Essays on Labor and Health Economics

by

Mallory Avery
Bachelor of Science, Tulane University, 2014
Master of Arts, University of Pittsburgh, 2016

Submitted to the Graduate Faculty of
the Dietrich School of Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh
2020

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Mallory Avery

It was defended on

April 28, 2020

and approved by

David Huffman, Department of Economics

Osea Giuntella, Department of Economics

Lise Vesterlund, Department of Economics

Sera Linardi, Graduate School of Public and International Affairs

Dissertation Director: David Huffman, Department of Economics

Essays on Labor and Health Economics

Mallory Avery, PhD

University of Pittsburgh, 2020

This dissertation consists of three essays on labor and health economics. Chapter 1 presents work from an experimental study the effect of affirmative action on beneficiaries', particularly women's, career progression. This study presents the first evidence that women face a cost in terms of future employment outcomes when successful in an environment with affirmative action, compared to one without affirmative action. In this particular case, this cost comes as a removal of a hiring advantage successful women have over successful men in environments without affirmative action. Chapter 2 presents work from an experimental study on the behavioral biases impacting sleep choice and the effect of incentives to sleep on sleep behavior. In this study, we use fitness trackers to monitor sleep of subjects who either have no incentives, or who are incentivized to sleep 7-9 hours per night and go to bed by between 10 pm and 1 am. We find that sophisticated hyperbolic discounting, self-serving bias and overconfidence, and risk aversion are important behavioral determinants of sleep behavior. We also find that incentives to sleep are effective in improving sleep choice, with some habit formation after the incentives are removed. Chapter 3 presents work from a study that examines the effect of community care, via the Community Mental Health Act of 1963 and the establishment of Community Mental Health Centers, on mental-illness-related mortality. We find substantial decreases in mortality from such causes of death among those demographic groups at most risk from those particular causes of death and among those with the fewest alternative sources of mental health care when a Community Mental Health Center is places within a county. We argue that these decreases in mortality are a caused by the Community Mental Health Centers.

Table of Contents

Preface xi

1.0 A Hidden Cost of Affirmative Action: Muddying Signals about Women's Ability 1

1.1 Introduction 1

1.2 Experimental Design 4

1.2.1 Employee Design 5

1.2.2 Employer Design 6

1.3 Theoretical Framework 8

1.4 Results 9

1.4.1 Hiring Outcomes 10

1.4.2 Analysis of Employer Beliefs as the Mechanism 11

1.4.3 Welfare and Counterfactual 12

1.5 Discussion and Conclusion 15

1.6 Figures 17

1.7 Tables 25

2.0 Incentives to Sleep: an Experimental Analysis of Sleep Choices (Co-Authors: Osea Giuntella and Peiran Jiao) 30

2.1 Introduction 30

2.2 Experimental Procedure, Design, and Data 35

2.2.1 Experimental Procedure 35

2.2.2 Experimental Design 36

2.3 Data 38

2.3.1 Measuring Sleep 38

2.3.2 Descriptive Analysis: Pre-Intervention Data 39

2.4 Behavioral Biases and Sleep Choice 41

2.4.1 Time Inconsistency and Demand for Commitment 41

2.4.2 Overconfidence 42

2.4.3 Risk Preferences 44

2.5 Incentives, Sleep Behavior, and Habit Formation 45

2.5.1 Post-Intervention 46

2.5.1.1 Habit Formation and Sleep with Fitbit Data 46

2.5.1.2 Follow-up Survey: Effects on Self-Reported Sleep, Health, and Human Capital 47
2.5.2 Incentives to Sleep and Time-Use Allocation 47
2.5.3 Additional Findings: Sleep Regularity, Structure and Size of the Incentives 49
2.6 Conclusion 49
2.7 Figures 52
2.8 Tables 64
3.0 The Mortality Effects of Community Mental Health Centers (Co-Author: Jessica LaVoice) 70
3.1 Introduction 70
3.2 Background Information 73
3.2.1 Effect of Mental Health and Mental Health Treatment 73
3.2.2 Evidence on Publicly-Provided Health Care 74
3.2.3 Historical Context of Community Mental Health Act 74
3.2.4 What Did CMHCS Do and Whom Did They Serve? 76
3.3 Data 76
3.4 Identification Strategy 78
3.4.1 The Expected Effects of CMHCs on Mortality Rates 80
3.5 Results 80
3.6 Robustness Checks 82
3.6.1 Event Study Framework 82
3.6.2 Boundary Analysis 83
3.7 Conclusion 84
3.8 Figures 86
3.9 Tables 92
4.0 Appendix 100
4.1 A: Theoretical Framework 100
4.1.1 Employee Decision Making Framework Without Affirmative Action 100
4.1.2 Employer Decision Making Framework 102
4.1.3 Employer Beliefs that Would Generate Hiring Advantage for Successful Women without Affirmative Action 103
4.1.4 Employee Decision Making Framework with Affirmative Action 104
4.1.5 Employer Beliefs that Would Remove Hiring Advantage for Successful Women with Affirmative Action 105
4.1.6 Alternative Employer Beliefs that Could Lead to Cost of Affirmative Action for Successful Women 106

4.2 B: Appendix Figures 108
4.3 C: Appendix Figures 115
4.4 D: Appendix Tables 119
4.5 E: Elicitation of Risk and Time Preferences 143
4.6 F: Appendix Figures and Tables 150
5.0 Bibliography 160

List of Tables

1.1 Employee Characteristics . . . 25
1.2 Employee Choices . . . 26
1.3 Employer Characteristics . . . 27
1.4 Employment Regressions . . . 28
1.5 Counterfactual Breakdown . . . 29
2.1 Time Preferences and Sleep Duration at Baseline . . . 64
2.2 Perceived Own and Other's Sleep Quality and Sleep Deprivation Risks . . . 65
2.3 Overconfidence and Sleep Duration (Self-reported vs Fitbit Data) . . . 65
2.4 Incentives and Sleep . . . 66
2.5 Incentives to Sleep and Other Outcomes (Fitbit Data) . . . 67
2.6 Correlations with Academic Achievement . . . 68
2.7 Incentives and Time Allocation . . . 69
3.1 Pre-treatment Differences in Demographics across Counties . . . 92
3.2 Priority Ranking and CMHC Rollout . . . 93
3.3 Predictors of CMHC Location . . . 93
3.4 Pre-treatment Mortality Trends . . . 94
3.5 Effects of CMHC on Mortality . . . 95
3.6 Sub-group Analysis by Age - White Population Only . . . 96
3.7 Sub-group Analysis by Age - Non-White Population Only . . . 97
3.8 Urban vs. Rural Sub-group Analysis - Non-White Subsample . . . 98
3.9 Boundary Analysis - Effects of CMHC on Mortality . . . 99
D.1 Summary of Treatments . . . 119
D.2 Baseline Characteristics and Attrition . . . 120
D.3 Comparisons of Sleep Measurements . . . 121
D.4 Summary Statistics, Baseline (Survey-based Metrics) . . . 122
D.5 Intention to Improve Sleep . . . 123
D.6 Correlations Between Sleep, Health, and Well-Being . . . 124
D.7 Demand for Commitment . . . 125
D.8 Present-Bias, Impatience, Overconfidence and Commitment . . . 126
D.9 Determinants of Bedtime Targets . . . 127

D.10 Predicting Achievement Rate 128
D.11 Commitment Devices and Target Achievement 129
D.12 Demand for Commitment and Sleep 130
D.13 Incentives and Sleep, Other Outcomes 131
D.14 Incentives, Bedtime, and Wake-Up Time 132
D.15 Baseline Characteristics and Sample Attrition in Follow-Up Survey 133
D.16 Follow-up Sleep Quality 134
D.17 Naps 135
D.18 Incentives and Sleep, Weekends 136
D.19 Baseline Characteristics and Sample Attrition in Time-Use Survey 137
D.20 Incentives to Sleep and Screen Time Near Bedtime 138
D.21 Incentives and Sleep Regularity 139
D.22 Incentives and Sleep: Timing of the Incentives 140
D.23 Incentives and Sleep: the Role of the Size of the Financial Incentive 141
D.24 Incentives and Sleep 142
F.1 Urban vs. Rural Sub-group Analysis - White Subsample 156
F.2 Boundary Analysis: Sub-group Analysis by Age - White Population Only 157
F.3 Boundary Analysis: Sub-group Analysis by Age - Non-White Population Only 158
F.4 Boundary Analysis: Urban vs. Rural Sub-group Analysis - Non-White Subsample 159

List of Figures

1.1 Employment by Employee Gender, Tournament Type, and Round Three Outcome. 17
1.2 Employer Beliefs by Employee Gender, Tournament Type, and Round Three Outcome 18
1.3 Simulated Probability of Employer Seeing Employee in Particular Gender – Tournament Outcome Category, by Tournament Type 19
1.4 Expected Value of Each Type of Employee, Based on Optimal Hiring Choices. 20
1.5 Probability of being Hired across Gender and Tournament Type 21
1.6 Probability an Individual is in the Top Half of Their Group, by Gender, Tournament Type, and Whether or not They Were Hired 22
1.7 Hiring Rates for Men and Women in the Top and Bottom Half of Their Group, by Tournament Type 23
1.8 Counterfactual Hiring Rates of Women Under Different Tournament Entry Choices and Hiring Probabilities 24
2.1 Design Illustration 52
2.2 Sleep Duration Before Intervention 53
2.3 Sleep Duration Before Intervention 54
2.4 Insufficient Sleep 55
2.5 Sleep Duration Over the Week 55
2.6 % of Subjects Choosing Bedtime Before 1 am. 56
2.7 % of Subjects Choosing Sleep Target >7 hours 57
2.8 Bedtime Targets and Pre-treatment Bedtimes 58
2.9 Bedtime Targets and Predicted Bedtime 59
2.10 Overconfidence and Beliefs 60
2.11 Achievement 61
2.12 % of Nights Bedtime Target was Met 62
2.13 % of Nights Sleep Duration Target was Met 63
3.1 Patient Care Episodes 86
3.2 Rollout of CMHCs 87
3.3 Age-Adjusted Mortality Summary Statistics. 88
3.4 Age-Adjusted Mortality Rates by Age 89
3.5 Event-Study: Suicide among Whites. 90
3.6 Event-Study: Homicide among Non-Whites. 91

B.1 Optimal Minus Actual Hiring Rates for Men and Women With and Without Affirmative Action, for Different Perceived Levels of Ability, CRRA Risk Averse Employees . . . 108
B.2 Optimal Minus Actual Hiring Rates for Men and Women With and Without Affirmative Action, for Different Perceived Levels of Ability, CRRA Risk Averse Employees . . . 109
B.3 Expected Value of Each Type of Employee, Based on Actual Hiring Choices . . . 110
B.4 The Probability of a Hired or Not Hired Individual being Female, Across Tournament Type. . . . 111
B.5 Probability of Employee Being in the Top Half of Their Group, by Tournament Type and Whether They Were Hired or Not . . . 112
B.6 Theoretical Framework Diagram - Discrete . . . 113
B.7 Theoretical Framework Diagram - Continuous . . . 114
C.1 Design Illustration: Biweekly Intervention . . . 115
C.2 Self-Reported Sleep Duration at Baseline . . . 116
C.3 Self-Reported Sleep Duration at Baseline and Risk Taking Behavior . . . 117
C.4 Timing of Survey Response and Bedtime Target Choice . . . 118
E.1 Choice List for Risk Preference 1 . . . 144
E.2 Choice List for Risk Preference 2 . . . 145
E.3 Choice List for Time Preference 1 . . . 146
E.4 Choice List for Time Preference 2 . . . 147
E.5 Choice List for Time Preference 3 . . . 148
E.6 Choice List for Time Preference 4 . . . 149
F.1 Age-Adjusted Mortality Summary Statistics . . . 150
F.2 Age-Adjusted Mortality by Urban/Rural . . . 151
F.3 Event-Study: Suicide among Young Whites . . . 151
F.4 Event-Study: Homicide among Young Adult Non-Whites . . . 152
F.5 Event-Study: Homicide among Adult Non-Whites . . . 153
F.6 Event-Study: Homicide among Elderly Non-Whites . . . 154
F.7 Event-Study: Alcohol Related Mortality among Adult Non-Whites . . . 155

Preface

This dissertation would never have been completed without the support of my family, my partner Greg Gaumond, and the network of friends I have made in graduate school, including and not limited to Jessica LaVoice, Rachel Landsman, Sijie Li, Neeraja Gupta, Kelly Hyde, Marissa Lepper, Bea Ahumada, Tianyi Wang, Domi Vamossy, and many, many others. I would also have never made it this far without the help of my advisor, David Huffman, who never made me feel stupid for doing something wrong, but instead taught me how recognize that something was wrong and how to do it better the next time. I additionally would like to thank my entire dissertation committee and the faculty of the Department of Economics at the University of Pittsburgh for their advice and suggestions to make these papers the best I could make them. Finally, I would like to commemorate and thank Werner Troesken, former member of my committee and faculty of the Department of Economics at the University of Pittsburgh, whose enthusiasm for research, positivity, and supportiveness I hope to emulate for others in recollection of how much it helped me.

1.0 A Hidden Cost of Affirmative Action: Muddying Signals about Women's Ability

Large gender gaps persist at the higher ends of the income distribution, despite gains in female representation in early career stages. This paper uses an experiment to study whether affirmative action, which has been used mainly in early career stages, could have a hidden cost. Specifically, by decreasing the strength of the signal about a woman's ability when she is successful in early career competitive environments, affirmative action could thereby inhibit her future career progression. I find that women who are successful in an early competitive stage without affirmative action have an advantage over successful men when it comes to hiring in a second stage. The implementation of affirmative action removes this advantage, leaving successful men and women with equal hiring probabilities. I find empirical support for employer beliefs as a key underlying mechanism. A welfare analysis shows that affirmative action has an overall positive effect for women in this experiment, but the welfare improvement would be substantially larger if there were not the cost in terms of muddying signals about women's ability.

1.1 Introduction

Gender-based affirmative action policies are intended to increase the representation of women in careers, and at career echelons, that have traditionally been male dominated.[1] For the most part, affirmative action policies are implemented early in the career track, such as in college admissions or for early-career jobs.[2] Such policies make it more likely that women are successful in the competitive selection processes associated with early career stages. While the entry of women into such fields has increased in recent decades, there continues to be a so-called "leaky pipeline" for female talent, as women do not advance as quickly or as far up the career ladder as men (Gang et al., 2003; Bertrand et al., 2010; Niels-Jakob Harbo et al., 2019). As a result, women remain underrepresented in the highest paying jobs and at the upper ends of the career track (Guvenen et al., 2014; Bailey and DiPrete, 2016; Blau and Kahn, 2017; Barth et al., 2017).

This paper tests for a potential hidden cost of affirmative action at early career stages, which could contribute to this leaky pipeline.[3] The cost can arise due to muddying signals about female ability. Suppose that, in the absence of

[1] While the focus of this paper will be affirmative action for women, the conclusions here would likely extend to other affirmative action policies, such as those that benefit racial minorities.

[2] For a history of affirmative action in the United States, see Libertella et al. (2007). An important exception to affirmative action being implemented primarily early in the career track is the use of gender quotas in corporate boards of directors. See Seierstad and Opsahl (2011), Matsa and Miller (2013), and Smith (2018) for conclusions from the Norwegian context. While all of these papers show that the quotas were successful in drawing women into the corporate boards, Seierstad and Opsahl (2011) find that increasing the proportion of women in corporate boards did not substantially increase the proportion of women in the highest positions of power, the chair of the board, and Smith (2018) points out that implementing these quotas drew in female board members with less experience than their male counterparts. These findings from corporate boards, though not in the same early career context that is the focus of my paper, are in line with my findings.

[3] Other explanations for gender gaps later in the career track that have been investigated include differences in ability, discrimination, and differential rewarding/punishing of men and women (e.g., Landsman, 2018).

affirmative action, employers believe that only the highest ability women will enter into the competitions that govern early career access (e.g. getting into an Ivy League school, getting a competitive first job). In this case, observing a woman who enters the competition and is successful may be interpreted as a particularly strong signal of her high ability, and successful women may actually enjoy an advantage over similarly successful men in terms of hiring and promotion.[4] Implementing affirmative action could change this dynamic, leading employers to believe that the average ability of women who enter and win the competition is lower, thereby muddying the signal associated with success and potentially eliminating or even reversing the hiring advantages for women. While the literature has speculated that this might be a cost of affirmative action (Noon, 2010; Antonovics and Sander, 2013), this cost has remained "hidden" since it is difficult to verify the cost's existence with extant data given the way affirmative action policies have traditionally been implemented, as well as the hypothesized underlying mechanisms. The ideal data would involve exogenous variation in the implementation of affirmative action as well as measures of employer beliefs.[5]

In this paper, I use a series of laboratory experiments to test for such a cost, exploiting the ability to exogenously vary the presence of an affirmative action policy in a controlled experimental setting. Specifically, I run a two-part experiment. First, subjects in the lab act as "candidate employees" by choosing whether to enter a competition either with or without affirmative action in place. Second, subjects on Amazon Mechanical Turk[6] act as "employers" and decide whether to hire the candidate employee after finding out the employee's gender and whether they entered and won the competition. Affirmative action in this case makes it easier for women to win the competition by ensuring that at least one out of two winners of the competition is a woman. A real-world analog of this environment is admissions to a prestigious university that acts as a prerequisite for getting better, higher paying jobs after graduation. Affirmative action in admissions to the university makes it more likely that women will get admitted, conditional on applying. However, when employers see a woman with a diploma from that university, they then also know that she was admitted under an affirmative action admissions policy. If the cost hypothesized by this paper is in place, employers will be less likely to hire a female graduate of that university compared to if there was no affirmative action.

I show in a simple theoretical framework, which mimics the experimental setting, that female competition winners will be more likely to be hired than male competition winners in the absence of affirmative action if employers believe that only particularly high ability women choose to enter the competition. I also show that with the implementation

[4]This argument is similar to what is found in Bohren et al. (2019), who find that women are rated more poorly than men on a math-focused question-answering website when they don't have any reviews, but if the women have many high reviews, they are rated better than men with a similar number of high reviews. It is also in line with the findings that women, conditional on reaching a certain position and having the same characteristics as a man has a higher likelihood of promotion (e.g., Lewis, 1986; Williams and Ceci, 2015; Ayalew et al., 2018). This argument assumes that there are other differences between the men and women that may prevent women from being promoted, which is clearly not true; discrimination, harassment, and greater home production responsibilities are all additional aspects in women not rising through their careers despite similar previous success to men. The cost of affirmative action proposed here is in addition to all of these other costs faced by women in the labor market.

[5]Negative perceptions of affirmative action beneficiaries' ability by peers, employers, and beneficiaries themselves has been identified and studied in the field of psychology for decades ((Garcia et al., 1981; Heilman et al., 1987, 1990, 1993, 1996, 1997; Heilman, 1997; Heilman and Alcott, 2001; Unzueta et al., 2010). However, this literature does not use incentivized measures and does not consider how this poorer perception affects employment outcomes.

[6]Amazon Mechanical Turk, or MTurk, is an online work-sharing platform. MTurk is used to connect workers on MTurk with academic experiments and surveys.

of affirmative action, employers may no longer believe that only the highest ability women chose to enter the competition. If this happens, the hiring advantage may be taken away, giving rise to the hidden cost of affirmative action.[7] However, both the hiring advantage for successful women without affirmative action and the loss of this advantage with affirmative action depends on employers believing that only the most able women enter without affirmative action, and that on average less able women enter with affirmative action. Whether employers will have beliefs with this structure, and will change their hiring patterns in response to their beliefs, is an empirical question.

Analyzing the results of the experiment, I find evidence of the hypothesized cost of affirmative action. In the absence of affirmative action, employers are 13% more likely to hire women who are successful compared to successful men. When affirmative action is implemented, successful women lose this hiring advantage and are hired 13% less than women successful without affirmative action. Because the hiring rate of successful men does not change with affirmative action, successful men and women are hired at the same rate when affirmative action is in place. Additionally, I find support for employer beliefs being the mechanism of hiring changes for women with the implementation of affirmative action. Employers believe that women who enter and win in the absence of affirmative action are of a higher ability than male winners. With affirmative action, this difference goes away. Furthermore, employers' beliefs predict their hiring decisions.

I conduct a welfare analysis to establish to what extent the identified cost reduces the welfare benefits women receive from affirmative action in this experiment. I find that the probability of a woman being hired increases with affirmative action, even with this cost in place. However, using a counterfactual simulation I can address how large the increase in hiring would have been for women if affirmative action did not decrease the signaling value of success for women. I find that, if one could eliminate this channel of decreased signaling value, the increase in hiring that women experience when moving to affirmative action would be triple what is found with that channel in place. In the conclusion, I address potential policy implications of the findings.

This paper contributes to the extensive literature in economics and other disciplines about the potential benefits and costs of affirmative action. Both empirical and experimental evidence shows that affirmative action is effective in inducing female and minority candidates to enter competitive environments, and the experimental literature shows that this does not result in a loss of efficiency[8] (Leonard, 1984; Ayres and Brooks, 2004; Rothstein and Yoon, 2008;

[7] These beliefs are reasonable given the previous literature on this subject. Previous experimental literature has shown that women are less likely to enter competitions than men (e.g., Gneezy et al., 2003; Niederle and Vesterlund, 2007; Croson and Gneezy, 2009) and that this gender gap in willingness to compete can explain some of the real-world gender gaps in employment outcomes (Kleinjans, 2009; Buser et al., 2014; Flory et al., 2014; Morin, 2015). This gender gap in willingness to compete has, at least in part, been attributed to women having lower confidence about their ability than men, even when they have the same actual ability (e.g., Niederle and Vesterlund, 2007), especially in stereotypically male tasks (Barber and Odean, 2001; Huang and Kisgen, 2013; Coffman, 2014; Sarsons and Xu, 2015; Coffman et al., 2019a,b). Gender gaps in willingness to compete have also been attributed to differences in preferences for competition across genders (e.g., Gneezy et al., 2003; Niederle and Vesterlund, 2007; Croson and Gneezy, 2009). Implementing affirmative action induces women to enter the competition by making it more likely that they will win, conditional on entering (Balafoutas and Sutter, 2012; Calsamiglia et al., 2013; Niederle et al., 2013).

[8] Efficiency in this environment is measured by the most able people entering the competition, and the least able people not entering the competition, regardless of gender or other demographic characteristics. One concern about affirmative action is that it draws in lower ability women at the cost of higher ability men, decreasing the average ability of the pool of entrants. The experimental literature has shown that this hypothesized inefficiency does not occur, and efficiency often actually increases with affirmative action by drawing high ability women into the competition (e.g., Balafoutas and Sutter, 2012; Niederle et al., 2013).

Balafoutas and Sutter, 2012; Calsamiglia et al., 2013; Kurlaender and Grodsky, 2013; Niederle et al., 2013; Baltrunaite et al., 2014; Cotton et al., 2014; Ibañez and Riener, 2018; Bagde et al., 2016; Beaurain and Masclet, 2016; Sutter et al., 2016; Besley et al., 2017; Maggian et al., 2017; Czibor and Dominguez Martinez, 2019).[9] However, a growing literature has recently pointed out potential negative impacts of affirmative action, including retaliation against beneficiaries by other competitors (Fallucchi and Quercia, 2018; Brown and Chowdhury, 2017; Leibbrandt et al., 2017; Petters et al., 2017; Banerjee et al., 2018) and the activation of stereotype threat resulting in decreased performance of high ability women (Bracha et al., 2019). My paper contributes to this literature by pointing out a potential cost of affirmative action that is specific to affirmative action implemented at early career stages, the weakening of signals about ability for successful women in later employment environments. To my knowledge, this paper is the first empirical study on affirmative action to consider how employers interpret information about an individual gleaned from an initial competitive environment and translate that information into hiring decisions, and how affirmative action changes their interpretation of this information and their hiring decisions.[10]

1.2 Experimental Design

This experiment took place in two parts. The first part was run in the lab and was used to collect the employee information to be used in the second part; the second part was run on Amazon Mechanical Turk (MTurk), a workshare platform commonly used for academic studies, where employers' responses to the employees' information was collected.[11] Both employers and employees were exposed to one treatment, standard or affirmative action, to minimize cognitive load and to prevent experimenter demand effects as much as possible.

[9] Kleinjans (2009), Buser et al. (2014), and Flory et al. (2014) show that the decision to compete in experimental tournament settings is predictive of selecting into competitive programs and careers outside the lab.

[10] This is, in part, due to the lack of literature showing how affirmative action affects any later outcomes after the affirmative action has been removed. Maggian et al. (2017) shows that affirmative action in an earlier round of a tournament does not induce beneficiaries to enter later rounds of the tournament that are not subject to affirmative action. Dianat et al. (2018) considers an experimental gift exchange market with statistical discrimination and anonymous re-matching between employers and employees. Their study establishes that a period of affirmative action benefiting the disadvantaged group does not prevent discrimination from resuming after the affirmative action has been removed, and that this return to discriminatory behavior can be explained by sticky beliefs. On the other hand, Miller and Segal (2012) and Miller (2017) find that temporary racial affirmative action policies in police forces and companies with federal contracts, respectively, do lead to increased representation of Black workers after the policy is removed, with the latter attributing the continued gains to improved hiring practices. These papers address how temporary affirmative action policies affect discrimination in entire labor markets across time. Unlike these papers, my paper considers how employers, who are not constrained by affirmative action, interpret information about a specific individual employee's ability generated in earlier competitive environments that may or may not have had affirmative action in place. To my knowledge, the only paper that addresses a somewhat similar point is Beaman et al. (2009), which finds that people whose elected official is a woman due to a gender quota, and who have no previous experience with gender quotas, rate their female official lower than those who either do not have a gender quota or those with previous gender quota experience. Also, they find that this is not due to observable differences in the female officials and is in spite of decreased gender stereotyping as a result of experience with gender quotas. While they do not attribute the negative perception of gender quota beneficiaries to the same mechanism that I evaluate, their results are on the whole consistent with the conclusions of this paper.

[11] Both employee and employer data were collected in the lab sessions, but I will primarily consider the employee data. The employer data was a pilot for the later employer sessions on MTurk and was used to ensure subjects understood the employers' decision-making environment. The employer pilot data is not used because it became clear that the employer subjects did not understand the decision-making environment as presented in the lab instructions. The instructions were then changed to make the design and the incentives more clear, and comprehension questions were used to ensure that the MTurk subjects did understand the instructions before moving forward with the study.

1.2.1 Employee Design

The employee part of the study took place at the Pittsburgh Experimental Economics Lab at the University of Pittsburgh. 78 subjects were recruited using SONA and were undergraduate students at the University of Pittsburgh.[12] The employee experimental design is based off of those of Niederle and Vesterlund (2007) and Niederle et al. (2013), with the primary difference of making the design between-, rather than within-, subject. Subjects were brought into the lab and randomly put into groups of 6, each comprised of three men and three women. These groups were based on seating location, with clusters of computer stations defining the group. While subjects could see the other people in their group and determine their gender in that way, the instructions also repeatedly pointed out that the groups were comprised of three men and three women. This was done to ensure that gender was similarly salient across both the standard and affirmative action treatments, rather than in just the affirmative action treatment.[13]

Subjects participated in three rounds of a sums task (Niederle and Vesterlund, 2007; Niederle et al., 2013). This task asked subjects to do sums of 5 two-digit numbers for five minutes. The problems were presented on the screen one at a time, and subjects received a new problem after submitting the previous one. Subjects also had a tally of the number they got correct and incorrect at the bottom of the screen. This was the only feedback subjects got between rounds, and provided a rough indicator of their own ability. I used this task because, despite their underlying performance being equal, men have been shown to be more confident and competitive than women in this task (Niederle and Vesterlund, 2007). This makes it plausible that employers may have the belief that men and women are equally able in the task but that men are more confident in their ability on the task than women. It has also been shown to have external validity as a proxy for choosing competitive, prestigious career tracks (Buser et al., 2014).

In the first round of this task, subjects got a piece rate of \$0.50 for each correct answer. In the second round, subjects were entered into a tournament against their group members. In this round, subjects got \$1.50 per correct answer if they were a winner, and no money otherwise. Winning was determined by the treatment and their score relative to the scores of the other subjects in their group. In the standard treatment, the two group members with the most correct answers were the winners. In the affirmative action treatment, the woman with the most of correct answers among the other women in her group was one winner, and the other winner was the remaining group member (minus the highest performing woman) with the most correct answers. This ensured that in the affirmative action tournament at least one woman and at most one man would be the winner.

In the third round of the task, subjects made a choice between the piece rate and tournament payment schemes. If subjects picked the piece rate, they earned \$0.50 for each correct answer as in the first round. If subjects picked the tournament, they faced the same tournament structure as in the second round, in which winners earned \$1.50 per correct answer and non-winners earned nothing. However, in the third round, the tournament was against the performance of the subject's group members in the previous round; meaning, if a subject chose to enter the tournament

[12] SONA allows researchers to invite pre-screened participants for research studies. See Greiner and Stephanides (2019) for details.

[13] Gender salience could lead to experimenter demand effects or stereotype threat for female participants (Bracha et al., 2019), so holding it constant across treatments is important for understanding how affirmative action affects subjects' choices.

in the third round, whether they won or lost depended on their performance in the third round compared to their group members' performance in the second round. The tournament outcomes in the third round were based on the round two scores of the entrants' group members so the entry decision would be based on beliefs about relative ability and preferences over risk and competition, rather than beliefs about which and how many other people would choose to enter the tournament. In the standard treatment, a tournament entrant won if their round three score was larger than at least four of their group members' round two scores. In the affirmative action treatment, a female tournament entrant won if their round three score was higher than both of their female group members' round two scores OR if their round three score was higher than at least four of their group members' round two scores. A male tournament entrant in the affirmative action treatment won if their round three score was higher than both of their male group members' round two scores AND if their round three score was higher than at least four of their group members' round two scores.

Subjects completed a demographic survey about their gender and year in school that would later be used in the resumes that employers saw. They also completed a BDM risk elicitation (Becker et al., 1964). Subjects were then paid for one randomly chosen round over the entire study and one subject per session was randomly chosen to be paid for the risk elicitation.[14]

1.2.2 Employer Design

The 180 subjects for the employers' portion of the study were recruited through MTurk and the experiment was implemented via Qualtrics.[15] Subject knew they were being recruited for an academic decision-making study, and that they would get a $0.50 payment plus a bonus payment of up to $15 based on their choices in the study.

Subjects were randomly assigned to either the standard or affirmative action treatment. They first received an extensive explanation of the decision-making environment for the employees in the lab sessions. This included a description of the sums task and the rules for how winners were chosen in the tournament, which varied by treatment. The summary of both treatments included similar salience of gender, such as reminding them that all groups were comprised of three men and three women.[16] After reading the summary, subjects had to answer comprehension questions about how winners were selected in order to continue with the study. They had two opportunities to answer the questions correctly, and had access to the summary if they wished to refer to it.

[14]The lab session also included an employment round, in which subjects labeled as employers decided whether or not to hire the employee, in a decision very similar to that faced by the employer subjects on MTurk, and a belief elicitation of both employers and employees. Employers reported their belief about the employee's ability, also similar to the beliefs elicited from the employers on MTurk. Employees reported their belief that they would be hired by the in-lab employers. This information is not used because of the small sample size collected from the lab and the concerns about the subjects' understanding of the employers' decision-making environment.

[15]Subjects were asked to report their level of engagement with the survey, and 20 subjects who reported being un-engaged were removed from the analysis. Their inclusion attenuates the results found here, due to them mostly picking the same value for every resume. After removing those subjects who did not pay attention, 180 employers remained, with 10 hiring and beliefs observations per employer. Although there have been concerns about the validity of experimental results found from MTurk, multiple studies have directly tested the validity of MTurk responses and found that MTurk subjects respond similarly to treatments and pay the same amount of or more attention to the experiment as compared to in-lab subjects (e.g., Horton et al., 2011; Hauser and Schwarz, 2016; Thomas and Clifford, 2017).

[16]Again, the salience of gender may lead to experimenter demand effects, so keeping its presence constant across treatments makes it possible to disentangle the effect of affirmative action from the effect of having gender be salient.

After completing the comprehension questions, subjects received instructions for the rest of the study. Subjects would get ten employee resumes, one at a time. Each resume would contain the employee's gender, year in school, whether they entered the tournament or not in round three, and, conditional on entering the tournament, whether they won. Then, after seeing an employee's resume, the subject made two decisions about that employee: a hiring decision and a belief elicitation about the employee's relative ability.

In each hiring decision subjects received \$3 and they could use any portion of that \$3 to buy some probability of hiring the employee. That is, if the subject paid \$0, they would get a 0% probability of hiring the employee; if they paid \$1.50, they would get a 50% probability of hiring the employee; and so on. In addition to what remained of the initial payment of \$3, they would receive another \$6 if they hired the employee and the employee was in the top three in their group in terms of the number of correct answers in round two. If the employee wasn't hired or the employee was hired but wasn't in the top three in their group in round two, the subject got no additional money on top of the remainder of the initial \$3.

Note that the information in the resume was about the employee's decision and outcome in round three but the payment for the subject is based on the employee's performance in round two, the mandatory tournament. This models a hiring environment in which the information employers get is an imperfect reflection of employees' actual performance, and is contaminated with the employees' choices about whether or not to enter a competitive environment.

Also note that while the information is about entrance and winning, the payment for the employer based on whether the employee is in the top three within their group, which may include those who did not win. The goal of this design element was to make sure subjects did not just automatically choose winners and nobody else. It was important that subjects understood in both treatments that those who entered and won in round three were not necessarily the same as those who performed well in round two.

After the hiring decision, subjects reported their beliefs about whether the employee was in the top three performers in their group in round two. This was incentivized using the Binarized Scoring Rule (Hossain and Okui, 2013), though subjects were not informed as to how these payments were exactly calculated: they knew that they could make \$0 or \$6, and that they had the best probability of getting the higher payment if they reported their honest beliefs. Subjects were provided with an email to contact for exact details about the payment rules if they desired; nobody asked for this information.

Subjects made these two decisions for ten employees, and were paid for one randomly chosen decision out of those twenty decisions. They also completed a risk elicitation via BDM (Becker et al., 1964) between \$3 for sure or some probability of \$6 with a one in thirty probability of being paid for the risk elicitation, and a demographic survey including an age range, gender, labor status, the subjects' dependence on MTurk for income, and attention questions. Finally, subjects were asked to comment on their decision-making process in a free-form response.

1.3 Theoretical Framework

In this section I summarize a simple theoretical framework. The purpose of the framework is to illustrate, in an environment that mimics my experiment, conditions under which the hypothesized cost of early stage affirmative action could arise. The framework identifies employer beliefs as the key mechanism, and the cost only arises if beliefs take a particular form. The empirical question is thus whether the cost in fact arises, and whether this can be linked to the hypothesized underlying mechanism. While this section summarizes the key conclusions of the theoretical framework, the details of this framework are described in Appendix 4.1.

In this framework, employees have to decide whether to enter a competition. There is a cost to entering the competition, and winners of the competition get a bonus. High ability employees are more likely to win the tournament than low ability employees, and the parameters of the tournament are such that if someone knows with certainty that they are high ability they should enter, and if someone knows with certainty that they are low ability they should not enter. However, in this environment, employees do not necessarily know if they are high or low ability, but they have beliefs about their ability level, i.e. confidence. Thus, there will be some confidence level threshold above which employees enter the competition, and below which they don't.

Employers have to decide whether or not to hire an employee. There is a cost to hiring the employee, and the employer gets a bonus only when they hire a high ability employee. Suppose the parameters are such that employers want to hire high ability employees, and not hire low ability employees. However, they do not know the employee's ability. What they do know is the employee's gender and whether the employee entered and won in the competition or not.

Consider a particular form of employer beliefs. Specifically, employers believe that male and female employees are equally likely to be high ability, but they also believe that men are more confident about their ability than women for a given true ability level.[17] In this case, for a given true ability level, men will be more likely to believe they are high ability than women. Thus, the confidence threshold for entering the tournament for men will happen at a lower actual ability level than for women. This means that employers believe the average ability of women who enter and win the tournament will be higher than for men who enter and win the tournament, and they will be more likely to hire a successful woman than a successful man for that reason.

Suppose, further, that affirmative action is put into place in the employees' competition. Specifically, a woman who enters becomes more likely to win the competition, and a man who enters becomes less likely to win the competition. Still, high ability employees are more likely to win the competition than low ability employees. This results in two different confidence thresholds, one for men and one for women, such that the confidence threshold for women is lower than that for men, and lower than it originally was before affirmative action.

With affirmative action, employers will believe that women of a lower average ability will enter and win the

[17] While this is modeled by men and women having different priors about their ability, it could also be modeled with men and women having the same priors but different updating rules when faced with information about their underlying ability (Mobius et al., 2011).

competition compared to without affirmative action, because they do not have to be as confident to enter. This decrease in belief about women's ability then results in lower hiring rates for women who enter and win the competition, compared to without affirmative action, giving rise to the hidden cost of affirmative action. Depending on how much the confidence threshold drops for women under affirmative action, as well as how much it increases for men, it may be the case that the actual average ability level of men and women who enter and win is the same, resulting in a total loss of the hiring advantage successful women had without affirmative action. With a large enough gap between the confidence threshold of men and women, there could be a reversal of the hiring advantage, with successful men getting hired more than successful women under affirmative action.

Previous empirical evidence suggests that the beliefs needed to give rise to the cost of affirmative action could be reasonable beliefs, although it is far from obvious that subjects in the role of employers will have such beliefs. Women are generally less confident than men, particularly in male-dominated settings, and are less likely to enter competitive environments (Barber and Odean, 2001; Gneezy et al., 2003; Niederle and Vesterlund, 2007; Croson and Gneezy, 2009; Huang and Kisgen, 2013). The task used in this paper is specifically one in which women have been shown to be less confident and less competitive despite equal performance (e.g Niederle and Vesterlund, 2007). The empirical evidence on the effect of affirmative action on the average ability of female (male) entrants is more mixed, with some evidence suggesting that the average ability of female (male) entrants increases (decreases), with other evidence suggesting it stays the same (e.g., Balafoutas and Sutter, 2012; Niederle et al., 2013). As such, it is an empirical question as to whether employers will have the beliefs required to generate this cost, and whether employer will translate these beliefs into the hiring cost hypothesized here.

