(0.431)	(0.031)	(0.030)	(0.030)	
		[0.258]	[0.095]	[0.570]	
Permanent contract	0.186	-0.056	-0.011	0.023	895
	(0.390)	(0.025)	(0.024)	(0.024)	
		[0.025]	[0.650]	[0.335]	
Temporary contract	0.264	-0.041	-0.034	0.009	895
	(0.442)	(0.032)	(0.030)	(0.031)	
		[0.198]	[0.264]	[0.781]	
Job search past 4w	0.024	-0.016	0.000	-0.002	904
	(0.155)	(0.011)	(0.009)	(0.008)	
		[0.131]	[1.000]	[0.807]	
Social participation	0.378	0.010	0.073	-0.027	904
	(0.486)	(0.035)	(0.037)	(0.037)	
		[0.771]	[0.052]	[0.467]	
Education past 6m	0.212	0.034	0.091	0.032	900
	(0.410)	(0.032)	(0.034)	(0.033)	
		[0.283]	[0.007]	[0.323]	
No. of members working	0.870	-0.155	-0.080	-0.002	904
	(0.823)	(0.051)	(0.052)	(0.052)	
		[0.003]	[0.122]	[0.976]	
At least one member working	0.638	-0.089	-0.066	-0.009	904
	(0.481)	(0.033)	(0.034)	(0.035)	
		[0.006]	[0.056]	[0.800]	
No. of members in education	0.533	0.058	0.069	0.006	904
	(0.806)	(0.054)	(0.054)	(0.053)	
		[0.282]	[0.203]	[0.917]	
At least one member in education	0.394	0.045	0.044	0.002	904
	(0.490)	(0.036)	(0.037)	(0.037)	
		[0.214]	[0.232]	[0.947]	

Note: OLS estimates of treatment effects on survey outcomes at endline (individual and household level). Outcome variables are listed on the left and described in detail in Table 5.C.2 in Appendix 5.C. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy, estimating Eq.(5.7.1). Column (3) and (4) present coefficients on the treatment dummy, estimating Eq.(5.7.2). We report robust standard errors in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4. Additional controls are a dummy variable indicating the gender of the main recipient, dummies for the neighborhood in which the household is located (ten neighborhoods), dummies for the type of household (four types), dummies for household composition regarding nationalities (three types), and dummies for the highest education level reached by any household member (three levels).

Table 5.E.9 Treatment Effects at Endline Using Logistic Regression.

	Control mean (SD) (1)	Treatment effect (2)	Activation policy (3)	Partial withdrawal (4)	N (5)
Working	0.473	0.589	0.758	1.203	901
	(0.500)	(0.184)	(0.191)	(0.196)	
		[0.004]	[0.146]	[0.346]	
Job search past 4w	0.024	0.390	0.823	0.824	694
	(0.155)	(0.575)	(0.752)	(0.820)	
		[0.102]	[0.796]	[0.814]	
Social participation	0.378	1.043	1.492	0.902	904
	(0.486)	(0.171)	(0.174)	(0.176)	
		[0.805]	[0.021]	[0.557]	
Education past 6m	0.212	1.221	1.671	1.224	900
	(0.410)	(0.197)	(0.191)	(0.194)	
		[0.310]	[0.007]	[0.299]	
At least one member working	0.638	0.618	0.720	0.939	904
	(0.481)	(0.181)	(0.185)	(0.188)	
		[0.008]	[0.076]	[0.738]	
At least one member in education	0.394	1.235	1.259	1.025	904
	(0.490)	(0.172)	(0.169)	(0.169)	
		[0.220]	[0.173]	[0.882]	

Note: Logistic regression estimates of treatment effects on survey outcomes at endline (individual and household level). Outcome variables are listed on the left and described in detail in Table 5.C.2 in Appendix 5.C. Column (1) reports control group means with standard deviations in parentheses. Column (2) shows coefficients on the treatment dummy and Column (3) and (4) coefficients on dummies indicating the respective treatment arm. We report coefficients in Odds Ratios. Robust standard errors are shown in parentheses and p-values in brackets. All models include randomization strata fixed effects and control for the survey mode, the respective baseline value, and the covariates listed in Table 5.4.

5.F Additional Figures

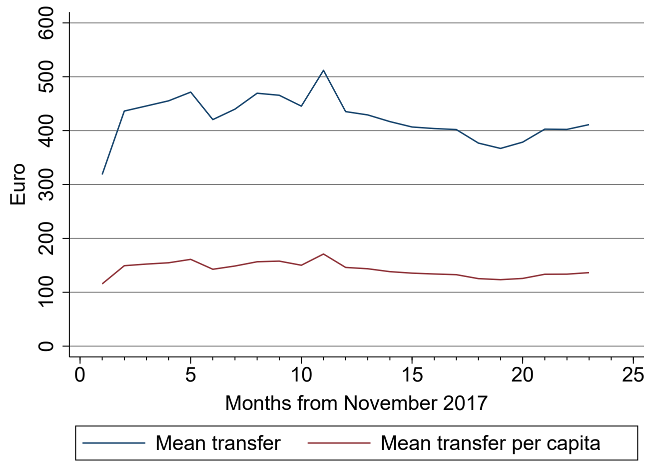


Figure 5.F.1 Mean Transfer and Mean Transfer per Capita per Treatment Month.

Note: Zero payments included.

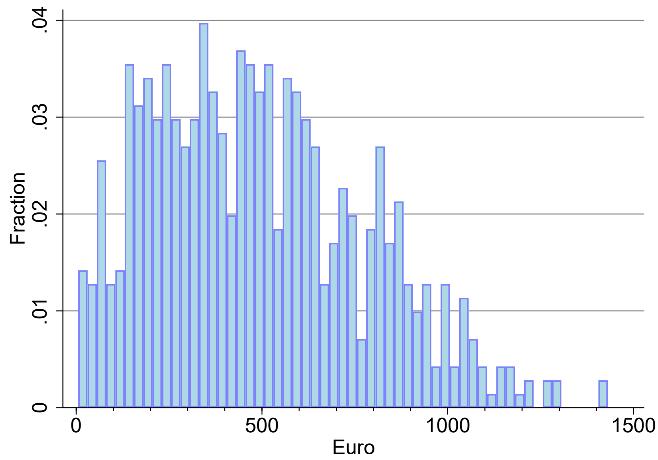


Figure 5.F.2 Distribution of Mean Monthly Transfers.

Note: Zero payments excluded.

CHAPTER 6

Conclusion

This dissertation addressed the following main research question: "What are promising directions for organizing income support?" I conducted four separate empirical investigations to collect answers to this question. Each investigation examined the effects of an existing or new policy feature on labor market and other relevant policy outcomes. All policy features studied have in common that they relate to activating benefit recipients and promoting their economic self-sufficiency. Examples of such features include earnings exemptions, job search requirements, and other activity-related criteria, such as participation in activation and training programs, as well as sanctions in the case of infractions. I focused on three different income support schemes: a minimum income scheme (social assistance), a social insurance scheme (unemployment insurance (UI) benefits), and a cash transfer program operating outside the existing social protection system. All studies employed (quasi-) experimental research designs.

In what follows, I briefly summarize each investigation and discuss the principal findings of each chapter in the light of the main research question (6.1). For a detailed discussion of research findings, I refer to the conclusion sections of the individual chapters. Next, I discuss the main contributions of this dissertation (6.2). To conclude this chapter and dissertation, I review the main limitations and provide suggestions for future research (6.3).

6.1 Summary and Discussion of Main Findings

6.1.1 Requirements Versus Autonomy: What Works in Social Assistance?

Chapter 2 aimed to answer the following research question: "What are the effects of implementing autonomy-enhancing regimes in social assistance?" Autonomy-enhancing regimes refer to regimes that give claimants a greater say in reemployment decisions. Together with my co-authors, I studied the effects of two such regimes in a randomized controlled trial (RCT) conducted with Dutch social assistance claimants. Claimants in a first treatment group were exempted from usual requirements, such as reporting job search efforts, accepting vacancy referrals, or meeting caseworkers. Consequently, monitoring and sanctioning were also eliminated. Claimants in a second treatment group remained subject to the usual obligations but received intensive counseling according to their needs and wishes.

The main findings of the chapter can be summarized in three parts. First, exempting claimants resulted in positive labor market effects on average, leading to higher chances of employment, self-sufficiency, and working under a permanent contract roughly 1.5 years later. Second, while the effects of exemption increased over time, the effects of counseling appeared sooner but stagnated and remained largely statistically insignificant in the longer term. Third, there was no evidence that any of the two treatments affected job search behavior, social participation, health, or well-being. However, survey results indicated gains in experienced autonomy among claimants assigned to exemption.

In sum, the results of Chapter 2 suggest that welfare claimants are quite certainly more successful in returning to work and becoming self-sufficient when making their own reemployment decisions and refraining from interaction with the authority tasked to supervise them. This finding contrasts with standard economic predictions and previous empirical evidence, both suggesting impaired labor market outcomes in the absence of control (see, e.g., Bolhaar et al., 2020; Johnson and Klepinger, 1994; Klepinger et al., 2002; McVicar, 2008, 2010). On the other hand, the finding appears consistent with theories postulating the benefits of autonomy and self-determined behavior (Deci and Ryan, 1985; Frey and Jegen, 2001) and empirical evidence on the positive relationship between experienced autonomy, job search, and reemployment chances (Koen et al., 2016).

The findings of Chapter 2 provide important insights into the design of income support. Recent welfare reforms in many countries have focused on increasing strictness to make a scheme more activating. The results presented in Chapter 2 suggest that a different direction may be worth considering. This direction gives claimants more agency over their strategies and activities to return to the labor market and become self-sufficient. In particular, limiting this agency to decisions over existing programs and instruments does not seem to produce favorable results. Finally, exemption appears to be an interesting direction not only in terms of outputs but also in terms of inputs. In contrast to a regime that requires monitoring and supervision, an exemption scheme can be expected to involve less cost and fewer administrative requirements.

6.1.2 Do Earnings Exemptions Stimulate Paid Work Among Welfare Claimants?

In addition to activity-related incentives, financial work incentives constitute an important design feature of income support schemes. Therefore, Chapter 3 addressed the following research question: “Do generous earnings exemptions stimulate welfare claimants to work?” Earnings exemptions refer to regulations that allow claimants to keep fractions of their income earned on top of benefits. Such regulations aim to increase labor market attachment and facilitate claimants’ transition toward full-time employment and self-sufficiency. Together with my co-authors, I studied the research question mentioned above using data from the same experiment subject to Chapter 2. Specifically, I examined whether allowing claimants to keep (i) a larger share of their income for (ii) a more extended period would affect chances of part-time and full-time employment.

The results of Chapter 3 indicated that a more generous earnings exemption stimulated part-time work and had positive effects on claimants’ income situation, on average. However, the policy change did not affect chances of full-time employment and exit from benefits. Finally, there was no evidence of an effect on welfare expenditures, suggesting that higher employment rates largely compensated higher costs in terms of foregone benefit reductions.

Taken together, the findings of Chapter 3 confirm previous evidence (see, e.g., Blank et al., 1999; Knoef and van Ours, 2016) suggesting that earnings exemptions may stimulate employment in addition to benefits. However, the findings indicate that the impact of even generous exemptions appears limited when it comes to full-time exits from welfare. Potentially, transition to full-time work remains hampered by factors such as higher reservation wages, decreased job search effort, low human capital accumulation, or signaling of low productivity. Finally, the absence of more substantial effects on part-time work could hint at administrative hurdles involved in claiming earnings exemptions.

Current debates about income support frequently mention expanded earnings exemptions as an important direction for policy reform. The findings of Chapter 3 suggest that this direction warrants clarity about specific policy goals. Expanding exemptions seem a promising strategy if the goal is to increase labor market attachment. Policies may fall short of expectations if the goal is to achieve independence from benefits altogether. Given that a large group of welfare claimants has limited prospects of working full-time, the former goal may dominate. In that case, an aspect that speaks in favor of broader exemptions is their little apparent impact on welfare expenditures.

6.1.3 The Effects of Sanctions and Reprimands in Unemployment Insurance

An important instrument to enforce compliance with requirements is benefit sanctions. Chapter 4 asked: “What are the effects of imposing benefit sanctions and reprimands on job seekers?” Together with my co-author, I studied this question in the Dutch UI benefit system, where infractions with job search requirements (four activities every four weeks) can result in a reprimand or a benefit cut of 25 percent for four months. I applied an instrumental variable (IV) approach to solve endogeneity problems resulting from selectiveness in imposing disciplinary measures. The empirical design relied on the quasi-random assignment of claimants to caseworkers, which systematically vary in their tendency to impose a certain measure. More specifically, I calculated caseworkers' stringency and used it as an instrument for actual treatment—an approach also referred to as *leniency design* (Cunningham, 2021).

The main findings of Chapter 4 are threefold. First, results showed that imposing a sanction promotes compliant behavior in the future while issuing a reprimand leaves probabilities of re-offense unchanged. Second, in contrast to previous evidence (see, e.g., van den Berg et al., 2004; Svarer, 2011), both measures appeared to leave job finding and exit from benefits unaffected. However, tentative evidence suggested that sanctions could harm earnings prospects in the long term. Lower earnings may be explained by sanctioned claimants accepting lower-quality jobs, as suggested by earlier studies (see, e.g., Arni et al., 2013). Third, both measures showed to stimulate the reporting of job search efforts, with reprimands resulting in larger effects.

Income support schemes of different types commonly rely on benefit sanctions to enforce compliance. The findings of Chapter 4 suggest that benefit sanctions can be a helpful tool in this regard. When outcomes are considered more broadly, however, the use of reprimands emerges as an alternative direction worth considering. Not only do reprimands show no adverse effects on long-term economic outcomes, they also appear to be more effective in stimulating (reported) job search efforts.

It is important to note that the chapter's findings cannot speak to the question whether a sanctioning regime should be replaced altogether. After all, sanctions may also have deterring effects. Examining these effects lay beyond the scope and possibilities of the investigation in Chapter 4. Hence, there is a reason why policymakers might want to retain sanctions as an enforcement mechanism. In that case, favoring reprimands over sanctions when actually imposing a disciplinary measure could achieve more favorable results. A layered sanctioning scheme may help reconcile different demands. It may be feasible, e.g., to issue reprimands in case of first-time infringements and a sanction in case of recurring violations.

6.1.4 The Labor Supply Effects of Generous and Unconditional Cash Support

While previous chapters investigated (changes in) policies of already existing support schemes, Chapter 5 took a step back and examined the implementation of a new scheme providing monetary transfers. The chapter addressed the following research question: "What are the labor supply effects of offering generous and unconditional income support?" Together with my co-authors, I studied a cash transfer targeting economically vulnerable households that was tested in an RCT in Barcelona (Spain). The cash transfer paid roughly half the local monthly minimum wage and did not include any activity-related criteria.

The main findings of Chapter 5 can be summarized in three parts. First, results showed persistent negative labor supply effects, while social participation and education were unaffected. Second, combining the transfer with a social activation plan appeared to aggravate negative employment effects; imposing lower benefit reduction rates alleviated negative effects. Third, a heterogeneity analysis indicated that negative employment effects occurred almost exclusively among recipients with children in the household. Hence, the overall effects may mask important differences in labor supply responses between households with and without care responsibilities. This last finding fits with evidence from some of the 1970s U.S. negative income tax experiments, indicating stronger reductions in work effort among single female household heads (Burtless, 1986; Robins, 1985).

The findings of Chapter 5 yield some important lessons for the design of income support. On the one hand, the results confirm the predictions offered by standard labor supply models: the income and substitution effects of a welfare program lead to lower labor supply on average. On the other hand, the results hint at an important potential trade-off: generous and unconditional income support could allow for substitution of labor for care duties. From a broader welfare perspective, this outcome may not be detrimental. If more parental time leads to better educational outcomes for children, e.g., policies of the type studied could improve the future earnings prospects of children and impede the transmission of intergenerational poverty. More parental time may also improve children's health and well-being. In sum, even though supporting households with unconditional transfers involves costs in the form of labor force withdrawal, such policies could pay off in the longer term and help achieve broader policy goals.

6.2 Main Contributions

This dissertation expands our understanding of which policies improve (and do not improve) the working of income support. Its main contributions can be summarized in three parts.

First, this dissertation features evidence on the effects of policies and policy features that so far remain understudied. For instance, policies that favor autonomy over control have, until now, barely been investigated. Related evidence is almost exclusively limited to contexts in which subjects were exposed to fewer requirements coincidentally (e.g., due to welfare office refurbishments) or to treatments that refrain from stressing opportunities for choice and self-direction. Likewise, most of the evidence on disciplinary measures imposed on claimants concerns sanctions. Little is known about the effectiveness of softer alternatives, such as reprimands. Finally, although unconditional cash support is well understood in the context of development aid, less is known about its effects elsewhere. Expanding the evidence base in the directions mentioned above is of great practical relevance, allowing for reconsideration of policy designs. In addition, the insights featured are of great scientific importance, opening up new avenues of research.

Second, this dissertation leverages empirical strategies underrepresented in the existing body of research. Experimental designs are well-established in research on labor market policies more generally and activation policies in particular. However, they still represent a relatively small share of the literature. Using RCTs to evaluate policy regimes and instruments in income support, I contribute to expanding the experimental evidence base in that field. In addition to experiments, I employ an IV design to identify causal effects. IV methods have played an important role in empirical economic research for many years. However, only a few studies have used IV methods to study labor market policies. More specifically, this dissertation features the first attempt to identify the causal effects of benefit sanctions using an IV approach. Notably, the growing importance of RCTs and instrumental variables in empirical economic research has lately received recognition by awarding forerunners in both strategies with the Nobel Memorial Prize in Economic Sciences.

Third, this dissertation comprises comprehensive evaluations in terms of the outcomes studied. Much of the existing work studying income support focuses on employment and exit probabilities exclusively. However, a more comprehensive perspective is needed to understand the broader effects of social welfare policies. Only recently do investigations respond to this demand and include a broader range of outcomes. Examples include effects on health and crime (Bolhaar et al., 2019), or reemployment quality (Arni et al., 2013). The studies compiled in this dissertation contribute to this development. In addition to labor market effects, I examined the impacts on health and well-being, social participation, education, job search, and compliance behavior. In doing so, this dissertation expands our understanding of broader policy impacts and helps uncover the potential trade-offs between different policy goals.

6.3 Main Limitations and Directions for Future Research

Naturally, there are several questions that this dissertation has to leave unanswered. In what follows, I will review some overarching limitations and discuss directions for future research. I refer to the individual chapters for a more detailed and topic-specific discussion.