1.4 Results

This section presents the results of the experiment. The focus of the analysis is on the results of the employer sessions, although the employees' characteristics and their choices are summarized in Tables 1.1 and 1.2.[18]

Employers were drawn from the MTurk worker population living in the United States. Other than the restriction that they must be 18 years of age or older, there were no restrictions on who could participate. The employer characteristics are described in Table 1.3. Employer characteristics are balanced across the two treatments except for age,

[18] For the most part, characteristics of the subjects across treatments are balanced and the choices of employee subjects follow with what would be expected from the previous literature, particularly for women - moving from an environment without affirmative action to an environment with it, women are more likely to enter and are more likely to win conditional on entering. Women who enter without affirmative action are more likely to be in the top half of their group in round two compared to men who enter, and that probability decreases with affirmative action. For men, we see that the move to affirmative action decreases the likelihood of winning and increases the likelihood of losing, but that it also *increases* the likelihood of entering. While possibly an artifact of the small sample from which these results were simulated, the increase in entry for men in the affirmative action treatment compared to the standard treatment could also be an important consideration in terms of the effect of affirmative action. Free-form responses to the question "why did you decide to enter or not enter the tournament" indicate that some men in the affirmative action saw the longer odds of winning the tournament as a challenge to overcome, leading to tournament entry. Analysis of the employee data, though not the focus of this exercise, revealed that the increase in tournament entry for men in the affirmative action tournament is primarily driven by those men with lower scores on the initial two rounds of the tournament. Results available upon request.

in which those who answered the standard tournament survey were approximately 3.6 years younger than those who answered the affirmative action survey.[19] Controlling for this age difference, however, does not affect the results found in this experiment.

Note that, moving forward, I will refer to the tournaments as the standard tournament and the affirmative action tournament. The standard tournament had the standard, gender neutral rule for determining winners, whereas the affirmative action tournament required that at least one winner was a woman. Also note that the experiment has an extra category for tournament outcomes compared to the theoretical framework, in that employer subjects observe whether a candidate employee entered the competition, regardless of whether they won or lost (the latter group being, e.g., people who applied to Harvard but were not admitted). In reality, employers do not generally have such information on losers, but the experiment allows me to observe how employers might condition hiring on entering but losing a competition. I will refer to those who entered and won the tournament as winners and those who entered and lost the tournament as losers. Those who did not enter the tournament will be referred to as non-entrants or as DNEs.

The results are split into three components: hiring outcomes; an analysis of employer beliefs as the mechanism; and welfare and counterfactual analysis.

1.4.1 Hiring Outcomes

Result 1: In the standard tournament, women who win are hired more than men who win.

Figure 1.1 demonstrates this result. When considering the resumes of male and female winners of the standard tournament, employers hire female winners 7.9 percentage points (13%, Wilcoxon rank-sum test p=0.05) more than male winners.[20] Additionally, from this figure we can see that losers and non-entrants of both genders in the standard tournament are hired less than those who enter and win.

Column 1 of Table 1.4 shows support for these conclusions using regressions which include a triple interaction of employee gender, tournament outcome, and tournament type, as well as controls for year in school (the only other information about the employee the employers saw on the resume). Controlling for the year in school of the employee and clustering standard errors at the employer level does not remove the hiring boost female winners get over male winners in the standard tournament.[21]

Result 2: Women who win the affirmative action tournament are hired at a lower rate than women who win the standard tournament.

Figure 1.1 and Column 1 of Table 1.4 also present this result. Women who win the affirmative action tournament are hired 8.9 percentage points (13%, p=0.01) less than women who win the standard tournament. This lowers the hiring rate of female winners down to the level of male winners. Women who win with affirmative action are only hired

[19] This difference is, to some extent, an artifact of how age was elicited in the survey. Subjects were asked to select an age range, rather than an exact age. To calculate the average age for each group, I used the middle of the selected age range as the subjects age. This artificially places more weight on later age groups with larger age ranges.

[20] From now on, all comparisons of means are from Wilcoxon Rank-Sum tests unless otherwise noted.

[21] These results also hold for Tobit regressions.

1.2 percentage points less than men who win with affirmative action, a small and statistically insignificant different difference (p=0.85). Men who win either with or without affirmative action are hired at almost exactly the same rate.

1.4.2 Analysis of Employer Beliefs as the Mechanism

Result 3: Differences in employer beliefs across the standard and affirmative action tournaments help explain the reduced hiring of female winners under affirmative action.

Figure 1.2 shows the average beliefs employers have about employees of every gender (men and women), tournament type (with or without affirmative action), and tournament outcome (winners, losers, and non-entrants). Here, we see that, without affirmative action, women who win the tournament are believed to be of higher ability than men who win, though these differences are smaller and less precisely estimated than those of the hiring decisions. Women who win the standard tournament are believed to be in the top half of their group 4.6 percentage points (6%, p=0.13) more than men who win the standard tournament. With affirmative action, female winners are believed to be in the top half 2.4 percentage points less (3%, p=0.38) than those women who won the standard tournament. Column 2 of Table 1.4 presents these results in regression format with controls for grade and clustering of standard errors at the employer level. Adding controls increases the precision of the estimates, and within this framework the increased beliefs about female winners' ability without affirmative action and the decreased beliefs about female winners' ability with affirmative action are significant at a 5% level.

Employers were asked to write a free-form response about their decision making process in regards to hiring. Answers were coded in terms of whether gender was specifically mention as being used in making the hiring decision and, where applicable, which gender was preferred by the employer. Only 31 employers stated clearly how they used gender in their decision making process, leading to a small sample. However, those employers in the treatment without affirmative action were 21.8 percentage points (108.3%, p=0.32) more likely to express a preference for hiring women over men compared to those employers in the affirmative action treatment. Some of these comments explicitly make the argument supporting the mechanism proposed here – that women who win are more likely to be of high ability than men who win, since women are typically less confident and competitive and are more likely to not enter the tournament even when they would likely be successful in it.

In additional analysis I explore whether employers' beliefs entirely explain the differences in hiring with and without affirmative action. Column 3 of Table 1.4 replicates column 1, but with a control for the employers' prediction about that employees' ability. Adding beliefs halves the magnitude of the hiring differences for female winners with and without affirmative action, and only remains significant for the female winners without affirmative action. This indicates that, while the belief measure does not explain all of the hiring differences in terms of magnitudes, the relationship is substantial in size. The fact that beliefs do not explain everything could reflect factors such as noise in the belief measure, or a role for other traits to influence hiring decisions, such as risk preferences.[22]

[22] Using the risk elicitation results and a CRRA utility function, I assigned each subject a coefficient of relative risk aversion. When I run the

1.4.3 Welfare and Counterfactual

To address welfare considerations, both for the employers and the employees, I use the employee data to understand how the probability of a high ability employee entering and winning the tournament is affected by gender and treatment. However, there is a small number of employee observations, which could result in skewed outcomes given that the outcomes of the small sample may not be a good approximation of the average behavior. Specifically, for each individual I know how they performed in the task compared to the individuals in their own group, but not how they well they would have performed compared to the average group. To address this issue, I run a simulation that gives each employee 10,000 random groups, and in each group finds that employee's probability of being in the top half of their group in round 2, whether they entered the tournament in that group, and whether they won that tournament conditional on entering.[23]

Result 4: Welfare for employers is not significantly changed by the implementation of affirmative action, though how much revenue employers can make from male vs. female employees does change.

To analyze how employers' welfare would be affected by the treatment, average probabilities of being in the top half of the group in round two, of entering round 3, and of winning round 3 conditional on entering were generated for each of the gender-tournament outcome-tournament type (with or without affirmative action) groups, as well as a count of how many subjects' iterations fell within that category. This provides an approximation of what employers should have expected from the employees. Figure 1.3 presents the likelihood of an employer seeing an individual of each gender in each tournament outcome category, across tournament types (i.e. the fractions sum to 1 within tournament type).

Figure 1.4 shows, conditional on seeing an employee of a certain gender, tournament outcome, and tournament type, the highest amount of revenue that could be made on that employee given optimal hiring practices.[24] Overall,

regression in Column 3 of Table 1.4 interacting the prediction with the coefficient of relative risk aversion, there is not a substantial change in the coefficient for the interaction of female and entering and winning (0.0517, p=0.099) and the coefficient for the interaction of female, entering and winning, and affirmative action (0.0423, p=0.344). Additionally, when using the CRRA coefficient and the employers beliefs, I can find the optimal hiring choice employers should have chosen by finding the decision (x^*) that solves

$$\underset{x\in[0,3]}{\text{argmax}} E[U(x)] = \frac{(\frac{x}{3} * p)(9-x)^(1-\rho)}{1-\rho} + \frac{(1-\frac{x}{3} * p)(3-x)^(1-\rho)}{1-\rho}$$

where p is the probability the employer beliefs the employee is in the top half of their group in round 2. In this case, I still find excess hiring of women believed to be high ability (probability of being high ability greater than 50%) in the standard competition compared to men believed to be high ability in both competition environments and women believed to be high ability in the affirmative action competition (see Figure B.1). When assuming employees are risk neutral, there is no excess hiring of women who are believed to be high ability in the standard competition compared to men believed to be high ability in both competition environments and women believed to be high ability in the affirmative action competition (see Figure B.2).

[23]For each employee I randomly assigned them a group formed from the other individuals in their treatment in such a way that the group would have three men and three women. I then found whether they were in the top half of their new group with their round 2 score and whether they would have won in round 3 (depending on the rules of their treatment). This iteration was done 10,000 times for each employee subject. If an employee chose to not enter the tournament in the experiment, it was assumed that they would not enter the tournament in any of these artificial groups, since the information they had in the experiment was the same as it would have been in any of the other potential groups.

[24]Optimal hiring practices are such that the employer hires the individual if their category's probability of being in the top half of their group is greater than or equal to one half, and does not hire them otherwise. These results betray another anomaly of the employee data – the employees that did not enter the tournament, particularly the male employees, were more likely to be high ability than the employees that entered and lost the tournament. This is in line with the evidence that ability and overconfidence are inversely related, with those who are high ability being more likely

there is not a substantial change in expected employer payment overall given optimal hiring. Thus, if an employer was to see employees with the underlying probability generated by the simulation, and hired optimally given the probability that an individual in that category was in the top half of their group, an employer in the treatment without affirmative action would on average earn \$3.43, whereas an employer in the treatment with affirmative action would on average earn \$3.50, an increase of only 2%. Looking to actual hiring practices, there is little difference in the conclusion for employee welfare: an employer in the treatment without affirmative action would on average earn \$3.14, whereas an employer in the treatment with affirmative action would on average earn \$3.13, a 0.3% decrease (see Figure B.3). In summary, employer welfare does not substantially change with the implementation of affirmative action in an earlier, information-producing stage, either under optimal or actual hiring practices.

Result 5: Welfare, measured as the probability of being hired, increases for female employees and decreases for male employees, with the implementation of affirmative action, leading to more equal gender representation among those who are hired.

To determine the welfare effects of affirmative action on employees, a similar simulation was run as in the last section for employer welfare, but this time the average hiring practice for each gender-tournament outcome-tournament type category was applied to each simulated individual. Then, for each simulated individual, they were marked as hired or not hired with the probability assigned by the average hiring rates found in the experiment. While employees were not incentivized to be hired in this experiment, for this analysis I will assume that being hired is positive and that a change in welfare is equivalent to a change in hiring rates.[25]

Figure 1.5 presents the probability of being hired for men and women across tournament types. Before affirmative action, men are 7.8 percentage points more likely to be hired than women. With affirmative action, this gap shrinks to 0.2 percentage points, resulting from a 6 percentage point drop in hiring for men and a 1.5 percentage point increase in hiring for women. This results in a much more equal distribution of gender in both the hired and not hired groups (see Figure B.4).[26]

Result 6: Affirmative action does not change efficiency, measured as the average ability of those who are hired. However, it does decrease the average ability of hired women and increase the average ability of hired men.

One concern about affirmative action is that it could result in a lower ability level of entrants and winners, which could then lead to a lower average ability of those being hired. In this sample, this does not seem to be the case, as affirmative action does not change the distribution of ability across both hired and not hired individuals (see Figure B.5). However, Figure 1.6 shows that, while in all cases the population of individuals that are hired are more likely

---

to be underconfident (and not enter) and those who are low ability being overconfident (and thus entering and losing) (e.g., Kruger and Dunning, 1999).

[25]Using hiring as a proxy for welfare in this environment extracts away from other factors, like the match of the individual to the job and the balance between entering and experiencing the tournament with the likelihood of getting a job, that may moderate the relationship between hiring and welfare.

[26]Given the equal proportion of men and women in our experimental population, this leads to a decrease in overall hiring rates of 2.2 percentage points. While it is unlikely that affirmative action would actually lead to an overall decrease in hiring outside the experimental setting, it may result in longer searches and more unfilled vacancies, as employers are less confident about their candidates' abilities.

to be high ability than the population of individuals that are not hired, the average ability level of women who are hired decreases, and the average ability level of men who are hired increases, with affirmative action. This would indicate that, among those people who are hired, the men will become on average more able than the women with the implementation of affirmative action.This could potentially generate or promote negative beliefs about women's relative ability compared to men, a consequence that could then reinforce statistical discrimination. Also, Figure 1.7 shows that it is not only high ability women (low ability men) becoming more (less) likely to be hired with affirmative action, but rather that all women (men) become more (less) likely to be hired regardless of ability level. Thus, there are people gaining from affirmative action that should not be, specifically low ability women, and people losing from affirmative action that should not be, specifically high ability men.

Result 7: If the signaling channel of being successful in the competitive environment were turned off, the effect of affirmative action on the probability a woman is hired would be tripled.

While I do find a net positive effect of affirmative action on hiring probability for women in this experiment, the initial results on the decreased hiring for successful women indicates that there is a hidden cost of affirmative action in this environment, specifically the loss of signaling for successful women about their ability. To quantify the magnitude of this cost, I provide a counterfactual showing what would happen if one could turn off the decrease in signal strength for successful women that happens with affirmative action without eliminating the increase in tournament entry generated by affirmative action. Specifically, I use the simulated distributions of tournament entry and success in the tournament used in the welfare calculations, and for each individuals I determine their probability of being employed both in their own treatment and in the other treatment. Thus, I can disentangle the positive effect of increased tournament entry from the negative effect of decreased hiring of successful women.

Figure 1.8 presents these counterfactuals, along with the actual standard and affirmative action outcomes found in the experiment (ST-ST and AA-AA, respectively). We can see that using the hiring probabilities of the standard treatment and the entry decisions from the affirmative action treatment increase women's hiring probability. To find the magnitude of the cost of affirmative action on hiring in this experiment, I consider the increase in hiring generated from moving from the standard treatment (ST-ST) to the counterfactual using the entry decisions of the affirmative action treatment and the hiring probabilities from the standard treatment (AA-ST), and compare that to the increase in hiring when going from the standard treatment (ST-ST) to the affirmative action treatment (AA-AA). The increase when moving from the standard treatment (ST-ST) to the counterfactual (AA-ST) is three times as large as when moving from the standard treatment to the affirmative action treatment (AA-AA). Thus, the cost of affirmative action in terms of muddying the signals about women's ability attenuates the benefit of affirmative action to a quarter of what it could be without that cost. Table 1.5 presents the breakdown of the counterfactual changes in hiring rates for women.[27]

[27] Alternatively, the decrease in expected hiring rates when going from the standard treatment (ST-ST) to a counterfactual with tournament entry from the standard tournament and hiring decisions from the affirmative action tournament (ST-AA) would emulate a situation in which a company or organization claims that it engages in affirmative action benefiting women, women do not believe this claim and thus do not respond with higher entry, but employers do believe it and respond with decreased hiring of those women that are still successful.

1.5 Discussion and Conclusion

Affirmative action has been shown to successfully draw beneficiaries into environments that they otherwise would not have entered. However, these affirmative action policies are typically implemented early in careers, primarily in undergraduate admissions and in initial hires into a company or division. In these types of situations, there may be a hidden cost to affirmative action, as it could weaken the signal associated with being successful in competitive environment for beneficiaries, making it less likely the beneficiary is then hired or promoted. In this paper, I analyze this concern using an experiment, and find that this is the case: women who are successful without affirmative action are hired 13% more than successful men, and this hiring advantage is taken away with the implementation of affirmative action. I also provide evidence that employer beliefs about employee ability is the driving mechanism behind these hiring changes, as employers also believe that women who are successful without affirmative action are more likely to be of high ability than men who are successful with or without affirmative action, or women who are successful with affirmative action.

Despite the cost I identify in this experiment, I find that women still have a net benefit from affirmative action, as enough women are drawn into the competitive environment to outweigh the decrease in hiring probability for successful women. However, if one were able to turn off the signaling channel of affirmative action decreasing employers' beliefs about successful women, the increase in welfare would have been quadruple what is found with the signaling channel in place. When trying to understand why the pipeline for female talent is leaky, such that improvements in female representation early in the career track have not filtered through to representation in later, more prestigious positions, the hidden cost of affirmative action that I have identified could be important.

The findings of this paper have potential policy implications for when and how to best implement affirmative action. The hidden cost of affirmative action tends to reduce the welfare benefits for women, ceteris paribus. Thus, when one considers implementing an affirmative action policy, one should consider not only the immediate effects of the policy but also the potential downstream effects later in the beneficiaries' careers. These results also raise the possibility that affirmative action should be continued throughout the entire career track in an attempt to prevent such hidden costs from harming beneficiaries when the affirmative action is discontinued.[28] Additionally, policies and technological advancements that increase the availability and reliance on quantitative information in hiring and promotion decisions, rather than subjective decision making that may be influenced by candidate gender, could substantially reduce the opportunity for gender to be a focal aspect of hiring and promotion decisions, decreasing the potential for affirmative action to then have a negative impact on future employment outcomes.

Alternatively, with advancements in technology and information availability, one could imagine a world in which the entire structure of hiring and promotions could be improved. Affirmative action, as it is currently constructed, is

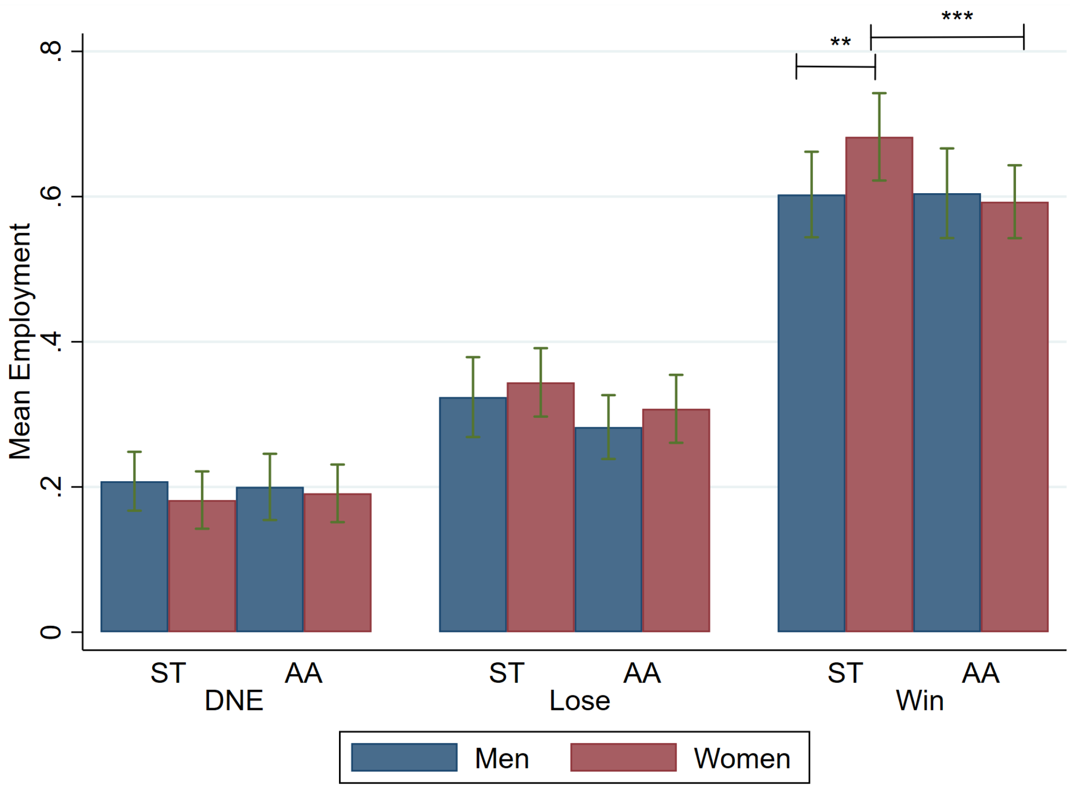
[28]However, as positions become more sparse, for example in the environment where there is only one of a particular position in a company, equal representation may be difficult to ensure and these hidden costs may then come back into play, though modifications can possibly be made such as ensuring that the gender of the person hired into that singular position changes every time the position needs to be filled.

effective at least in part because it encourages women to apply to positions or promotions to which they otherwise would not have been confident or competitive enough to apply for. An alternative to affirmative action as it is currently formulated could be a policy that moves away from voluntary entry into competition and towards a more meritocratic method of opting qualified candidates into a competition unless they choose not to compete.[29] In a world with more detailed and readily available information, employers may be able to identify suitable candidates without relying on an application structure that requires women to put themselves forward, obviating the issue of lower confidence and competitiveness leading to lower rates of entry or application. A system of assuming everyone within an organization applies to a certain position and algorithmically determining a short-list of candidates based on the requirements of the job and the employees' data could give women the encouragement they need to then continue with the application process once they have been short-listed, thus improving representation without the current affirmative action policy that carries with it a hidden cost.

[29] He et al. (2019) finds this to be a more gender-neutral competition structure; while a traditional opt-in structure leads to significantly more men than women entering the tournament, an opt-out structure leads to equal tournament entry for men and women. Additionally, Erkal et al. (2019) finds that an opt-out structure decreases the gender gap in willingness to be considered for a leadership role that occurs with a traditional opt-in structure.

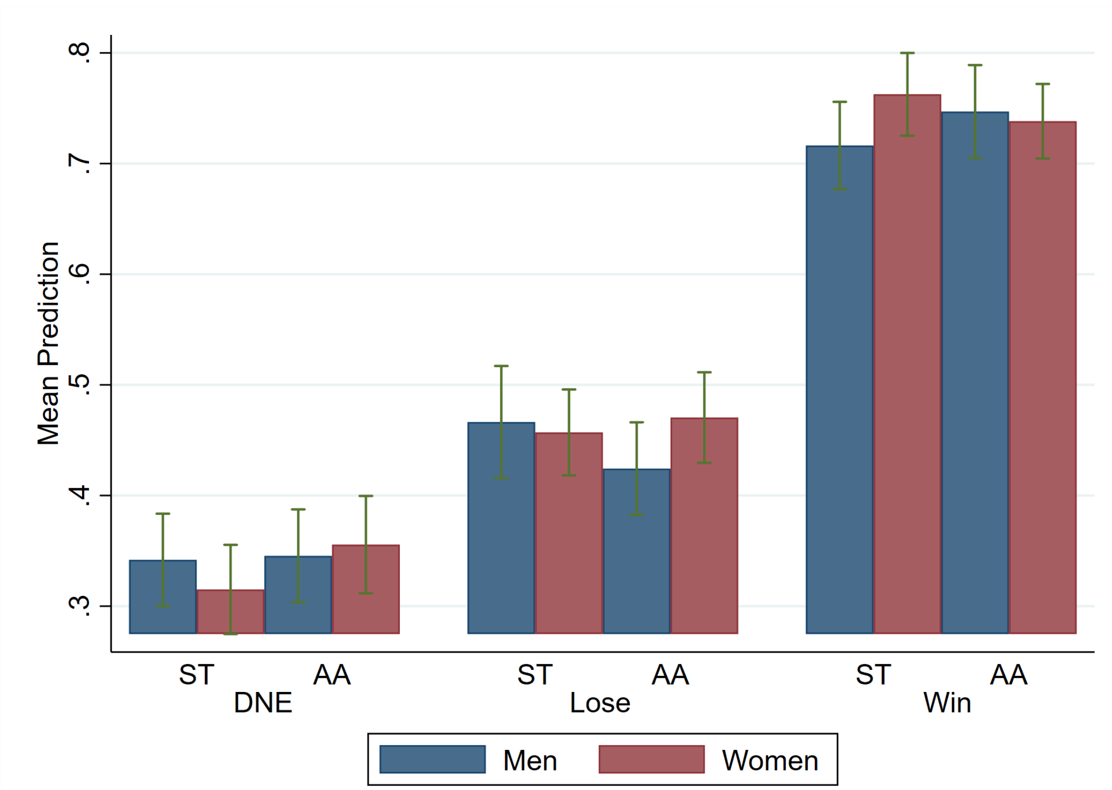
1.6 Figures

Figure 1.1: Employment by Employee Gender, Tournament Type, and Round Three Outcome



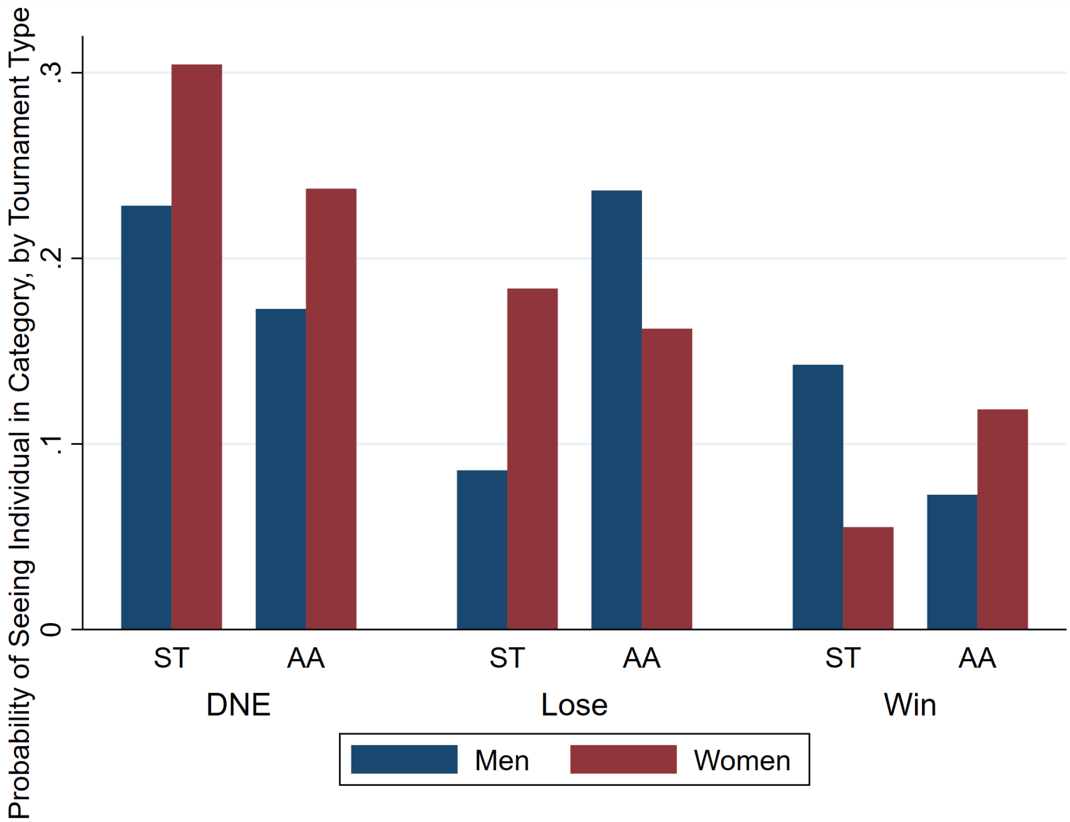
Notes - *** p<0.01, ** p<0.05, * p<0.1 based on Wilcoxon Rank-Sum Test. This figure presents the mean probability an employee was hired given their gender, treatment, and tournament outcome. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament; DNE=did not enter.

Figure 1.2: Employer Beliefs by Employee Gender, Tournament Type, and Round Three Outcome



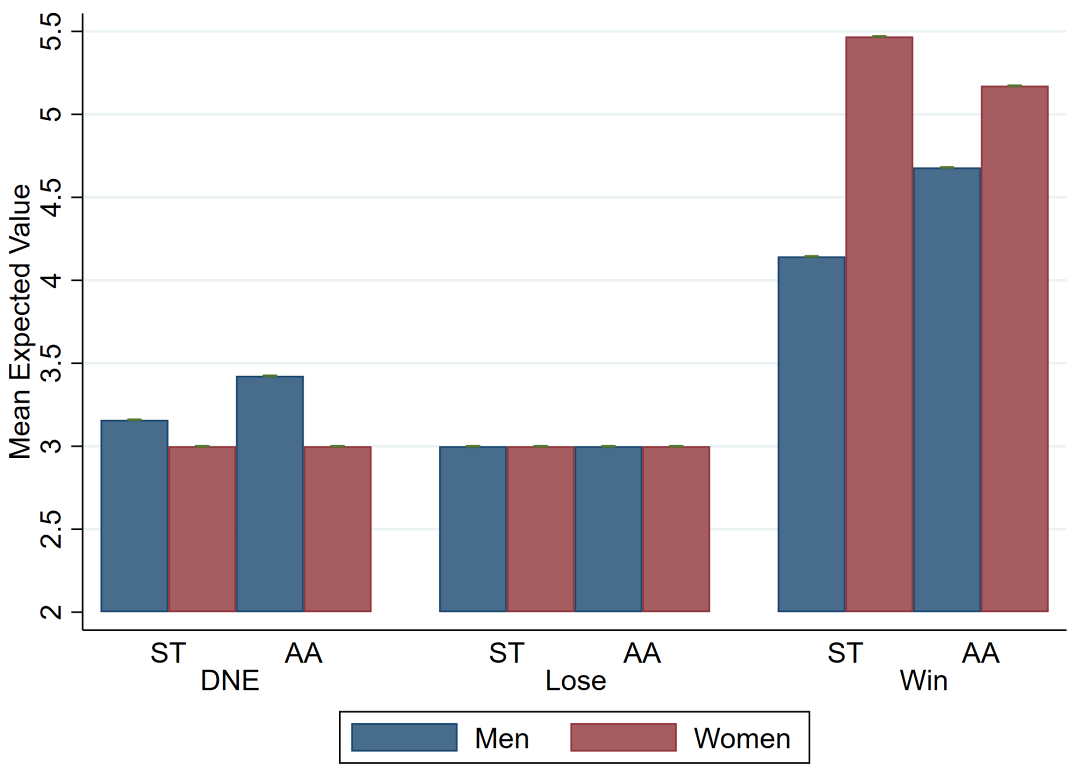
Notes - *** p<0.01, ** p<0.05, * p<0.1 based on Wilcoxon Rank-Sum Test. This figure presents the mean prediction about an employee's probability of being in the top three performers in their group in round 2 given their gender, treatment, and tournament outcome. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament; DNE=did not enter.

Figure 1.3: Simulated Probability of Employer Seeing Employee in Particular Gender – Tournament Outcome Category, by Tournament Type



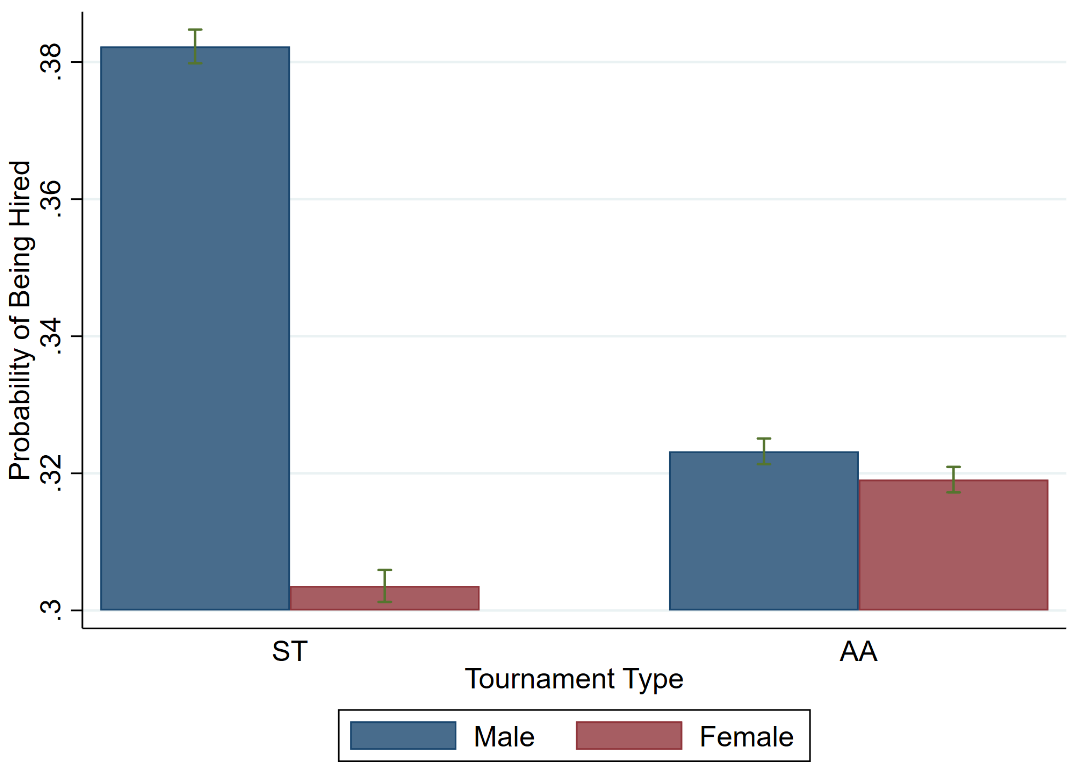
Notes - This figure provides the probability that an employee fell into a particular gender, tournament type, and tournament outcome category based on a simulation with 10,000 iterations for each employee. See text for more details. ST=Standard tournament; AA= affirmative action tournament; DNE=did not enter.

Figure 1.4: Expected Value of Each Type of Employee, Based on Optimal Hiring Choices



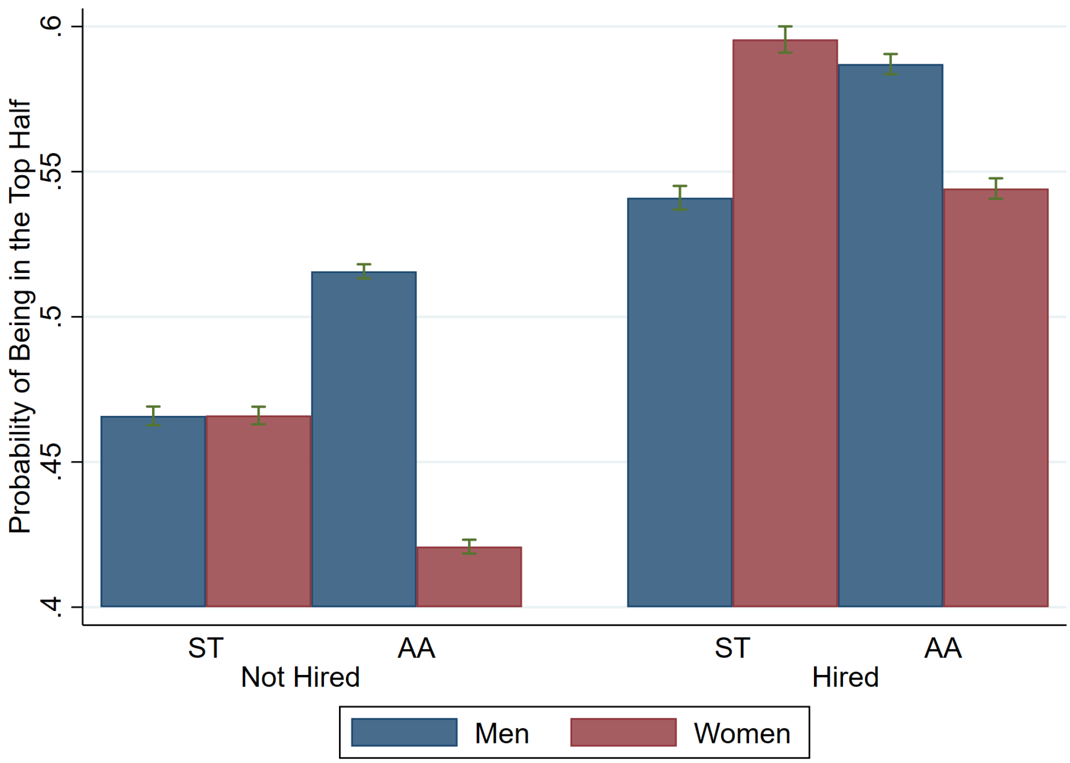
Notes - This figure presents the expected value for an employer when facing an employee of a particular gender, tournament type, and tournament outcome category, if the employer was to hire optimally. Hiring optimally is defined as hiring employees from categories with more than 50% of entries being in the top half of their group in round 2, and not hiring employees from categories with less than 50% of entries being in the top half of their group in round 2. ST=Standard tournament; AA= affirmative action tournament; DNE=did not enter.

Figure 1.5: Probability of being Hired across Gender and Tournament Type



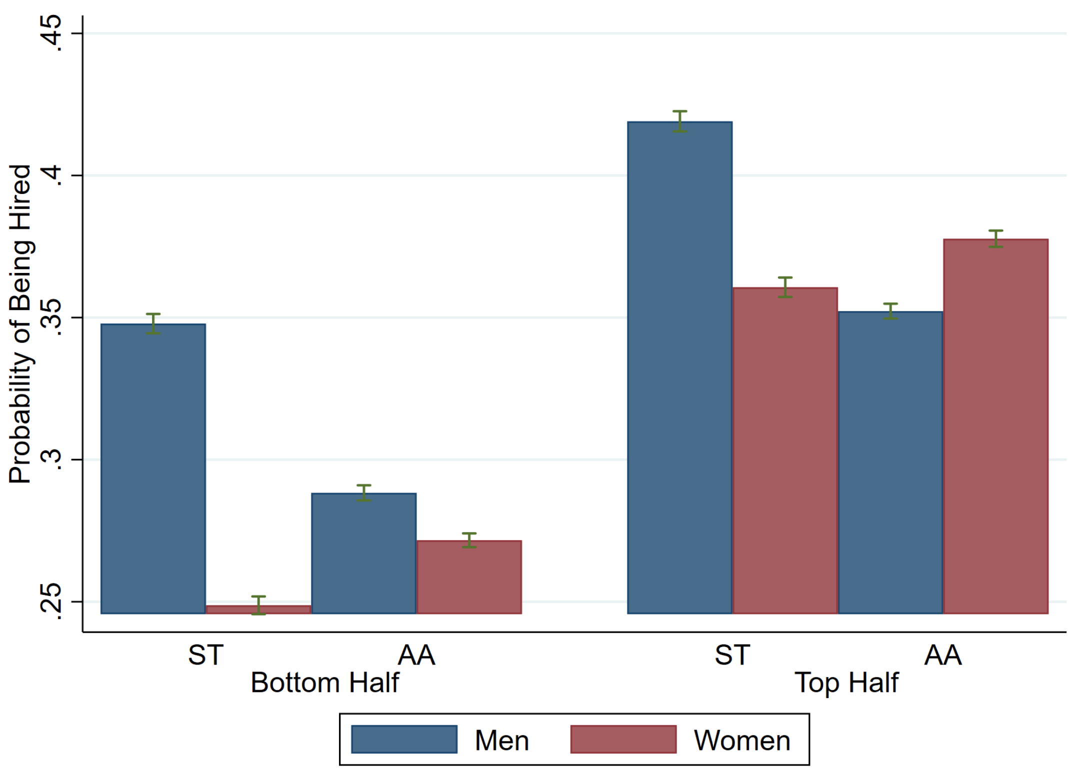
Notes - This figure presents the probability that an employee is hired across gender and tournament type from a simulation of each employee 10,000 times. See text for details about the simulation. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament.