Three of the four chapters included in this dissertation use data from an RCT. While the randomized exposure to treatment alleviates concerns about internal validity, RCT designs often raise questions about external validity (i.e., does the causal relationship hold elsewhere?). More specifically, *elsewhere* may refer to other situations, populations, temporal settings, or measures. Without a doubt, the evidence presented in Chapter 2, 3, and 5 is subject to this limitation. It remains an open question, e.g., whether the autonomy-enhancing social assistance regimes studied in Chapter 2 produce the same results among a population of new claimants, which do not experience a switch from one regime to the other. Similarly, the financial work incentives studied in Chapter 3 may work differently in times of economic downturn. In Chapter 5, it remains unclear, e.g., whether the transfer program has similar effects in other target areas or among a sample of randomly chosen households.

However, some of the interventions featured in the chapters mentioned above have rarely been investigated before. Consequently, the RCTs included in this dissertation may best be described as studies *exploring* causal relations rather than confirming the broader application of the interventions tested. List (2020) classifies this type of studies as WAVE1. While studies of WAVE1 focus on "producing first tests of theory or establishing initial causality" (List, 2020, p.43), WAVE2 studies "broaden the exploration of boundary conditions, and replicate" (List, 2020, p.44). Following this taxonomy, it is WAVE2 studies that should be prioritized in research following-up on the results presented. Part of that effort may include multi-site trials. As for Chapter 2 and 3, such efforts have already been initiated. Together with a group of researchers from different Dutch universities and the Netherlands Bureau for Economic Policy Analysis (CPB), I am currently compiling and comparing data from comparable experimental studies in the Netherlands.[1] Important for the design of WAVE2 studies will also be the inclusion of samples more representative of the target population.

The distinction between WAVE1 and WAVE2 studies also relates to a second limitation of the work presented in this dissertation. Although the included studies offer solid indications of *what* happens when implementing certain treatments, the question of *why* things happen largely remains subject to speculation. In Chapter 2, e.g., I find evidence that social assistance claimants exempted from usual compliance requirements report higher levels of experienced autonomy in their reemployment decisions. However, if and how these experiences translated into job search behavior and successful job finding remains an open question. The findings of Chapter 5 suggest stronger negative effects on labor supply among transfer recipients with children. However, it remains unclear whether these disparate effects derive from the substitution of labor for care duties or from other potential drivers.

While WAVE1 studies provide initial evidence, WAVE2 studies "dig deeper into mechanisms" (List, 2020, p.44). It is highly recommendable for these studies to invest in collecting detailed data. For instance, obtaining access to comprehensive administrative data might free up space for more extensive instruments in complementary surveys. Regarding Chapter 4, in which I studied the effects of benefit sanctions and reprimands, additional data collection efforts have already been initiated. Together with the Dutch Employee Insurance Agency (UWV), I have begun to gather more detailed information on the search behavior of job seekers. I expect that this information will allow for a better understanding of how sanctions and reprimands translated into the outcomes observed. In addition to detailed data, including multiple treatment arms to better study program variations and conducting comprehensive heterogeneity analyses may constitute other ways of digging deeper into mechanisms. Without a doubt, both strategies require the recruitment of larger samples than the ones featured in this dissertation.

A third limitation encountered throughout this dissertation concerns statistical power. In Chapter 2 and 3, e.g., the limited sample size of the RCT on which the chapters are based frustrated the detection of small treatment effects. In Chapter 4, limited data availability for key outcomes of interest restricted the size of our estimation sample and—in combination with the empirical strategy chosen—led to

[1] These studies are described in local policy reports, which are only available in Dutch. For more information, see Betkó et al. (2020), Edzes et al. (2020), Gramberg and de Swart (2020), and Muffels et al. (2020a,b).

less precision than desired. Needless to say, it is recommended that future research includes larger samples. Not only will larger samples lead to more precise estimates, but they also allow for more comprehensive investigations of effect heterogeneity (as put forward in the previous paragraph) and more confident conclusions if multiple hypotheses are tested.

Two aspects, in particular, may be vital in obtaining sufficient sample sizes in the future. I mention those aspects due to the lessons learned by conducting the two RCTs included in this dissertation. First, it is highly recommended to avoid attempts by partner organizations to expand the number of treatment conditions. While the goal of partner organizations to gain insight on a multitude of interventions is understandable, each additional treatment condition increases the sample size requirements. Second, it is advisable to prioritize the harmonization of research designs if trials on multiple sites are in the works. Although local trial partners understandably strive for unique study designs, harmonization may lead to larger samples and higher statistical power. As mentioned above, harmonization also benefits efforts to generalize the results of the study.

Still, there may be research settings that do not allow for (adequately sized) RCTs. Therefore, it will remain important to identify opportunities for applying quasi-experimental techniques. Settings in which administrative procedures involve a (quasi-) random allocation of cases or individuals to decision makers, such as the setting in Chapter 4, may prove fruitful. Other natural sources of exogenous variation in treatment, such as cutoff points of phased roll-outs, should also continue to be leveraged in the future.

Bibliography

Abbring, J.H., van den Berg, G.J., 2003. The Nonparametric Identification of Treatment Effects in Duration Models. Econometrica 71, 1491–1517.

Abbring, J.H., van den Berg, G.J., van Ours, J.C., 2005. The Effect of Unemployment Insurance Sanctions on the Transition Rate from Unemployment to Employment. Economic Journal 115, 602–630.

Aizer, A., Doyle, J.J., 2015. Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges. Quarterly Journal of Economics 130, 759–803.

Akee, R.K.Q., Copeland, W.E., Keeler, G., Angold, A., Costello, E.J., 2010. Parents' Incomes and Children's Outcomes: A Quasi-Experiment Using Transfer Payments from Casino Profits. American Economic Journal: Applied Economics 2, 86–115.

Angrist, J.D., Imbens, G.W., 1994. Identification and Estimation of Local Average Treatment Effects. Econometrica 62, 467–475.

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association 91, 444–455.

Arni, P., van den Berg, G.J., Lalive, R., 2022. Treatment Versus Regime Effects of Carrots and Sticks. Journal of Business & Economic Statistics 40, 111–127.

Arni, P., Lalive, R., van Ours, J.C., 2013. How Effective Are Unemployment Benefit Sanctions? Looking Beyond Unemployment Exit. Journal of Applied Econometrics 28, 1153–1178.

Arni, P., Schiprowski, A., 2019. Job search requirements, effort provision and labor market outcomes. Journal of Public Economics 169, 65–88.

Ashenfelter, O., Plant, M.W., 1990. Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs. Journal of Labor Economics 8, S396–S415.

Autor, D., Kostøl, A., Mogstad, M., Setzler, B., 2019. Disability Benefits, Consumption Insurance, and Household Labor Supply. American Economic Review 109, 2613–2654.

Baird, S., Ferreira, F.H.G., Özler, B., Woolcock, M., 2014. Conditional, Unconditional and Everything in Between: A Systematic Review of the Effects of Cash Transfer Programs on Schooling Outcomes. Journal of Development Effectiveness 6, 1–43.

Banerjee, A.V., Hanna, R., Kreindler, G.E., Olken, B.A., 2017. Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs. The World Bank Research Observer 32, 155–184.

Bargain, O., Doorley, K., 2011. Caught in the trap? Welfare's disincentive and the labor supply of single men. Journal of Public Economics 95, 1096–1110.

Bastagli, F., Hagen-Zanker, J., Harman, L., Barca, V., Sturge, G., Schmidt, T., 2016. Cash transfers: what does the evidence say? A rigorous review of impacts and the role of design and implementation features. Overseas Development Institute, London.

van den Berg, G.J., Hofmann, B., Uhlendorff, A., 2019. Evaluating Vacancy Referrals and the Roles of Sanctions and Sickness Absence. Economic Journal 129, 3292–3322.

van den Berg, G.J., van der Klaauw, B., 2006. Counseling and Monitoring of Unemployed Workers: Theory and Evidence from a Controlled Social Experiment. International Economic Review 47, 895–936.

van den Berg, G.J., van der Klaauw, B., van Ours, J.C., 2004. Punitive Sanctions and the Transition Rate from Welfare to Work. Journal of Labor Economics 22, 211–241.

van den Berg, G.J., Vikström, J., 2014. Monitoring Job Offer Decisions, Punishments, Exit to Work, and Job Quality. Scandinavian Journal of Economics 116, 284–334.

Betkó, J., Spierings, N., Gesthuizen, M., Scheepers, P., 2020. Rapportage experiment Participatiewet gemeente Nijmegen. Radboud Universiteit, Nijmegen.

Bhuller, M., Dahl, G.B., Løken, K.V., Mogstad, M., 2020. Incarceration, Recidivism, and Employment. Journal of Political Economy 128, 1269–1324.

Black, D.A., Smith, J.A., Berger, M.C., Noel, B.J., 2003. Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System. American Economic Review 93, 1313–1327.

Blank, R.M., Card, D., Robins, P.K., 1999. Financial Incentives for Increasing Work and Income among Low-Income Families. NBER Working Paper No. 6998.

Bloom, D., Michalopoulos, C., 2001. How Welfare and Work Policies Affect Employment and Income: A Synthesis of Research. Manpower Demonstration Research Corp., New York.

Bohmann, S., Fiedler, S., Kasy, M., Schupp, J., Schwerter, F., 2021. Experimental evaluation of a Basic Income Pilot in Germany. AEA RCT Registry, May 27. doi:`https://doi.org/10.1257/rct.7734-1.0`.

Bolhaar, J., Ketel, N., van der Klaauw, B., 2019. Job Search Periods for Welfare Applicants: Evidence from a Randomized Experiment. American Economic Journal: Applied Economics 11, 92–125.

Bolhaar, J., Ketel, N., van der Klaauw, B., 2020. Caseworker's discretion and the effectiveness of welfare-to-work programs. Journal of Public Economics 183, 1–19.

Boone, J., Sadrieh, A., van Ours, J.C., 2009. Experiments on unemployment benefit sanctions and job search behavior. European Economic Review 53, 937–951.

Brewer, M., Duncan, A., Shephard, A., Suárez, M.J., 2006. Did working families' tax credit work? The impact of in-work support on labour supply in Great Britain. Labour Economics 13, 699–720.

Bruhn, M., McKenzie, D., 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. American Economic Journal: Applied Economics 1, 200–232.

Burtless, G., 1986. The Work Response to a Guaranteed Income: A Survey of Experimental Evidence. Conference Series [Proceedings], Federal Reserve Bank of Boston 30, 22–59.

Busk, H., 2016. Sanctions and the exit from unemployment in two different benefit schemes. Labour Economics 42, 159–176.

Cairo, S., Mahlstedt, R., 2021. Screening, Deterrence, or Both? Work Requirements and the Labor Market Performance of Welfare Recipients. mimeo.

Caliendo, M., Künn, S., Uhlendorff, A., 2016. Earnings exemptions for unemployed workers: The relationship between marginal employment, unemployment duration and job quality. Labour Economics 42, 177–193.

Card, D., Kluve, J., Weber, A., 2010. Active Labour Market Policy Evaluations: A Meta-Analysis. Economic Journal 120, F452–F477.

Card, D., Kluve, J., Weber, A., 2017. What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. Journal of the European Economic Association 16, 894–931.

Card, D., Robins, P.K., 1996. Do Financial Incentives Encourage Welfare Recipients to Work? Evidence from a Randomized Evaluation of the Self-sufficiency Project. NBER Working Paper No. 5701.

Carril, A., 2017. Dealing with misfits in random treatment assignment. The Stata Journal 17, 652–667.

Commissie Regulering van Werk, 2020. In wat voor land willen wij werken? Naar een nieuw ontwerp voor de regulering van werk. Eindrapport van de Commissie Regulering van Werk.

Cunningham, S., 2021. Causal Inference: The Mixtape. Yale University Press, New Haven & London.

Deci, E.L., Ryan, R.M., 1985. Intrinsic Motivation and Self-Determination in Human Behavior. Plenum, New York.

Divosa, 2015. Divosa-monitor factsheet (2015-II): Parttime werk in de bijstand. Divosa, Utrecht.

Divosa, 2021. Divosa Benchmark Werk & Inkomen, Jaarrapportage 2020. Divosa, Utrecht.

Dolton, P., O'Neill, D., 1996. Unemployment Duration and the Restart Effect: Some Experimental Evidence. Economic Journal 106, 387–400.

Dolton, P., O'Neill, D., 2002. The Long-Run Effects of Unemployment Monitoring and Work-Search Programs: Experimental Evidence from the United Kingdom. Journal of Labor Economics 20, 381–403.

Edzes, A., Rijnks, R., Kloosterman, K., Venhorst, V., 2020. Bijstand op Maat, Beleidsrapport, URSI-onderzoeksrapport 366. Rijsuniversiteit Groningen, Groningen.

Eleveld, A., Kampen, T., Arts, J., 2020. Welfare to Work in Contemporary European Welfare States: Legal, Sociological and Philosophical Perspectives on Justice and Domination. Policy Press, Bristol.

Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games and Economic Behavior 54, 293–315.

Filges, T., Smedslund, G., Due Knudsen, A.S., Klint Jørgensen, A.M., 2015. Active Labor Market Programme Participation for Unemployment Insurance Recipients: A Systematic Review. Campbell Systematic Reviews 2015:2.

Fredriksson, P., Holmlund, B., 2006. Improving Incentives in Unemployment Insurance: A Review of Recent Research. Journal of Economic Surveys 20, 357–386.

French, E., Song, J., 2014. The Effect of Disability Insurance Receipt on Labor Supply. American Economic Journal: Economic Policy 6, 291–337.

Frey, B.S., Jegen, R., 2001. Motivation Crowding Theory. Journal of Economic Surveys 15, 589–611.

Gaudet, F.J., Harris, G.S., St. John, C.W., 1933. Individual Differences in the Sentencing Tendencies of Judges. Journal of Criminal Law and Criminology 23, 811–818.

Geerdsen, L.P., 2006. Is there a Threat Effect of Labour Market Programmes? A Study of ALMP in the Danish UI System. Economic Journal 116, 738–750.

Gerards, R., Welters, R., 2021. Does eliminating benefit eligibility requirements improve unemployed job search and labour market outcomes? Applied Economics Letters.

Gerber, A.S., Green, D.P., 2012. Field Experiments: Design, Analysis, and Interpretation. W. W. Norton & Company, New York.

Gorter, C., Kalb, G.R.J., 1996. Estimating the Effect of Counseling and Monitoring the Unemployed Using a Job Search Model. Journal of Human Resources 31, 590–610.

Gramberg, P., de Swart, J., 2020. Wat werkt op weg naar werk? Eindrapport experiment Participatiewet gemeente Deventer. Saxion Hogeschool, Enschede.

Graversen, B.K., van Ours, J.C., 2008. How to help unemployed find jobs quickly: Experimental evidence from a mandatory activation program. Journal of Public Economics 92, 2020–2035.

Greenberg, D.H., Michalopoulos, C.H., Robins, P.K., Wood, R.H., 1995. Making work pay for welfare recipients. Contemporary Economic Policy 13, 39–52.

Grogger, J., 2001. The Effects of Time Limits and Other Policy Changes on Welfare Use, Work, and Income Among Female-Headed Families. NBER Working Paper No. 8153.

Handa, S., Daidone, S., Peterman, A., Davis, B., Pereira, A., Palermo, T., Yablonski, J., 2018. Myth-Busting? Confronting Six Common Perceptions about Unconditional Cash Transfers as a Poverty Reduction Strategy in Africa. The World Bank Research Observer 33, 259–298.

Haushofer, J., Shapiro, J., 2016. The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. Quarterly Journal of Economics 131, 1973–2042.

Heckman, J.J., Lalonde, R.J., Smith, J.A., 1999. The Economics and Econometrics of Active Labor Market Programs, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. Elsevier, North Holland. volume 3A, pp. 1865–2097.

Heß, S., 2017. Randomization Inference with Stata: A Guide and Software. The Stata Journal 17, 630–651.

Hillmann, K., Hohenleitner, I., 2015. Impact of Welfare Sanctions on Employment Entry and Exit from Labor Force—Evidence from German Survey Data. HWWI Research Paper No. 168.

Hoff, S., Jehoel-Gijsbers, G., 2003. De uitkering van de baan. Reïntegratie van uitkeringsontvangers: ontwikkelingen in de periode 1992-2002. Sociaal en Cultureel Planbureau, Den Haag.

Horowitz, J.L., Manski, C.F., 2000. Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data. Journal of the American Statistical Association 95, 77–84.

Hämäläinen, K., Verho, J., Kanninen, O., 2021. Removing Welfare Traps: Employment Responses in the Finnish Basic Income Experiment. EconPol Working Paper No. 61.

Immervoll, H., Knotz, C.M., 2018. How Demanding Are Activation Requirements For Jobseekers. OECD Social, Employment and Migration Working Paper No. 215.

Immervoll, H., Pearson, M., 2009. A Good Time for Making Work Pay? Taking Stock of In-Work Benefits and Related Measures across the OECD. OECD Social, Employment and Migration Working Paper No. 81.

Johnson, T.R., Klepinger, D.H., 1994. Experimental Evidence on Unemployment Insurance Work-Search Policies. Journal of Human Resources 29, 695–717.

Jones, D., Marinescu, I., 2018. The Labor Market Impacts of Universal and Permanent Cash Transfers: Evidence from the Alaska Permanent Fund. NBER Working Paper No. 24312.

Jones, D., Molitor, D., Reif, J., 2019. What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study. Quarterly Journal of Economics 134, 1747–1791.

Kasy, M., Lehner, L., 2021. Employing the unemployed of Marienthal: Evaluation of a guaranteed job program. AEA RCT Registry, February 08. doi:`https://doi.org/10.1257/rct.6706-1.1`.

van der Klaauw, B., van Ours, J.C., 2013. Carrot and Stick: How Re-Employment Bonuses and Benefit Sanctions Affect Exit Rates from Welfare. Journal of Applied Econometrics 28, 275–296.

van der Klaauw, B., te Voortwis, A., de Vos, R., Willems, I., 2008. Waarschuwingen voor werklozen even effectief als sancties. Economische Statistische Berichten 93, 724–726.

Klepinger, D.H., Johnson, T.R., Joesch, J.M., 2002. Effects of Unemployment Insurance Work-Search Requirements: The Maryland Experiment. Industrial and Labor Relations Review 56, 3–22.

Kling, J.R., 2006. Incarceration Length, Employment, and Earnings. American Economic Review 96, 863–876.

Kluve, J., 2010. The effectiveness of European active labor market programs. Labour Economics 17, 904–918.

Knoef, M., van Ours, J.C., 2016. How to stimulate single mothers on welfare to find a job: evidence from a policy experiment. Journal of Population Economics 29, 1025–1061.

Knotz, C.M., 2018. A rising workfare state? Unemployment benefit conditionality in 21 OECD countries, 1980–2012. Journal of International and Comparative Social Policy 34, 91–108.