Figure 1.6: Probability an Individual is in the Top Half of Their Group, by Gender, Tournament Type, and Whether or not They Were Hired



Notes - This figure presents the probability that an employee is in the top half of their group across gender, tournament type, and whether they are hired from a simulation of each employee 10,000 times. See text for details about the simulation. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament.

Figure 1.7: Hiring Rates for Men and Women in the Top and Bottom Half of Their Group, by Tournament Type



Notes - This figure presents the probability that an employee is hired across gender, tournament type, and whether they are in the top half of their group in terms of performance in round 2 from a simulation of each employee 10,000 times. See text for details about the simulation. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament.

Figure 1.8: Counterfactual Hiring Rates of Women Under Different Tournament Entry Choices and Hiring Probabilities

Notes - This figure presents the probability that a female employee is hired based on the entry decisions of a particular treatment (first label) and the hiring probability of a particular treatment (second label). Thus, ST-AA shows the hiring probability of women when they enter the tournament and are successful at the rates indicated by the simulation for the standard treatment, while they are hired with the probabilities found in the affirmative action treatment. See text for details about the simulation. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament.

1.7 Tables

Table 1.1: Employee Characteristics

Variable	ST	AA	Diff(ST-AA)
Female	0.50	0.50	0.00
	(0.09)	(0.07)	(0.12)
Age	20.03	20.15	0.15
	(0.25)	(0.24)	(0.37)
% Nat. Sci.	0.33	0.48	-0.15
Major	(0.09)	(0.07)	(0.12)
% Soc. Sci.	0.37	0.38	-0.01
Major	(0.09)	(0.07)	(0.11)
Year in	2.67	2.44	0.23
School	(0.23)	(0.18)	(0.29)
Risk	10.50	9.48	1.02
	(0.56)	(0.44)	(0.71)
PR Number	7.73	8.10	-0.37
Correct	(0.52)	(0.44)	(0.69)
Tournament	0.53	0.60	-0.07
Entry	(0.09)	(0.07)	(0.12)
N	30	48	78

Notes - *** $p<0.01$, ** $p<0.05$, * $p<0.1$ This table presents summary statistics for employees across tournament type. Year in school ranges from 0 for freshman to 3 for senior and above. Risk ranges from 1 for most risk loving to 15 for most risk averse (8=risk neutral).

Table 1.2: Employee Choices

	ST		AA		Diff(ST-AA)		Diff(M-W)	
	Men	Women	Men	Women	Men	Women	ST	AA
PR Num.	8.13	7.33	8.33	7.88	-0.20	-0.54	0.80	0.46
Correct	(0.72)	(0.77)	(0.68)	(0.56)	(1.03)	(0.04)	(1.05)	(0.88)
Tourn. Num.	9.20	9.53	10.58	9.58	-1.38	-0.05	-0.33	1.00
Correct	(0.76)	(0.71)	(0.62)	(0.68)	(0.98)	(1.03)	(1.04)	(0.92)
Tourn. Top	0.53	0.46	0.58	0.42	-0.05	0.05	0.07	0.17
Half	(0.13)	(0.13)	(0.10)	(0.10)	(0.17)	(0.17)	(0.19)	(0.15)
Tourn. Entry	0.60	0.47	0.66	0.54	-0.07	-0.08	0.13	0.13
	(0.13)	(0.13)	(0.10)	(0.10)	(0.16)	(0.17)	(0.19)	(0.14)
Tourn. Num.	9.00	10.43	10.63	9.62	-1.63	0.81	-1.43	1.01
Correct|Entry	(1.21)	(1.00)	(0.81)	(1.20)	(1.41)	(1.80)	(1.64)	(1.40)
Tourn. Top	0.44	0.57	0.56	0.31	-0.12	0.26	-0.13	0.25
Half|Entry	(0.18)	(0.20)	(0.13)	(0.13)	(0.22)	(0.23)	(0.27)	(0.19)
Prob(hire)	58.60	50.93	65.63	54.13	-7.03	-3.19	7.67	11.50
	(4.50)	(7.88)	(4.94)	(4.97)	(7.20)	(8.83)	(9.07)	(7.00)
Prob(hire)	55.00	58.71	74.38	60.00	-19.38**	-1.29	-3.71	14.38*
—Entry	(5.66)	(10.66)	(4.67)	(7.23)	(7.55)	(12.53)	(11.28)	(8.32)
N	15	15	24	24	39	39	30	48

Notes - *** p<0.01, ** p<0.05, * p<0.1 This table summarizes employee choices across tournament type. Prob(hire) and Prob(hire)|Entry are the belief of the employee that they would be hired by their matched employer, either unconditionally or conditional on having entered the tournament.

Table 1.3: Employer Characteristics

Variable	ST	AA	Diff(ST-AA)
Female	0.58	0.52	0.06
	(0.52)	(0.54)	(0.07)
Age (approx.)	33.02	36.59	-3.57**
	(0.82)	(1.12)	(1.38)
Risk	10.85	10.97	-0.12
	(0.32)	(0.31)	(0.45)
% College	0.56	0.66	-0.10
or higher	(0.05)	(0.05)	(0.07)
% Employed	0.77	0.75	0.02
	(0.04)	(0.05)	(0.06)
% Student	0.05	0.06	-0.01
	(0.02)	(0.02)	(0.03)
N	92	88	120

Notes - *** $p<0.01$, ** $p<0.05$, * $p<0.1$ This table summarizes employee choices across tournament type. Risk ranges from 1 for most risk loving to 15 for most risk averse (8=risk neutral).

Table 1.4: Employment Regressions

	(1) Employment	(2) Prediction	(3) Employment
Female	-0.0255	-0.0264	-0.00788
	(0.0235)	(0.0230)	(0.0154)
Enter-Lose	0.113***	0.122***	0.0316
	(0.0345)	(0.0317)	(0.0280)
Enter-Win	0.381***	0.363***	0.139***
	(0.0432)	(0.0387)	(0.0362)
FemaleXEnter-Lose	0.0496	0.0201	0.0362
	(0.0374)	(0.0282)	(0.0297)
FemaleXEnter-Win	0.105***	0.0725**	0.0564*
	(0.0401)	(0.0332)	(0.0317)
AA	-0.00740	0.00385	-0.00996
	(0.0438)	(0.0408)	(0.0333)
FemaleXAA	0.0167	0.0367	-0.00773
	(0.0340)	(0.0336)	(0.0239)
Enter-LoseXAA	-0.0295	-0.0424	-0.00125
	(0.0505)	(0.0460)	(0.0391)
Enter-WinX	0.0208	0.0363	-0.0339
	(0.0647)	(0.0544)	(0.0497)
FemaleXEnter-LoseXAA	-0.0171	0.0147	-0.0268
	(0.0499)	(0.0435)	(0.0384)
FemaleXEnter-WinXAA	-0.105*	-0.0894**	-0.0455
	(0.0542)	(0.0445)	(0.0450)
Prediction			0.666***
			(0.0423)
Constant	0.178***	0.318***	-0.334
	(0.0301)	(0.0290)	(0.0218)
N	1800	1800	1800
R-Sq	0.262	0.296	0.507

Notes - *** $p<0.01$, ** $p<0.05$, * $p<0.1$ This table presents OLS regressions with controls for grade of employee, standard errors clustered at the employer level.

Table 1.5: Counterfactual Breakdown

Tournament Entry	Hiring Decision ST	AA
ST	- -	-1.97 pp (-6%)
AA	+4.06 pp (+13%)	+1.29pp (+4%)

Notes - This table provides the probability of a woman being hired when varying the tournament entry behavior and hiring behavior independently, compared to the probability of a woman being hired in the standard treatment.

2.0 Incentives to Sleep: an Experimental Analysis of Sleep Choices (Co-Authors: Osea Giuntella and Peiran Jiao)

Sleep deprivation is prevalent in modern societies leading to negative health and economic consequences. However, we know little about why people decide to sleep less than the recommended number of hours. This study investigates the mechanisms affecting sleep choice and explores whether commitment devices and monetary incentives can be used to promote healthier sleep habits. Toward this end, we conducted a field experiment with college students, providing them incentives to sleep, and collected data from wearable activity trackers, surveys, and time-use diaries. Our results are consistent with sophisticated time-inconsistent preferences and overconfidence. The subjects in the treatment group responded to the monetary incentives by significantly increasing the likelihood of sleeping between 7 and 9 hours (+19%). We uncover evidence of demand for commitment. Overall, 63% of our subjects were sophisticated enough to take up commitment, and commitment improved sleep for the less overconfident among them. Using time-use diaries, we show that during the intervention, there was a reduction in screen time near bedtime (-48%) among treated subjects. Individuals in the treatment group were less likely to report insufficient sleep than at baseline even after removal of the incentive (-16%), which is consistent with habit formation. Finally, our treatment also had positive (albeit small) effects on health and academic outcomes.

2.1 Introduction

Sleep deprivation is an emerging public health challenge. According to the Center for Disease Control and Prevention, more than a third of American adults sleep less than the recommended minimum of seven hours (Liu, 2016). Some scholars consider it the most prevalent risky behavior in modern societies and evidence suggests that in many countries people may be sleeping between one and two hours less than what their ancestors used to sleep a hundred years ago (Roenneberg, 2013). Growing evidence documents the causal effects of sleep deprivation on chronic diseases, health, cognitive skills, decision making, human capital, and productivity (Luyster et al., 2012; Giuntella and Mazzonna, 2019; Giuntella et al., 2017; Jin et al., 2015; McKenna et al., 2007; Hafner et al., 2017; Heissel and Norris, 2018; Gibson and Shrader, 2018). Firms, athletes, and military training programs increasingly recognize how sleep deprivation can impair performance.[1]

Despite sleep being increasingly recognized as a fundamental contributor to health and human capital and despite economists' interest in time-use (Becker, 1965; Aguiar and Hurst, 2007; Aguiar et al., 2013; Hamermesh, 2019), sleep

[1]Recently Aetna, an American managed health care company, introduced incentives to increase workers' sleep (see https://www.cnbc.com/2016/04/05/why-aetnas-ceo-pays-workers-up-to-500-to-sleep.html). Concern has been raised regarding sleep deprivation among NBA players (see https://www.espn.com/nba/story/_/id/27767289/dirty-little-secret-everybody-knows-about). Finally, sleep with physical activity and nutrition is also one of the three pillars of the army performance triad (see https://armymedicine.health.mil/Performance-Triad).

behavior has received little attention in the economic literature. Given that we spend approximately a third of our time –one of our scarcest resources– sleeping, and given the substantial economic and health impacts of sleep deprivation, sleep behavior should be an object of natural interest to economists (Mullainathan, 2014). However, most economic models analyzing time allocation regard sleeping as a pre-determined and homogeneous constraint on time allocation. While for some individuals sleep duration and quality are influenced by medical conditions (insomnia, sleep apnea etc.), for most individuals bedtime and sleep duration are a choice. Individuals may optimally allocate less time to sleep and delay their bedtime (or anticipate their wake-up time) to work longer or enjoy more leisure. And indeed, the few pioneering studies analyzing sleep choice have assumed individuals choose hours of sleep optimally (Biddle and Hamermesh, 1990). Yet, according to the Royal Philips global sleep survey, 8 in 10 adults worldwide want to improve their sleep and a poll from YouGov suggests that, while 89% of Americans would like to sleep for 7 hours or more each night, more than 40% report to sleep less than that.[2]

Delaying bedtime may have immediate benefits (i.e., the utility from watching a further episode of a TV series, or working an extra hour), but delayed costs (i.e., the lack of energy or alertness following a night of poor sleep). This suggests sleep decisions may be characterized by dynamic inconsistency, and there may be scope for incentives and commitment devices to promote optimal behavior (O'Donoghue and Rabin, 2003; O'Donoghue et al., 2006; O'Donoghue and Rabin, 2015). Daily experience with time-inconsistent behavior may increase demand for commitment (Laibson, 1997; Rabin et al., 1999; DellaVigna and Malmendier, 2006; Laibson, 2015; Schilbach, 2019). Individuals who are sophisticated about their time inconsistency may restrict their future choice set without receiving any form of compensation (e.g., Ashraf et al., 2006; Dupas and Robinson, 2013; Kaur et al., 2015; Toussaert, 2018) or even at a cost (e.g., Casaburi and Macchiavello, 2019; Milkman et al., 2013).

This study investigates sleep choice and the role of commitment devices and monetary incentives to promote healthier sleep habits. We conducted a field experiment among college students and collected data from wearable activity trackers, surveys, and time-use diaries. Eliciting preferences and randomizing incentives to go to bed earlier and sleep longer, we shed light on the role of present bias, overconfidence, commitment, and habit formation in sleep.

The effectiveness of incentives and commitment devices to promote optimal choices in the presence of self-control problems has been analyzed in the context of other health behaviors such as alcohol consumption, unhealthy eating, and exercise (O'Donoghue and Rabin, 2006; Charness and Gneezy, 2009; Acland and Levy, 2015; Volpp et al., 2009; Just and Price, 2013; Royer et al., 2015). Yet, sleep is a particularly interesting domain in which to investigate the prevalence and persistence of behavioral biases. It is an activity that people engage in every day, and about which they have received repeated feedback throughout their lives. Thus, sleep is a domain wherein demand for commitment might be highly relevant, to the extent that some individuals become more aware of their time inconsistency problem due to experience and feedback (Laibson, 1997, 2015; O'Donoghue and Rabin, 1999, 2001). Sleep is also an interesting domain in which to study overconfidence. If people are overconfident about sleep, this is a demonstration that

[2] See https://www.usa.philips.com/c-e/smartsleep/campaign/world-sleep-day.html and https://today.yougov.com/topics/health/articles-reports/2019/03/13/sleep-habits-americans-survey-poll

such bias can persist, even in the face of extensive experience and feedback (Huffman et al., 2018). If individuals have biased recall of own sleep and fail to correct their overconfidence in the face of continuous information, this may be consistent with a motivated-belief perspective (Bénabou and Tirole, 2016), suggesting that incentives may mitigate the role of motivated reasoning (Zimmermann, 2019).

While sleep deprivation is a problem for many age groups, there are several reasons for sleep deprivation and sleep choice being of particular interest for college students. First, time management is a major challenge among college students transitioning from high school and home habits to campus life (Misra and McKean, 2000; Trockel et al., 2000). Second, sleep deprivation among college students is increasingly becoming a reason for concern. According to recent statistics published in a report of the National Institute of Health (Hershner and Chervin, 2014), more than 70% of college students sleep less than eight hours a day, 60% say they are "dragging, tired, or sleepy" at least three days a week, and more than 80% say loss of sleep affects their academic performance. Third, sleep deprivation and poor sleep quality has been associated with various aspects of undergraduate mental health (Milojevich and Lukowski, 2016), including symptoms of psychological distress, anxiety, attention deficit, and depression problems (McEwen, 2006; Kahn-Greene et al., 2007).[3] Fourth, college is also a crucial phase to shape one's lifestyle and habits (Buboltz et al., 2001). Indeed, Giuntella et al. (2019), who investigate the age-sleep profile, document that during college years, sleep duration markedly declines before reaching a minimum in the early forties. Fifth, college students are a group that is physically healthier, with fewer social and familial constraints and with more time flexibility, suggesting that this is an appropriate group for our experimental study of sleep choice. Additionally, understanding the behavioral mechanisms behind sleep choice within this population may help design educational programs and interventions aimed at improving sleep duration and quality, with non-negligible effects on students' mental health and with potential long-lasting effects on both habits and health.

We recruited 319 participants at the University of Oxford (163 subjects) and the University of Pittsburgh (156 subjects). The subjects were given wearable devices (Fitbit) to collect data on their sleep, physical activity, and heart rate for 8 weeks. In the incentive treatments, subjects set bedtime and sleep duration targets for themselves each Monday of the three treatment weeks and were rewarded for each night (Monday through Thursday) that both targets were achieved based on Fitbit data. We elicited subjects' time and risk preferences in the lab, and integrated the data collected from wearable devices with weekly surveys, time-use diaries, and a follow-up survey conducted three months after the end of the experiment to examine how behavioral biases, such as present bias and overconfidence, affect sleep choice.

We uncover evidence that the subjects voluntarily opted for commitment devices in the form of more demanding targets and dominated incentive schemes. Our findings are consistent with sophisticated time inconsistency and over-confidence as key behavioral mechanisms underlying poor sleep choices. In total, 63% of our subjects took up some

[3]This is of particular concern, given that depression, anxiety, and suicide rates are rising among US college students (Liu et al., 2019; Mortier et al., 2018; Eisenberg et al., 2013). Reetz et al. (2014) report that 95% of college counseling center directors said that the number of students with significant psychological problems is a growing concern in their center or on campus. Anxiety was found to be the top concern among college students (41.6%), followed by depression (36.4%).

form of commitment. More present-biased subjects reported less sleep at baseline and were more likely to take up commitment devices (+28%). Among present-biased individuals, commitment devices reduced insufficient sleep by at least 25%. Meanwhile, many subjects were overconfident about their achievement rates, over-remembered their own bedtime and sleep duration, over-placed their own sleep duration and quality among peers, and understated personal risk associated with sleep deprivation relative to the risk they predicted for peers. Overconfident subjects were more likely to be sleep deprived at baseline and selected overly optimistic targets. Present-biased individuals were more likely to achieve their targets if they were less overconfident.

Our monetary incentives were effective in improving sleep behavior. The participants responded to monetary incentives by sleeping longer. They were 19% more likely to sleep the recommended number of hours (between 7 and 9) and 23% less likely to sleep less than 6 hours. This finding is robust to the inclusion of individual fixed effects, accounting for time-invariant individual heterogeneity. Furthermore, we document a persistent improvement in sleep. Even after the intervention was removed, the subjects were 16% less likely to sleep less than 6 hours. Our intervention also had effects on sleep regularity, reducing sleep, bedtime, and (more weakly) wake-up time variance. Additionally, as sleep deprivation has been linked to detrimental effects on health and human capital, we explore the potential indirect effects of our intervention on health and academic achievement. We find suggestive evidence that our intervention improved heart rate efficiency, physical activity, and self-reported health, although the effects are relatively small. There is also evidence of positve effects on academic achievement.[4]

Given that the subjects positively responded to our incentive to sleep, a natural question is how the subjects re-allocated their time to achieve their targets. To address this question, we collected time-use diaries before, during, and after the intervention, and examined how individuals in the treatment group allocated their time when receiving incentives to go to bed earlier and sleep longer. We find no evidence of significant changes in time spent on studying, working, personal care activities, exercising, or socializing. The only activity that systematically and significantly declined during the intervention was screen time (watching TV, videos, and so on). Interestingly, we show that among those who complied with the treatment, evening screen time (after 8 pm) declined by 48% during the intervention with respect to baseline, and by about 28% after the incentive was removed. We see these results as particularly noteworthy given the growing evidence that digital temptations and the use of blue light technologies near bedtime severely impair sleep (Billari et al., 2018; Nie and Hillygus, 2002; Twenge et al., 2017). Consistent with the evidence that repetition of behavior, such as following fixed routines, increases habit formation (e.g., Wood and Neal, 2007; Lally et al., 2010), adjusting activities before bedtime may help develop better sleep habits.

We directly relate to recent studies analyzing the effects of wearable technology on sleep and health behavior (Jakicic et al., 2016; Patel et al., 2015). Handel and Kolstad (2017) exploit a large-scale intervention in a firm to randomize subjects into treatments to improve sleep and exercise through planning. They find small effects of accessing planning

[4]Our evidence on the effects of sleep on health and academic achievement adds to the growing literature analyzing sleep behavior and its effects on human capital and health (Gibson and Shrader, 2018; Giuntella and Mazzonna, 2019; Giuntella et al., 2017; Jin et al., 2015; Heissel and Norris, 2018; Jagnani, 2018). While these studies used quasi-experimental variation, we exploit a unique field experiment setting, which provides us greater control, to identify the relationship between more sleep and better health and academic outcomes.

tools. Our findings suggest in the presence of persistent behavioral biases that the introduction of monetary incentives and commitment devices may be more effective than using planning tools alone. Bessone et al. (2018) randomize incentives to sleep longer to analyze the effects on labor market productivity and health in a developing country, finding little evidence of an impact of sleep on short-run economic outcomes, but significant effects of naps on attention and well-being. While their main goal was to induce exogenous variation in sleep (and naps) to assess its effects on human capital and productivity, our study focuses directly on the mechanisms behind sleep choice. Furthermore, the differences in the contexts, sleep conditions (e.g., quality of mattress, noise), and samples are likely to explain the different results found in the two studies when examining the effects of sleep on health and human capital. Finally, by following individuals for eight weeks and surveying them three months after the end of the experiment, we are the first to examine persistence and habit formation effects in the context of sleep decisions. We find that sleep changes are persistent, suggesting that temporary incentives could lead to long-run lifestyle changes in the sleep domain.[5]

Our study also contributes to the literature analyzing demand for commitment and the effectiveness of commitment devices (see, e.g., Bryan et al., 2010; Kremer et al., 2019; Schilbach, 2019, for a review). To date, little evidence exists in the sleep domain regarding the effectiveness of commitment devices in improving sleep habits.[6] The evidence on the effectiveness of commitment devices is mixed (Laibson, 2015, 2018), suggesting that uncertainty could undermine the demand for commitment (Laibson, 2015), and that, unless subjects are sufficiently sophisticated, commitment devices may be welfare reducing (Bai et al., 2017). Furthermore, in a recent work, Carrera et al. (2019b) show that commitment contract take-up may reflect, at least in part, demand effects or "noisy valuation" when there is substantial uncertainty about the desirability of an activity, even if subjects are time consistent. However, the continuous experience and immediate feedback that characterize sleep behavior suggest commitment devices may be more effective in this domain. Our experiment provides a relatively soft commitment device in the form of setting bedtime and sleep duration targets, at the cost of forgone rewards.[7] Additionally, we elicit time preferences in incentivized tasks and then measure sophistication in subjects' overconfidence about own performance. We find evidence not only for sufficiently sophisticated individuals taking up commitment but also for the positive effects of commitment on behavior.

The rest of the paper is organized as follows. Section 2 describes the experimental procedure and the design of our intervention. The data are presented in Section 3. In Section 4, we discuss the role of present bias, overconfidence, and risk preferences in the sleep domain. In Section 5, we present the results of our randomized experiment, discuss the effectiveness and persistence of incentives to sleep, and their effects on time allocation, academic outcomes, and

[5] Breig et al. (2018) also consider sleep in a study using wearable devices. However, their main focus is task allocation. In a 2-week experiment, they randomize feedback on subject's time allocation and explore how that affects their time use in the following week. Their findings show the role for over optimism in time allocation decisions. Our focus is instead sleep, and we conduct an eight-week field experiment to analyze the effects of randomized incentives to sleep, their effects on time use, and shed light on the role of demand for commitment, overconfidence, and habit formation in the sleep domain.

[6] Previous studies have examined the effectiveness of commitment devices in various contexts, including saving decisions, alcohol, smoking, and exercising (Schilbach, 2019; Augenblick et al., 2015; Royer et al., 2015; Giné et al., 2010; Ashraf et al., 2006; Kaur et al., 2015). Some of these studies support the idea that commitment devices may help sophisticated agents with present bias mitigate their future self-control problems (Schilbach, 2019; Kaur et al., 2015; Ashraf et al., 2006).

[7] Previous evidence also suggests that softer commitments may work better than hard commitments (Dupas and Robinson, 2013).

health. Concluding remarks are provided in Section 6.

2.2 Experimental Procedure, Design, and Data

2.2.1 Experimental Procedure

The experiment was conducted at the Centre for Experimental Social Sciences (CESS) in Nuffield College, Oxford, UK and at the Pittsburgh Experimental Economics Laboratory (PEEL) at the University of Pittsburgh.[8] The experimental procedure was approved by the Central University Research Ethics Committee of the University of Oxford, the ethical review committee of CESS, and the University of Pittsburgh Institutional Review Board. All subjects provided informed consent before participating in the experiment. The participants were given sufficient information about the nature and tasks of the experiment, although it was not specified that the focus of the study was sleep.[9] During the experiment, the participants were free to withdraw any time without penalty.

The results reported in this paper were derived from five waves of experimental sessions. The first three waves were conducted in Oxford: the first from October to December 2016; the second from April to June 2017; and the third from October to December 2017. These periods correspond to the Michaelmas Terms of the 2016–17 and 2017–18 and the Trinity Term of the 2016–17 academic years at the University of Oxford, respectively.[10] The fourth and fifth waves were conducted at the University of Pittsburgh between mid-January and mid-March during the Spring semester of the 2017–18 academic year and between mid-September and mid-November during the Fall semester of 2018–19 academic year, respectively. The experiment was first advertised in the University of Oxford and on the campus of the Oxford Brookes University in the Oxford waves (1–3) and on the University of Pittsburgh campus in the Pittsburgh waves (4–5). Interested participants then signed up on our recruiting website. The participants in all the waves were recruited through the Online Recruiting System for Economic Experiments (Greiner, 2015) at CESS and the SONA online management system at PEEL, respectively.

Each wave was conducted over eight weeks (which in Oxford coincided with the length of the academic term). Recruitment occurred a week before the beginning of the experiment (Week 0). In Week 1, the subjects were invited to the lab for an experimental session and were given a Fitbit Charge HR device. We collected baseline data from Fitbit devices for the first two weeks. Experimental surveys and treatments started on Monday morning of Week 3, and all participants' Fitbit data were monitored through Week 8. On Friday of Week 8, the participants returned the devices and received final payments. A show-up fee of GBP 4 ($\approx$ USD 5.3) in Oxford or USD \$6 in Pittsburgh was given both in the Week 1 lab session and when they returned the Fitbit, regardless of their performance in the experiment. Among

[8] In all our estimates, we include dummies for whether the subject was recruited in Oxford or in Pittsburgh.

[9] We framed it as a study about the use of wearable devices.

[10] Including dummies for each wave of the experiment does not affect our results.

subjects who successfully completed all parts of the experiment, a lottery was drawn, and 3% of them could each win a reward of GBP 150 ($\approx$ USD 199) in Oxford or USD $200 in Pittsburgh.

During the lab session in Week 1, subjects were given an oral description of the experiment, including the exclusion criteria, before their consent was sought. This lab session was divided into three parts. The first part was an incentivized elicitation of risk and time preferences using multiple price lists, each comparing two options. The subjects needed to make one choice for each list: at which row they would switch from choosing Option A to choosing Option B. We elicited risk preferences using two price lists, each comparing a fixed lottery with various certainty amounts. We elicited time preferences using four price lists, each comparing different sooner payments with a fixed future payment. We varied both the size and timing (immediately or in 4 weeks) of the sooner payments as well as the gap between the sooner and later payments (4 or 8 weeks). Finally, one choice was chosen from each preference elicitation to determine payments. The risk payments were made at the end of the lab session. All time payments were made in the form of a gift card sent to participants' email address. To jointly elicit time and risk preferences using multiple price lists, we adopt a similar method to those in, for instance, Tanaka et al. (2010), Falk et al. (2016), and the Double Multiple Price List in Andreoni and Sprenger (2012).[11] For details, please see Appendix B.

The second part of the lab session involved several survey items, which elicited details on subjects' demographics, health conditions, cognition, lifestyle, health behaviors, and physical activity. Additionally, a survey measure of domain-specific risk attitudes (Weber et al., 2002) was implemented, which included a health domain. A part of the survey was specifically about sleep. We asked participants about their sleep habits before the experiment and their general knowledge about the negative consequences of bad sleep habits. We then let them read a short paragraph on the medical evidence of the negative consequences, after which we asked them to evaluate the probability of suffering negative health effects due to sleep deprivation for themselves and for others. The questions about self and others were kept distant from each other and were framed in different ways to encourage subjects to think about the questions independently.

In the third part of the lab session, each participant was given a Fitbit Charge HR device, registered for a Fitbit account, which was linked with Fitabase for data collection. The device was then synchronized with the account.[12] They were asked to wear the device as much as possible including during sleep, to charge and to synchronize the device regularly, and to return the device on or soon after the Friday of Week 8.

2.2.2 Experimental Design

Subjects were randomly assigned to a control condition or one of the three different treatment conditions. In the control group, participants were asked only to wear the device, allowing their Fitbit data to be recorded, and to respond to surveys during the experimental period. Control group participants received two types of surveys. One

[11] These are based on risk preference elicitation in Holt and Laury (2002) and time preference elicitation in Harrison et al. (2002).
[12] Fitabase is a paid service that collected Fitbit data from our subjects.

was a Weekly Survey, sent on the Monday of each week, which asked subjects about their health and sleep activity in the previous week.[13] We also surveyed subjects on their time use. On two randomly chosen mornings of each week, subjects were asked to fill in diaries recalling what they did during the previous day in 30-minute intervals. Participants could select the activity for each time slot from a drop down menu of categories (e.g., sleeping, grooming, watching TV, surfing the Internet, playing games, working, studying, preparing meals or snacks, eating or drinking, cleaning, laundry, grocery shopping, attending religious service, hanging out with friends, paying bills, exercising, commuting [bus/train], commuting [walk or bike]). Subjects were permitted to not respond if they felt uncomfortable.

In addition to the surveys completed by the control group, subjects in the three treatment conditions also completed sleep incentive surveys in the treatment weeks: as part of the Weekly Survey, treated participants were asked to choose a bedtime target (between 10 pm and 1 am) and a sleep duration target (between 7 and 9 hours) for Monday through Thursday nights of the current week and received incentives for achieving the targets.[14]

The three treatments varied in the timing of the incentivized weeks and the form of incentives. In Treatment 1 (Incentive-Weekly), the treatment weeks were Weeks 3, 4, and 5. Figure 2.1 illustrates the timeline of our main intervention. We used gain/loss framing: each week, these subjects were told that they would be rewarded GBP 10 (USD 15 in Pittsburgh) for participation in the following week. Rewards and punishments were added to this amount. Each reward was GBP 2.5 (USD 3.75) and each punishment was also GBP 2.5 (USD 3.75), so that the largest gain for achieving targets on all four nights was GBP 20 (USD 30). The subjects would achieve their target by complying with both bedtime and sleep duration targets, measured by Fitbit data, on a given day. A failure was to miss either target on a given day. We also provided feedback on performance in the previous week and asked subjects to predict their own performance for all remaining treated weeks. Depending on the size of the prediction reward, Treatment 1 (Incentive-Weekly) can be further divided into 2 arms: For approximately 40% of the subjects, only one prediction was finally chosen, and a correct prediction was rewarded GBP 2 (USD 3); for the remaining 60%, one prediction was chosen from each of the 3 treatment weeks, and each correct prediction was rewarded GBP 2 (USD 3).

We then tested two slight modifications to this treatment to see how subjects would respond to changes in frequency and structure of the incentives (Carrera et al., 2019a) (Treatment 2, Incentive-Biweekly); and in the size of the monetary incentive (Treatment 3, small incentive-biweekly). Treatment 2 (Incentive-Biweekly) was the same as 1, except that the sleep incentives were given biweekly, in weeks 3, 5, and 7 (see Figure C.1). In Treatment 3 (Small Incentive-Biweekly), the incentives were given biweekly as in 2, but we did not use gain/loss framing in the incentives; in other words, there was no initial endowment in each week. Instead, subjects could choose between two contracts. The first one was a reward of GBP 2.5 for each night the target was met, and there was no punishment. Therefore, meeting the target on all nights of a week could lead to a total reward of GBP 10. The alternative contract would not only involve the same reward for meeting the goals but also penalize unmet goals. The punishment for each failed night was GBP

[13] Subjects in the control group did not know others were paid.

[14] The sleep duration targets were set between 7 and 9 hours to reflect the recommended number of hours of sleep; see Cappuccio et al. (2010).

2.[15]

In all treatment groups, rewards and punishments were added to their payments on the day they returned the device. One of the 3 treated weeks was selected for each subject to determine payment for their sleep performance. Table D.1 summarizes our treatments.

2.3 Data

A total of 359 participants volunteered for the experiment, and 319 of them generated usable data; 40 subjects (11%) either felt uncomfortable wearing a Fitbit or dropped out due to other reasons. We check the sensitivity of our results to the inclusion of individuals who dropped out but generated some usable data. We find no evidence of significant association between compliance with the treatment and the likelihood of dropping out before week 8. Furthermore, withdrawing from the experiment does not correlate significantly with baseline characteristics of the subjects (Table D.2).

Among the 319 remaining participants, 107 were in the control condition, 104 in the weekly Treatment 1 (Incentive Weekly), 76 in the biweekly Treatment 2, and 32 in the weak incentive Treatment 3. We provide the full questionnaires of the surveys conducted during the experiment in the Online Appendix C.

2.3.1 Measuring Sleep

Measuring sleep is challenging. Previous studies have shown that self-reported measures of sleep, whether based on time-use diaries or survey questions, are prone to severe measurement errors. Self-reports tend to overestimate sleep duration compared to objective measures (Lauderdale et al., 2008b). Time-use diaries may also be subject to overestimation bias, as often, the activity lexicon associated with sleeping includes transition states (e.g., falling asleep) (Basner et al., 2007). Personal wearable devices (such as Fitbit) have been used to study health behavior (e.g., Handel and Kolstad, 2017). Concerns have also been raised regarding the ability of Fitbit devices to provide an accurate measurement of sleep, although some medical studies (e.g., Lee et al., 2017) find wearable activity trackers that detect heart rate perform fairly well in terms of tracking sleep compared to actigraphy, the more sophisticated method used in medical studies (Beattie et al., 2017).[16]

We contribute to the methodological discussion on sleep measurement (e.g., Lauderdale et al., 2008a), by comparing information on sleep obtained from three of the main sources used in the literature—wearable devices, time-use diaries, and self-reported sleep in surveys. We identify substantial disparities in sleep measurements obtained using

[15]This treatment was done in Oxford only. Subjects were paid GBP 8 for returning the device in addition to the show-up fee in Week 8. One treated week was randomly chosen and any loss was deducted from this amount.

[16]Beattie et al. (2017) suggest that Fitbit heart rate-tracking devices accurately track light, REM, and deep sleep stages (see also http://www.sleepreviewmag.com/2017/06/study-shows-fitbit-heart-rate-tracking-devices-accurately-track-light-rem-deep-sleep-stages/.)

these three methods, which partially reflect the distance between beliefs and actual behavior. We have (1) the bedtime, wake-up time, and sleep duration as collected by the Fitbit devices; (2) self-reported information on sleep habits and quality before and during the experiment collected in surveys; and (3) sleep measured through our time-use diaries. Therefore, we can directly compare these three different measures of sleep. Additionally, Fitbit offers limited but useful information about sleep quality through sleep efficiency–the fraction of time spent asleep while in bed–and the number of sleep episodes per night.

Table D.3 compares the different measures of sleep obtained using Fitbit devices, survey data, and time-use diaries. On average, subjects reported 8.15 hours of sleep in time-use diaries and 7.07 hours of sleep for the previous week in self-reported surveys. Thus, time-use data tend to significantly overestimate time allocated to sleep, while self-reported sleep duration is only a few minutes longer than the average sleep duration measured by Fitbit devices (7.02 hours during the week). Further, according to time-use data, only 7% of the subjects reported sleeping less than 6 hours, while the survey-based measure indicated 10% of the subjects slept less than 6 hours—closer to but still significantly smaller than the 23% recorded by Fitbit devices during the school week. These results were also consistent with the overestimation by the subjects of own sleep duration in the first-day survey. As research on sleep choice, its determinants, and its effects advances, understanding the extent to which each of these methods captures both pure sleep duration and biased beliefs will be crucial in identifying best practices in sleep measurement.

2.3.2 Descriptive Analysis: Pre-Intervention Data

Table D.4 reports summary statistics for subjects at baseline. This information was collected in the lab on the first day of the experiment. Subjects were 59% male, with an average age of 21.54 (min: 18; max: 45; median age: 21) .[17] Of our respondents, 58% were White, 22% Asian, 9% Black, and 11% other minorities.

We measured subjects' health, well-being, and sleep behavior before the intervention. Subjects were relatively healthy. Only 11% reported poor health status. The average BMI in the sample was 24 (min: 15.5; max: 47.0), with only 5% obesity rate (BMI>30), and 24% overweight status (BMI>25); 23% of the subjects had ever smoked, but 61% of those subjects quit smoking; 26% reported drinking more than once per week.

However, self-reported mental health problem was a cause for some concern in this group. While 45% of the sample reported feeling depressed rarely or never, 36% reported that they had felt depressed 1–2 days over the last week, 15% reported occasional feelings of depression (3–4 days per week), and 4% reported feelings of depression most of the time (5–7 days per week). Moreover, 6% of the sample reported feeling completely satisfied with their life; 44% considered themselves very satisfied; 42% somewhat satisfied; and 9% not satisfied or not satisfied at all.

According to the survey results of sleep patterns at baseline, subjects sleep an average of 7 hours and 15 minutes (Min: 4; Max: 10) each night during the month before the experiment, with women sleeping 15 minutes longer on

[17] One part-time student was aged 45. Excluding this observation from the analysis does not affect the results.

average— consistent with what has been found in time-use studies (see Hamermesh (2019)).[18] Most subjects reported an ideal sleep of 8 hours (7.97 on average), and 97% of subjects considered it ideal to sleep more than 7 hours. Yet, 46% reported sleeping less than 7 hours on an average night during term (see Figure C.2). Subjects reported falling asleep during the day on 3.79 days over the last month and a quality of sleep of 6.61 on a 1–10 scale. At baseline, 17.7% (19.3%) of subjects expressed that they were definitely willing to improve their sleep by sleeping longer (going to bed earlier); 43% (41%) stated they were probably willing to; the rest were either unwilling to improve or did not know how to (Table D.5).