Koen, J., van Vianen, A.E.M., van Hooft, E.A.J., Klehe, U.C., 2016. How experienced autonomy can improve job seekers' motivation, job search, and chance of finding reemployment. Journal of Vocational Behavior 95-96, 31–44.

Lachowska, M., Meral, M., Woodbury, S.A., 2016. Effects of the unemployment insurance work test on long-term employment outcomes. Labour Economics 41, 246–265.

Lalive, R., Zweimüller, J., van Ours, J.C., 2005. The Effect of Benefit Sanctions on the Duration of Unemployment. Journal of the European Economic Association 3, 1386–1417.

Lee, D.S., 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. Review of Economic Studies 76, 1071–1102.

Lemieux, T., Milligan, K., 2008. Incentive effects of social assistance: A regression discontinuity approach. Journal of Econometrics 142, 807–828.

List, J.A., 2020. Non est Disputandum de Generalizability? A Glimpse into The External Validity Trial. NBER Working Paper No. 27535.

Machin, S., Marie, O., 2006. Crime and benefit sanctions. Portuguese Economic Journal 5, 149–165.

Maestas, N., Mullen, K.J., Strand, A., 2013. Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. American Economic Review 103, 1797–1829.

Maibom, J., Rosholm, M., Svarer, M., 2017. Experimental Evidence on the Effects of Early Meetings and Activation. Scandinavian Journal of Economics 119, 541–570.

Mani, A., Mullainathan, S., Shafir, E., Zhao, J., 2013. Poverty Impedes Cognitive Function. Science 341, 976–980.

Manning, A., 2009. You can't always get what you want: The impact of the UK Jobseeker's Allowance. Labour Economics 16, 239–250.

Marinescu, I., 2017. No Strings Attached: The Behavioral Effects of U.S. Unconditional Cash. The Roosevelt Institute.

Markussen, S., Røed, K., 2016. Leaving Poverty Behind? The Effects of Generous Income Support Paired with Activation. American Economic Journal: Economic Policy 8, 180–211.

Martin, J.P., 2015. Activation and Active Labour Market Policies in OECD Countries: Stylized Facts and Evidence on their Effectiveness. IZA Journal of Labor Policy 4, 1–29.

Matsudaira, J.D., Blank, R.M., 2014. The Impact of Earnings Disregards on the Behavior of Low-Income Families. Journal of Policy Analysis and Management 33, 7–35.

McKenzie, D., 2012. Beyond baseline and follow-up: The case for more T in experiments. Journal of Development Economics 99, 210–221.

McVicar, D., 2008. Job search monitoring intensity, unemployment exit and job entry: Quasi-experimental evidence from the UK. Labour Economics 15, 1451–1468.

McVicar, D., 2010. Does Job Search Monitoring Intensity Affect Unemployment? Evidence from Northern Ireland. Economica 77, 296–313.

McVicar, D., 2020. The impact of monitoring and sanctioning on unemployment exit and job-finding rates. IZA World of Labor 49.

Meyer, B.D., 1995. Lessons from the U.S. Unemployment Insurance Experiments. Journal of Economic Literature 33, 91–131.

Meyer, B.D., Rosenbaum, D.T., 2001. Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers. Quarterly Journal of Economics 116, 1063–1114.

Michalopoulos, C., Robins, P.K., Card, D., 2005. When financial work incentives pay for themselves: evidence from a randomized social experiment for welfare recipients. Journal of Public Economics 89, 5–29.

Moffitt, R.A., 1992. Incentive Effects of the U.S. Welfare System: A Review. Journal of Economic Literature 30, 1–61.

Moffitt, R.A., 1996. The Effect of Employment and Training Programs on Entry and Exit from the Welfare Caseload. Journal of Policy Analysis and Management 15, 32–50.

Moffitt, R.A., 2002. Welfare programs and labor supply, in: Auerbach, A.J., Feldstein, M. (Eds.), Handbook of Public Economics. Elsevier, North Holland. volume 4, pp. 2393–2430.

Mosthaf, A., Schank, T., Schwarz, S., 2021. Do Supplementary Jobs for Welfare Recipients Increase the Chance of Welfare Exit? Evidence from Germany. IZA Discussion Paper No. 14268.

Muffels, R., Blom-Stam, K., van Wanrooij, S., 2020a. Vertrouwensexperiment Tilburg: werkt het en waarom wel of niet? Tilburg University/Reflect, Tilburg.

Muffels, R., Blom-Stam, K., van Wanrooij, S., 2020b. Vertrouwensexperiment Wageningen: werkt het en waarom wel of niet? Tilburg University/Reflect, Wageningen.

Mullainathan, S., Shafir, E., 2013. Scarcity – The True Cost of Not Having Enough. Penguin Books, London.

OECD, 2015. OECD Employment Outlook 2015. OECD, Paris.

OECD, 2018. Benefits an Wages: 2018 Policy Tables. OECD, Paris.

OECD, 2021a. Average annual wages. OECD Employment and Labour Market Statistics (database). doi:`https://doi.org/10.1787/data-00571-en`.

OECD, 2021b. Benefits and wages: Net replacement rates in unemployment. OECD Social and Welfare Statistics (database). doi:`https://doi.org/10.1787/705b0a38-en`.

OECD, 2021c. Building inclusive labour markets: Active labour market policies for the most vulnerable groups. OECD Policy Responses to Coronavirus (COVID-19).

OECD, 2021d. PPPs and exchange rates. OECD National Accounts Statistics (database). doi:`https://doi.org/10.1787/data-00004-en`.

OECD, 2021e. Social Expenditure: Aggregated data (Edition 2020). OECD Social and Welfare Statistics (database). doi:`https://doi.org/10.1787/d3a45935-en`.

Pega, F., Liu, S.Y., Walter, S., Pabayo, R., Saith, R., Lhachimi, S.K., 2017. Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low- and middle-income countries. Cochrane Database of Systematic Reviews Issue 11.

Petrongolo, B., 2009. The long-term effects of job search requirements: Evidence from the UK JSA reform. Journal of Public Economics 93, 1234–1253.

Robins, P.K., 1985. A Comparison of the Labor Supply Findings from the Four Negative Income Tax Experiments. Journal of Human Resources 20, 567–582.

Romer, D., 2020. In Praise of Confidence Intervals. AEA Papers and Proceedings 110, 55–60.

Rosholm, M., 2008. Experimental Evidence on the Nature of the Danish Employment Miracle. IZA Discussion Paper No. 3620.

Rosholm, M., Svarer, M., 2008. The Threat Effect of Active Labour Market Programmes. Scandinavian Journal of Economics 110, 385–401.

Sampat, B., Williams, H., 2019. How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome. American Economic Review 109, 203–236.

Schoeni, R., Blank, R., 2000. What has Welfare Reform Accomplished? Impacts on Welfare Participation, Employment, Income, Poverty, and Family Structure. NBER Working Paper No. 7627.

Shah, A.K., Mullainathan, S., Shafir, E., 2012. Some Consequences of Having Too Little. Science 338, 682–685.

Svarer, M., 2011. The Effect of Sanctions on Exit from Unemployment: Evidence from Denmark. Economica 78, 751–778.

Todeschini, F., Sabes-Figuera, R., 2019. Barcelona city council welfare programme: Impact evaluation results. Ivàlua, Barcelona.

UWV, 2017. UWV Jaarverslag 2017. UWV, Amsterdam.

Venn, D., 2012. Eligibility Criteria for Unemployment Benefits: Quantitative Indicators for OECD and EU Countries. OECD Social, Employment and Migration Working Paper No. 131.

Verlaat, T., Bijleveld, E., Meeldijk, A., Berkhout, P., 2021. Sancties voor burgers: een bruikbaar instrument voor gedragsverandering? Boom criminologie, Den Haag.

Verlaat, T., de Kruijk, M., Rosenkranz, S., Groot, L., Sanders, M., 2020a. Onderzoek Weten wat werkt: samen werken aan een betere bijstand, Eindrapport. Universiteit Utrecht, Utrecht.

Verlaat, T., Meeldijk, A., Bijleveld, E., 2020b. The Effect of Benefit Sanctions in Unemployment Insurance. OSF Registry, September 1. doi:https://doi.org/10.17605/OSF.IO/92RJK.

Verlaat, T., Rosenkranz, S., Groot, L., Sanders, M., 2020c. What works – study into the effects of four different social welfare schemes. AEA RCT Registry, July 10. doi:https://doi.org/10.1257/rct.3592-6.0.

Vethaak, H., van der Klaauw, B., 2021. Empirical Evaluation of Broader Job Search Requirements for Unemployed Workers. mimeo.

Ware, Jr., J.E., Sherbourne, C.D., 1992. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. Medical Care 30, 473–483.

Westfall, P.H., Young, S.S., 1993. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. John Wiley & Sons, Hoboken.

Wetenschappelijke Raad voor het Regeringsbeleid, 2020. Het betere werk. De nieuwe maatschappelijke opdracht, WRR-Rapport 102. Wetenschappelijke Raad voor het Regeringsbeleid, Den Haag.

Young, A., 2019. Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. Quarterly Journal of Economics 134, 557–598.

Nederlandstalige Samenvatting

Deze dissertatie behandelt de volgende overkoepelende onderzoeksvraag: “Wat zijn veelbelovende richtingen in beleid om inkomensondersteuning te organiseren?” Ik voer vier afzonderlijke empirische studies uit om antwoorden op deze vraag te verzamelen. In elk onderzoek bestudeer ik de effecten van bestaand of nieuw beleid op arbeidsmarkt- en andere relevante uitkomsten. De onderzochte beleidsinstrumenten hebben gemeen dat ze het doel hebben uitkeringsgerechtigden te activeren en hun economische zelfredzaamheid te bevorderen. Voorbeelden zijn een vrijlating van bijverdiensten naast de uitkering, verplichtingen om actief naar werk te zoeken of deel te nemen aan activerings- en opleidingsprogramma’s, en financiële sancties in het geval van regelovertredingen.