Fitbit data of sleep before the intervention are plotted in Figure 2.2. Most people on most days slept between 6 and 9 hours, with subjects in Pittsburgh (dashed line) sleeping less than those in Oxford (solid line). On an average night, in the first 2 weeks before the experiment, subjects in Pittsburgh slept approximately 6 hours and 45 minutes, while subjects in Oxford slept 7 hours and 20 minutes. Women in our sample slept on average 7 hours, men 6 hours and 50 minutes (difference not statistically significant), but at baseline, women were significantly less likely to report sleeping less than 7 hours (-7% with respect to the mean). The gender difference in sleep duration is consistent with previous studies (Hamermesh, 2019). Figure 2.3 documents the cumulative distribution of sleep hours. On an average night of the week, 70% of the time individuals slept less than 8 hours, 47% less than 7 hours, 25% less than 6 hours, and 12% less than 5 hours (see Figure 2.4). Sleep duration was highly irregular—the standard deviation was 2 hours—varying substantially throughout the week, with subjects sleeping significantly less during the week than on weekends (see Figure 2.5). Subjects compensated during the weekend for some of their lost sleep hours during the week, wherein approximately 47% of the subjects slept less than 7 hours in the first 2 weeks, while during the weekend the fraction of individuals sleeping less than 7 hours of sleep declined to 39%.

We also document the association between insufficient sleep and self-reported measures of health and well-being at baseline using self-reported data drawn from the survey conducted on the first day of the experiment. Individuals who report sleeping between 7 and 9 hours were more likely to report good health status (+6% with respect to the mean; p-value$<$0.01); they were also 6 percentage points less likely to be obese (p-value$<$0.05) and overweight (-48% with respect to the mean; p-value$<$0.001); 55 percentage points less likely to report feelings of depression (p-value$<$0.05); and more likely to be satisfied with life (+56% with respect to the mean; p-value$<$0.001) (see Table D.6 for details). Individuals who were identified as more likely to take risks were also more likely to sleep less (see Figure C.3).

[18]The question asks “During the past month, how many hours of actual sleep did you get at night (average hours for one night)? (This may be shorter than the number of hours you spend in bed.)”

2.4 Behavioral Biases and Sleep Choice

2.4.1 Time Inconsistency and Demand for Commitment

Several aspects of our participants' behavior were consistent with sophisticated time inconsistency. We correlated our measures of subjects' time preference with baseline sleep patterns and performance in the experiment. In Appendix B, we describe in detail how we built our measure of present bias and impatience. The results are reported in Table 2.1. Columns 1–3 report estimates based on self-reported sleep in the survey conducted on day 1. Columns 4–6 report estimates based on the first two weeks of data collected from Fitbit devices. While estimates are not precise due to the small sample size, we find that before intervention, present-biased subjects were more likely to be sleep deprived. In particular, at baseline, individuals in the top quartile of our measure of present bias were 10% (11%) more (less) likely to report sleeping less than 7 hours (between 7 and 9 hours, columns 1–3). Fitbit data reveal even larger differences (columns 4–6). Present-biased subjects were 19% more likely to sleep less than 7 hours with respect to the mean and 21% less likely to sleep between 7 and 9 hours, sleeping on average 12 minutes less per night. The relationship between sleep duration and impatience appears to be less clear (see columns 4–6, Panel B).

Our experiment included two features that allowed us to directly observe the demand for commitment consistent with sophisticated hyperbolic discounting models. First, in all intervention groups, we asked subjects to choose bedtime targets between 10 pm and 1 am and sleep duration targets between 7 and 9 hours. An agent with standard preferences would maximize rewards by choosing the least binding targets, namely 1 am and 7 hours. By contrast, choosing more restrictive targets is equivalent to disciplining one's future behavior and can serve as a commitment device. Second, in Treatment 3 (Small Incentive-Biweekly), we asked subjects to choose between a contract that only rewards successes and a contract that not only rewards successes to the same extent but also punishes failures. To maximize monetary payoff, an agent with standard preferences would choose the former, whereas an agent who demands commitment would choose the latter (e.g., Kaur et al., 2015). An agent with naive time-inconsistent preferences may mistakenly predict that her/his future self will achieve all targets and thus be indifferent between having and not having a commitment device, whereas a sophisticated agent may anticipate her/his future time-inconsistent behavior and would actively demand for a commitment device even at some cost. The "cost" of the commitment device in our setting is the forgone reward, or explicit punishment in some cases.

We uncover some interesting evidence of demand for commitment. Despite 1 am being a dominant choice for bedtime target, in 50% of the weeks, subjects in the treatment group chose bedtime targets earlier than 1 am (Figure 2.6). Moreover, 60% of the subjects chose bedtime targets earlier than 1 am in at least 1 week; 24% chose bedtime targets earlier than 1 am in all 3 treatment weeks (Table D.7, column 1). Similarly, despite 7 hours being a dominant choice for sleep duration target, in approximately 48% of the subject-week observations in the treatment group, bedtime targets longer than 7 hours were chosen (Figure 2.7). Moreover, 60% of the subjects chose sleep duration targets longer than 7 hours in at least 1 week, and 19% chose sleep duration targets longer than 7 hours in all 3 treatment weeks (Table D.7, column 2). These results are comparable in magnitude to those of Schilbach (2019), who find that one-third to half of study participants chose sobriety incentives over unconditional payments, even when this choice implied a cost in terms of forgone payments. Furthermore, consistent with demand for commitment, in Treatment 3 (Biweekly-Small), in approximately 10% of the subject-week observations, the contract with punishment was chosen (Table D.7, column 3). A total of 13% of the subjects in this treatment chose a contract with punishment. Subjects choosing a dominated contract were also significantly more likely to choose a dominated bedtime target (+50%).

Present-biased individuals were more likely than other subjects to take up a commitment device (Table D.8, Panel A). They were 25% more likely to choose a bedtime target before 1 am (column 1) and 6% more likely to choose a sleep duration target longer than 7 hours (column 2). Overall, present-biased subjects were 22% more likely to commit to at least one dominated target (column 3). Similarly, impatient individuals were more likely than other subjects to choose a bedtime target before 1 am (+26%, column 1 Panel B).

Time-inconsistent subjects may be more likely to choose more demanding sleep targets earlier in the day, when the cost of last night's bad sleep choice is still felt. Yet later in the day, when the desire to watch another episode of a TV series sets in, they may be more likely to choose less restrictive targets. We exploit variation in the time the surveys were answered and targets chosen by the subjects.[19] While only 35% of the subject chose the least binding bedtime (1 am) when responding to the survey before noon, among those responding after noon, 53% of the subjects chose the least binding target. Among subjects responding after 6 pm, those choosing 1 am as bedtime target increases up to 64% (see Figure C.4). We also find that, in a continuous measure, people who responded later in the day set later bedtime targets. Additionally, we also find evidence that the later the average actual bedtime the week before, the earlier the bedtime target set by subjects, suggesting their sophistication and willingness to improve (see Table D.9).

Figure 2.8 shows that subjects with later bedtimes at baseline (as measured by Fitbit devices in the first two weeks) were more likely to select earlier bedtime targets. Consistent with the hypothesis that sophisticated time-inconsistent preferences may be an important factor behind sleep choice behavior, we found that the behavior of opting for the commitment device was correlated with subjects' predicted bedtime (elicited before the target selection). Subjects who expected to go to bed later set earlier bedtime targets than their predicted bedtime (Figure 2.9), which could reflect subjects with sophistication wanting a commitment device.

2.4.2 Overconfidence

As mentioned earlier, our evidence suggests that overconfidence contributes to explaining individual sleep choices. First, the data drawn from the survey conducted on the first day of the experiment reveal that subjects systematically reported longer sleep durations, better sleep quality, and lower risks associated with sleep than what they considered the average for persons of the same age (see Table 2.2). Consistent with overconfidence, the majority of subjects (62%) believed they sleep longer than the average person of the same age. Individuals reported sleeping 20 minutes longer

[19] We varied the timing of surveys throughout the experiment, although we could not fully control the timing of the answers.

than the average for people of their age. Similarly, 58% of the subjects thought that their sleep quality was better than that of the average person of their age, with 25% of the subjects rating their sleep quality 2 points higher than average on a 1–10 scale.

This biased recall of own sleep in face of repeated feedback is consistent with motivated beliefs, rather than individuals fully integrating information to update beliefs about themselves. We are also able to evaluate whether subjects change their self-reported sleep duration when provided with information from the Fitbit. If subjects do not update their self-reported sleep duration with this additional information, this evidence may be consistent with motivated beliefs. Indeed, when looking at the control group subjects, we see no evidence of significant differences in self-reported sleep duration before and after being provided with the Fitbit and the Fitbit's information, supportive evidence that they are not fully incorporating information into their beliefs. If anything subjects in the control group reported longer sleep duration at the end of the experiment than in the survey conducted on the first day in the lab, but the differences were small and non-significant (p-value=0.44). Among overconfident subjects in the control group, the difference between self-reported and fitbit-measured sleep duration remained significant and substantially unchanged throughout the experiment. These results suggest that despite the feedback available through the wearable device, subjects kept overestimating sleep duration. Furthermore, we show that the higher the distance between self-reported sleep and actual sleep as measured by fitbits, the higher the number of hours subjects would predict to sleep in the next week (see Figure 2.10). In particular, a one standard deviation increase in the difference between self-reported sleep hours for the previous night and the sleep measured by the fitbit device, would be associated with a .26 standard deviation increase in the number of hours a subject would predict to sleep on a typical night of the following week.

We find also evidence of overconfidence with respect to the perceived risks of sleep deprivation: 66% of the subjects estimated for themselves a lower risk of detrimental consequences of sleep deprivation (loss of alertness, weight gain, insomnia, cold, arterial stiffening). In particular, 82% of them thought that others would have a higher likelihood of losing alertness as a consequence of sleep deprivation, with an average of 30-percentage-point higher risk estimated for other individuals of the same age group. Similarly, approximately 65% assessed a higher likelihood for others of the same age group (than themselves) to gain weight and to have insomnia as a result of sleep deprivation. In contrast, differences in the perceived risk of self and others suffering a cold or arterial disease induced by sleep loss were less pronounced.

Comparing Fitbit data and self-reported data on sleep duration, we also find evidence that individuals sleeping less than 7 hours were significantly more likely to overestimate their sleep duration, suggesting that overconfidence may be an important factor behind insufficient sleep. As mentioned above, subjects tended to overestimate the duration of own sleep and, consistent with previous evidence, time-use data were particularly prone to this bias (see Table D.3).

Using these data, we built an index of overconfidence along these different dimensions. In practice, we summed the overconfidence measures in a single index and defined as overconfident those individuals in the upper quartile of the index. Splitting individuals in this way, overconfident subjects were less likely to report insufficient sleep at baseline based on self-reported data, but more likely to be sleep deprived based on Fitbit data before treatment (Table 2.3). In other words, while individuals who were overconfident about sleep reported longer sleep duration at baseline, these subjects were also sleeping significantly less than the rest of the sample based on Fitbit data. We did not find significant differences in their likelihood to take up commitment devices (Table D.8, Panel C). However, on average, overconfident individuals chose sleep duration targets that were 1 hour longer than their sleep at baseline, while the rest of the subjects, on average, selected targets that were 8 minutes longer than their sleep at baseline. The difference between the sleep duration target and the usual sleep was approximately 52 minutes longer for overconfident subjects (p-value=0.001). In other words, while overconfident subjects were equally likely to choose dominated targets, given that their bedtime at baseline was significantly later and their baseline sleep duration was significantly shorter, they took up overly optimistic sleep duration and bedtime targets.

Furthermore, as mentioned above, among present-biased individuals, overconfident subjects were less likely to achieve targets, and commitment devices were not effective (possibly even welfare diminishing) for them, consistent with Bai et al. (2017). While the differences are not precisely estimated, we find that overconfident subjects with present bias were 12% less likely to sleep the recommended number of hours (p-value=0.27).

Participants were also asked to predict the likelihood that they would achieve their chosen target in each of the following treated weeks. Correct predictions were rewarded. Table D.10 shows that individuals tended to overpredict their likelihood of achieving the targets. Predictions do not seem to be improving over time: while subjects were revising their predictions down from week to week, they were also increasingly falling short of their targets as the study proceeded. In the first treated week, 62% of the subjects were too optimistic about the number of nights they could achieve; in the second (third) week of treatment 61% (71%) of the subjects were too optimistic (Figure 2.11-2.13). The decreasing achievement rate may be partially explained by increasing demands on time as the semester proceeds. While students might recognize that this is happening, they may be consistently underestimating how the demands on their time will change. Interestingly, we find no evidence of increasing prediction incentive affecting prediction accuracy.

As mentioned above, choosing a dominated target (or contract) was associated with a higher success rate (see Tables D.11 and D.12). However, we find no evidence that choosing dominated targets improved sleep duration among present-biased individuals who were also classified as overconfident (coef. 0.13, p-value=.62), while the effect is large and significant among the rest of the sample (coef. 0.28, p-value=0.009).

Overall, these results, albeit not all precisely estimated, appear consistent with sophisticated time inconsistency.

2.4.3 Risk Preferences

In Table 2.1 (Panel C), we explore the correlation between subjects' risk aversion and their average sleep duration as estimated during the first lab session. Risk-averse individuals reported longer sleep duration, were less likely to report less than 7 hours of sleep (column 2), and more likely to sleep between 7 and 9 hours (column 3). Overall, Fitbit data confirm these qualitative associations, although the magnitude of the estimates is somehow smaller.

Risk-averse individuals were also less likely (-23%) to choose a demanding target (earlier than 1 am, column 1 of Table D.8, Panel D) and less likely to choose a sleep duration target longer than 7 hours (-10%, column 2). Interestingly, subjects choosing a dominated contract tended to have low risk aversion. In anything, risk-averse individuals were 8% more likely to meet their target (p-value=0.27).

2.5 Incentives, Sleep Behavior, and Habit Formation

The commitment devices and monetary incentives were effective. Subjects met their targets approximately 48% of the time (see Figures 2.11, 2.12, and 2.13). Overall, female subjects were 8% more likely to meet their targets compared to their male counterparts, as female subjects met their targets on at least 49% of the treatment nights while men met their targets only on 45% of those nights.

Subjects who chose dominated bedtimes ended up with better sleep outcomes. When choosing a dominated bedtime (earlier than 1 am), subjects were 14% more likely to achieve the target than those choosing 1 am as a bedtime target (Table D.11). Similarly, subjects choosing a dominated sleep duration target (longer than 7 hours) were more likely to achieve it. Overall, choosing a more demanding target was associated with higher success rates. Subjects choosing a more demanding bedtime (sleep duration) were 13% (20%) less likely to sleep less than 7 hours and those choosing both a demanding bedtime and a demanding sleep duration target were 26% less likely to sleep less than 7 hours (Table D.12, columns 1–3 and 5–7). Similarly, subjects choosing a dominated contract were less likely to report insufficient sleep during the treatment weeks (columns 4 and 8), although the latter result is not precisely estimated due to the small sample size of Treatment 3 (Small Incentive-Biweekly). It is worth noting that all these estimates restrict the sample to the intervention weeks while including controls for insufficient sleep at baseline, partially mitigating concerns of selection bias.

Table 2.4 shows our main regression results.[20] Relative to control, we find that subjects receiving monetary incentives in Treatments 1 (Incentive-Weekly) and 2 (Incentive-Biweekly) were 19% more likely to sleep the recommended number of hours (between 7 and 9 hours, see Cappuccio et al. (2010)) (column 1). This result holds with the inclusion of individual fixed effects (column 2): accounting for persistent individual heterogeneity, the coefficient reduces by 42%, but still indicates an economically and statistically significant effect of the treatment on the likelihood of sleeping between 7 and 9 hours (+11% with respect to the mean). In columns 3 and 4, we examine the effects of treatment on a metric of insufficient sleep (sleeping less than 6 hours). When receiving the monetary incentive, individuals were 23% less likely to sleep less than 6 hours with respect to the average in the sample (column 3), and this effect holds even with the inclusion of individual fixed effects (column 4). Specifically, during treatment, individuals were 12%

[20] We first pool Treatments 1 (Incentive-Weekly) and 2 (Incentive-Biweekly) and then document the heterogeneous effects of the treatments later in the text.

less likely to sleep less than 6 hours. The results tend in the same direction when analyzing alternative dichotomic outcomes for sleeping less than 7 or 5 hours (see Table D.13). On average, incentives increased sleep duration by 6–12 minutes. Individuals spent, on average, 10 minutes more in bed, 8 of which were minutes spent asleep. Regarding the nights on which subjects complied with the incentives, individuals in the treatment group slept 22 minutes longer than those in the control group.

The results on sleep duration are largely driven by earlier bedtimes. When receiving the monetary incentive, the subjects' bedtimes were moved earlier by approximately 20 minutes, while the average wake-up time did not change significantly (see Table D.14). Restricting the sample to the nights individuals reported at least 4 hours and less than 10 hours of sleep, the results are substantially unchanged and, in fact, more precisely estimated, suggesting the main results are not driven by extreme values (the results are available upon request). These effects survived even after removal of the monetary incentive (See subsection 2.5.1). Interestingly, we find no evidence that sleeping more on incentivized nights (Monday–Thursday) crowded out sleep on non-incentivized nights during the intervention. In fact, subjects in the treatment group were also more likely to sleep the recommended number of hours during weekends.

2.5.1 Post-Intervention

Our experiment had two post-intervention periods. The first was within the 8 experimental weeks, and thus, we still had data drawn from the wearable devices. The second part occurred 3 months after the experimental period ended, and consisted of a follow-up survey to additionally investigate the effect of our treatments after the experiment.

2.5.1.1 Habit Formation and Sleep with Fitbit Data We first explore the first part of the post-intervention period. In Table 2.4, we find evidence that the effects of monetary incentives on sleep persist to some extent in the weeks following the termination of treatment. After removing the monetary incentive, subjects in the treatment group were 9% more likely to sleep between 7 and 9 hours, although these results are not precisely estimated (column 1) and are not robust to the inclusion of individual fixed effects (column 2). Yet, when focusing on the left tail of the sleep distribution, we find significant and sizable effects when removing the financial incentive (column 3)[21]. While the coefficients are marginally smaller, the effects hold even after including individual fixed effects (column 4). In fact, after the removal of the incentive, the effect was even larger (+16% with respect to the mean). Using the natural logarithm of sleep, we find that even after the removal of incentives, treated subjects' sleep was 2% longer than at baseline. The difference in magnitude between the treatment and post-treatment effect is comparable with recent evidence on habit formation effects when using financial incentives to promote exercising (Carrera et al., 2019a).

Examining other outcomes drawn from the Fitbit data (Table 2.5), we find no evidence of significant effects of treatment on the efficiency of sleep (columns 1–2), measured as the ratio between sleep duration and time spent in bed (including time awake). There is some weak evidence that treated subjects were more likely to have an efficient

[21] Approximately 25% of subjects reported sleeping less than 6 hours at baseline.

resting heart rate, defined as a resting heart rate in the lowest 25th percentile of those reported in the first 2 weeks of the experiment and before the start of the intervention (columns 3–4). Finally, treated subjects spent less time in sedentary activities (columns 5–6) in their wakeful time during treatment weeks (-9 minutes per day). The magnitude of these effects is relatively small. We find no evidence of any significant effect on the number of steps walked.

2.5.1.2 Follow-up Survey: Effects on Self-Reported Sleep, Health, and Human Capital In the second part of the post-intervention period, we sent subjects a follow-up survey that included questions about health, sleep, and academic performance three months after the experiment. The response rate to our follow-up survey was 46%, and thus, the results should be interpreted with some caution. However, there is little evidence of systematic selection when examining the baseline characteristics of those who did not respond to the follow-up survey (see Table D.15). Subjects not responding to this survey tended to be older and were more likely to be African-American than the respondents. Nonetheless, most characteristics are not significantly different between the two samples.

Subjects receiving any incentive during the experiment had significantly higher sleep quality 3 months after the end of the experiment compared to those not treated (see table D.16)[22]. Additionally, those who had a higher achievement rate for their bedtime targets also reported better sleep quality. There is little evidence of changes in self-reported health, but treated subjects were 1.4 percentage points less likely to report very poor health status.

Finally, we investigated the qualitative effects of our intervention on academic performance (see Table 2.6). Sleep duration and regularity were found to be directly related to grade changes—those who slept longer, slept more regularly between 7 and 9 hours, and were less likely to sleep less than 6 hours experienced larger grade increases than those who did not. Having greater variance in sleep was associated with decreases in self-reported percentile rank. Being a part of any treatment group is associated with a 6.3-point increase in percentile rank with respect to one's own percentile rank before the experiment. Having a higher rate of compliance with the treatment, through meeting the target, was also associated with an increase in the letter grade—those who met the target more than 50% of the time had a 0.162-point greater increase in their letter grades than those who met their target less than 50% of the time.

2.5.2 Incentives to Sleep and Time-Use Allocation

A natural question is whether and how the allocation of time changed in response to our intervention. Individuals may compensate insufficient sleep at night by napping during the day or by sleeping longer during the weekend. Other studies find significant effects of naps on productivity and well-being (Bessone et al., 2018; Monk et al., 2001). We investigated whether our intervention affected the time allocated to naps. Only 5% of the subjects reported any sleep lasting less than 2 hours between 7 am and 7 pm during weekdays. Although nap duration is negatively correlated

[22] Because different waves of the experiment had different follow-up sleep questions, their answers are made into z-scores in order to be comparable across waves. Waves 1 through 4 had questions about the number of days without enough sleep, the number of days the subject nodded off, the percent of time their bedtime was before midnight, and the percent of time the subject slept more than 7 hours. Wave 4 also had a question about the number of sleep hours per night. Wave 5 only had questions on the number of sleep hours and whether their sleep habits had improved since the period before experiment.

with sleep duration at night (-0.13) and individuals sleeping between 7 and 9 hours are significantly less likely to report any naps (-3.89%), we find no evidence that our intervention systematically affected the likelihood of taking a nap and the nap duration (see Table D.17, columns 1–4). Thus, unsurprisingly, the results are substantially unchanged when we include controls for napping behavior (columns 5–6). We also find no evidence of subjects changing their weekend sleep duration during the intervention in response to the longer sleep duration induced by the incentives during the week. In fact, during the three weeks of the intervention, treated subjects were sleeping longer also during weekend (Table D.18). Although the effects are less precisely estimated than when analyzing the treated nights, the point estimates are not statistically different. Overall, these results are consistent with habit formation.

The subjects may also reallocate their time devoted to other activities when receiving incentives to go to bed earlier and to sleep longer. Using time-use diaries, we directly examine the effects of our incentives on individual time allocation. Time-use data are available for approximately 72% of the participants, and thus, results should be interpreted with some caution. The subjects not responding to the time-use surveys were younger, more likely to be Blacks, and were 10% more likely to report less than 7 hours of sleep during a typical night of the term, although this difference is only marginally significant (see Table D.19).

As mentioned above, consistent with previous evidence (Lauderdale et al., 2008a), we find that individuals tend to overestimate sleep when using time-use diaries. Indeed, there is no evidence that the subjects sleep longer during treatment when using time-use data and examining the likelihood of reporting between 7 and 9 hours of sleep (Table 2.7, columns 1–3). However, individuals do report significantly lower likelihood of sleeping less than 6 hours (-66% with respect to the mean, column 3). When examining other activities, we find no significant evidence that the increase in sleep duration was associated with a change in time spent studying, working, on personal care activities, exercising, relaxing, hanging around with friends or on the Internet, although we may be have not sufficient statistical power to identify some of these effects. Interestingly, the only activity that is systematically and significantly less likely to be reported under the intervention is "watching TV videos" (column 5, panel B). Indeed, for those who complied with the treatment, screen time after 8 pm declined by 13 minutes (see column 1 of Table D.20), equivalent to a 48% reduction with respect to the average screen time (45 minutes) observed in the sample. Among those who achieved the target at least half of the times, the coefficient decreases by 40% after the incentive is removed, but it is still economically and statistically significant. We find similar results when considering the likelihood of spending any amount of time on TV, Internet, or video games. Those achieving the target during the intervention are 12.5 percentage points less likely to report any screen time. This is equivalent to 33% of the sample mean. After the incentive is removed, the subjects who achieved the target on most nights are 8 percentage points less likely to report any screen time (a 20% effect with respect to the mean).

These results are consistent with recent research suggesting that screen time near bedtime is associated with lower sleep duration (Billari et al., 2018; Nie and Hillygus, 2002; Twenge et al., 2017).

2.5.3 Additional Findings: Sleep Regularity, Structure and Size of the Incentives

This section reports some additional findings regarding the effect of our intervention. Interestingly, the monetary incentives affected the regularity of sleep, bedtime, and waking time, reducing their variance (Table D.21). However, these effects did not persist after the removal of the incentive.

We do not find statistically significant differences when examining the role of the frequency and the structure of the incentives (Table D.22). In fact, the weekly incentive has stronger post-treatment effects, although these differences are not precisely estimated, and thus, should be interpreted with caution.[23]

Finally, we explore the role of incentive size. Using a smaller monetary incentive and eliminating loss framing leads to effects that are smaller and non-precisely estimated. In particular, the effects of weaker incentives on the likelihood of reporting between 7 and 9 hours are about 50% lower and non-significant (Table D.23). Furthermore, while the effect of larger monetary incentives survives the inclusion of individual fixed effects, the point estimate of the weak incentive treatment is close to zero. Unsurprisingly, given the lack of in-treatment effects, we find no evidence of post-treatment effect. However, consistent with what we found earlier, the effects are larger and more precisely estimated when focusing on the likelihood of sleeping less than 6 hours. Pooling all the treatments (1–3) in one, we substantially confirm the main results (see Table D.24) while increasing the precision of the point estimates as the sample size increases.

2.6 Conclusion

Statistics reveal that many individuals sleep less than the recommended number of hours. There are several factors affecting individuals' sleep choices. Understanding how to improve health habits is crucial in designing policies aimed at promoting healthier behavior. As pointed out by Charness and Gneezy (2009), people tend to underestimate the impact of current actions on future utility and discount the future too much. Our evidence suggests that this tendency also characterizes sleep behavior. The prevalence and persistence of behavioral biases in the sleep domain is particularly interesting given the repeated feedback individuals receive on sleep throughout their lives (Huffman et al., 2018).

We studied sleep choice, and whether commitment devices as well as monetary incentives can improve sleep behavior among students. We find supportive evidence for sophisticated time-inconsistent preferences in sleep choice. The subjects in our experiment chose commitment devices even if this meant a lower monetary reward in expectation. Present-biased subjects were more likely to be sleep deprived at baseline, but many of them committed to dominated bedtime or sleep duration targets. Subjects choosing more demanding targets were also more likely to achieve them,

[23] In the biweekly treatment, we regard as post-treatment any week after the first week of treatment during which subjects did not receive a monetary incentive. Using an alternative definition and focusing only on the last week of the experiment (week 8), we find similar results.

with the exception of those who were classified as overconfident. Indeed, many subjects tended to be overconfident in their own sleep duration and quality and were more optimistic about themselves than about others when assessing the risks associated with insufficient sleep. Overconfident individuals were more likely to be sleep deprived at baseline and more likely to select overly optimistic targets, and thus less likely to achieve them. Risk aversion was associated with better sleep and a higher likelihood of achieving target during the intervention.

Our incentives improved sleep behavior and led to some habit formation effects, with subjects in the treatment groups sleeping longer even after the incentives were removed. Furthermore, monetary incentives increased sleep regularity, reducing the variance of bedtime, wake-up time, and sleep duration. Finally, we show that incentives to sleep may also have positive effects on academic outcomes, although these results are at best suggestive and further research is needed to establish this finding. When receiving incentives to sleep longer, individuals significantly reduced screen time (watching TV and videos), while time spent with friends, working, or studying were not affected. Overall, these results give us a more nuanced understanding of sleep choice. Despite many economic models regarding sleep as an exogenous and homogeneous constraint on time, we provide evidence that behavioral biases play an important role and affect the heterogeneity of choice.

Our findings suggest that dynamic inconsistency and overconfidence can persist in the face of extensive experience and feedback. Thus, interventions only based on information (i.e. educational programs on sleep hygiene or fatigue management) may not be effective in the presence of these behavioral biases. Self-control problems may lead to procrastination with subjects repeatedly placing higher weight on immediate outcomes, and constantly delaying the start of good sleep habits. Also, people with motivated beliefs may be able to suppress the recall of objective feedback challenging their self-image, so that the simple provision of information may be ineffective in correcting misperceptions. Yet, to the extent subjects become more aware of their time inconsistent preferences due to the repeated feedback, sleep is also a domain where demand for commitment may be relevant and commitment devices effective. We show that appropriate incentives can be used to improve an individual's sleep behavior. Incentives to go to bed earlier and to sleep longer sleep were effective, suggesting that there is a cost to sleep, either in effort or in alternative uses of time, which can be compensated with a monetary payment.

Our findings also suggest that commitment devices and incentive structures may be more effective than planning tools at improving sleep behavior (Handel and Kolstad, 2017), and that temporary interventions, as those adopted by some companies, may have persistent effects, particularly when individuals lack a commitment device in natural settings. Providing incentives and commitment devices may help time inconsistent and overconfident individuals improve their sleep habits. Incentives and commitment devices may promote better sleep behaviors among subjects with self-control problems in the form of a time-inconsistent taste for immediate gratification (O'Donoghue et al., 2006), and with overconfidence as a result of motivated beliefs resilient to repeated information (Bénabou and Tirole, 2016). Incentives may also mitigate the role of motivated reasoning (Zimmermann, 2019). At the same time, our results imply that interventions that help individuals form routines conducive to healthy sleep habits (i.e., reduced screen time) may have longer-lasting effect.

The results on academic achievement, self-reported health, and heart rate efficiency support the growing evidence suggesting that sleep is a fundamental input for human capital and health. Taken together, the evidence on the behavioral factors behind sleep choice and the direct effects of sleep on health and productivity indicates the importance of not treating sleep as a mere time constraint, as well as the need to account for its direct effects on productivity of waking hours. The potential effects of our intervention on post-treatment sleep behavior, health outcomes, and human capital suggest the significance of further research along this line. Future research efforts exploiting larger samples may shed further light on the human capital and health effects of interventions aimed at improving sleep duration and quality. Future studies could also explore the relative effectiveness of non-monetary incentives and alternative commitment devices in nudging individuals into healthier and persistent sleep habits.

2.7 Figures

Figure 2.1: Design Illustration

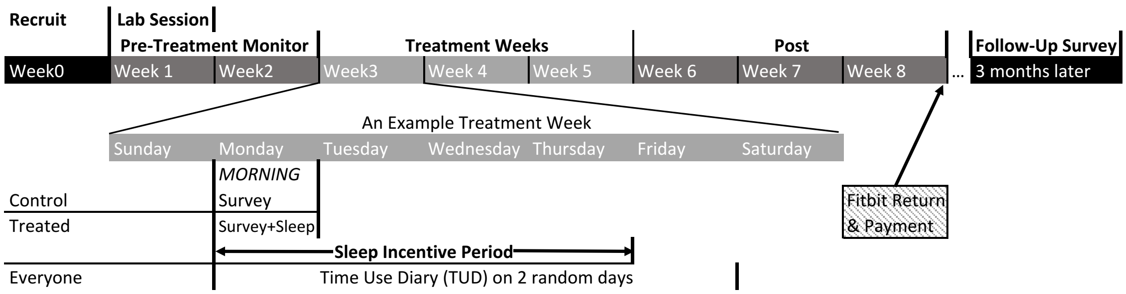
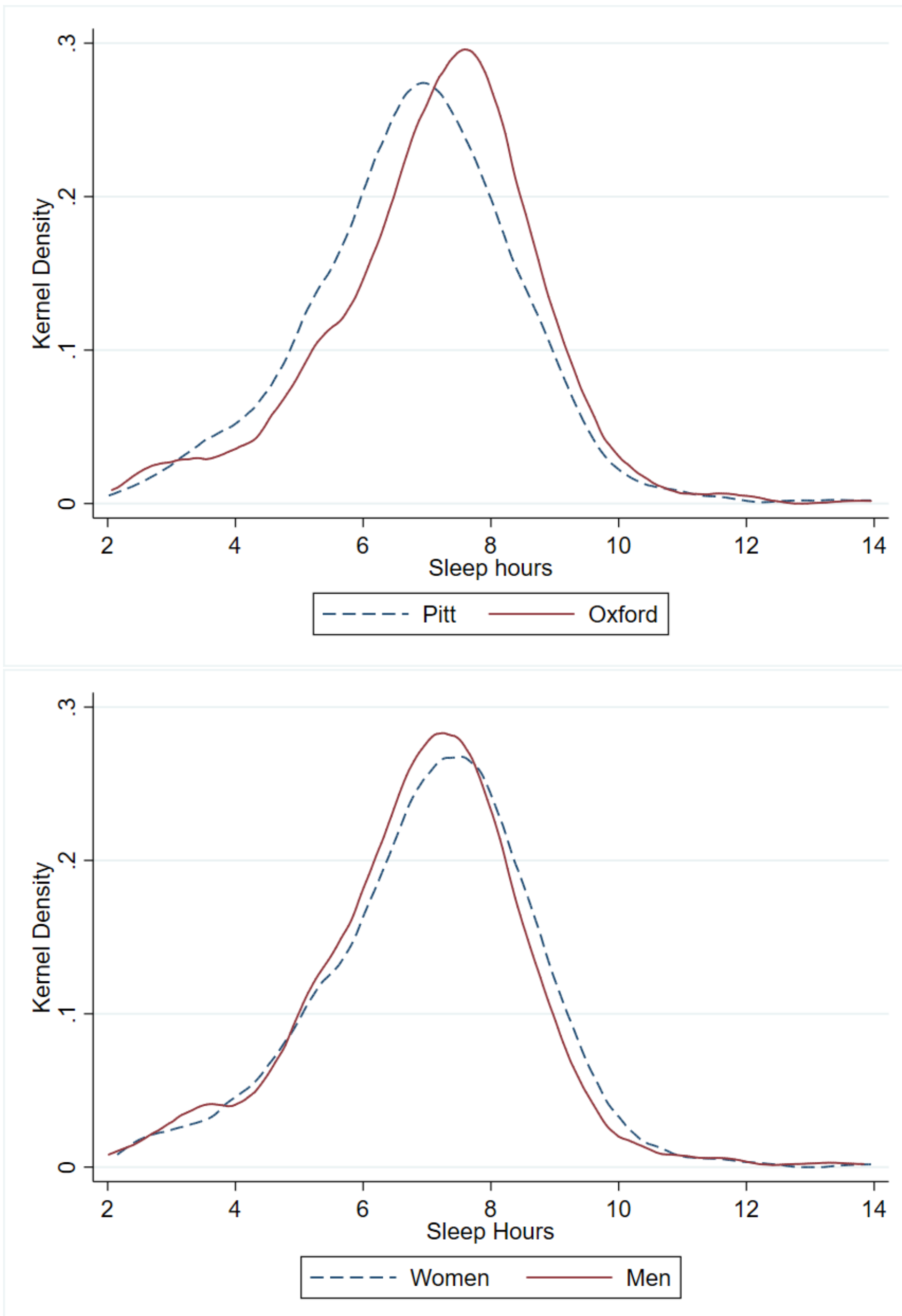
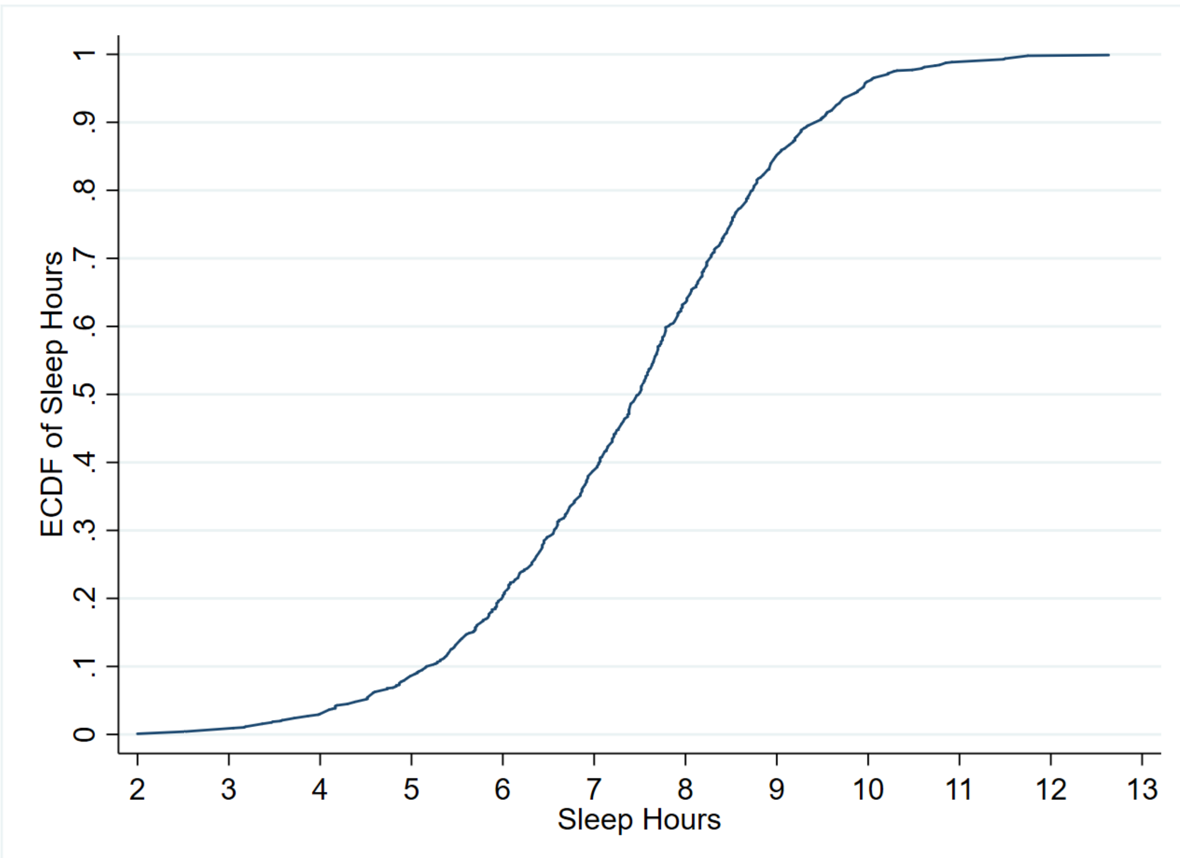


Figure 2.2: Sleep Duration Before Intervention



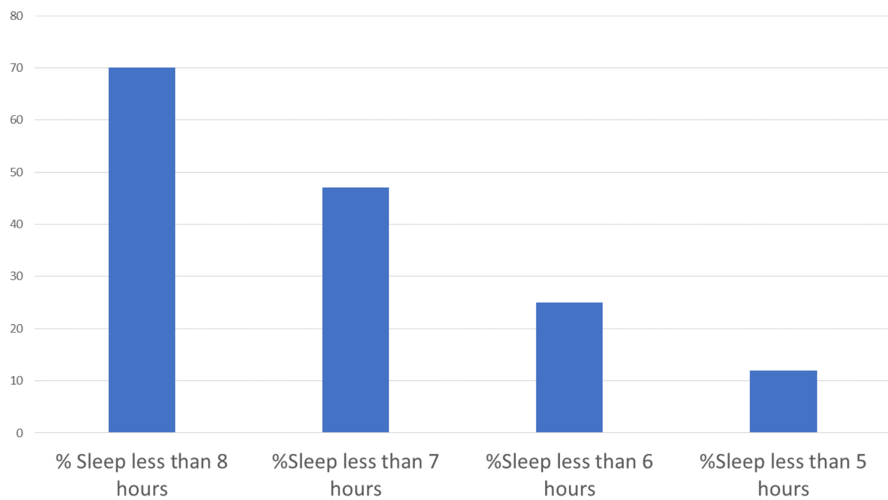
Notes - Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.