Mijn onderzoek richt zich op drie verschillende financiële regelingen: een minimaregeling (bijstand), een sociale verzekering (WW-uitkering), en een *cash-transfer*-programma dat buiten het bestaande sociale zekerheidsstelsel opereert. Ik bestudeer de eerste twee regelingen in een Nederlandse context, terwijl de laatste regeling in Spanje is ingevoerd. De drie regelingen hebben gemeen dat ze tijdelijke steun verlenen en als doel hebben de economische zelfredzaamheid van de uitkeringsgerechtigde op zo kort mogelijke termijn te herstellen. Arbeidsmarktgerelateerde uitkomsten, met name een (duurzame) arbeidsinschakeling, spelen daarom in alle vier de studies een prominente rol. Inkomensondersteuning kan echter ook andere effecten hebben, bijvoorbeeld op de gezondheid en het welzijn van uitkeringsgerechtigden, of hun onderwijsdeelname. Het is belangrijk om ook deze uitkomsten in beschouwing te nemen, om een beeld van de brede effecten van beleidsinstrumenten en de afwegingen die verschillende soorten beleid met zich mee kunnen brengen te kunnen schetsen. Daarom bekijk ik in de meeste van mijn studies uitkomsten in verschillende domeinen.

Om arbeidsmarktgerelateerde uitkomsten te meten, verzamel ik administratieve gegevens bij het Centraal Bureau voor de Statistiek (CBS) en bij het Spaanse ministerie van Arbeid en Sociale Economie. Voor andere uitkomsten grijp ik vooral terug op enquêtes. Waar mogelijk maak ik gebruik van gevalideerde en veelgebruikte enquête-instrumenten. In gevallen waarin ik over informatie uit zowel enquêtes als administratieve bronnen beschik, vergelijk ik de resultaten. Dit om vertekende resultaten als gevolg van sociaal wenselijke of onjuiste antwoorden uit te sluiten.

Alle studies in deze dissertatie zijn gebaseerd op (quasi-)experimentele onderzoekstechnieken. Specifiek gebruik ik twee verschillende empirische strategieën om de causale effecten van de bestudeerde programma's en beleidsinstrumenten vast te stellen. Ten eerste gebruik ik een onderzoeksdesign dat in het Engels ook wel een *randomized controlled trial* (RCT) wordt genoemd. Hierin worden personen willekeurig toegewezen aan interventiegroepen, die een per groep verschillende behandeling krijgen, of aan een controlegroep die het status-quo regime volgt. Ten tweede maak ik gebruik van een situatie waarin exogene variatie in behandeling van nature voorkomt. In deze situatie lopen personen verschillende kans om een bepaalde behandeling te krijgen, omdat zij (quasi-)willekeurig worden toegewezen aan beslissingsbevoegde met verschillende neigingen om een behandeling op te leggen. Ik gebruik deze situatie voor een instrumentele-variabele-benadering (IV).

Als geheel vergroot deze dissertatie het begrip van welk beleid de werking van inkomensondersteuning verbetert (en welk niet). De dissertatie vult hierbij een rijke empirische literatuur aan. Deze literatuur reikt van evaluaties van specifieke beleidsonderdelen, zoals actief arbeidsmarktbeleid (zie, bijvoorbeeld, Card et al., 2017; Filges et al., 2015; Kluve, 2010, voor een overzicht), uitkeringssancties (zie McVicar, 2020, voor een overzicht) en verschillende regels en verplichtingen (zie, bijvoorbeeld, Arni and Schiprowski, 2019; Cairo and Mahlstedt, 2021), tot studies van hele programma's, zoals de experimenten met een negatieve inkomstenbelasting in de Verenigde Staten in de jaren 1970 (zie Burtless, 1986, voor een beschrijving van de experimenten). De studies in deze dissertatie bieden op verschillende vlakken een aanvulling op deze literatuur. Voor een gedetailleerde bespreking van de relevante literatuur en bijdragen verwijs ik de lezer naar de respectievelijke hoofdstukken. Overkoepelende bijdragen van deze dissertatie zijn, onder meer, de evaluatie van beleidsinstrumenten die tot dusver onderbelicht zijn gebleven, het gebruik van empirische strategieën die ondervertegenwoordigd zijn in de bestaande literatuur en het uitbreiden van kennis met betrekking tot de bredere effecten van beleid.

In elk van de onderstaande paragrafen vat ik een van de vier studies samen en bespreek ik per studie de belangrijkste bevindingen in het licht van de hoofdonderzoeksvraag.

Verplichtingen versus autonomie: wat werkt in de bijstand?

In hoofdstuk 2 bestudeer ik de effecten van twee alternatieve regimes in de bijstand.[1] Beide regimes verminderen bepaalde uitkeringsvoorwaarden en versterken de zeggenschap van de bijstandsgerechtigde bij beslissingen over de terugkeer naar de arbeidsmarkt. Bijstandsgerechtigden die onder het eerste regime vielen, kregen een ontheffing van alle gemeentelijke arbeids- en re-integratieverplichtingen. Hierbij horen onder andere het actief zoeken naar werk, het aanvaarden van betaald werk, het bijwonen van afspraken met een consulent en deelname aan opleidings- en

[1]De resultaten van dit hoofdstuk (en die van hoofdstuk 3) zijn ook te vinden in een Nederlandstalig onderzoeksrapport (Verlaat et al., 2020a). Het rapport is beschikbaar via `https://dspace.library.uu.nl/handle/1874/395951`.

activeringsprogramma's. Ook ging de gemeente bijstandsgerechtigden in het kader van de ontheffing niet langer monitoren. Aangezien men nog steeds moest voldoen aan een inkomenstoets, was de uitkering niet geheel onvoorwaardelijk. Onder het tweede regime ontvingen bijstandsgerechtigden een intensieve één-op-één begeleiding van een vaste consulent. Bij de invulling van het begeleidingsprogramma werd rekening gehouden met de wensen van de bijstandsgerechtigde wat betreft de hulp en steun die zouden worden verleend.

Samen met mijn coauteurs evalueer ik bovengenoemde twee alternatieve regimes in een veldexperiment met 752 bijstandsgerechtigden in Utrecht. Het veldexperiment vond plaats tussen juni 2018 en december 2019. De belangrijkste bevindingen van het hoofdstuk kunnen in drie delen worden samengevat. Ten eerste blijkt de ontheffing van regels en verplichtingen gemiddeld genomen positieve arbeidsmarkteffecten te hebben. De resultaten laten zien dat de kansen op werk, op zelfredzaamheid en op werk met een vast contract in deze groep aan het einde van het experiment ruwweg 1,5 keer hoger liggen. Ten tweede nemen de effecten van ontheffing in de loop van de tijd toe, terwijl de effecten van intensievere begeleiding eerder optreden maar op de langere termijn grotendeels stagneren. Ten derde vind ik geen bewijs dat een van de twee behandelingen een effect heeft op werkzoekgedrag, sociale participatie, gezondheid of welzijn. Enquêteresultaten wijzen echter op een toename van de ervaren autonomie bij bijstandsgerechtigden met een ontheffing.

Samengevat suggereren de resultaten van hoofdstuk 2 dat bijstandsgerechtigden niet minder succes hebben bij het terugkeren naar betaald werk en het zelfvoorzienend worden als ze hun eigen beslissingen over re-integratie mogen nemen. Deels levert een ontheffing zelfs betere resultaten op, bijvoorbeeld als het gaat om werk met een vast contract. De bevindingen van dit hoofdstuk leveren dan ook belangrijke inzichten op voor de vormgeving van inkomensondersteuning. Recente hervormingen van sociale voorzieningen, met het doel regelingen meer activerend te maken, betekenen in veel landen een verhoging van de striktheid van het stelsel. De in hoofdstuk 2 gepresenteerde resultaten suggereren dat een andere koers het overwegen waard kan zijn. Binnen die andere koers krijgen bijstandsgerechtigden meer zeggenschap over hun strategieën en activiteiten gericht op arbeidsdeelname en uitstroom uit de uitkering. De resultaten in hoofdstuk 2 suggereren verder dat het geen gunstige resultaten oplevert als deze zeggenschap wordt beperkt tot beslissingen over bestaande programma's en instrumenten. Ten slotte lijken hervormingen richting ontheffing interessant voor overheden, niet alleen wat *outputs* betreft, maar ook qua *inputs*: in tegenstelling tot een regeling die monitoring en toezicht vereist, zal een ontheffingsregeling naar verwachting minder kosten en minder administratieve lasten met zich meebrengen.

Stimuleren inkomstenvrijlatingen betaald werk onder bijstandsgerechtigden?

Naast gedragsgerelateerde regels en verplichtingen vormen financiële prikkels een belangrijk onderdeel van financiële ondersteuningsregelingen. In hoofdstuk 3 richt ik me daarom op de volgende onderzoeksvraag: “Worden bijstandsgerechtigden door een genereuze inkomstenvrijlating gestimuleerd om te werken?” Een inkomstenvrijlating is een regeling die bijstandsgerechtigden toestaat een deel van hun looninkomen te behouden als zij werken naast de uitkering. Dergelijke regelingen hebben als doel de arbeidsparticipatie van bijstandsgerechtigden te bevorderen en de terugkeer naar voltijd werk en zelfredzaamheid makkelijker te maken. Samen met mijn coauteurs bestudeer ik bovenstaande onderzoeksvraag met data uit hetzelfde veldexperiment als dat van hoofdstuk 2. Specifiek onderzoek ik of het vrijlaten van (i) een hoger bedrag gedurende (ii) een langere periode een effect heeft op de kans om deeltijd of voltijd te werken.