Figure 2.3: Sleep Duration Before Intervention



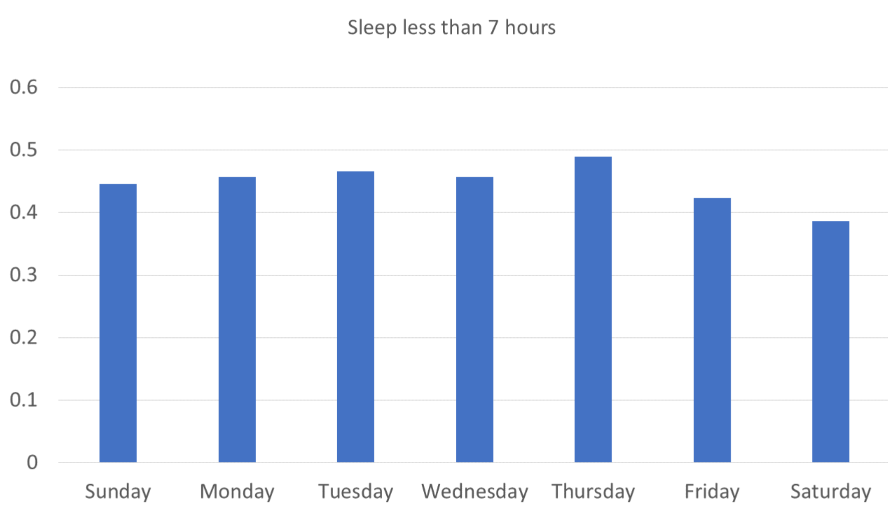
Notes - Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.

Figure 2.4: Insufficient Sleep



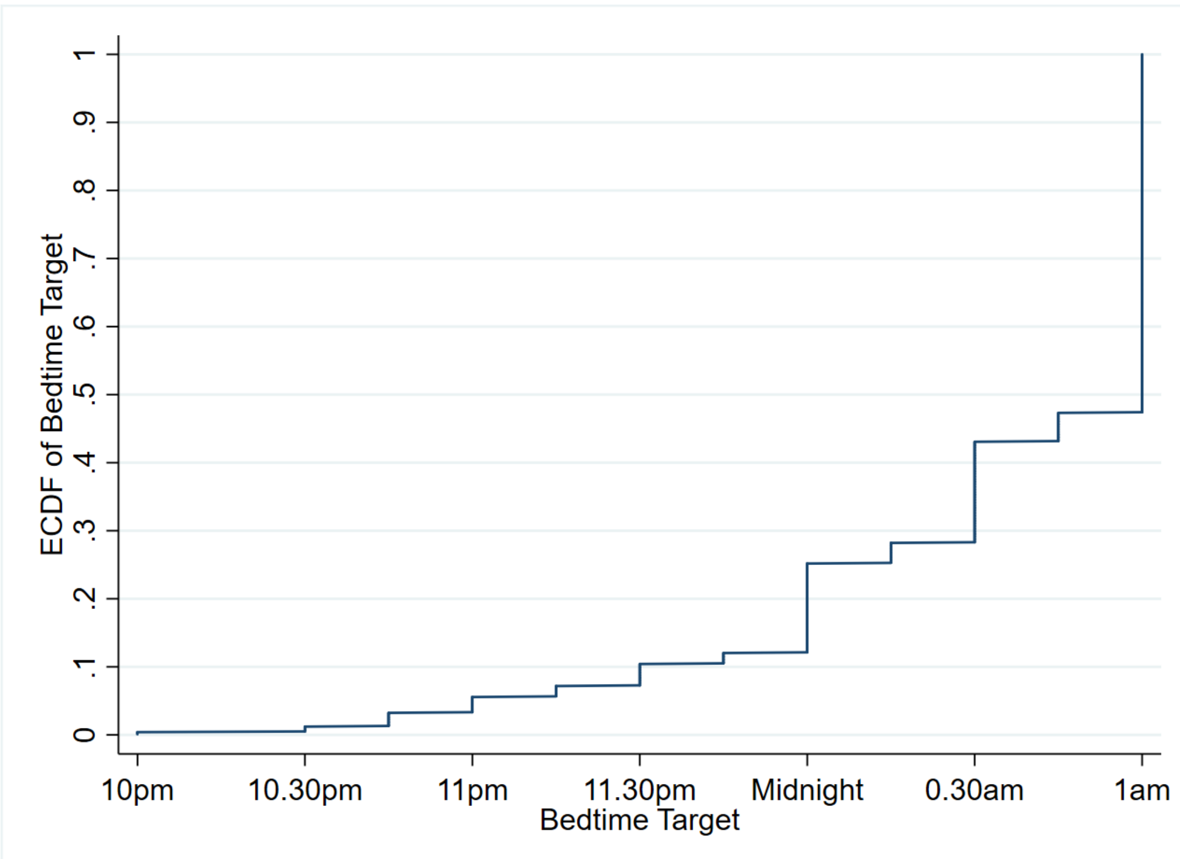
Notes - Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.

Figure 2.5: Sleep Duration Over the Week



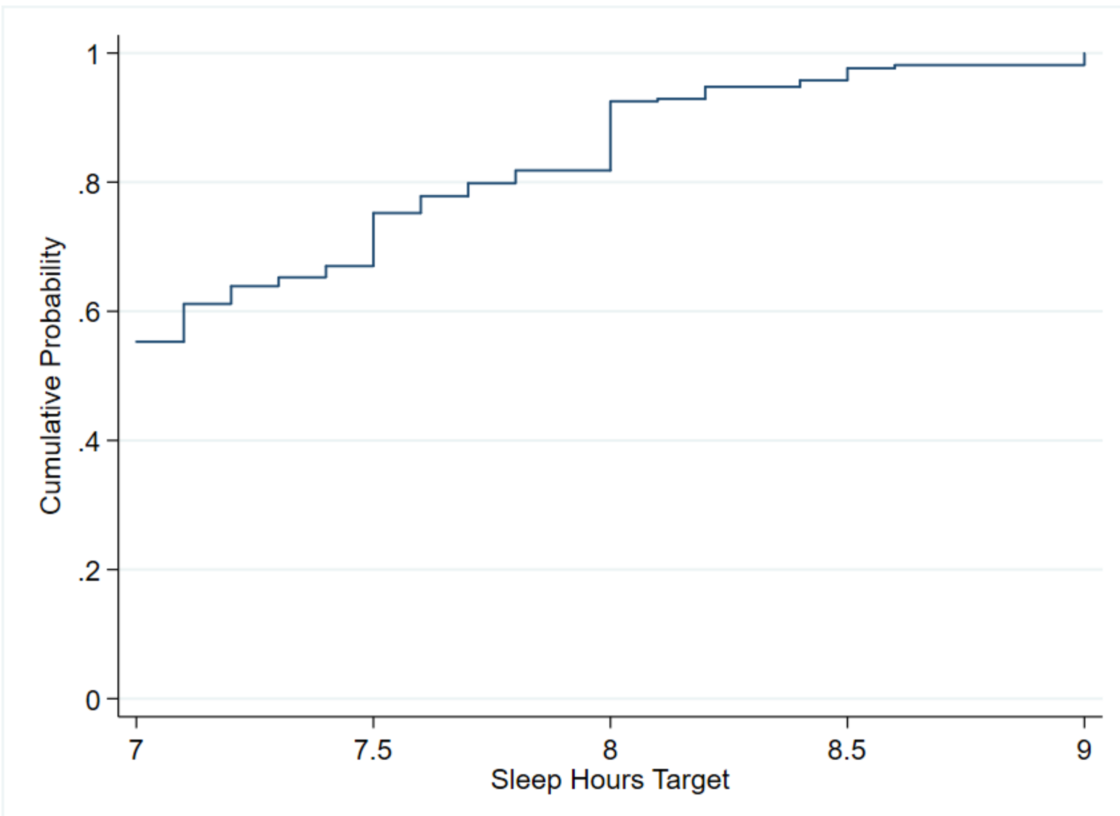
Notes - Data are drawn from the Fitbit devices during week 1 and 2 of the experiment before starting the intervention.

Figure 2.6: % of Subjects Choosing Bedtime Before 1 am



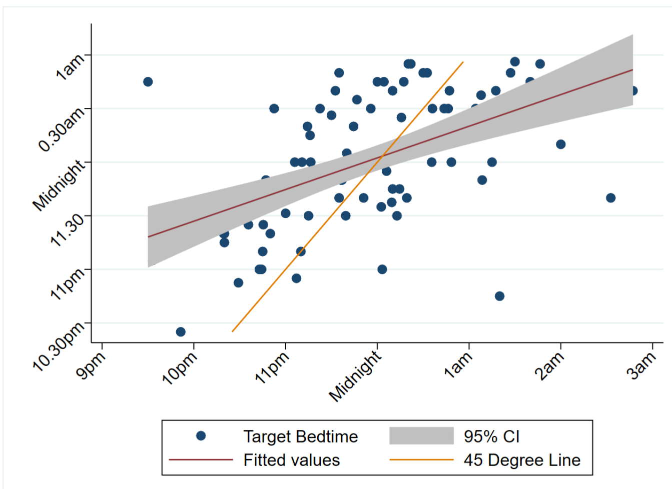
Notes - Data are drawn from the weeks of intervention.

Figure 2.7: % of Subjects Choosing Sleep Target >7 hours



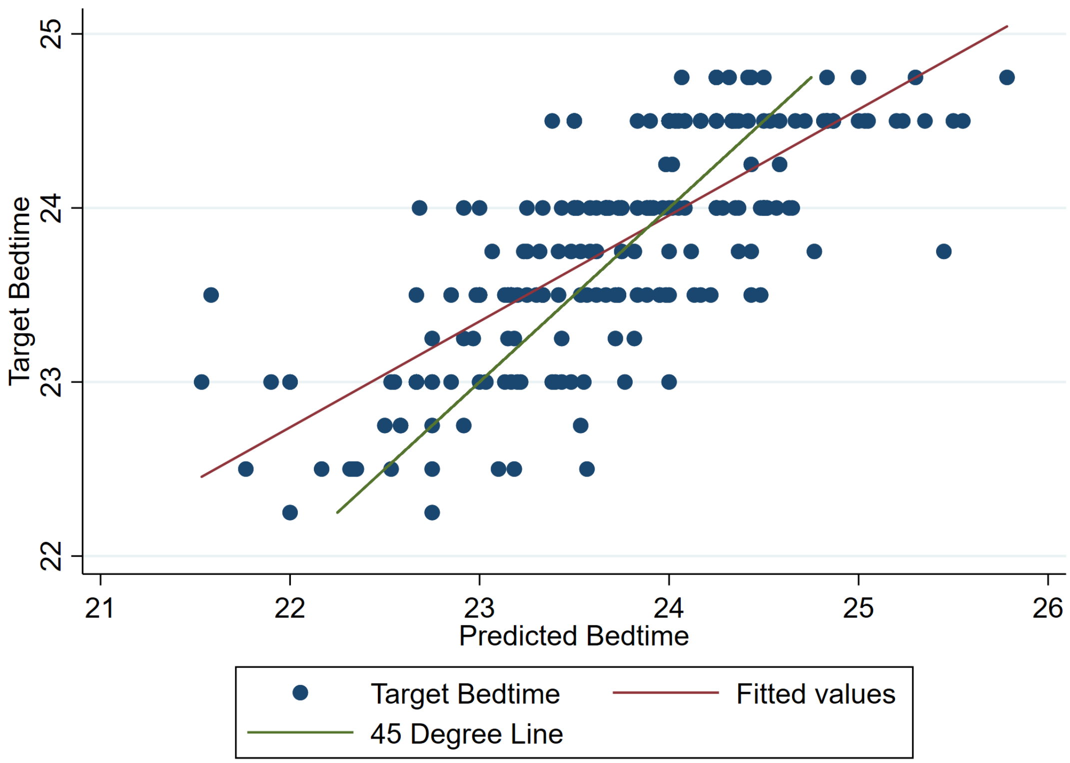
Notes - Data are drawn from the weeks of intervention.

Figure 2.8: Bedtime Targets and Pre-treatment Bedtimes



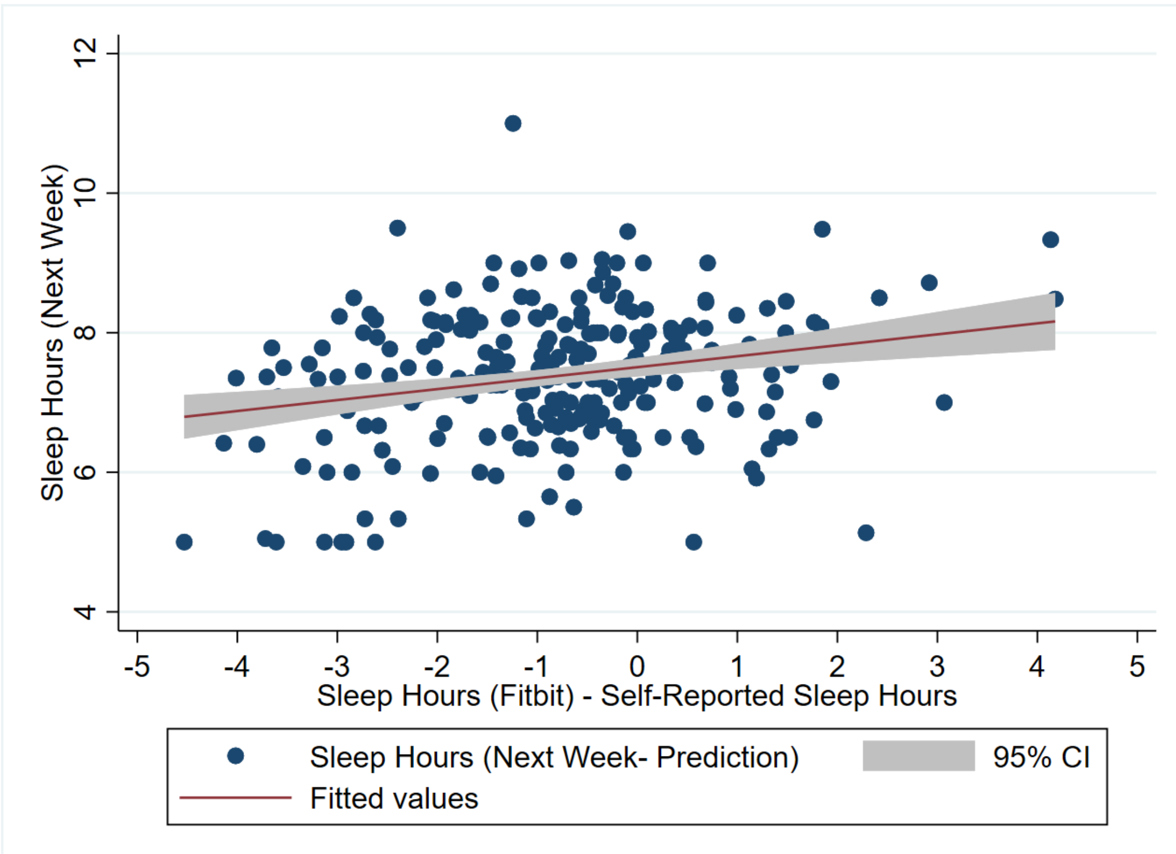
Notes - The figure presents average pre-treatment bedtime and target bedtime in the first treatment week by subjects. Bounded observations have been removed to alleviate bounding concerns.

Figure 2.9: Bedtime Targets and Predicted Bedtime



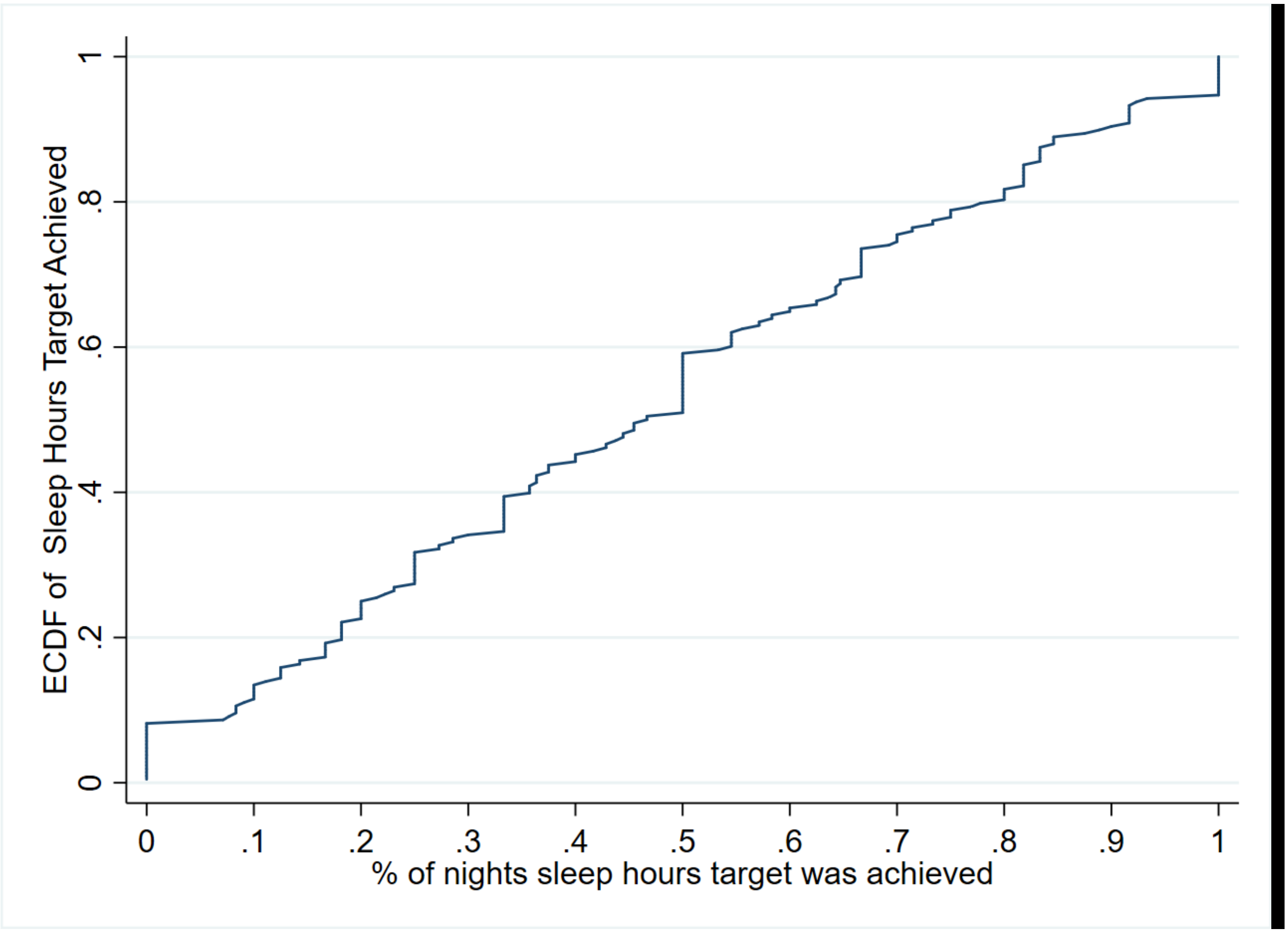
Notes - The figure presents predicted and target bedtimes by subjects. Bounded observations have been removed to alleviate bounding concerns.

Figure 2.10: Overconfidence and Beliefs



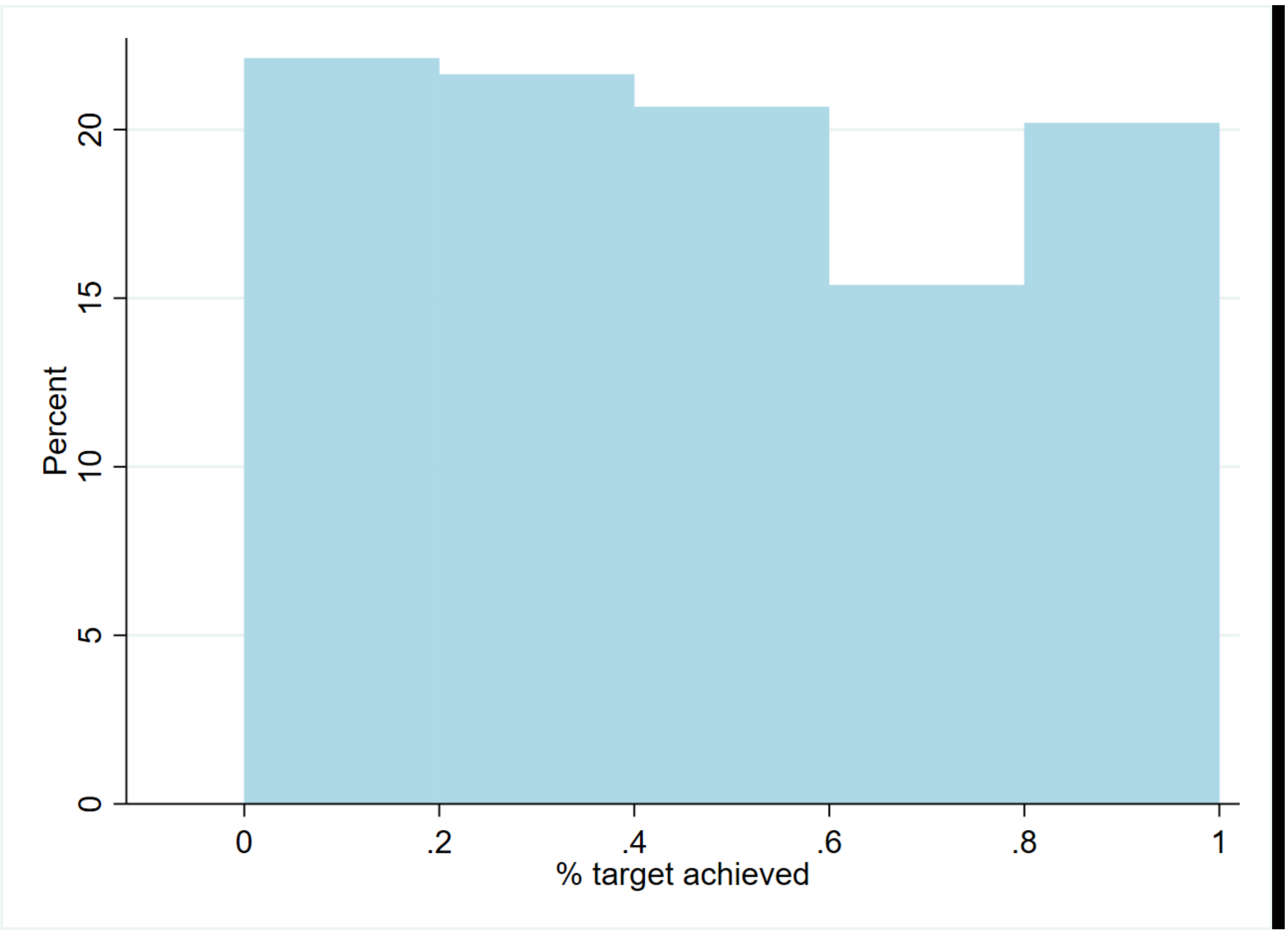
Notes - The figure plots the relationship between how many hours subjects predict to sleep in the following week and the difference between self-reported sleep and sleep as measured by fitbits the night before the survey.

Figure 2.11: Achievement



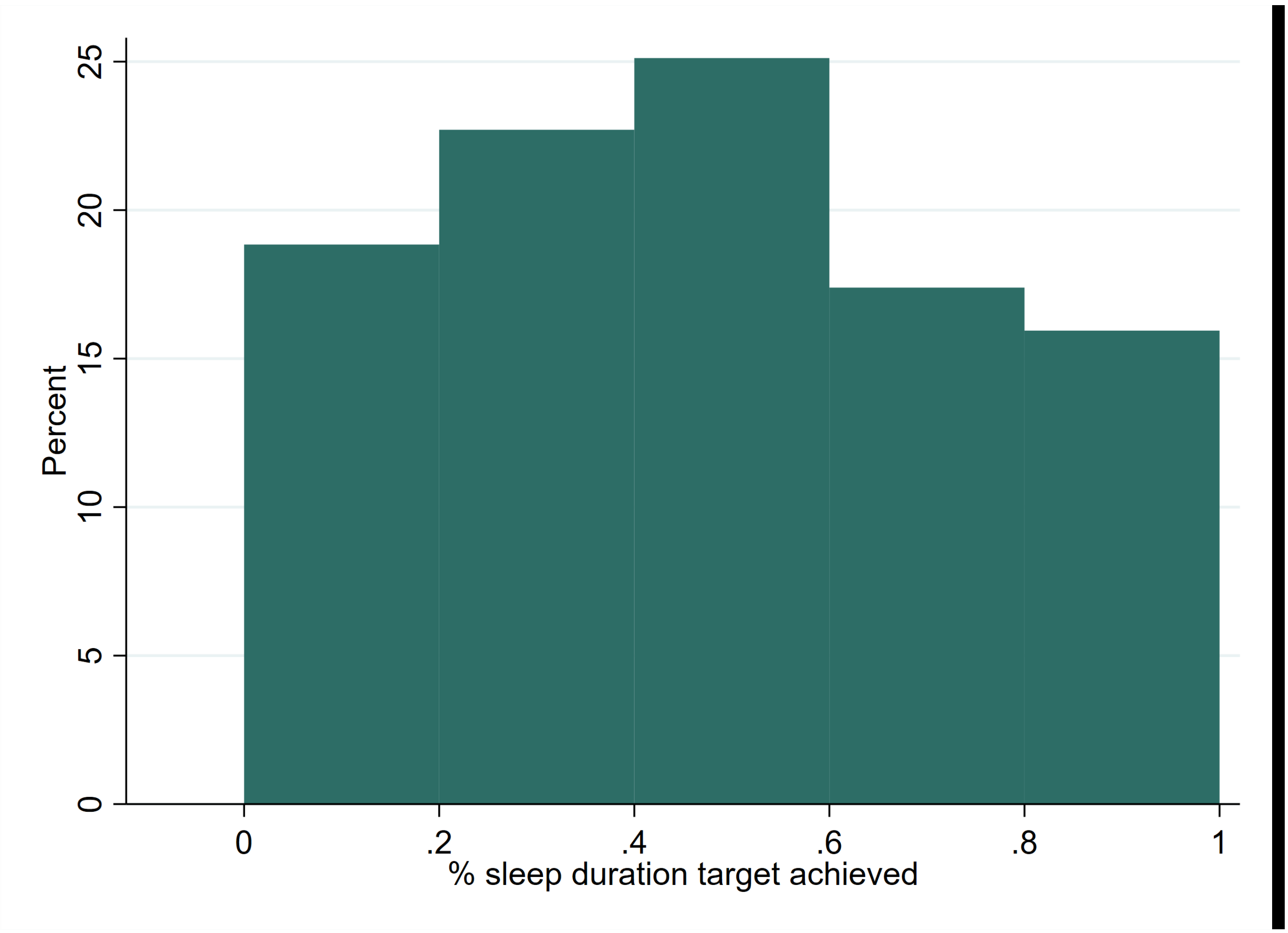
Notes - Data are drawn from the weeks of intervention.

Figure 2.12: % of Nights Bedtime Target was Met



Notes - Data are drawn from the weeks of intervention.

Figure 2.13: % of Nights Sleep Duration Target was Met



Notes - Data are drawn from the weeks of intervention.

2.8 Tables

Table 2.1: Time Preferences and Sleep Duration at Baseline

	(1)	(2)	(3)	(4)	(5)	(6)
	Self-reported (Day 1 Survey)			Actual Sleep (Fitbit)		
	Sleep hours	Sleep<7hrs	7≤Sleep≤9	Sleep hours	Sleep<7hrs	7≤Sleep≤9
Panel A: Present-Bias						
Present-biased	-0.1067	0.0438	-0.0566	-0.2914	0.0929	-0.1034
	(0.142)	(0.069)	(0.069)	(0.284)	(0.069)	(0.068)
Observations	319	319	319	319	319	319
Mean of Dep. Var.	6.845	0.458	0.522	7.078	0.465	0.468
Std.Dev. of Dep. Var.	0.984	0.499	0.500	1.979	0.500	0.500
Panel B: Impatience						
Impatient	0.2374	0.0140	-0.0381	0.1315	-0.0352	0.0099
	(0.269)	(0.066)	(0.067)	(0.243)	(0.066)	(0.066)
Observations	319	319	319	319	319	319
Mean of Dep. Var.	6.895	0.462	0.516	6.895	0.500	0.465
Std.Dev. of Dep. Var.	1.380	0.499	0.501	1.667	0.501	0.500
Panel C: Risk Aversion						
Risk Averse	0.2873**	-0.1507**	0.1747**	0.2043	-0.0878	0.0970
	(0.134)	(0.073)	(0.073)	(0.341)	(0.076)	(0.077)
Observations	319	319	319	319	319	319
Mean of Dep. Var.	6.845	0.458	0.522	7.078	0.465	0.468
Std.Dev. of Dep. Var.	0.984	0.499	0.500	1.979	0.500	0.500

Notes - Data are drawn from the first-day survey (columns 1-3) and the Fitbit data for the first two weeks of the experiment before intervention (columns 4-6).

Table 2.2: Perceived Own and Other's Sleep Quality and Sleep Deprivation Risks

	Own		Others		Difference	
Variables	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
Sleep quality (1-10)	6.67	1.58	6.08	1.35	0.58	1.73
Sleep duration	6.92	0.91	6.60	0.97	0.31	1.24
Sleep deprivation risks for:						
Mental alertness (1-100)	25.96	12.80	59.73	24.26	-33.93	24.76
Weight gain (1-100)	39.20	24.54	51.17	22.40	-12.00	22.95
Insomnia (1-100)	23.10	17.86	35.32	21.95	-12.72	19.88
Getting a cold (1-100)	37.84	23.50	45.46	25.01	-7.88	20.86
Arterial (1-100)	30.65	21.98	34.51	22.16	-3.34	18.47
Average risk	31.81	13.02	45.40	16.78	-13.72	13.74
Observations	319	319	319	319	319	319

Notes - We report averages and standard deviations obtained from our day 1 of the experiment survey.

Table 2.3: Overconfidence and Sleep Duration (Self-reported vs Fitbit Data)

	Self-reported (Day 1 Survey)			Actual Sleep (Fitbit)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Sleep hours	Sleep<7hrs	7≤Sleep≤9	Sleep Hours	Sleep<7hrs	7≤Sleep≤9
Overconfident	0.8867***	-0.3441***	0.3695***	-1.1449***	0.2824***	-0.2179***
	(0.109)	(0.059)	(0.059)	(0.259)	(0.066)	(0.066)
Observations	319	319	319	319	319	319
Mean of Dep. Var.	6.845	0.458	0.522	7.078	0.465	0.468
Std.Dev. of Dep. Var.	0.984	0.499	0.500	1.979	0.500	0.500

Notes - Data are drawn from the first-day survey (columns 1-3) and the Fitbit data for the first two weeks of the experiment before intervention (columns 4-6).

Table 2.4: Incentives and Sleep

	(1)	(2)	(3)	(4)
VARIABLES	7<Sleep<9		Sleep<6 hours	
Treatment	0.0850***	0.0493***	-0.0584***	-0.0316*
	(0.024)	(0.018)	(0.022)	(0.016)
Post-Treatment	0.0418	0.0053	-0.0589*	-0.0422**
	(0.038)	(0.024)	(0.032)	(0.021)
Individual fixed effects		YES		YES
Observations	7,690	7,690	7,690	7,690
Individuals	280	280	280	280
Mean of Dep. Var.	0.453	0.453	0.250	0.250
Std.Dev. of Dep. Var.	0.498	0.498	0.433	0.433

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table 2.5: Incentives to Sleep and Other Outcomes (Fitbit Data)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Sleep Efficiency		Low Resting Heart Rate		Sedentary Minutes	
Treatment	0.1984	-0.2502	0.0392	0.0145	-14.2803**	-9.1444*
	(0.575)	(0.373)	(0.028)	(0.010)	(6.811)	(4.860)
Post-Treatment	0.6906	-0.3316	0.0848*	0.0026	-1.3966	5.2014
	(0.701)	(0.285)	(0.045)	(0.022)	(10.765)	(7.249)
Individual fixed effects		YES		YES		YES
Observations	7,690	7,690	7,690	7,690	7,690	7,690
Mean of Dep. Var.	92.65	92.65	0.203	0.203	720.3	720.3
Std.Dev. of Dep. Var.	8.459	8.459	0.403	0.403	143.7	143.7

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2, 4, and 6 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table 2.6: Correlations with Academic Achievement

	(1) Letter Grade Change	(2) Percentile Change
Sleep measures:		
7<Sleep<9	0.710***	8.421*
	(0.183)	(4.636)
Sleep<6	-0.557***	-14.49***
	(0.171)	(4.361)
Sleep Duration	0.0753**	1.654*
	(0.035)	(0.884)
SD of Sleep	-0.0619	-2.681*
	(0.0647)	(1.559)
Incentives:		
Any Treatment	-0.0459	6.304**
	(0.115)	(3.069)
Large Treatment	-0.0626	8.738**
	(0.127)	(3.636)
Compliance:		
Achievement Rate	0.422***	3.386
	(0.151)	(3.023)
High Achiever	0.162*	2.627
	(0.084)	(2.384)

Notes - Each cell reports the raw correlation between

Table 2.7: Incentives and Time Allocation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A	Sleep (hours)	7< Time Use Sleep < 9	6< Time Use Sleep	Total Study hours	Total Work hours	Total Care hours	Total Exercise hours
Treatment	0.041	-0.004	-0.045**	0.055	-0.106	-0.013	-0.022
	-0.144	(0.044)	(0.022)	(0.206)	(0.153)	(0.087)	(0.053)
Post-Treatment	-0.06	-0.066	-0.029	-0.103	0.107	-0.065	-0.062
	0.155	(0.045)	(0.025)	(0.246)	(0.170)	(0.100)	(0.072)
Observations	1,363	1,363	1,363	1,363	1,363	1,363	1,363
Number of id	215	215	215	215	215	215	215
Mean of Dep. Var.	8.212	0.591	0.0602	4.864	1.982	2.900	0.425
Std.Dev. of Dep. Var.	1.727	0.492	0.238	3.340	2.823	1.329	0.773
Panel B	Total Relaxing hours	Total Other hours	Total Social hours	Total TV&Internet hours	Total TV hours	Total Internet hours	Total Gaming hours
Treatment	0.051	-0.008	0.165	-0.114	-0.175*	0.112	-0.052
	(0.171)	(0.094)	(0.130)	(0.140)	(0.105)	(0.092)	(0.066)
Post-Treatment	0.244	-0.188	0.338**	-0.094	-0.328***	0.340***	-0.107**
	(0.223)	(0.119)	(0.167)	(0.166)	(0.117)	(0.126)	(0.050)
Observations	1,363	1,363	1,363	1,363	1,363	1,363	1,363
Number of id	215	215	215	215	215	215	215
Mean of Dep. Var.	4.066	1.550	1.800	2.266	0.858	1.145	0.263
Std.Dev. of Dep. Var.	2.511	1.530	2.065	2.076	1.376	1.569	0.848

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). All estimate include individual fixed effects. Standard errors are clustered at the individual level.

3.0 The Mortality Effects of Community Mental Health Centers (Co-Author: Jessica LaVoice)

The Community Mental Health Act of 1963 established Community Mental Health Centers (CMHCs) across the country with the goal of providing continuous, comprehensive, community-oriented care to people suffering from mental illness. In this paper, we construct a novel dataset documenting the rollout of CMHCs from 1971 to 1981 to identify the effect of implementing a CMHC on county level mortality rates, focusing on causes of death related to mental illness. We find evidence that CMHCs reduced suicide rates among whites between the ages of 15 and 24 by 4%. CMHCs were particularly effective in reducing deaths from homicide and alcohol in the non-white population, with non-whites experiencing a 5% decline in homicide rates and non-whites age 45 to 64 experiencing an 11% decline in deaths caused by alcohol. The effect on mortality for non-white people is focused in rural areas. These results suggest CMHCs were effective in reducing mental illness related mortality, particularly in populations with the greatest need and least access to alternative forms of treatment.

3.1 Introduction

Mental illness is a worldwide concern with negative impacts at both the individual and societal levels. As of 2016, 1 in 5 adults lives with a mental illness in the United States, and 1 in 25 lives with a mental illness serious enough to make them unable to actively participate in one or more parts of their lives (Abuse and Administration, 2018). In 2001 the World Health Organization estimated that serious mental illness cost the United States $193.2 billion each year in lost earnings alone and identified depression as the leading global cause of disability (2001). The negative impacts for the individual and society include lower employment rates and earnings, higher likelihood of violent or criminal behavior, and homelessness (Frank and McGuire, 2000). Between the personal distress caused by the symptoms of mental illness and the negative individual and societal effects of mental illness as describe above, the treatment and management of mental illness is of growing concern to national and international health organizations.

While certain forms of treatment have been shown to be effective at ameliorating some of these negative effects (Mintz et al., 1992; Zhang et al., 1999b,a; Lang, 2013), it is often difficult for people to access effective treatment. The private market for mental health care suffers from inefficiencies and inequities that makes it unable to cope with growing mental health care needs. Insurance providers have been wont to provide adequate coverage for mental health needs due to concerns about overuse and treatment efficacy (McGuire, 1981; Keeler et al., 1988; Mechanic, 2014) and insufficient mental health resources, such as psychiatric beds and medical providers, prevent patients who are seeking help from being able to get it (Mechanic, 2002; Sipe et al., 2015). Decreased earning ability, negative stigmas, and the decreased ability to make decisions associated with (serious) mental illness prevent those who need mental healthcare from accepting it (McGuire, 1981; Mechanic, 2002; Rowan et al., 2013; Mechanic, 2014). Furthermore, inequalities in access to care arise along income level and insurance coverage, both of which are negatively associated with the severity of mental illness (Schlesinger and Dorwart, 1984; Mechanic, 2002; Rowan et al., 2013; Mechanic, 2014). These inequalities also exist along urban/rural lines, with rural areas having fewer available services and patients in rural areas being less likely to be able to get treatment for mental illness from a specialized mental health care provider (Blank et al., 1995; Mechanic, 2014; Sipe et al., 2015). Between these inefficiencies and the positive externalities associated with treatment, mental health care is a prime candidate for government intervention.

Publicly provided community care provides an potential solution to these concerns with private market provision of mental health treatment, and the World Health Organization recommends community care as the gold standard worldwide (2001). Community care, as opposed to institutionalization, allows patients to live and work within the community while receiving outpatient treatment services. There is currently little evidence about the efficacy of community care.[1] The United States had it's first experiment with community care in the 1950s through 1980s with the passage of the Community Mental Health Act (CMHA) and the establishment of Community Mental Health Centers (CMHCs).[2] The goal of this program was to shut down the mental institutions that were providing mental health care to the seriously mentally ill, and to replace them with a system of community care.

In this paper, we utilize the United States' experiment with CMHCs to analyze the efficacy of publicly provided community care in terms of mental illness related mortality. We use the rollout of CMHCs over the course of 10 years - 1971 to 1981 - to identify the effect of implementing a CMHC in a county on mortality rates. To do this we construct a novel data set of the county-level location of CMHCs every two years within this time-frame to establish which counties ever received a CMHC and when. Also, we constructed a novel data set of state priority rankings, in which states ranked counties in terms of need for mental health resources. While the construction of CMHCs were intended to follow these priority rankings, we show that these rankings do not predict which communities ever got CMHCs, nor do they predict which communities got CMHCs earlier rather than later. The lack of correlation between state rankings and treatment assists in the identification of treatment effects, and we take additional steps to address concerns regarding potential correlations between pre-existing mortality trends and selection into treatment. Due to the structure of our data and the randomness in receiving a CMHC, we are able to causally identify the effect of receiving a CMHC through a generalized difference-in-difference strategy.

While this paper is the first to quantitatively analyze the effect of community care on mental illness-related mortality, the CMHA has received plenty of qualitative scholarly attention. The conclusions of this work have been overwhelmingly negative. This program has long been regarded as a failure, charged with the "general abandonment" of the formerly institutionalized mentally ill and disabled population (Rose, 1979) and "the creation of large popula-

[1] Community care is associated with less stigmatization of the mentally ill (Link and Cullen, 1986; Boyd et al., 2010) and better post-hospitalization outcomes for the mentally ill compared to no follow-up treatment (Stein and Test, 1980). This form of care has been found to be more cost effective than institutionalization (Weisbrod et al., 1980; Test and Stein, 1980; Weisbrod, 1983) based on cost of care and the usage decisions of patients.

[2] See Morrissey and Goldman (1986) for an account of the history of mental health treatment in the United States.

tions of 'homeless, deranged people' " (Grob, 1994). This literature highlights two main reasons why the CMHA was considered unsuccessful: the program was never fully funded and far fewer CMHCs were built than were projected to be needed (Rose, 1979; Grob, 1994); and CMHCs were neither designed nor incentivized to treat the seriously mentally ill that were being forcibly released from mental institutions (Gronfein, 1985; Grob, 1994).[3] Because the creation of CMHCs was wrapped up with the deinstitutionalization of the mentally ill and seen as an alternative to mental institutions, their inability to treat everyone that was deinstitutionalized led to the conclusion that they were ineffective and a failed policy experiment.[4] However, CMHCs ultimately served a different population, those that were not mentally ill enough to warrant institutionalization but who could not otherwise access treatment in their communities. Therefore, the goal of this paper is to consider the efficacy of community care as provided by CMHCs, rather than the success or failure of the Community Mental Health Act as a whole. Our paper is the first to isolate the effect of community care from the overarching failure of the CMHA outside of the political context of the United States' first attempt at publicly provided community care.

Despite the negative conclusions from the previous literature on the CMHA, this paper shows that CMHCs were effective in reducing mental health related mortality for those who were most at risk. We find evidence that CMHCs reduced suicide rates among whites between the ages of 15 and 24 by 4%. CMHCs were particularly effective in reducing deaths from homicide and alcohol in the non-white population, with non-whites experiencing a 5% decline in homicide rates and non-white adults experiencing an 11% decline in deaths caused by alcohol. The effect on mortality for non-white people is focused in rural areas. These results suggest CMHCs were effective in reducing mental illness related mortality, especially in populations with the greatest need and least access to alternative forms of treatment.

Due to the negative associations of CMHCs with the problems of deinstitutionalization and growing financial burden of mental illness on the national government, the CMHA was effectively disbanded in 1981 and replaced with mental health block grants to the states. Back of the envelope calculations, however, show that if the program was extended to the full 2000 CMHCs anticipated to be required to fully support the entire US population and were operational through the 1980s, 828 fewer deaths from suicide, 792 fewer deaths from homicide, and 234 fewer deaths from alcohol-related mortality would have occurred during the 1980s. These results only measure improvements in mortality; we would also expect improvements in other areas correlated with these improvements in mortality, such as improved quality of life, increased employment, increased earnings, decreased homelessness, and decreased

[3]This is due to the fact that the treatment of serious mental illness is much more costly and requires more resources and specialized training compared to the treatment of non-serious mental illness, and extra funding was not provided for providing treatment to the seriously mentally ill. In effect, the government policy treated every mental illness as equally costly, which then incentivized CMHCs to focus on non-serious, less expensive mental illnesses. Additionally, the community care setting could not provide day-to-day services, such as housing, basic care, and medication management, that patients in mental hospitals received and that this patient body needed assistance with in order to function in daily life (on the Homeless, 1990).

[4]By the beginning of our sample period, over half of the deinstitutionalization that would occur in this period had already happened (Gronfein, 1985). Additionally, because most mental health patients were at state mental health hospitals, deinstitutionalization can be seen as a state-wide effect - patients from all around a state would be sent to the state's mental health hospital, and upon release would be sent to another institution, such as a nursing home, or back to their home communities.

incarceration. As the United States and other countries address the growing need for mental health care, publicly provided community care is a potential solution that should be considered.

3.2 Background Information

3.2.1 Effect of Mental Health and Mental Health Treatment

Mental illness is associated with a host of costly problems for the individual and society. Mental illness has been found to cause increased unemployment and impoverishment (Hamilton et al., 1997), decreased labor market participation (Mullahy and Sindelar, 1993), decreased work hours and income conditional on employment (Benham and Benham, 1982; Bartel and Taubman, 1986; Ettner et al., 1997; Marcotte and Wilcox-Gok, 2003), higher absenteeism (French and Zarkin, 1998), more transitions into and out of the labor market (Roan Gresenz and Sturm, 2004), performance deficits (Lerner and Henke, 2008), and even decreased coworker performance (Ettner et al., 2011). The total economic burden of serious mental illness in 2002 was estimated to be \$317.6 billion, with \$193.2 billion due to lost earnings (Insel, 2008). Mental illness is also associated with decreased educational attainment (Currie and Stabile, 2006; Fletcher, 2008; Eisenberg et al., 2009), decreased social connectedness and problems with social relationships (Kirk, 1974; Bartel and Taubman, 1986; Kelleher et al., 1994; Kessler et al., 1998; Teitler and Reichman, 2008), violent and criminal behavior (Link et al., 1992; Steadman et al., 1998), incarceration and homelessness (Jemelka et al., 1989; Harcourt, 2011; Raphael and Stoll, 2013), decreased quality of life (Saarni et al., 2007), and excess mortality (Brown et al., 2000; Saha et al., 2007; McGrath et al., 2008; Druss et al., 2011; Thornicroft, 2011), especially for the seriously mentally ill. While this list provides a plethora of outcomes that CMHCs could have affected, we focus on mortality due to this being the most extreme and substantial concern related to mental illness.

There is limited evidence evaluating the impact that mental health treatment has on economic outcomes. The treatment of depression using medication and psychotherapy has been found to decrease depressive symptoms and return depressed people to work (Mintz et al., 1992), making treatment either cost-neutral or beneficial based on the increased earnings and the cost of treatment (Zhang et al., 1999b,a). The effect of changing access to treatment is even less well understood. Lang (2013) finds that laws requiring that health insurance include mental health benefits at parity with physical health benefits lead to a decrease in the suicide rate by 5%. However, these benefits are restricted to the subset of the population with health insurance, which is likely not universal in the population of the mentally ill due to the negative employment and earnings effects of mental illness. This will attenuate the potential effect that increased access could have on mental health related mortality. Because CMHCs were available to everyone regardless of income or insurance coverage, we will be able to get a better grasp of how increases in access would affect mortality for the entire population.

3.2.2 Evidence on Publicly-Provided Health Care

While the public provision of mental health care has a history mostly limited to the CMHCs, there has been more attempts to provide (physical) health care through public provision. A recent federally funded program to provide free breast and cervical cancer screenings, analyzed in Bitler and Carpenter (2019), increased the probability of being screened by about 3 to 6 percentage points for eligible women. In a more historical context, Bailey and Goodman-Bacon (2015) analyze another community health program: Community Health Centers (CHCs). Compared to CMHCs, CHCs provided care in a similar way and to a similar population, but the care they provided was for physical, rather than mental, health. They find a 2% decrease in mortality for people ages 50 and over, with effects concentrated in cerebrovascular diseases. Additionally, they find that the timing of the roll out of CHCs had little to do with underlying need or planned development, citing the "great administrative confusion" that led to CHCs being approved and developed almost at random. We show that the roll out of CMHCs was also inconsistent with prior plans and priority rankings.

3.2.3 Historical Context of Community Mental Health Act

During the early 1960s, community-based care was considered a "bold new approach" to the treatment of the mentally ill. To stimulate the usage of community care techniques, President Kennedy signed the Community Mental Health Centers Construction Act of 1963. This act provided a three-year authorization for grants totaling $150 million to fund the development and construction of Community Mental Health Centers (CMHCs) across the country.

This act mandated community focused treatment centers be regionally planned and oriented toward prevention. To be eligible for federal funds, states had to first submit a comprehensive plan to the Department of Health, Education, and Welfare (HEW). This state plan was required to designate an agency to administer the plan, as well as an advisory council with broad representation. State mental health planning reports outlined the condition of mental health services available at the time of compilation, the mid-1960s, and contained recommendations for improvements.

The state plan was also required to develop a proposal for the construction of community mental health centers. The construction plan defined and prioritized catchment areas serving a population of 75,000 to 200,000 people. When determining the priority of each catchment area, states were required to target especially needy populations and, as such, priority was largely determined by demographics correlated with the need for mental health services including median family income, the infant mortality rate, and alcoholism rates. Priority of catchment areas was also impacted by the current availability of mental health resources in each community, such as construction projects approved in previous years and the number of psychiatrists in an area.

For a project to be approved, it was required to have priority over other projects within the state. Funds were to be allocated to applicants in areas of greatest unfilled need and in the order of area priority, meaning the neediest catchment areas would be the first to get CMHCs and other areas could only receive CMHCs after those neediest areas were served. However, a 1971 report issued to Congress by the Comptroller General of the United States reviewed this process and found evidence of considerable discrepancy between funding guidelines and actual spending. For example, they cite that, although centers were supposed to be funded by need as specified in state plans, California and Florida had been funding centers with little regard to prioritization (Kenig, 1992).

Under President Johnson in 1965, amendments expanded the original legislation to include funding for staff. This funding took the form of staffing grants that lasted for 51 months but declined over the life of the grant using a sliding scale.[5] The construction and staffing grant were a "seed money" mechanism to encourage the development of community focused centers that would eventually be funded by third parties, such as patient fees, local and state funds, and fundraising. After 51 months, centers were expected to have generated adequate alternative funds. To qualify for staffing grants, centers were required to offer the following five services: inpatient services, outpatient services, partial hospitalization, emergency services, and consultation/education programs (Naierman et al., 1978).

In 1967, the CMHC construction grants were extended for three more years and staffing grants for an additional two years (Kenig, 1992). In the years that followed, it became apparent that centers would not be able to acquire adequate funds to replace federal funding by the end of 51 months and eligibility for staffing grants was extended to eight years. A more generous sliding scale was introduced in high poverty catchment areas that left as much as 70 percent of the initial grant in the last year, compared to 30 percent for non-poverty centers. By 1975, in an attempt to force centers to obtain higher levels of alternative funding earlier in their development, a new sliding scale was introduced that maintained high initial levels for both poverty and non-poverty centers, but funds declined at a faster rate.[6] Lastly, recognizing that the preventative functions of centers might be the first to be eliminated as funding diminished, the law provided the only permanent grant mechanism for consultation and education services.

The policies that mostly interacted with the CMHA and the efficacy of CMHCs were implemented at the national level. The Social Security Disability Insurance (SSDI) program implemented in 1956 and Supplemental Security Income for the Aged, the Disabled, and the Blind (SSI) program implemented in 1972, which provided income support to those whose age or disability made them unable to hold a job, gave credence to the idea that the formerly institutionalized mentally ill could survive in their communities, even without a job, hastening states' deinstitutionalization efforts (Grob, 1994). Additionally, the implementation of Medicaid and focus of greater funds towards the Medicaid program made state hospitalization economically infeasible; therefore, there was a shift towards nursing homes for the aged and those otherwise unable to care for themselves, and community care for others (Gronfein, 1985; Grob, 1994). Both of these policies, though cited as motivating deinstitutionalization, are not varying at the county level and the impacts can be interpreted as state-wide effects.

[5]This bill authorized $73.5 million dollars for three years (although funding could be spread over fifty-one months to ensure that new centers receiving grants in the second and third year of the program would have full funding). The sliding scale begin with 75 percent cost coverage and decreased to 30 percent (Grob, 1994).

[6]Funds decreased to a 30 percent federal contribution in the eighth year for poverty centers and a 25 percent federal contribution in non-poverty centers. Another provision of this amendment provided "distress" grants to some of the older centers that failed in finding adequate alternative funding. These grants were limited to a total of three years and mandated that 7 additional services be provided by the center.

Under the Carter administration, a new President's Commission on Mental Health was appointed to revisit the nations mental health needs and services. The outcome of the commission's work was a short-lived piece of legislation, the Mental Health Systems Act, which was passed in 1981, replacing the earlier Community Mental Health Centers Construction Act with a newly developed approach to providing mental health services.[7] However, the 1981 Omnibus Budget Reconciliation Act repealed most of the previous mental health legislation, including the CMHC Act and the Mental Health Systems Act, in favor of Alcohol, Drug Abuse, and Mental Health Block Grants to states. These block grants consolidated funding for services related to mental health, alcoholism, and drug abuse into a block grant starting in fiscal year 1982. Most CMHCs initially funded prior to 1982 received some portion of each State's allotment for as many years as they would have been eligible for basic staffing or operations support when first funded. However, the amount of the award to each center was not guaranteed.

By the end of the program in 1981, a total of $2,659.3 million was spent on CMHCs.[8] Despite the goal of having mental health care coverage nationwide through the establishment of over 2000 centers, only 781 CMHCs ever existed.

3.2.4 What Did CMHCS Do and Whom Did They Serve?

While these centers were originally viewed as an alternative to mental hospitals, most centers devoted their attention to the less severely mentally ill by offering preventative services, counseling and crisis interventions; ultimately, the centers served a drastically different purpose than originally intended (Grob, 1994). Figure 3.1 shows the number of patient care episodes reported annually from 1971 to 1975. By 1971, just under 300 centers existed that treated a total of 797,000 patient care episodes. By 1975, over 500 centers existed that treated over 1,961,000 patient care episodes. In general, most episodes were treated with outpatient services, although inpatient and partial services were also provided. While we don't have individual level information about patients, statistical notes compiled by the National Institute of Mental Health indicate that, in 1975, about 42% of patients where under the age of 25 and 39% were between the ages of 25 and 45.[9] We also know that whites made up a majority of patients in CMHCs, although the ratio of non-white to white patients is larger than the ratio of non-whites to whites within the population.

3.3 Data

The data used in this project was compiled from various sources. The locations of CMHCs were identified by digitizing the Mental Health Directories and the Directories of Federally Funded Community Mental Health Centers.

[7] Rather than providing twelve services in one center, the Systems act funded a phased-in system of services and rather than providing direct federal funding, the system called for providing money to states for distribution through individual state departments of mental health.

[8] $1,552.3 million was spent on staffing and construction grants combined and the rest on other various grants including distress grants for centers that could not obtain alternative funding and consultation and education grants which were the only "permanent" grants that did not decline on a sliding scale.

[9] Furthermore, females made up about 53% of the patient pool, although, males made up 53% of the patients under the age of 25.

These directories document the addresses of CMHCs and were published every other year from 1971-1981. The variation in centers across subsequent editions of these publications allows us to document the roll-out of community mental health centers nationwide over the 1970s[10]. The locations and roll-out of CMHCs can be seen geographically in Figure 3.2, which shows the first year a CMHC was established in a county, with the earlier centers shown with a lighter color.

We supplement these directories with data obtained from State Mental Health Planning reports, which were required before a county could be granted funding for community mental health centers. These planning reports defined catchment areas serving 75,000 to 200,000 people and, in a subsample of states, aggregated multiple counties into one planning area. Each planning area was ranked according to relative need[11]. We collect and digitize planning area and relative rankings for each state that defined planning areas using county boundaries.

Our primary outcome variables are age-adjusted mortality rates. Mortality data was obtained from the Multiple Cause of Death Vital Statistics published by the National Center for Health Statistics. This data contains the universe of civilian deaths reported by cause, age, and the decedent's county residence. We compute age-adjusted mortality rates from 1969 to 1988 using annual county population estimates from the Surveillance, Epidemiology, and End Results Program.

We consider the following causes of death due to their relationship in the literature to mental illness: suicide, homicide, and deaths caused by alcoholism.[12] We expect mental health treatment to affect homicide rates for two reasons: violent and criminal behavior is associated with untreated mental illness (Link et al., 1992; Torrey, 1994; Steadman et al., 1998), so greater access to treatment would likely decrease these behaviors and thus homicide; and mental illness, particularly severe mental illness, make people more vulnerable to being victims of all types of crimes, including homicide (Torrey, 1997; Hiday et al., 1999; Hiroeh et al., 2001; Teplin et al., 2005; Maniglio, 2009). We use an alternative measure of suicide which combines suicide deaths with accidental gun deaths; in the time-frame considered here, mental illness and suicide were still stigmatized in much of the country, and coroners would often mislabel suicides as accidental deaths to protect the family from the knowledge of the real cause of death or from the scrutiny of the community.

The age-adjusted mortality measures for our outcome variable of interest are shown graphically in Figure 3.3. The suicide rate remains at about 16 deaths per every 100,000 people over our time period of interest, with our alternative suicide definition trending similarly around 18 deaths per every 100,000 people. The homicide rate remains at about 8 deaths per every 100,000 people although it drops to about 6 deaths per every 100,000 around 1984. Deaths caused by alcoholism remain stable at about 2 deaths per every 100,000 over our sample period.

[10] Between 1966 and 1971 there is data that includes community mental health centers, though it is unclear from the documentation whether these centers were a part of the CMHA and were expected to follow the guidelines as such. In 1971 and later there are mental health locations called community mental health centers that were not included under the official category of federally funded community mental health centers, throwing doubt on whether those CMHCs in pre-1971 documents actually are CMHCs as defined and funded by the federal government.

[11] These reports were maintained by the National Institute of Mental Health and are available at the National Archives in Maryland.

[12] We do not include deaths caused by drug dependence because, as seen in Figure F.1, there were very few deaths caused by drug dependence during our sample period.

Lastly, we use county level controls such as educational attainment, labor force participation, and income measures from the decennial censuses. We use 1960 and 1970 data and linear interpolate values for non-census years. Table 3.1 shows averages of our variables of interest in 1960 in counties that would eventually receive a CMHC and counties that would not. We see that counties receiving a center had higher population rates and tended to be more urban. Counties that would eventually receive a CMHC had slightly higher educational attainment and a slightly higher labor force participation rate. Due to differences along these dimensions, we include linear trends of percent less than high school education, percent with high school education, percent of the population living in an urban area (split into 5 categories), unemployment rate, labor force participation rate as controls in each of our main specifications.

3.4 Identification Strategy

Our primary identifying assumption is that, if any difference between treated and untreated counties existed, CMHCs were more likely to be placed in counties with higher levels of pre-treatment mental health mortality rates or relative need. Thus, any selection bias would mitigate our results toward zero. This allows us to identify a lower-bound on the effects of CMHC on age-adjusted mortality by directly comparing counties that received a center to counties that did not.

We support this identifying assumption in multiple ways. First, we utilize state preliminary planning reports that rank catchment areas by relative need to test if centers were being placed in areas based on predetermined need.[13] These rankings were established by mental health professionals familiar with the need of mental health services across the state. While each state had its own ranking algorithm, states were required to target especially needy populations and, as such, priority rankings were largely determined by demographics correlated with the need for mental health services including median family income, the infant mortality rate, and alcoholism rates. We calculate z-scores for priority ranking for each state to determine if higher rankings correspond with an increased likelihood of receiving a CMHC or with the timing of CMHC rollout. Table 3.2 presents the results of this analysis and shows that having a higher priority ranking does not increase the probability that a county would receive a center or, conditional on receiving a center, that higher priority areas would receive centers sooner. This is consistent with a 1971 report issued to Congress by the Comptroller General of the United States that found evidence of considerable discrepancy between funding guidelines and actual spending (Kenig, 1992).

Furthermore, we show the extent to which mortality rates predict if a county would eventually receive a CMHC and the order in which counties received CMHCs. We regress pre-treatment age-adjusted mortality rates of interest on

[13] We have these planning reports for 29 states: Alabama, Arkansas, Delaware, Florida, Idaho, Indiana, Kansas, Louisiana, Maryland, Mississippi, Missouri, Montana, Nebraska, New Jersey, New Mexico, New York, North Carolina, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, and Wisconsin. In some instances, particularly in rural areas, centers were designed to treat multiple counties. Due to the large geographic catchment areas, we find it likely that CMHCs were not accessible to everyone they were meant to target. Thus, our preferred specification uses counties to characterize treatment status.

a binary variable which equals 1 if a county would eventually receive a CMHC using a probit regression model. The results of this regression are presented in Column 1 of Table 3.3. Only homicide rates predict treatment, with higher mortality counties being more likely to receive a center and to receive a center earlier. In Column 2 of Table 3.3 we present the results of the regression of year of CMHC opening in a county, conditional on ever getting a CMHC, on pre-treatment age-adjusted mortality rates. Here, again, we see that only homicide rates predict roll-out, with higher homicide rates predicting receiving a CMHC earlier. However, these pre-trends on homicide are entirely driven by homicides of white people, while our results are primarily on homicides of non-white people, so we are less concerned about the effect this relationship will have on the interpretability of our results.

Lastly, we show that pre-treatment mortality trends are not statistically different between counties that would eventually receive a center and those that would not. We show this result by estimating the following equation

$$y_{cst} = \alpha + \beta_1 Year + \beta_2 CMHC_c + \beta_3 Year * CMHC_c + \epsilon_{cst} \tag{3.1}$$

where y_{cst} is an outcome for county c in state s for year t and $CMHC_c$ is a binary variable indicating if county c ever received a CMHC. The coefficient of interest is β_3 which identifies any differential trends in age-adjusted mortality between treat and untreated counties during the pretreatment period. These results are presented in Table 3.4 which shows that β_3 is statistically and economically insignificant across all specifications. Additionally, in section 3.6.1 we will use an event study framework to show the differential pre-trends for our outcomes of interest.

Given that centers appear to be constructed with little regard to pre-existing mental health mortality trends or relative need as determined by mental health professionals within each state, we identify the effect of receiving a CMHC on county level measures of age-adjusted mortality using a generalized difference-in-difference empirical specification. Our outcomes of interest include age-adjusted mental health related mortality, including suicides, homicides, and deaths caused by alcohol dependence. We estimate the following regression equation:

$$y_{cst} = \alpha + \beta CMHC_{ct} + \theta_c + \gamma_t + \lambda_s * t + X_{ct} + \epsilon_{cst} \tag{3.2}$$

where y_{cst} is an outcome for county c in state s for year t, $CMHC_{ct}$ is a binary variable indicating if county c received a CMHC by year t, θ_c are county fixed effects, γ_t are year fixed effects, $\lambda_s * t$ is a state specific linear time trend, and X_{ct} is a vector of controls.[14] The coefficient of interest is β which estimates the average intention to treat effect. We also use subgroup analysis to explore if CMHCs were effective in mitigating mental health related mortality within certain demographics of the population, including race and age.

Mortality may be serially correlated within a county across years so we cluster our standard errors at the county level.[15] However, there is evidence that clustering standard errors (and even robust standard errors) will result in too-

[14]Controls include linear trends of percent less than high school education, percent with high school education, urban category, unemployment rate, labor force participation rate.

[15]See Bertrand et al. (2004) for more information about serial correlation and see Abadie et al. (2017) for more information about clustering standard errors.

conservative standard error estimates in this environment.[16] We present clustered standard errors and highlight that we are likely being over conservative in our estimation strategy.

3.4.1 The Expected Effects of CMHCs on Mortality Rates

The expected effects of CMHCs on mortality rates depends both on the incidence of causes CMHCs might prevent and the effectiveness/availability of CMHCs' care relative to alternatives. Figure 3.4 show age-adjusted mortality trends for different subgroups of the population. This figure shows our outcome variables of interest broken down by the following age groups: 15-24 years old (young), 25-44 (young adult), 45-64 (adult), and 65 and older (elderly). We see that suicide rates are highest among people over the age of 45, although rates are increasing for the young population over our sample period. Furthermore, the young adult population is the most likely to be murdered and the adult population has the highest rate of deaths caused by alcoholism. Given the differences in the baseline mortality rates of different demographic groups across each cause of death, we expect the effects of access to a CMHCs to differ along these dimensions.

We know that a disproportionate number of patients of CMHCs were between the ages of 15 and 24. Furthermore, educational and prevention outreach programs, consisting mostly of outreach to schools, was the only permanent funding provided to centers. Thus, we would expect to see a greater impact on the younger segment of the population.

It is unclear, ex-ante, whether CMHCs will have a greater impact in urban or rural counties. Centers in urban counties may have been more accessible, since rural counties are often large and have less extensive public transportation, leading to greater effects for urban counties. However, rural counties probably didn't have as many other mental health resources as urban areas did, meaning that a CMHC entering a rural community could have a greater effect. This conjecture is supported by statistical notes compiled by the National Institute of Mental Health: admissions to rural CMHCs were less likely to have had any prior mental health treatment compared to admissions to more urban centers, and those that had treatment were more likely to have had it from a non-mental health professional, such as a primary care doctor (Bachrach, 1974). Additionally, more current evidence shows that closures of (general) hospitals in rural areas causes a larger increase in mortality than closures in urban areas (Gujral and Basu, 2019), pointing towards rural areas as being more strongly affected by the opening or closing of new health care facilities.

3.5 Results

Our first set of results are shown in Table 3.5. Each entry presents the results for a different regression, with the outcome variables corresponding to suicide rates, homicide rates, and deaths caused by alcoholism. These results

[16] See Abadie et al. (2017) for more details about why clustering may be too-conservative. Furthermore, see Goodman-Bacon for a discussion about the interpretation of generalized Difference-in-Difference coefficients.

indicate that, on average, CMHCs did not mitigate suicide rates or deaths caused by alcoholism. We do find evidence that CMHCs caused a decrease in the homicide rate by approximately 4%.[17] Due to the differing baseline mortality rates across demographic groups for our causes of death, we analyze subsamples of the population and find that CMHCs were particularly effective on certain demographic groups. Panel B and C explore mortality rates for the white and non-white population respectively. We see that the decrease in homicide rates are primarily driven by the non-white population.

We explore the impact of CMHCs across different age and racial groups in Table 3.6 and Table 3.7. Table 3.6 focuses on different age groups within the white population. Each panel of this table shows the regression results for different segments of the population. Panel A limits our sample to only deaths among 15-24 year olds, Panel B to 25-44 year olds, Panel C to 45-64 years and Panel D to those who died at an age of 65 or older. Panel A shows how CMHCs decreased suicide rates of the young white population by approximately 4%. We find no effects of CMHCs on homicide rate or deaths caused by alcohol across any ages in the white population. Thus, while there is no statistical impact of CMHCs on county level mental health mortality for those over the age of 25, CMHCs were particularly effective of reducing suicide rates among the younger white population. This age group made up the highest patient share among CMHCs and was likely to benefit from low or no-cost treatment, providing further evidence that CMHCs were effective at decreasing the morality rates of people visiting centers.

Table 3.7 presents the same specifications for different age groups within the non-white population. While we see little effect on suicide rates among the non-white population, we see a decline in homicide rates among those over the age of 25, particularly among 25-44 year olds who experienced a 6% decline in homicide rates and those 65 and older who experienced an 11% decline. We also see a decline in deaths caused by alcohol among the adult non-white population. Having a community mental health center in one's county reduced the number of deaths caused by alcohol dependence among this population by 11%.

In addition to differential effects based on demographic characteristics, we could expect that urban and rural counties could have different outcomes.[18] Table 3.8 shows the results for non-white mortality are primarily driven by substantial decreases in homicide and alcoholism-related mortality rates in rural counties. Homicide rates among non-whites decreased by 13% in rural counties, compared to a statistically insignificant 1% in urban counties. Additionally deaths caused by alcoholism declined by 28% among non-whites in rural counties compared to a statistically insignificant 1% decline in urban counties. This is despite there being similar pre-treatment rates of homicide and

[17] Another policy that could interact with the outcomes of the CMHA and CMHCs is community policing. Community policing is a policing method that focuses on having the police present and interconnected with their community, with increased focus on maintaining order in the community, rather than just solving crimes. While there were some, mostly urban, trials with community policing during our sample period, community policing did not become wide-spread until the 1990s and has not been found to be effective in reducing crime (Cordner, 2014). Especially since most of the positive effects of CMHCs will be focused in non-urban areas and there is no evidence of substantial overlap between CMHC implementation and community policing policies being implemented in communities, it seems unlikely that these decreases in homicide are due to community policing.

[18] Urban and rural counties are defined as having above and below the median percent of the population in the county living in an urban area, respectively

alcoholism-related deaths for non-whites in urban and rural counties.[19]

3.6 Robustness Checks

In this section, we will consider two robustness checks to evaluate the impact of CMHCs on mental illness related mortality: an event study framework and a pseudo-boundary analysis.

3.6.1 Event Study Framework

In an event study framework, we estimate the impact of receiving a CMHC by leveraging the randomness in the timing of the roll-out of CMHCs across space. In this framework, the identification of effects is entirely from the sample of counties that would ever receive a CMHC, though the impacts of controls are estimated using never-treated counties. Additionally, the event study framework allows us to visualize whether there differing pre-trends by treatment.

We estimate the following regression equation:

$$y_{cst} = \alpha + \theta_c + \gamma_t + \lambda_s * t + X_{ct} + \sum_{y=-10}^{-2} \pi_y CMHC_c \mathbf{1}(t - T_c^* = y) + \sum_{y=0}^{15} \tau_y CMHC_c \mathbf{1}(t - T_c^* = y) + \epsilon_{cst} \quad (3.3)$$

where y_{cst} is an outcome for county c in state s for year t, $CMHC_c$ is a binary variable indicating if county c ever received a CMHC, θ_c are county fixed effects, γ_t are year fixed effects, $\lambda_s * t$ is a state specific linear time trend, and X_{ct} is a vector of controls. The indicator function $\mathbf{1}(t - T_c^* = y)$ indicate years away from T_c^*, the year the CMHC was implemented in that county. Years more than 10 before the implementation of a CMHC or more than 15 years after the implementation of a CMHC are represented with dummy variables. The coefficients π_y and τ_y provide estimates of the effect of getting a CMHC in the years prior to and after the CMHC was built in that community, respectively. In this environment, null values on the estimates of π_y would indicate no differential pre-trends and negative values on the estimates of τ_y would represent the treatment effects y years after treatment. Due to the relatively low incidence rate and the volatile nature of our outcomes of interest, we aggregate data into the following bins: more than 10 years before; 7 to 10 years before; 3 to 6 years before; 0 to 3 years after; 4 to 7 years after; 8 to 11 years after; 12 to 15 years after; and more than 15 years after. The category of 1 to 2 years before is omitted as the baseline year.

Figures 3.5 and 3.6 provide the event study graphs for our two main results: the decrease in suicides of whites and the decrease in homicides for non-whites. These results are consistent with the results found earlier, specifically that the implementation of CMHCs led to decreases in mortality from suicide for whites and homicide for non-whites. Additionally, we can argue that, prior to the implementation of the CMHCs, conditions were similar to or worsening

[19] Figure F.2 in the appendix graphs our mortality rates by the full population across urban and rural counties. Also note that we do not see any difference across urban and rural mortality rates within the white population. See Table 3.8 in the appendix.

compared to places that did not get CMHCs. As a robustness check, this analysis allays concerns about selection on unobservables. Furthermore, with this analysis we can consider how the impact of CMHCs evolved over time. These results indicate that the impact of CMHCs was fairly immediate and constant over the time period, neither ramping-up to full capacity nor deteriorating substantially from the initial improvement. Event study graphs for suicide for your whites, homicide for young adult, adult, and elderly non-whites, and alcohol related mortality for adult non-whites are appendix Figures F.3-F.7.

3.6.2 Boundary Analysis

In this section, we will provide a pseudo-boundary analysis. We will do this by limiting the sample to all of the counties that ever received a CMHC and all of the counties adjacent to that initial group. By doing this, we are potentially focusing on counties that are more similar to each other on unobservables, dealing with potential selection effects. This brings our sample to 527 counties that ever got a CMHC and 1,537 adjacent counties that never got a CMHC.

To do this pseudo-boundary analysis, we will estimate Equation 2 with this smaller sample. We are not directly comparing the treated counties' outcomes to their adjacent un-treated counties' outcomes. Instead, we are comparing the treated counties' outcomes to the outcomes of any county not yet treated that is adjacent to an ever-treated county. This is important because, across a given boundary, there is possibly selection on which county is chosen to be treated, most likely due to urbanity. However, since in this analysis we still control for the set of covariates listed in Equation 2, we are comparing the outcomes of treated counties to the counties adjacent to other treated counties that have similar observable characteristics, particularly urbanity.

Table 3.9 presents the boundary analysis results for the total population, the white population, and the non-white population. While only the impact of CMHCs on non-white homicide remains statistically significant, the results in this table generally are larger in magnitude, though also less precise. This decrease in precision is likely due to the substantial decrease in the number of observations. However, the general conclusions of this table remain the same as with the full sample, as described in Table 3.5.

Tables F.2, F.3, and F.4 replicate Tables 3.6, 3.7, and 3.8, respectively, on this boundary analysis sample. The results are substantively the same, though in general both the magnitudes and standard errors are larger. The increase in standard errors is likely due to smaller sample size, whereas the increase in magnitudes suggest that any selection on unobservables that occurred actually attenuated the effects of CMHCs on mortality.

3.7 Conclusion

Mental illness is a pervasive and growing problem in the United States and around the world (Organization, 2001; Abuse and Administration, 2018). The World Health Organization's suggested solution is for community care, despite a lack of evidence as to its efficacy in managing mental health. The United States' experiment with community care in the 1950s through 1980s provides a context within which to test whether community care could be an effective means of handling the mental health problem in the United States and the rest of the world.

Since CMHCs were not equipped to treat the seriously mentally ill who were being deinstitutionalized during this time and less than half of the planned centers were ever built, Community Mental Health Centers have gone down in history as being completely ineffectual for treating the mentally ill. However, our results suggest that this is not the case.

We use the roll-out of CMHCs from 1971 to 1981 to identify the effect of implementing a CMHC in a community on mortality rates. While priority rankings were created by states to establish which areas had the greatest need for a CMHC, we show that these rankings do not predict which communities ever got CMHCs, nor do they predict the rollout of CMHCs. This apparent lack of correlation between state priority rankings and treatment allows us to directly compare treated and untreated counties in identifying the effect of a CMHC on county level mortality rates. We further show that our pre-treatment outcome variables of interest had similar levels and were experiencing similar trends before treatment.

We find evidence that CMHCs reduced suicide rates among whites between the ages of 15 and 24 by 4%. CMHCs were particularly effective in reducing deaths from homicide and alcohol in the non-white population, with non-whites experiencing a 5% decline in homicide rates and non-white adults experiencing an 11% decline in deaths caused by alcohol. The effect on mortality for non-white people is focused in rural areas. These results suggest CMHCs were effective in reducing mental illness related mortality, especially in populations with the greatest need and least access to alternative forms of treatment.

Death is the most extreme outcome of unmanaged mental illness; other outcomes, such as unemployment, labor force participation rates, poverty, and divorce rates, all contribute to the decreased quality of life for those suffering from mental illness. We see our results as a lower bound on the overall impact of CMHCs on the mentally ill and their communities, and find that, just considering this lower bound, we see substantial improvements in outcomes due to the implementation of CMHCs.

While CMHCs were implemented in a particular context, being in the United States and the results of a federal policy with numerous deployment issues, the effects that they had can be informative to our more general understanding of community care. In particular, it seems that community care is most likely to work in environments that are low in alternative mental health services, as was the case with rural counties in this paper, and for people who have the highest rates of mental illness-related issues. Additionally, they may be effective at helping not only the mentally ill, but also their communities, as the results on homicide rates suggest. While it is outside the scope of this paper to compare community care with other forms of care, the results suggest that community care does provide some benefits for the most extreme outcomes, and cautiously indicate that community care is worth considering as a viable option for public mental health.