Uit de resultaten van hoofdstuk 3 blijkt dat een ruimere vrijlatingsregeling deeltijdwerk kan stimuleren en gemiddeld genomen een positief effect heeft op de inkomenssituatie van bijstandsgerechtigden. De beleidswijziging heeft echter geen effect op de kans om voltijd te werken en volledig uit de uitkering te stromen. Ten slotte zijn er geen aanwijzingen voor een effect op uitkeringslasten. Deze laatste bevinding wijst erop dat een hogere arbeidsparticipatie, en de daardoor ontstane lagere uitkeringslasten, grotendeels de hogere kosten als gevolg van niet-gekorte uitkeringen compenseren.

Over het geheel genomen, bevestigen de bevindingen van hoofdstuk 3 eerder wetenschappelijk bewijs dat inkomstenvrijlatingen werk naast de uitkering kunnen stimuleren (zie, bijvoorbeeld, Knoef and van Ours, 2016). De resultaten wijzen er echter ook op dat zelfs een genereuze vrijlatingsregeling slechts beperkt effect heeft op voltijd werk en volledige uitstroom uit de uitkering. Mogelijk wordt de overgang naar voltijd werk binnen een vrijlatingsregeling belemmerd door factoren zoals een hoger reserveloon, minder inspanning om werk te vinden, geringe accumulatie van menselijk kapitaal of signalen van lage productiviteit. Dat ook de effecten op deeltijdwerk beperkt zijn, zou erop kunnen wijzen dat er administratieve hindernissen zijn die werken naast de uitkering bemoeilijken.

In de huidige discussies over inkomensondersteuning worden ruimere vrijlatingsregelingen vaak genoemd als een belangrijke richting voor beleidshervormingen. De bevindingen van hoofdstuk 3 wijzen erop dat deze richting duidelijke beleidsdoelstellingen vergt. Een ruimere inkomstenvrijlating lijkt een veelbelovende strategie als het doel is om werk naast de uitkering stimuleren. Het beleid kan echter tekortschieten als het doel is om volledige onafhankelijkheid van de uitkering te bevorderen. Aangezien een grote groep bijstandsgerechtigden beperkte vooruitzichten heeft om voltijd te werken, kan het eerste doel—werk naast de uitkering—van meer belang zijn. In dat geval spreekt voor een ruimere vrijlatingsregeling dat deze weinig invloed lijkt te hebben op de uitkeringslasten.

De effecten van sancties en waarschuwingen in de WW

Uitkeringssancties zijn een belangrijk beleidsinstrument om de naleving van regels en verplichtingen te bevorderen. In hoofdstuk 4 staat daarom de volgende onderzoeksvraag centraal: “Wat zijn de effecten van het opleggen van uitkeringssancties en waarschuwingen aan werkzoekenden?" Samen met mijn coauteur bestudeer ik deze vraag in een Nederlandse context, waar een overtreding van de sollicitatieplicht in de WW (vier werkzoekactiviteiten per vier weken) kan leiden tot een waarschuwing of een verlaging van de uitkering met 25 procent gedurende vier maanden.[2] In dit hoofdstuk pas ik een instrumentele-variabele-benadering (IV) toe om causale effecten vast te stellen. Ik zet deze methode in met het doel om endogeniteitsproblemen op te lossen die het gevolg zijn van selectiviteit bij het opleggen van maatregelen. De empirische opzet stoelt op de (quasi-)willekeurige toewijzing van WW-gerechtigden die de sollicitatieplicht hebben overtreden, aan behandelaars van UWV, die systematisch verschillen in hun neiging om een bepaalde maatregel op te leggen. Specifiek bereken ik de strengheid van de behandelaars en gebruik ik deze als een instrumentele variabele voor de feitelijk ontvangen behandeling. Deze aanpak wordt ook wel *leniency design* genoemd (Cunningham, 2021).

De belangrijkste bevindingen van hoofdstuk 4 zijn drieledig. Ten eerste blijkt dat het opleggen van een sanctie het toekomstige nalevingsgedrag bevordert, terwijl het geven van een waarschuwing de kans op recidive onveranderd laat. Ten tweede lijkt het erop dat beide maatregelen, in tegenstelling tot eerder empirisch bewijs, geen effect hebben op de kans om werk te vinden en uit de uitkering te stromen. Er zijn echter aanwijzingen dat sancties een nadelig effect kunnen hebben op de inkomensontwikkeling op de langere termijn. Lagere inkomsten in de toekomst kunnen mogelijk worden verklaard doordat gesanctioneerde WW-gerechtigden banen van mindere kwaliteit (bijvoorbeeld minder goed betaald of minder stabiel) aanvaarden, zoals door eerdere studies wordt gesuggereerd (zie, bijvoorbeeld, Arni et al., 2013). Ten derde blijken beide maatregelen het rapporteren van werkzoekactiviteiten te stimuleren, waarbij waarschuwingen een groter effect sorteren.

Regelingen voor inkomensondersteuning maken vaak gebruik van uitkeringssancties om nalevingsgedrag te bevorderen. De bevindingen van hoofdstuk 4 suggereren dat uitkeringssancties in dat opzicht een nuttig instrument kunnen zijn. Wanneer de resultaten echter breder worden bekeken, lijken waarschuwingen een te overwegen alternatief te vormen. Niet alleen blijken waarschuwingen geen nadelige gevolgen te hebben voor de economische uitkomsten van WW-gerechtigden op lange termijn, zij lijken ook effectiever te zijn in het stimuleren van (gemelde) inspanningen om werk te zoeken.

Het is belangrijk om op te merken dat de bevindingen van hoofdstuk 4 geen uitsluitsel geven over de vraag of een sanctieregeling helemaal moet worden vervangen. Sancties kunnen immers ook afschrikkende effecten hebben. Deze afschrikkende effecten heb ik niet binnen het kader van het hoofdstuk kunnen onderzoeken. Er kunnen dus goede redenen zijn waarom beleidsmakers sancties zouden willen behouden. In dat geval zou het verkiezen van waarschuwingen boven sancties bij het daadwerkelijk opleggen van een maatregel voor gunstigere resultaten kunnen zorgen. Een gelaagde sanctieregeling kan dan helpen om verschillende eisen met elkaar

[2] De resultaten van dit hoofdstuk zijn ook te vinden in een Nederlandstalig onderzoeksrapport (Verlaat et al., 2021). Het rapport is beschikbaar via `https://www.handhavingengedrag.nl/onderzoeken/sancties-voor-burgers`.

te verzoenen. Het is bijvoorbeeld denkbaar om bij de eerste overtreding een waarschuwing op te leggen en bij herhaalde overtredingen voor een sanctie te kiezen.

De arbeidsaanbodeffecten van genereuze en onvoorwaardelijke inkomensondersteuning

Terwijl ik in de vorige hoofdstukken de effecten van (mogelijke) beleidsveranderingen van bestaande regelingen heb onderzocht, doe ik in hoofdstuk 5 een stap terug en richt ik me op de invoering van een nieuwe regeling. Het hoofdstuk behandelt de volgende onderzoeksvraag: “Wat zijn de arbeidsaanbodeffecten van genereuze en onvoorwaardelijke inkomensondersteuning?” Om deze vraag te beantwoorden bestudeer ik samen met mijn twee coauteurs een *cash-transfer*-programma gericht op economisch kwetsbare huishoudens. Het programma werd getest in een grootschalig veldexperiment met 1,200 huishoudens in Barcelona (Spanje). Het veldexperiment startte in december 2017 en duurde bijna twee jaar. De hoogte van de maandelijkse overdracht was gemiddeld €492, ruwweg de helft van het lokale maandelijkse minimumloon. De betaling was niet gekoppeld aan gedragsgerelateerde regels of verplichtingen.

De belangrijkste bevindingen van hoofdstuk 5 kunnen in drie delen worden samengevat. Ten eerste blijkt het *cash-transfer*-programma een persistent negatief effect te hebben op de kans om te werken, terwijl er geen effecten zijn op sociale participatie en het volgen van onderwijs. Ten tweede blijken de negatieve arbeidsaanbodeffecten groter te zijn als de *transfer* wordt gecombineerd met een sociaal activeringsbeleid. Als de *transfer* minder snel wordt afgebouwd bij aanvullende inkomsten zijn de negatieve arbeidsaanbodeffecten juist kleiner. Ten derde blijkt uit een heterogeniteitsanalyse dat de negatieve effecten zich bijna uitsluitend voordoen bij huishoudens met kinderen. De gemiddelde effecten lijken dus belangrijke verschillen in de arbeidsaanbodreacties tussen huishoudens met en zonder zorgtaken te maskeren. Deze laatste bevinding sluit aan bij de resultaten van enkele van de Amerikaanse experimenten met negatieve inkomstenbelastingen uit de jaren 1970. De respectievelijke resultaten wezen op een sterkere vermindering van de arbeidsinspanning bij alleenstaande vrouwelijke gezinshoofden (Burtless, 1986; Robins, 1985).

Uit de bevindingen van hoofdstuk 5 kunnen enkele belangrijke lessen worden getrokken voor de vormgeving van inkomensondersteuning. Enerzijds bevestigen de resultaten de voorspellingen van standaard arbeidsaanbodmodellen: de inkomens- en substitutie-effecten van inkomensondersteuning leiden gemiddeld tot een lager arbeidsaanbod. Anderzijds duiden de resultaten op een belangrijke potentiële wisselwerking: genereuze en onvoorwaardelijke ondersteuning kan mogelijk leiden tot substitutie van arbeid door zorgtaken. Vanuit een breder welvaartsperspectief is dat resultaat wellicht gunstig. Meer tijd voor ouderschapstaken zou bijvoorbeeld tot betere onderwijsresultaten van kinderen kunnen leiden, waardoor de inkomensvooruitzichten van kinderen verbeteren en de intergenerationele overdracht van armoede teruggedrongen kan worden. Meer tijd voor ouderschap zou ook de gezondheid en het welzijn van kinderen kunnen verbeteren. Kortom, hoewel genereuze en onvoorwaardelijke inkomensondersteuning kosten met zich meebrengt in de vorm van terugtrekking uit de arbeidsmarkt, kan dergelijk beleid op de langere termijn lonend zijn en bijdragen tot de verwezenlijking van bredere beleidsdoelstellingen.

Curriculum Vitae

Timo Verlaat was born in Fürth (Germany) in 1989. He completed a Bachelor's program in Economics, Political Science, and Sociology at Zeppelin University in Friedrichshafen (Germany) before moving to the Netherlands for a two-year Research Master program in Multidisciplinary Economics at Utrecht University School of Economics (U.S.E.). After graduating from the program in 2016, Timo obtained a doctoral scholarship from the Dutch Research Council (NWO) and became a PhD Candidate at U.S.E. During his PhD, he was a visiting researcher at the Catalan Institute of Public Policy Evaluation (Ivàlua) in Barcelona (Spain) and conducted research projects in collaboration with the Municipality of Utrecht and the Dutch Employee Insurance Agency (UWV). Timo has presented his research at various international conferences and seminars, including the 2016 Conference of the European Network for Social Policy Analysis (ESPAnet), the 2020 Advances with Field Experiments Conference (AFE), and the 2021 Conference of the Society of Labor Economists (SOLE). As part of his appointment at U.S.E., Timo taught the course Behavioral Economics and Public Policy (graduate level) and was a teaching assistant in Econometrics (undergraduate level). As of October 2021, Timo works as a researcher at the Netherlands Bureau for Economic Policy Analysis (CPB).

U.S.E. Dissertation Series

USE 001 **Bastian Westbrock** (2010): *Inter-firm networks: economic and sociological perspectives.*

USE 002 **Yi Zhang** (2011): *Institutions and International Investments: Evidence from China and Other Emerging Markets.*

USE 003 **Ryan van Lamoen** (2011): *The Relationship between Competition and Innovation Measuring Innovation and Causality.*

USE 004 **Martijn Dröes** (2011): *House Price Uncertainty in the Dutch Owner-Occupied Housing Market.*

USE 005 **Thomas van Huizen** (2012): *Behavioural Assumptions in Labour Economics: Analysing Social Security Reforms and Labour Market Transitions.*

USE 006 **Martijn Boerman**s (2012): *International Entrepreneurship and Enterprise Development.*

USE 007 **Joras Ferwerda** (2012): *The Multidisciplinary Economics of Money Laundering.*

USE 008 **Federico D'Onofrio** (2013): *Observing the country: a history of Italian agricultural economics, 1900-1930.*

USE 009 **Saraï Sapulete** (2013): *Works Council Effectiveness: Determinants and Outcomes.*

USE 010 **Britta Hoyer** (2013): *Network Formation under the Threat of Disruption.*

USE 011 **Coen Rigtering** (2013): *Entrepreneurial Orientation: Multilevel Analysis and Consequences.*

USE 012 **Beate Cesinger** (2013): *Context and Complexity of International Entrepreneurship as a Field of Research.*

USE 013 **Jan de Dreu** (2013): *Empirical essays on the governance of financial institutions.*

USE 014 **Lu Zhang** (2013): *Industrial Specialization: Determinants, Processes and Consequences.*

USE 015 **Matthias Filser** (2013): *Strategic Issues in Entrepreneurship and Family Business Research.*

USE 016 **Mikko Pohjola** (2013): *A Compilation of Studies on Innovation in Firms: Capabilities, Strategies, and Performance.*

USE 017 **Han-Hsin Chang** (2013): *Heterogeneity in Development.*

USE 018 **Suzanne Heijnen** (2014): *Analyses of sickness absence.*

USE 019 **Mark Kattenberg** (2014): *The Economics of Social Housing: Implications for Welfare, Consumption, and Labor Market Composition.*

USE 020 **Daniel Possenriede** (2014): *The Economics of Temporal and Locational Flexibility of Work.*

USE 021 **Dirk Gerritsen** (2014): *The Relevance of Security Analyst Opinions for Investment Decisions.*

USE 022 **Shiwei Hu** (2014): *Development in China and Africa.*

USE 023 **Saara Tamminen** (2014): *Heterogeneous Firms, Mark-Ups, and Income Inequality.*

USE 024 **Marcel van den Berg** (2014): *Does Internationalization Foster Firm Performance?*

USE 025 **Emre Akgündüz** (2014): *Analyzing maternal employment and child care quality.*

USE 026 **Jasper Lukkezen** (2014): *From Debt Crisis to Sovereign Risk.*

USE 027 **Vesile Kutlu** (2015): *Essays on Subjective Survival Probabilities, Consumption, and Retirement Decisions.*

USE 028 **Brigitte Crooijmans** (2015): *Leiden fusies tot efficiëntere woningcorporaties? Een exploratieve studie naar schaalvoordelen in de sociale huisvesting.*

USE 029 **Andrej Svorenčík** (2015): *The Experimental Turn in Economics: a History of Experimental Economics.*

USE 030 **Secil Danakol** (2015): *Foreign Direct Investment, Foreign Aid and Domestic Entrepreneurship.*

USE 031 **Ioana Deleanu** (2015): *Anti-Money Laundering Efforts: Failures, Fixes and the Future.*

USE 032 **Jaap Oude Mulders** (2016): *Organizations, managers, and the employment of older workers after retirement.*

USE 033 **Malka de Castro Campos** (2016): *Private Consumption-Savings Behavior and Macroeconomic Imbalances.*

USE 034 **Tahereh Rezai Khavas** (2016): *Fairness concerns and cooperation in context.*

USE 035 **Joyce Delnoy** (2016): *Auctions with Competing Sellers and Behavioral Bidders.*

USE 036 **Krista Bruns** (2017): *Emergence and Diffusion of Institutions and their Effect on Economic Growth.*

USE 037 **Daan van der Linde** (2017): *Democracies under Rising Inequality: New Tests of the Redistributive Thesis.*

USE 038 **Swantje Falcke** (2017): *On the move: Analyzing immigration determinants and immigrant outcomes.*

USE 039 **Joep Steegmans** (2017): *House Prices and Household Mobility in The Netherlands: Empirical Analyses of Financial Characteristics of the Household.*

USE 040 **Najmeh Rezaei Khavas** (2017): *Essays in Information Economics.*

USE 041 **Maryam Imanpour** (2017): *The Role of Social Networks for Combating Money Laundering.*

USE 042 **Ye Li** (2018): *Hydrogen Infrastructure Decisions through a Real Option Lens.*

USE 043 **Li Lin** (2018): *Leadership across cultural contexts.*

USE 044 **Werner Liebregts** (2018): *Hidden entrepreneurship: Multilevel analyses of the determinants and consequences of entrepreneurial employee activity.*

USE 045 **Ian Koetsier** (2018): *Government debt: The economic consequences of natural disasters and pension funds' herding.*

USE 046 **Jordy Meekes** (2019): *Local Labour Markets, Job Displacement And Agglomeration Economies.*

USE 047 **Timur Pasch** (2019): *Essays On The Design Of The Management Accounting System: Determinants, Components And Effects.*

USE 048 **Jeroen Content** (2019): *The role of relatedness and entrepreneurship in regional economic development.*

USE 049 **Franziska Heinicke** (2019): *Essays on self-image and preferences for honesty.*

USE 050 **Rebean Al-silefanee** (2019): *Entrepreneurship and Private Sector Development: The Case of Kurdistan Region of Iraq.*

USE 051 **Markus Meinzer** (2019): *Countering cross-border tax evasion and avoidance: An assessment of OECD policy design from 2008 to 2018.*

USE 052 **Zornitza Kambourova** (2019): *Women's Adverse Health Events and Labor Market Participation.*

USE 053 **Tim van der Valk** (2019): *Household finance in France and the Netherlands 1960-2000: An evolutionary approach.*

USE 054 **Milena Dinkova** (2019): *Brace yourselves, Pension is coming: Consumption, financial literacy and tailored pension communication.*

USE 055 **Lisa Dumhs** (2019): *Finding the right job: School-to-work transitions of vocational students in the Netherlands.*

USE 056 **Dea Tusha** (2020): *FDI spillovers in developing countries: channels, conditions, challenges.*

USE 057 **Jingyang Liu** (2020): *Money and credit dynamics in the euro area.*

USE 058 **An Duong** (2020): *Financial integration, trade, and productivity.*

USE 059 **Katharina Weddige-Haaf** (2021): *Real and Financial Asymmetries in the Euro Area.*

USE 060 **Peter Gerbrands** (2021): *Tax Dynamics and Money Laundering. Simulating Policy Reforms in a Complex System.*