3.8 Figures

Figure 3.1: Patient Care Episodes

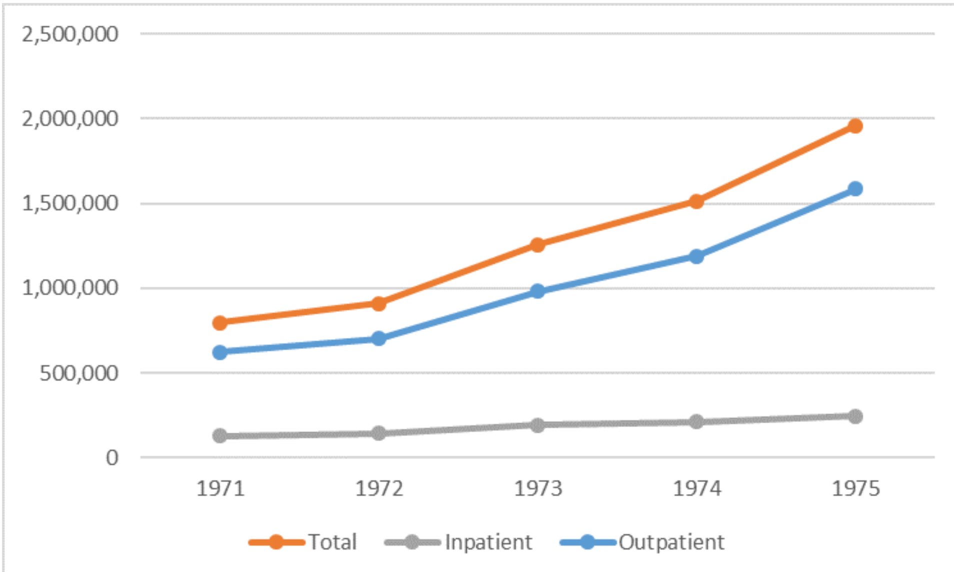


Figure 3.2: Rollout of CMHCs

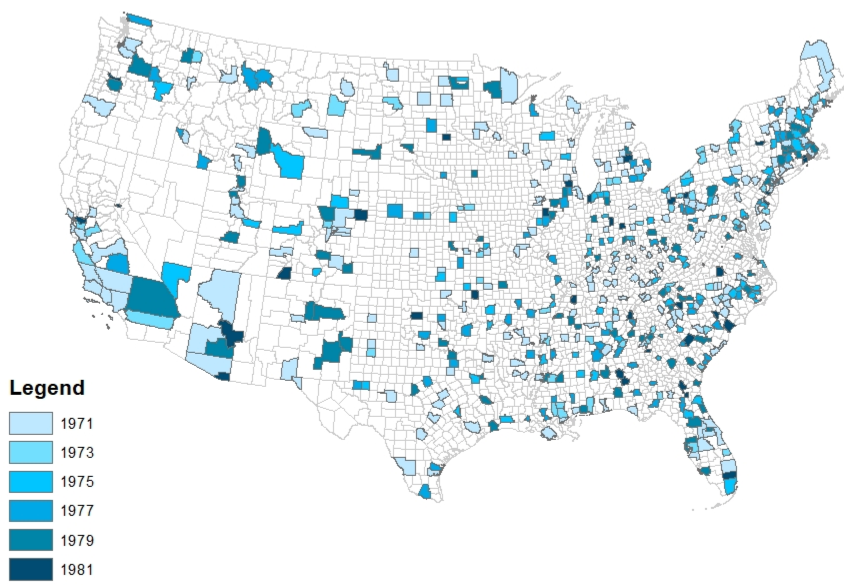


Figure 3.3: Age-Adjusted Mortality Summary Statistics

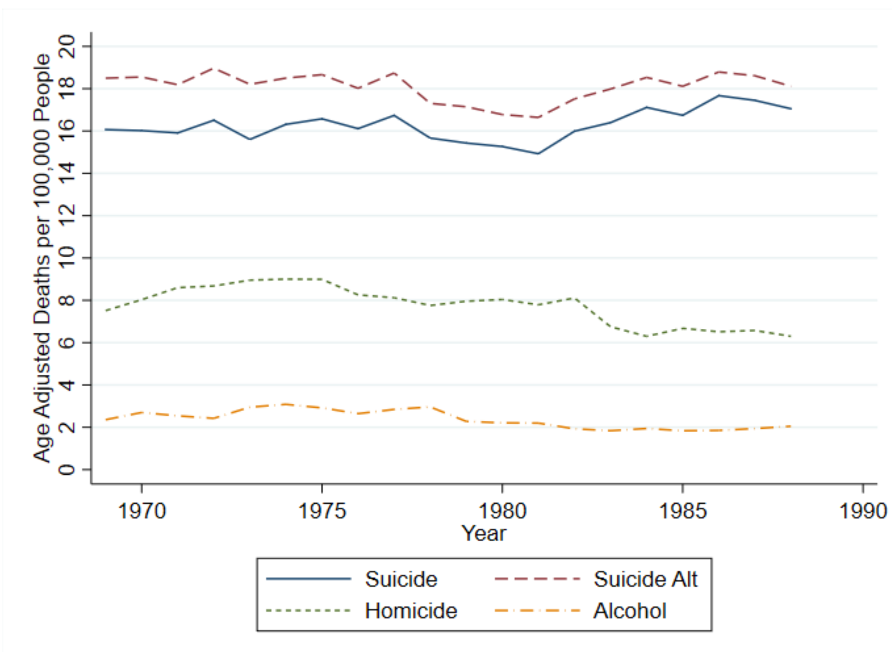
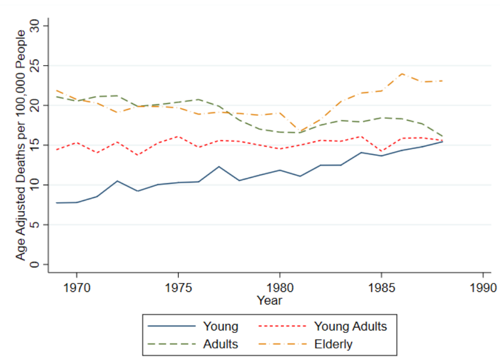
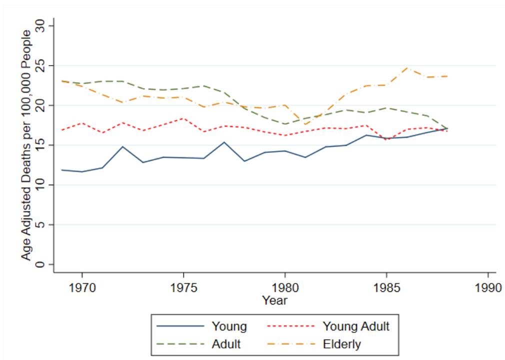


Figure 3.4: Age-Adjusted Mortality Rates by Age

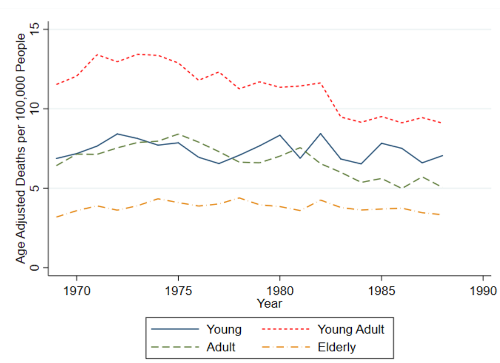
(a) Suicides



(b) Suicide Alt



(c) Homicide



(d) Alcohol

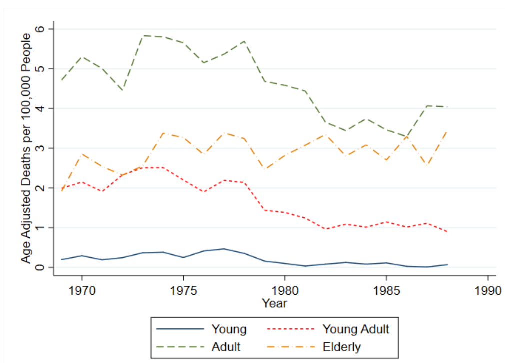


Figure 3.5: Event-Study: Suicide among Whites

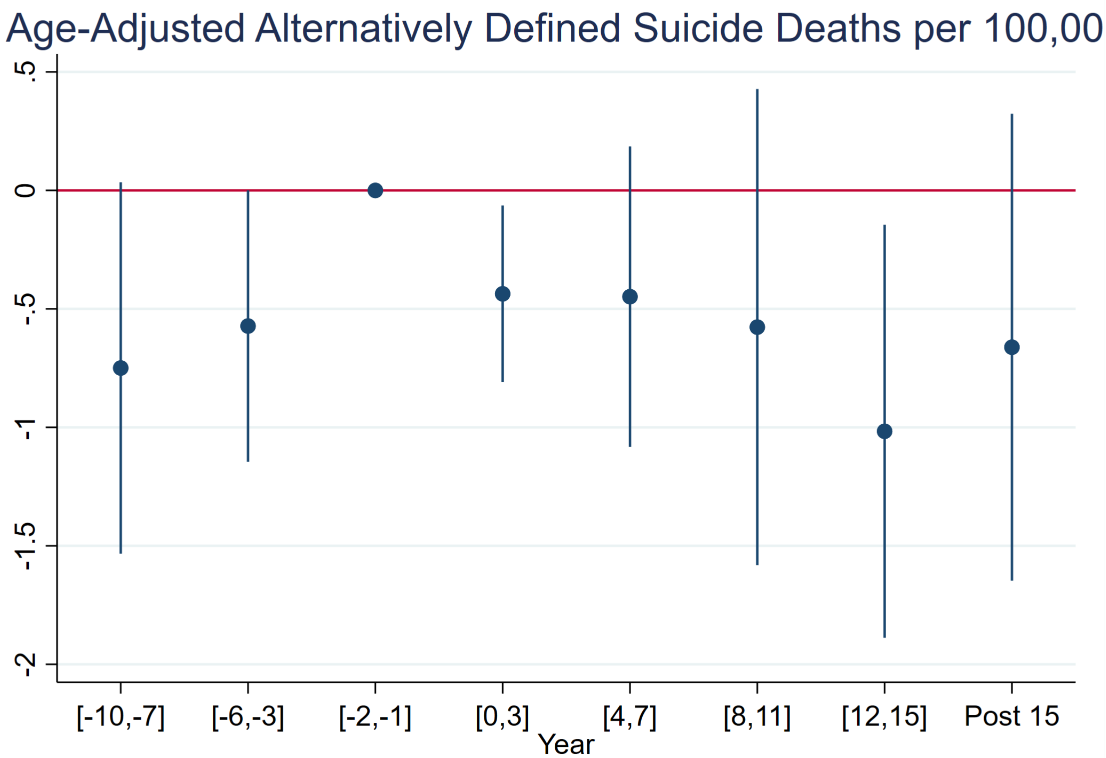
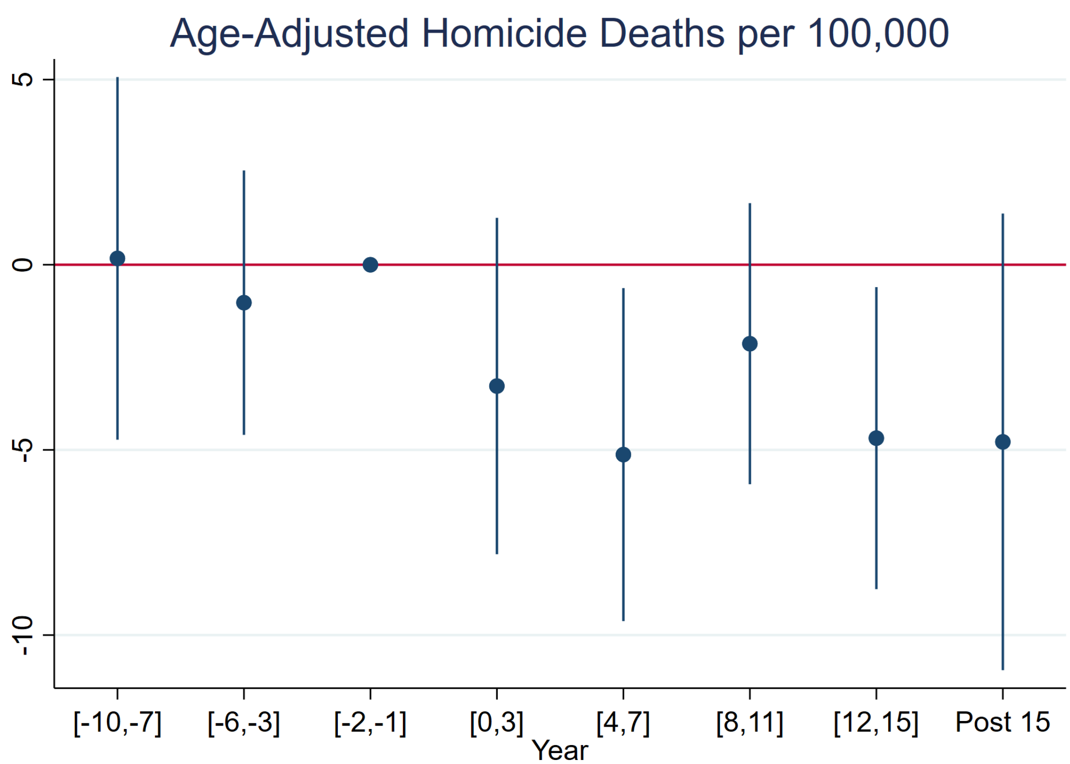


Figure 3.6: Event-Study: Homicide among Non-Whites



3.9 Tables

Table 3.1: Pre-treatment Differences in Demographics across Counties

	No CMHC	Gets CMHC	Difference
Population	21170	170757	-149586***
	(971)	(18601)	(8818)
Percent Urban	0.29	0.65	-0.37***
	(0.01)	(0.01)	(0.01)
Percent Less than HS	0.38	0.30	0.08***
	(0.00)	(0.00)	(0.01)
Percent HS	0.55	0.60	-0.04***
	(0.00)	(0.00)	(0.00)
Percent College	0.07	0.10	-0.04***
	(0.00)	(0.00)	(0.00)
LFPR	0.53	0.56	-0.03***
	(0.00)	(0.00)	(0.00)
Unemployment Rate	0.05	0.04	0.00
	(0.00)	(0.00)	(0.00)
Number of Counties	2503	531	

Notes- Standard errors are in parentheses. $*p < .10$, $**p < .05$, $***p < .01$.

Table 3.2: Priority Ranking and CMHC Rollout

	Ever Gets CMHC	Rollout
Rank	-0.064	0.060
	(0.064)	(0.194)
Observations	365	235
R^2	0.000	0.000

Notes- Standard errors are in parentheses. *$p < .10$, **$p < .05$, ***$p < .01$. A higher rank means you have greater need.

Table 3.3: Predictors of CMHC Location

	Ever gets CMHC	Rollout
Suicide	-0.000	-0.017
	(0.000)	(0.015)
Homicide	0.002***	-0.026**
	(0.000)	(0.015)
Alcohol	0.001	-0.019
	(0.001)	(0.028)

Notes- Standard errors are in parentheses. *$p < .10$, **$p < .05$, ***$p < .01$. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate.

Table 3.4: Pre-treatment Mortality Trends

	Suicide	Homicide	Alcohol
Year	-0.116	0.516	0.415*
	(0.504)	(0.345)	(0.181)
CMCHind	-1902.5	-9.45	838.6
	(2370.7)	(1626.3)	(853.8)
Interaction	0.965	-0.006	-0.425
	(1.204)	(0.826)	(0.434)
Observations	6068	6068	6068
R^2	0.001	0.007	0.003

Notes- Standard errors are in parentheses. $^*p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Table 3.5: Effects of CMHC on Mortality

	(1) Suicide	(2) Homicide	(3) Alcohol
Panel A: Total Population			
CMHC	-0.007	-0.038*	0.009
	(0.010)	(0.023)	(0.047)
Panel B: White Population			
CMHC	-0.007	-0.006	0.022
	(0.011)	(0.020)	(0.052)
Panel C: Non-white Population			
CMHC	-0.001	-0.053*	-0.019
	(0.031)	(0.031)	(0.053)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	60316	60316	60316

Notes- Standard errors clustered at the county level are in parenthesis *$p < .10$, **$p < .05$, ***$p < .01$. Years included: 1969-1988. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate. Regressions are weighted by county population.

Table 3.6: Sub-group Analysis by Age - White Population Only

	(1) Suicide	(2) Homicide	(3) Alcohol
Panel A: Young			
CMHC	-0.036*	0.006	0.014
	(0.022)	(0.031)	(0.169)
Panel B: Young Adult			
CMHC	0.019	-0.008	0.011
	(0.015)	(0.024)	(0.070)
Panel C: Adult			
CMHC	-0.008	0.020	0.024
	(0.014)	(0.032)	(0.055)
Panel D: Elderly			
CMHC	-0.024	-0.053	-0.020
	(0.020)	(0.044)	(0.067)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes- Standard errors clustered at the county level are in parenthesis *$p < .10$, **$p < .05$, ***$p < .01$. Years included: 1969-1988. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate. Regressions are weighted by county population.

Table 3.7: Sub-group Analysis by Age - Non-White Population Only

	(1) Suicide	(2) Homicide	(3) Alcohol
Panel A: Young			
CMHC	0.035	-0.051	0.291
	(0.050)	(0.052)	(0.212)
Panel B: Young Adult			
CMHC	-0.058	-0.058*	0.060
	(0.044)	(0.031)	(0.067)
Panel C: Adult			
CMHC	0.055	-0.041	-0.115*
	(0.072)	(0.032)	(0.070)
Panel D: Elderly			
CMHC	-0.019	-0.112*	0.039
	(0.089)	(0.062)	(0.120)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes- Standard errors clustered at the county level are in parenthesis $*p < .10$, $**p < .05$, $***p < .01$. Years included: 1969-1988. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate. Regressions are weighted by county population.

Table 3.8: Urban vs. Rural Sub-group Analysis - Non-White Subsample

	(1)	(2)	(3)
	Suicide	Homicide	Alcohol
Panel A: Urban CMHC			
CMHC	-0.006	-0.009	-0.007
	(0.032)	(0.020)	(0.056)
Panel B: Rural CMHCs			
CMHC	0.048	-0.133*	-0.324*
	(0.115)	(0.068)	(0.162)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes- Standard errors clustered at the county level are in parenthesis $*p < .10$, $**p < .05$, $***p < .01$. Years included: 1969-1988. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate. Regressions are weighted by county population.

Table 3.9: Boundary Analysis - Effects of CMHC on Mortality

	(1)	(2)	(3)
	Suicide	Homicide	Alcohol
Panel A: Total Population			
CMHC	-0.255	-0.203	0.012
	(0.190)	(0.315)	(0.149)
Panel B: White Population			
CMHC	-0.269	0.092	0.038
	(0.201)	(0.151)	(0.135)
Panel C: Non-white Population			
CMHC	-0.178	-3.478*	-0.324
	(0.398)	(1.923)	(0.422)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Observations	41041	41041	41041

Notes- Standard errors clustered at the county level are in parenthesis *$p < .10$, **$p < .05$, ***$p < .01$. Years included: 1969-1988. Limited to Boundary Analysis sample. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate. Regressions are weighted by county population.

4.0 Appendix

4.1 A: Theoretical Framework

In this section I outline a simple theoretical framework describing an employment environment in which employers get noisy, competition-driven information about an employee's ability. In doing this, I can understand what types of beliefs employers would need to have about employee decision making to generate the hidden cost outlined in this paper. First, I lay out the basic employee decision environment without affirmative action. Then, I go through the employers' decision-making environment, including the conditions under which they will give female competition winners a hiring advantage. Then, I add affirmative action and show how the employees' decision environment changes. Then, I show how affirmative action's effects on the employee's decision environment affects the employers' beliefs and hiring decisions, resulting in female competition winners losing their previous hiring advantage. Finally, I'll discuss how alternative models of employees' behavior (or employers' beliefs about employees' behavior) would result in different employment outcomes.[1]

4.1.1 Employee Decision Making Framework Without Affirmative Action

Suppose employees can be divided in two ways. The first is by gender: they can be male or female (though any other division based on minority status may be used). This will be denoted as their gender $g \in \{M, F\}$. The second by ability (θ): they can be either low ability (θ_L) or high ability (θ_H). Suppose the probability that an individual is high ability is α_g, where $0 < \alpha_g < 1$.[2]

Suppose also that employees can have gender-specific beliefs about their probability of being high ability, ($\hat{\alpha_g} \in \{M, F\}$), measuring their degree of confidence, which may or may not be equal to their actual probability of being high ability. This is a population-wide gender-specific belief, in the sense that all people of that gender believe that, ex-ante, they have that likelihood of being high ability. This could result from self-stereotypes about what genders are good at what tasks (see Coffman, 2014; Coffman et al., 2019a,b).

Employees get one of three signals about their ability, as generated by a random process that depends on their

[1] This paper's focus is the decision-making process of the employer. To address this aspect of affirmative action in a controlled way, employees do not know about or take into account the hiring decision when making their competition entry decision in both the model and the experiment. While it isn't unbelievable that people, particularly teenagers making college decisions or recent graduates trying to find their first job out of college, are short-sighted enough to not take into account future hiring decisions, this could be an important aspect to understand the competition entry decisions of individuals when faced with affirmative action, when those decisions will later be interpreted by employers (related to the argument in Antonovics and Sander (2013)). Given the conclusions of this paper, future work should expand into the employees' response to the employers' later decision stage.

[2] In this framework I use a single ability (actual probability of being high ability) and belief (believed probability of being high ability) parameter for each gender, rather than a continuum of ability and beliefs. However, a continuum of these measures would result in directionally similar outcomes. I will point out where possible how the outcomes will be different.

underlying ability level.[3] The signals are high (S_H), medium (S_M), and low (S_L). For both men and women, signals are generated in the same way, such that high ability people only get high or medium signals, and low ability people only get medium or low signals. More specifically:

$$P(S = S_H|\theta = \theta_H) = \beta_H$$

$$P(S = S_M|\theta = \theta_H) = 1 - \beta_H$$

$$P(S = S_L|\theta = \theta_H) = 0$$

$$P(S = S_H|\theta = \theta_L) = 0$$

$$P(S = S_M|\theta = \theta_L) = \beta_L$$

$$P(S = S_L|\theta = \theta_L) = 1 - \beta_L$$

$$\frac{1}{2} < \beta_H < 1$$

Employees use Bayes rule to update their gender-specific prior beliefs with their signals in order to get their posterior belief of their probability of being high ability:

$$\hat{P}(\theta = \theta_H|S = S_j, g) = \frac{P(S = S_j|\theta = \theta_H) * \hat{\alpha_g}}{P(S = S_j|\theta = \theta_H) * \hat{\alpha_g} + P(S = S_j|\theta = \theta_L) * (1 - \hat{\alpha_g})}$$

where $g \in \{M, F\}$ and $j \in \{H, M, L\}$. If someone gets a high or low signal, they know they are high or low ability, respectively, with certainty. However, an individual's belief about their ability with a medium signal depends on their gender-specific prior belief about their likelihood of being high ability.

Employees are then faced with a decision about whether or not to enter a tournament. If they enter the tournament, they can either win or lose. Conditional on entering, the probability that they win depends on if their ability is high or low:

$$P(win|enter) = \begin{cases} \frac{2}{3}\gamma & if \quad \theta = \theta_H \\ \frac{2}{3}(1-\gamma) & if \quad \theta = \theta_L \end{cases}$$

where γ is the noise parameter such that $\frac{1}{2} < \gamma < 1$, and the $\frac{2}{3}$ reflects the fact that, of the three people who are in the top half of their group in that round of the competition, only 2 of them will win the tournament in this experimental formulation. This allows people who are high ability to lose with some positive probability, and conversely people who are low ability could also win with a positive probability. However winning is still an informative signal about their underlying ability. Payoffs, denoted by pi, depend on whether the individual enters and, conditional on entering, whether the individual won:

$$\pi = \begin{cases} \pi_w - C & if \quad enter + win \\ -C & if \quad enter + lose \\ 0 & if \quad DNE \end{cases}$$

[3] In the experiment this comes through the information about the number of questions they got correct in the first and second rounds of the task. The model simplifies the signal structure to 3, rather than a continuum of, options. In a more general sense, these could be grades students get in high school before deciding on what college to apply to or whether to apply for STEM programs, or grades students get in college before deciding on what jobs to apply for.

where DNE means ”does not enter”. This reflects that there is a cost to entering the tournament, such as an effort cost, that is not incurred if one does not enter.

Because people who get high signals must be high ability and people who get low signals must be low ability, regardless of gender, we can establish their expected value of entering the tournament:

$$E[\pi(enter|S = S_H)] = \frac{2}{3}\gamma\pi_w - C$$

$$E[\pi(enter|S = S_L)] = \frac{2}{3}(1-\gamma)\pi_w - C$$

However, if someone gets a middle signal, they don't know whether they are high or low ability, so their expected value of entering depends on their posterior probability of being high ability:

$$E[\hat{\pi}(enter|S = S_M, g)] = \frac{2}{3}[\gamma\hat{P}(\theta = \theta_H|S = S_M, g) + (1-\gamma)\hat{P}(\theta = \theta_L|S = S_M, g)]\pi_w - C$$

As such, there will be a threshold on posterior probability of being high ability, above which employees enter and below which employees do not enter.

4.1.2 Employer Decision Making Framework

In this environment, employers receive information about a potential employee's gender and whether the employee won the tournament or not, and decide how much they want to pay (out of an endowment) to hire the employee. The employers want to hire high ability employees. They know about the tournament structure for employees, and that employees can have gender specific probability of being high ability and gender specific probability of prior beliefs that they are high ability.

More specifically, employers get \$3 as an endowment, and they can pay any amount of that money (called \$x) to buy some probability of hiring the employee, equal to $P_hire = x/3$ where $0 \leq x \leq 3$. Employers get a bonus of \$6 if they successfully hire the employee and the employee is high ability ($\theta = \theta_H$); otherwise they get no additional money. Employers have a utility function $U(.)$ such that $U'(.) \geq 0$ and $U''(.) \leq 0$.

Employers want to maximize their expected utility over the value x:

$$\underset{x\in[0,3]}{\text{argmax}} E[U] = \frac{x}{3}P(\theta = \theta_H|outcome, g)U(9-x) + (1 - \frac{x}{3}P(\theta = \theta_H|outcome, g))U(3-x)$$

Recall that the two options for tournament outcome are enter and win (w) or not (nw), with not winning coming from either not entering or from entering and losing. While finding a closed form of x*, where x* is the utility maximizing value of x, isn't possible, because $\frac{\partial^2 U}{\partial x \partial P(\theta=\theta_H|outcome,g)} > 0$ we know that $\frac{dx^*}{dP(\theta=\theta_H|outcome,g)} \geq 0$, or that x* is non-decreasing in the probability the employee is high ability via Topkis' theorem of supermodularity. This allows us to compare hiring rates for different groups of employees through the probabilities of them being high ability conditional on their gender and tournament outcome.

4.1.3 Employer Beliefs that Would Generate Hiring Advantage for Successful Women without Affirmative Action

In this section, I will describe the types of beliefs employers could have about employee beliefs and behavior that would generate the hiring advantage being hypothesized for successful women in the absence of affirmative action. However, while most of the beliefs proposed here will be supported by empirical data, it is an empirical question as to whether subjects in the role of employees will have these beliefs.

Suppose employers believe that men and women are of equal ability, but have different priors about being high ability in a particular task. Specifically, $\alpha_M = \alpha_F = \alpha$ but $\hat{\alpha_F} < \hat{\alpha_M}$ showing that men are more confident in the task than women.[4] As such, when employees get medium signals about their ability, men will have a higher posterior belief that they are high ability than women: $\hat{P}(\theta = \theta_H | S = S_M, M) > \hat{P}(\theta = \theta_H | S = S_M, F)$.[5]

When making entry decisions, we will assume that the parameters of the model are such that those individuals who know they are high ability with certainty (because they received high signals) enter the tournament, and those that know they are low ability with certainty (because they received low signals) don't enter the tournament. This is possible as long as the benefit winning the tournament is high enough compared to the cost of entering ($\frac{2}{3}\pi_w > C$) and the noise parameter is large enough, in that it is more likely that a high ability person wins and a less likely that a low ability person wins ($\gamma > \frac{3C}{2\pi_w}$). Also, the employers could believe that the parameters are such that men who get medium signals enter the tournament, and women who get medium signals do not enter the tournament: $E[\hat{\pi}(enter|S = S_M, M)] > 0 > E[\hat{\pi}(enter|S = S_M, F)]$. This is true under an additional set of parameter restrictions ensuring that the probability of getting a medium signal when high ability is low enough ($\beta_L < \frac{1}{2}$), and the cost tournament entry and the benefit of winning the tournament take a certain shape ($\pi_w > 1, C < \frac{1}{3}$).[6] In this case, employers will form certain predictions about the ability of men and women who won the tournament vs. those that did not win the tournament (either because they lost or did not enter).

Prediction 1: Employers who see women who win the tournament will believe they are more likely to be high ability and will be more likely to hire them compared to men who win the tournament.

When faced with employees that enter and win the tournament without affirmative action, employers are more likely to believe a woman is high ability compared to a man. Since only high ability women enter the tournament, they can be certain a female entrant is high ability. Thus, $P(\theta = \theta_H | w, F) = 1$. Since some men who are low ability enter the tournament, they must have a probability of being high ability that is less than one, so $P(\theta = \theta_H | w, M) < P(\theta =$

[4] While not true in every task, men are found to be more confident in the task used in this experiment (Niederle and Vesterlund, 2007), and in many other tasks found in work places dominated by men and considered relatively prestigious (Barber and Odean, 2001; Huang and Kisgen, 2013; Coffman, 2014; Sarsons and Xu, 2015; Coffman et al., 2019a,b).

[5] This difference in posterior beliefs between men and women could also be generated with the same priors but differences in updating, such that men are more likely to update their beliefs up compared to women, or women are more likely to update their beliefs down compared to men (e.g., Mobius et al., 2011).

[6] This could also be generated by making the cost of competing (C) higher for women than for men, which may also be a reasonable way to model this given the evidence of gaps in competitiveness unrelated to confidence (e.g., Niederle and Vesterlund, 2007; Croson and Gneezy, 2009).

$\theta_H|w, F)$, leading to $P(hire|w, M) \leq P(hire|w, F)$.[7]

Prediction 2: It is unclear how employer beliefs and hiring probabilities will compare between men and women who do not win the tournament (either through losing or not entering).

When faced with an employee that does not win the tournament, it is unclear how employers will respond to men compared to women. The probability that a woman is high ability given that she did not win can be larger than, equal to, or smaller than the probability that a man is high ability given that he did not win, depending on how likely a high ability individual knows that they are high ability. Specifically, as the likelihood a high ability person knows they are high ability (β_H) increases, the probability that a woman who does not win is low ability decreases without affecting the probability that a man who does not win is low ability. This occurs because if women who are high ability are likely to know, they're more like to have entered, meaning the women who did not enter are more likely to be low ability. The stronger β_H is a signal of ability, the less likely high ability women will end up not entering the tournament. Thus, at high (low) values of β_H, women who do not win will be hired less (more) than men who do not win.

4.1.4 Employee Decision Making Framework with Affirmative Action

In essence, an affirmative action policy changes the probability that an individual is likely to be admitted or hired based on their gender – the probability a female applicant is successful goes up, and the probability that a male applicant is successful goes down. This is reflected in a change to the probabilities of winning the tournament, given entry, for men and women:

$$P(win|enter) = \begin{cases} P_g\gamma & if \quad \theta = \theta_H \\ P_g(1-\gamma) & if \quad \theta = \theta_L \end{cases} \quad for \quad g \in \{M, F\}$$

were $0 < P_M < \frac{2}{3} < P_F < 1$. This now changes the expected value of entering for men and women, given the signals they receive:

$$E[\pi(enter|S = S_H, g)] = P_g\gamma\pi_w - C$$

$$E[\pi(enter|S = S_H, g)] = P_g[\gamma\hat{P}(\theta = \theta_H|S = S_M, g) + (1-\gamma)\hat{P}(\theta = \theta_L|S = S_M, g)]\pi_w - C$$

$$E[\pi(enter|S = S_L, g)] = P_g(1-\gamma)\pi_w - C$$

In essence, affirmative action replaces the singular threshold on employee posterior beliefs with two, gender specific thresholds where the threshold for women is at a lower posterior than the threshold for men. This means that the women who enter the tournament with affirmative action will be, on average, less confident than the men who enter

[7]If the probability an individual believed they were high ability could take a continuous form such that the distribution of prior beliefs for men had greater weight on the high end of the distribution than for women, then the belief that a woman that won was high ability would still be greater than the belief that a man that won was high ability, but it would not equal 1 as is the case here. Also, this result, both in the model and the experiment, is of a flavor similar to Bohren et al. (2019), which finds that women who post answers to math questions in an online forum initially face discrimination, but when they have accumulated a series of positive evaluations their responses are then favored over men's of an equal evaluation level. They attribute this to the general belief that women must be of a superior quality to reach the same level as a man in this male-type environment, similar to the conclusion being drawn by employers here.

the tournament. However, depending on how their confidence relates to their true probability of being high ability, the women who enter may be more able, as able, or less able on average than men who enter.

4.1.5 Employer Beliefs that Would Remove Hiring Advantage for Successful Women with Affirmative Action

When faced with affirmative action in the employees' stage, employers recognize the minimum posterior belief for women to enter the tournament decreases below the original threshold belief, and also below the minimum posterior belief for men to enter the tournament.

Prediction 3: As long as the posterior threshold for women with affirmative action falls below the posterior belief women have about their ability with medium signals, employers will believe that women who enter and win with affirmative action are of lower ability than women who enter and win without affirmative action, and will thus hire them less.

Suppose employers believe that the posterior threshold for women with affirmative action falls below the posterior belief women with medium signals. This happens so long as $\hat{P}(\theta = \theta_H | S = S_M, F) > \frac{3C}{2}$. Thus, employers will believe that women who enter and win in the affirmative action are of a lower ability than those women who enter and win without affirmative action, because women who enter the tournament are no longer guaranteed to be high ability: $P(\theta = \theta_H, w, F, AA) < P(\theta = \theta_H, w, F, ST)$. This generates the cost of affirmative action, in that the signal of entering and winning for a woman's ability is weaker with affirmative action than without, leading to lower hiring probability of women who win the tournament with affirmative action: $P(hire, w, F, AA) \leq P(hire, w, F, ST)$.

Prediction 4: If the posterior threshold for men with affirmative action still falls below the posterior belief men have about their ability with medium signals, employers will have the same beliefs about them and hire them at the same rate as without affirmative action, and will hire them with the same probability as women with affirmative action.

Suppose the change in the threshold for men with affirmative action is such that it still falls below the posterior belief that men have after receiving a medium signal. This occurs as long as $P_M > \frac{2C}{\pi_w}$. If this is the case, employers do not change their beliefs about men who win, nor do they hire them with a different probability, with the implementation of affirmative action. Furthermore, they will hire the men who win at the same rate as they hire women who win with affirmative action:

$$P(\theta = \theta_H | w, M, AA) = P(\theta = \theta_H | w, M, ST) = P(\theta = \theta_H | w, F, AA) < P(\theta = \theta_H | w, F, ST)$$

Thus, not only does affirmative action lower the probability a woman who enters the tournament is high ability and lower the hiring rates of women, but it lowers them to the exact same rate as men. The conditions required for this exact result can be found in Figure B.6.[8] This results in the following hiring probability comparisons:

$$P(hire | w, M, AA) = P(hire | w, M, ST) = P(hire | w, F, AA) \leq P(hire | w, F, ST)$$

[8] A diagram of the continuous version can be seen in Figure B.7.

Prediction 5: Employers will believe men who don't win the tournament with affirmative action, either from not entering or from losing, are higher ability than the men who don't win the tournament without affirmative action, and the women who don't win the tournament with affirmative action. Their beliefs about women who don't win the tournament without affirmative action are ambiguous.

The probability that an individual that entered the tournament lost is directly related to the probability of winning the tournament if the individual is in the high ability group. Because men are less likely to win the tournament in the affirmative action than in the standard tournament ($P_M < \frac{2}{3}$) and women are more likely to win the tournament in the affirmative action tournament ($\frac{2}{3} < P_F$), we have that $P(\theta = theta_H|nw, F, AA) < P(\theta = \theta_H|nw, M, ST) < P(\theta = \theta_H|nw, M, AA)$ and $P(hire|nw, F, AA) \leq P(hire|nw, M, ST) \leq P(hire|nw, M, AA)$. The comparison between these probabilities and the probability a woman who does not win the tournament is high ability is less clear, given that the average ability of tournament entrants for women without affirmative action is different from that of these other three groups. Thus, the comparison between the probability a woman who does not win is high ability without affirmative action depends on the magnitude of β_H, with a higher value of β_H resulting in a lower probability that women who do not win the tournament are high ability. Depending on the value of β_H, this probability can be anywhere along the continuum of probabilities found for men with and without affirmative action, and women with affirmative action, meaning that the probability that woman is hired can fall anywhere amongst or beyond the probabilities of other groups. However, the probability that an individual is high ability after not winning the tournament is lower than the probability an individual is high ability after winning the tournament for all gender-tournament type comparisons.

4.1.6 Alternative Employer Beliefs that Could Lead to Cost of Affirmative Action for Successful Women

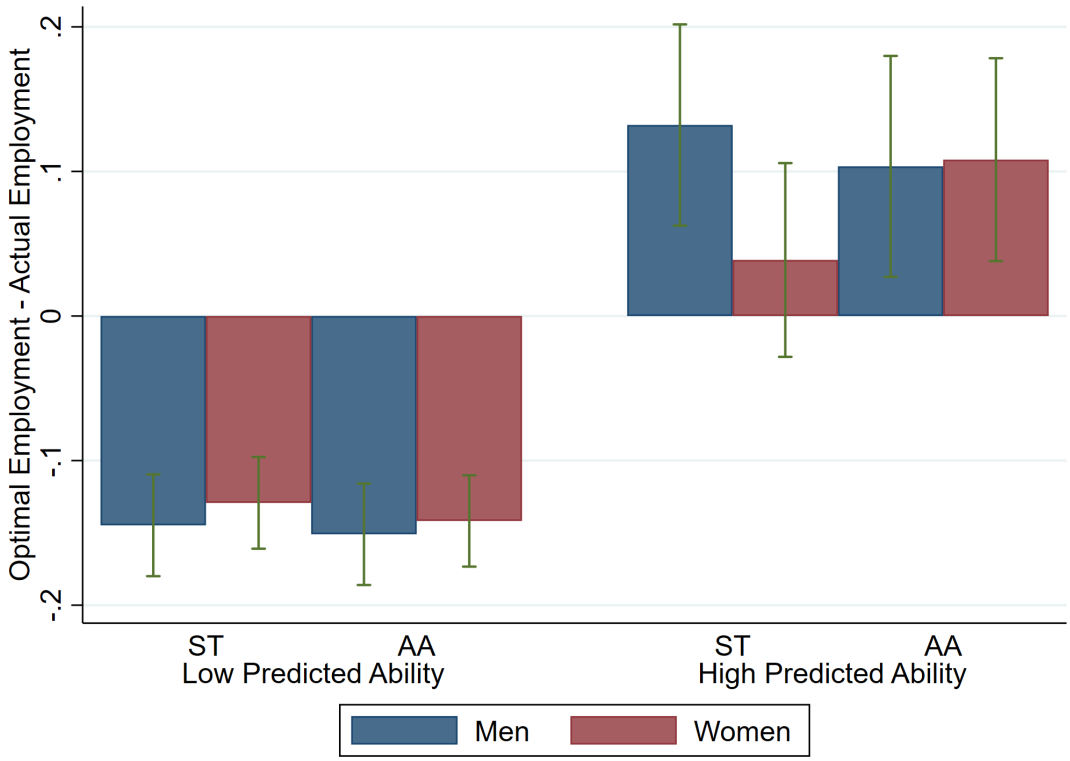
A cost of affirmative action in terms of beliefs about successful women's ability and their probability of being hired occurs so long as affirmative action decreases the threshold on posterior beliefs that women must have in such a way that women of a lower average ability enter the tournament. While, in the above case, it was generated through women having lower prior beliefs about their ability compared to men, this section points out some alternative cases in which this hidden cost could be generated, such as when men and women have the same prior beliefs, or they have different prior beliefs but those beliefs are correct.

Suppose, alternatively, that employers believe that men and women have the same prior beliefs, resulting in them having the same posterior beliefs after receiving a certain signal. In this case, employers would hire men and women that enter and win at the same rate, and they would hire men and women that enter and lose at the same rate. Thus, there would be no advantage given to women who win the tournament in terms of hiring. The effect of the implementation of affirmative action would depend on how the thresholds moved in relation to the posterior beliefs of the employees. If the threshold for women moved below the posterior of a lower signal group, there could still feasibly be a cost of affirmative action as women of a lower confidence level could then start entering the tournament. In this case, men who win would be hired more with affirmative action than women who win.

Consider alternatively the situation in which employers believe men and women are of different abilities, with men having a higher probability of being high ability than women reflective of $\hat{\alpha_M} > \hat{\alpha_F}$. Then, the different responses to the signals would be reflective of different ability distributions, rather than different confidence levels (though men and women would have different posteriors, they would be correct). In this case, you could still imagine a case in which only the women who know they are high ability enter (having gotten a high signal), whereas some men who are unsure also enter (having gotten either a high or medium signal). In this case, women who win will still have a hiring advantage over men who win. However, if the parameters of affirmative action are such that the thresholds are similar to those described in the above sections, then when women who are uncertain are drawn into the competition (women start to enter when they have both a high and medium signal), then women who enter the tournament will be much less likely to be high ability than the men who enter the tournament, leading them to be hired at a lower rate. Thus, in this case, one can also get a cost of affirmative action due to hiring.

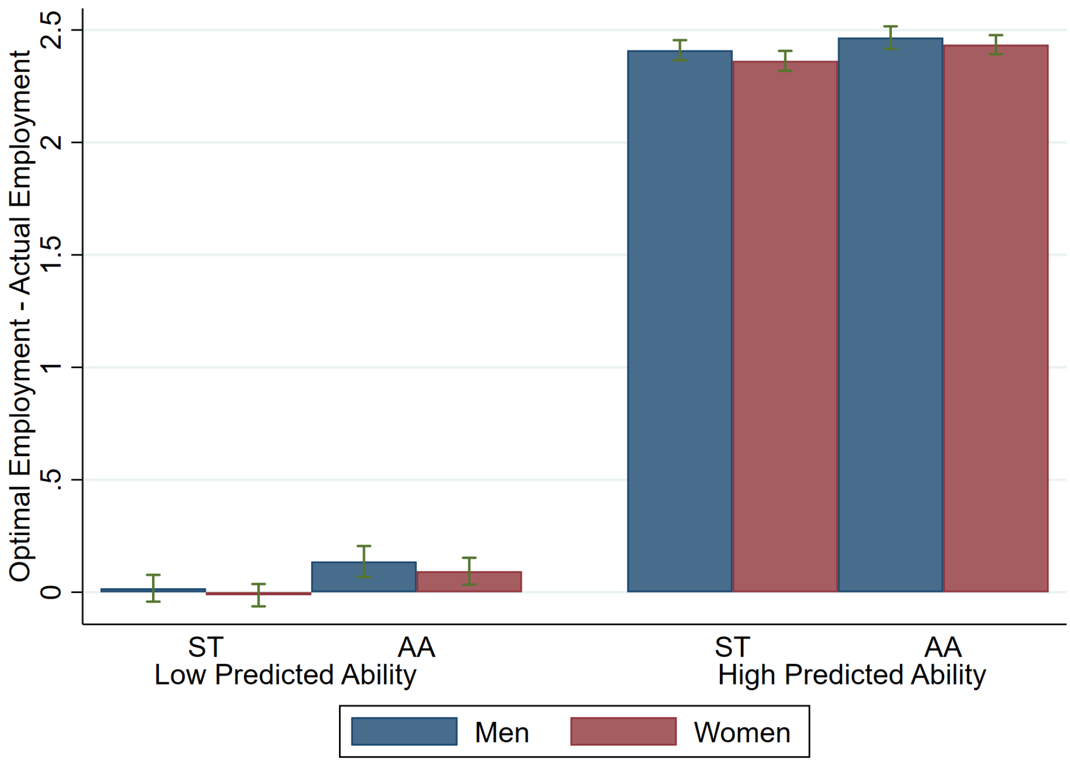
4.2 B: Appendix Figures

Figure B.1: Optimal Minus Actual Hiring Rates for Men and Women With and Without Affirmative Action, for Different Perceived Levels of Ability, CRRA Risk Averse Employees



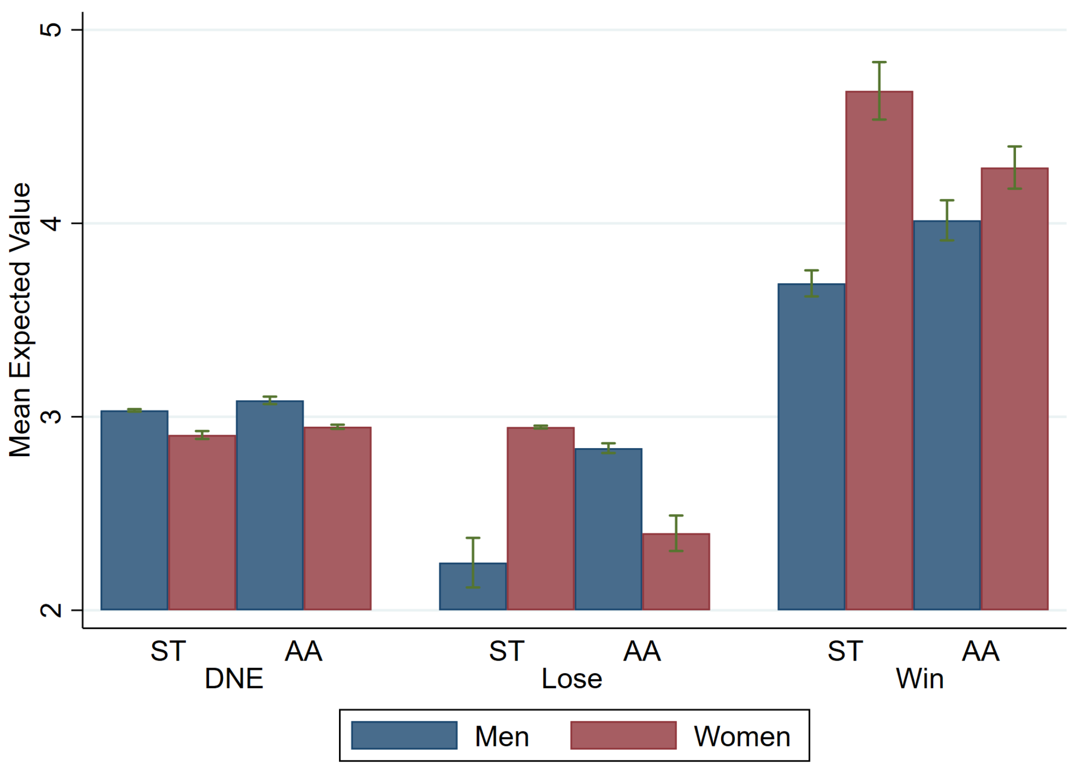
Notes - This figure presents mean gap between the optimal and actual employment choice for employees given the beliefs employers have about their ability and the employers' risk aversion parameters, split by the gender of the employee, the treatment, and whether the employer believed the employee was above 50% likely to be in the top half of their group in terms of performance on round 2 (high predicted ability) or below 50% (low predicted ability). Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament.

Figure B.2: Optimal Minus Actual Hiring Rates for Men and Women With and Without Affirmative Action, for Different Perceived Levels of Ability, CRRA Risk Averse Employees



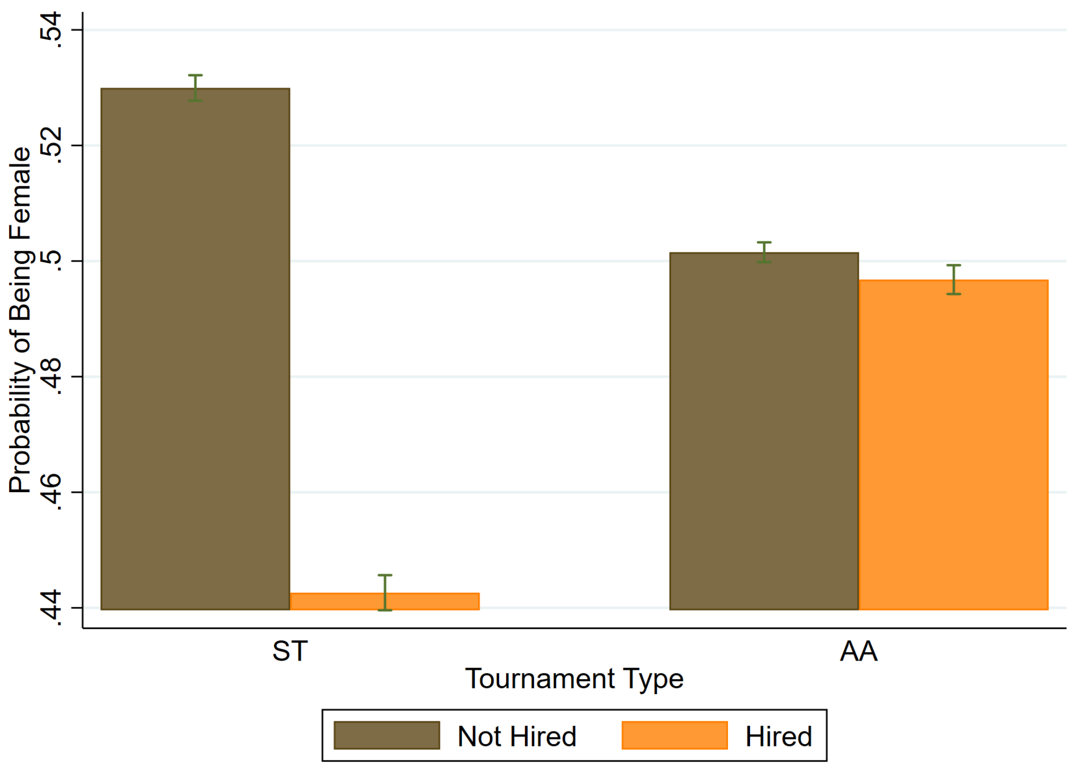
Notes - This figure presents mean gap between the optimal and actual employment choice for employees given the beliefs employers have about their ability and assuming the employees are risk neutral, split by the gender of the employee, the treatment, and whether the employer believed the employee was above 50% likely to be in the top half of their group in terms of performance on round 2 (high predicted ability) or below 50% (low predicted ability). Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament.

Figure B.3: Expected Value of Each Type of Employee, Based on Actual Hiring Choices



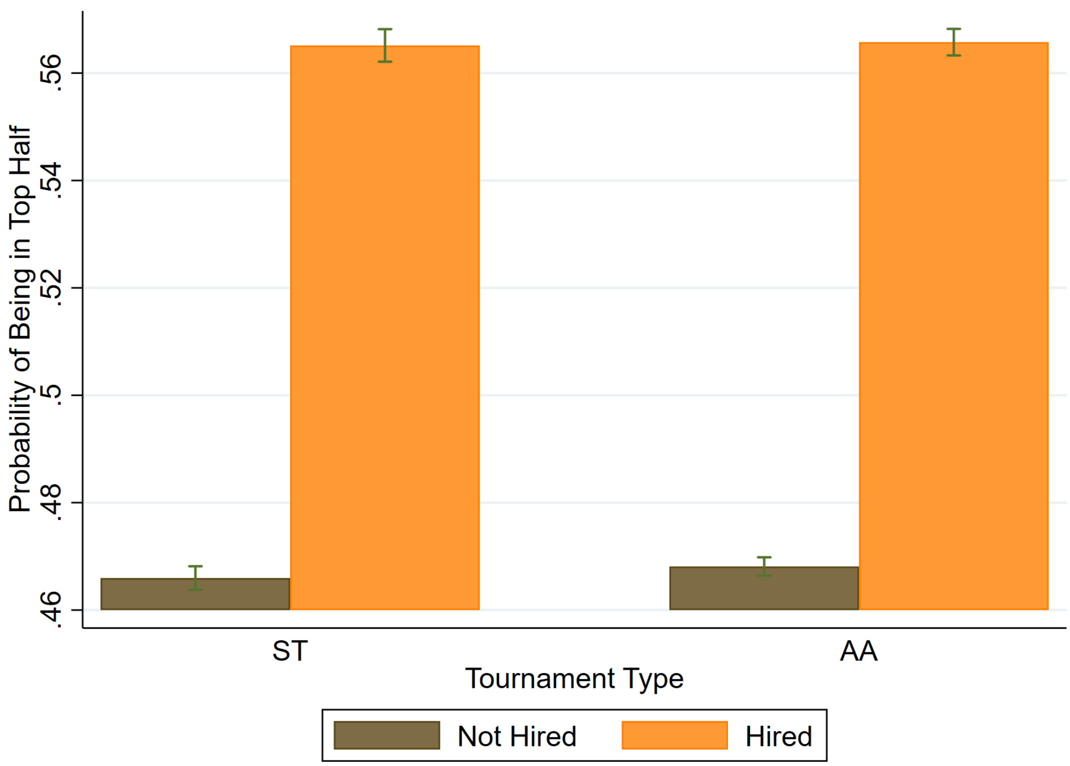
Notes - This figure presents the expected value for an employer when facing an employee of a particular gender, tournament type, and tournament outcome category, given actual hiring rates in the experiment. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament; DNE=did not enter.

Figure B.4: The Probability of a Hired or Not Hired Individual being Female, Across Tournament Type



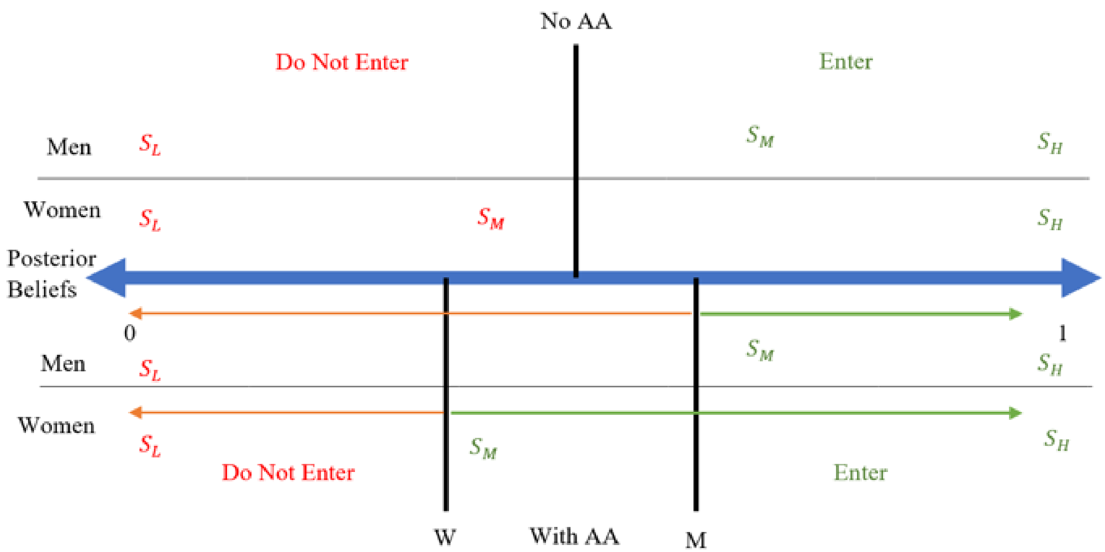
Notes - This figure presents the probability that an employee who is hired or not hired is female across tournament type from a simulation of each employee 10,000 times. See text for details about the simulation. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament.

Figure B.5: Probability of Employee Being in the Top Half of Their Group, by Tournament Type and Whether They Were Hired or Not



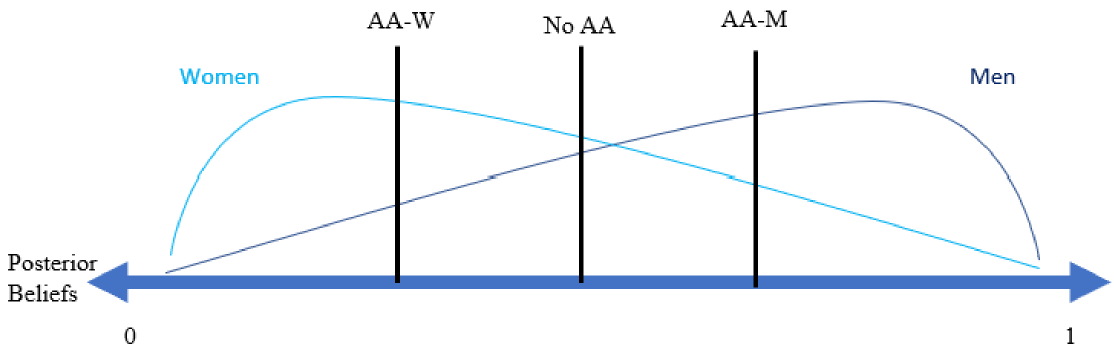
Notes - This figure presents the probability that an employee who is hired or not hired is in the top half of their group in terms of performance in round 2 across tournament type from a simulation of each employee 10,000 times. See text for details about the simulation. Bars represent 95% confidence intervals. ST=Standard tournament; AA= affirmative action tournament.

Figure B.6: Theoretical Framework Diagram - Discrete



Notes - This diagram represents employer beliefs that generate a cost of affirmative action through decreased signaling power for and decreased hiring of female employees who enter and win the tournament. The blue line presents the continuum of posterior beliefs that an individual could have, with 0 probability of being high ability on the far left and certainty of being high ability on the far right. S_L, S_M, and S_H are the posterior beliefs of individuals with low, medium, and high signals respectively. Both without affirmative action (top panel) and with affirmative action (bottom panel), the posteriors for men and women are presented for each of the three signals. Posterior values in red denote a group that does not enter the tournament; posterior values in green denote a group that does enter the tournament. Without affirmative action, the one vertical black line represents a hypothetical threshold on posteriors that would generate men with medium and high signals entering the tournament, with only women with high signals entering. The two vertical black lines with affirmative action represent the two, gender specific thresholds on posterior beliefs that occur with affirmative action, with the posterior for women (labeled W) to the left of the posterior for men (labeled M). These thresholds are in the hypothetical position such that, with affirmative action, both men and women enter the tournament with both high and medium signals. See figure B.7 for a diagram with continuous posteriors.

Figure B.7: Theoretical Framework Diagram - Continuous



Notes - See the notes for figure B.6 for a general description of the diagram's structure. In this diagram, the distribution of posterior beliefs is continuous, with the posterior for men (navy) being primarily to the right of the posterior for women (aqua). If we assume that the true distribution of ability is equal between men and women, we can have that, for certain distributional shapes, without affirmative action the average ability of men to the right of the no affirmative action threshold is less than the average ability of women to the right of the no affirmative action threshold, so men who enter the tournament are of lower ability than the women who enter. Then, with affirmative action, the threshold for women lowers, and women of a lower posterior are willing to enter, thus the average ability of women with affirmative action will drop. Depending on the shape of the distributions and the exact thresholds, it may be that the actual average ability of men and women who enter after affirmative action is put into place will be the same, as reflects the case described in discrete form in Appendix 4.1.5.

4.3 C: Appendix Figures

Figure C.1: Design Illustration: Biweekly Intervention

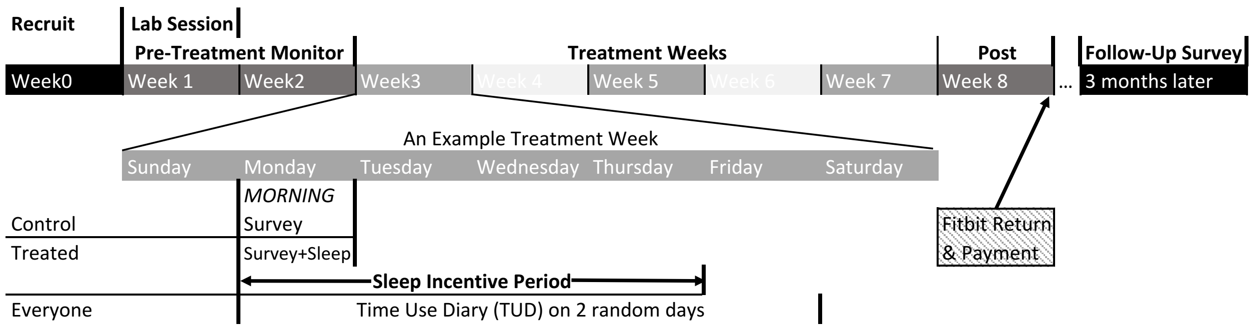
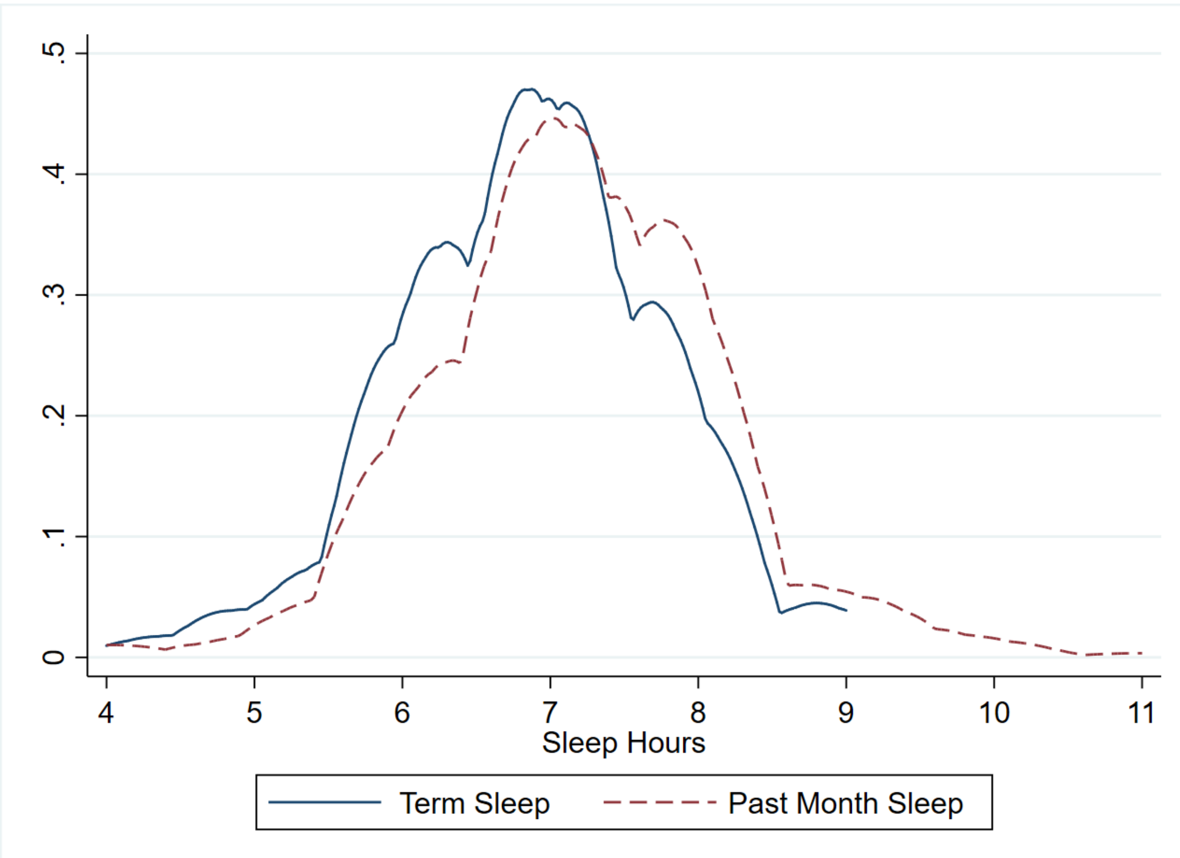
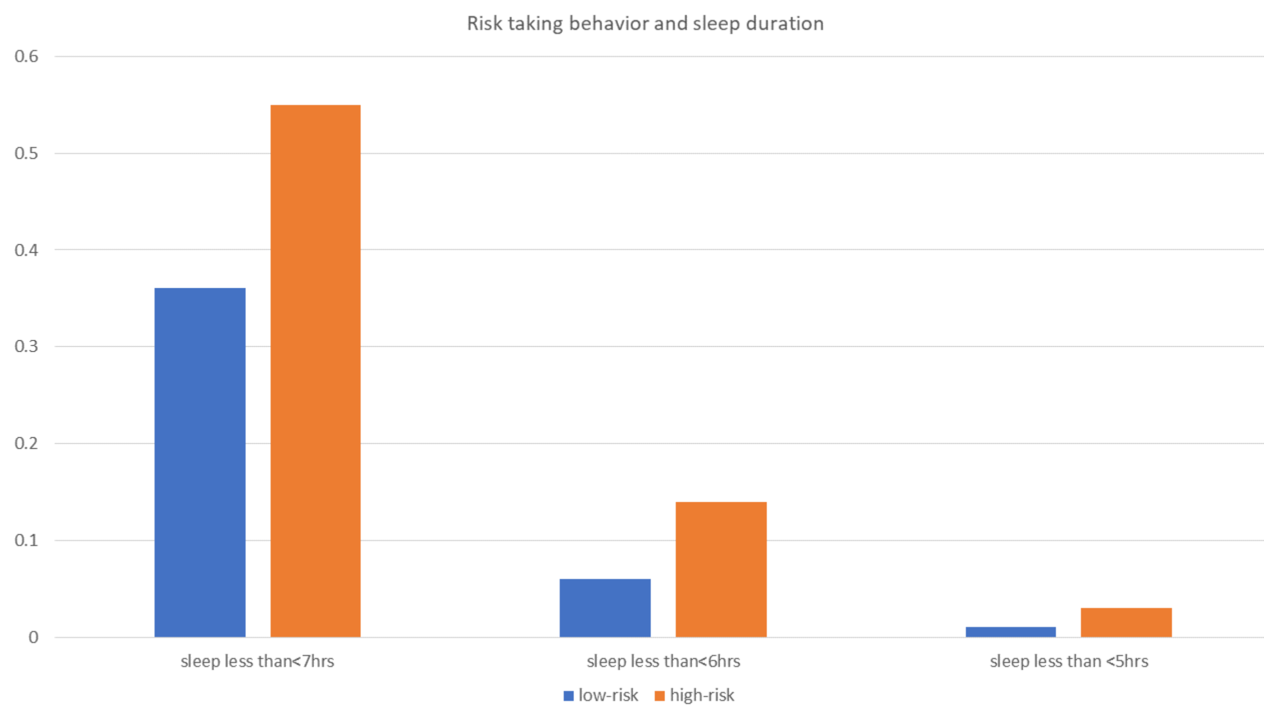


Figure C.2: Self-Reported Sleep Duration at Baseline



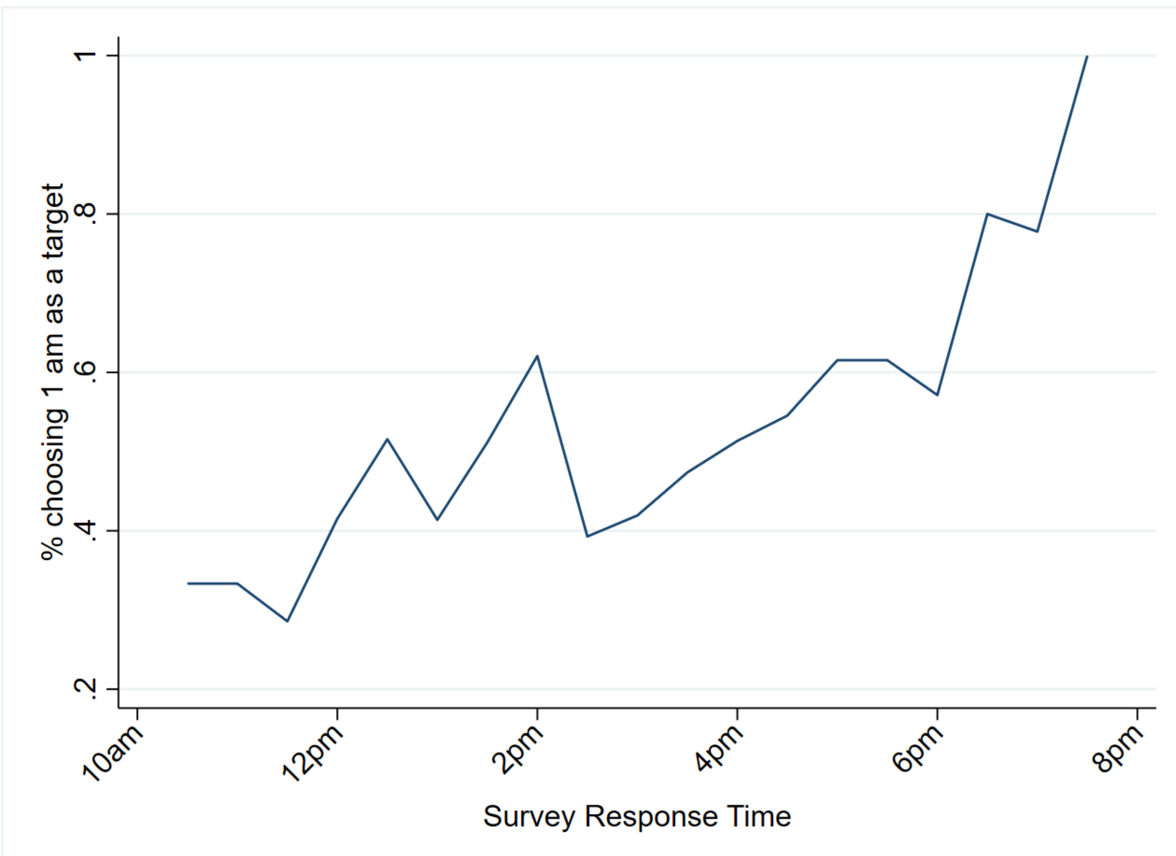
Notes - The figure reports self-reported sleep duration during term (solid) and over the month preceding the survey (dashed) from the Day 1 Survey, which occurred at the beginning of the term.

Figure C.3: Self-Reported Sleep Duration at Baseline and Risk Taking Behavior



Notes - The figure reports the share of individuals self-reporting sleeping less than 7, 6, and 5 hours, respectively, in the Day 1 Survey among low-risk individuals (blue) and high-risk individuals (orange).

Figure C.4: Timing of Survey Response and Bedtime Target Choice



Notes - The figure reports the share of individuals choosing the least binding bedtime target (1 am) by the timing of the survey (10 am-8 pm).This data is drawn from the weekly surveys during intervention weeks.

4.4 D: Appendix Tables

Table D.1: Summary of Treatments

Treatment	Wave	Location	Time	Incentive	Prediction Reward
Control	1	Oxford	Oct-Dec 2016	None	No
Treatment	1	Oxford	Oct-Dec 2016	Biweekly, Weak	No
Treatment	2	Oxford	Apr-Jun 2017	Weekly, Strong	Yes, 1
Treatment	3	Oxford	Oct-Dec 2017	Weekly, Strong	Yes, 3
Treatment	4	Pittsburgh	Jan-Mar 2018	Biweekly, Strong	Yes, 1
Control	5	Pittsburgh	Sep-Nov 2018	None	No
Treatment	5	Pittsburgh	Sep-Nov 2018	Weekly, Strong	Yes, 3

Notes - The table above describes the location, timing and incentive structure used in the different waves of the experiment.

Table D.2: Baseline Characteristics and Attrition

Dep. Var.	Female	Age	White	Black	Asian	Other	Last month sleep
Withdrew from the experiment	0.100	0.272	-0.102	0.053	-0.027	0.076	-0.182
	(0.079)	(0.486)	(0.079)	(0.045)	(0.067)	(0.050)	(0.155)
Observations	359	359	359	359	359	359	359
Dep. Var.	Sleep during term	Slee $<$ 7 hrs during term	Ever smoked	Ideal sleep hours	BMI	Overweight	Obese
Withdrew from the experiment	0.093	-0.116	0.091	-0.167	-1.041	-0.071	0.012
	(0.212)	(0.079)	(0.067)	(0.127)	(1.564)	(0.070)	(0.039)
Observations	359	359	359	359	359	359	359

Notes - Data are drawn by the Day 1 Survey. Each column reports a univariate regression estimate of the dependent variable (baseline characteristics) on a dummy indicating whether the individual withdrew from the experiment.

Table D.3: Comparisons of Sleep Measurements

	Sleep Duration	$7 \leq \text{Sleep} \leq 9$	$\text{Sleep} < 6$
Fitbit	7.02	0.47	0.23
	(1.76)	(0.50)	(0.42)
Self-Reported	7.07	0.61	0.10
	(1.08)	(0.49)	(0.31)
Time Use	8.15	0.59	0.07
	(1.74)	(0.49)	(0.25)

Notes - This table compares averages of our three different measures of sleep collected before our intervention started. Standard deviations are reported in parentheses. The first row (Fitbit) reports the sleep measures derived from the Fitbit data. The second row (Self-Reported) reports the sleep measures elicited in Day 1 Survey. The third row (Time Use) reports the sleep measures based on the time use surveys.

Table D.4: Summary Statistics, Baseline (Survey-based Metrics)

Variable	Mean	Std. Dev.	Variable	Mean	Std. Dev
Demographics			**Depression symptoms**		
Female	0.41	0.49	Rarely	0.44	0.49
Age	21.54	3.90	5-7 days	0.039	0.19
White	0.58	0.49	1-2 days	0.36	0.48
Asian	0.22	0.42	3-4 days	0.15	0.35
Black	0.09	0.28	**Life satisfaction**		
Other	0.11	0.31	Completely satisfied	0.06	0.24
Health and Behaviors			Very satisfied	0.44	0.49
Poor health	0.11	0.31	Somewhat satisfied	0.41	0.49
Weight (kg)	69.44	15.39	Not satisfied (or not at all)	0.09	0.27
Height (cm)	171.46	12.14	**Sleep**		
BMI	23.97	9.72	Sleep last month	7.17	.97
Obese	0.05	0.22	Sleep during term	6.90	1.32
Overweight	0.24	0.43	Less than 7 hrs sleep	0.46	0.50
Ever smoked	0.23	0.41	Ideal sleep	7.97	.78
Drinks (more than once per week)	0.26	.44	Sleep quality (1-10)	6.62	1.61
			# days falling asleep	3.79	4.94
			# days not rested	10.51	6.86

Notes - Summary statistics are drawn from the Day 1 Survey.

Table D.5: Intention to Improve Sleep

	Wants to improve sleep duration	Wants to improve bedtime
	%	%
Definitely yes	17.77	19.34
Probably yes	43.39	41.56
Might or might not	22.73	22.22
Probably not	14.88	15.64
Definitely not	1.24	1.23
Observations	359	359

Notes - Data are drawn from Day 1 Survey.

Table D.6: Correlations Between Sleep, Health, and Well-Being

VARIABLES	Good Health	Obese	Overweight	Depressed	Satisfied
$7 \leq$ Sleep ≤ 9	0.057*	-0.064**	-0.123***	-0.054**	0.259***
	(0.034)	(0.026)	(0.046)	(0.021)	(0.051)
Observations	359	359	359	359	359
R-squared	0.008	0.018	0.020	0.018	0.067
Mean of Dep. Var.	0.880	0.0616	0.252	0.0410	0.489
Std.Dev. of Dep. Var.	0.326	0.241	0.435	0.199	0.501
Sleep Less than 7hrs	-0.030	0.066**	0.120***	0.057***	-0.246***
	(0.033)	(0.026)	(0.046)	(0.022)	(0.051)
Observations	359	359	359	v	359
Mean of Dep. Var.	0.890	0.0616	0.252	0.0414	0.494
Std.Dev. of Dep. Var.	0.314	0.241	0.435	0.200	0.501

Notes - Data are drawn from Day 1 Survey. For this analysis, we used the self-reported measure of sleep duration obtained in the survey.

Table D.7: Demand for Commitment

	Bedtime target earlier than 1am	Sleep target longer than 7 hours	Dominated contract
At least 1 week	60%	60%	13%
All weeks	24%	19%	10%

Notes - The sample in columns 1-2 is restricted to subjects receiving monetary incentives in treatment weeks (N=207). The sample in column 3 is restricted to subjects receiving monetary incentives in treatment weeks in treatment 3 (Biweekly-Small) (N=32).

Table D.8: Present-Bias, Impatience, Overconfidence and Commitment

	(1)	(2)	(3)
	Before 1 am	More than 7hrs	Either
Panel A: Present Bias			
Present-Biased	0.1400*	0.0353	0.1665***
	(0.078)	(0.082)	(0.061)
Observations	207	207	207
Mean of Dep. Var.	0.595	0.590	0.745
Std.Dev. of Dep. Var.	0.492	0.493	0.437
Panel B: Impatience			
Impatient	0.1498*	0.0416	0.0423
	(0.079)	(0.083)	(0.072)
Observations	207	207	207
R-squared	0.016	0.001	0.002
Mean of Dep. Var.	0.595	0.590	0.745
Std.Dev. of Dep. Var.	0.492	0.493	0.437
Panel C: Overconfidence			
Overconfidence	-0.0470	0.0187	-0.0603
	(0.086)	(0.085)	(0.078)
Observations	207	207	207
Mean of Dep. Var.	0.595	0.590	0.745
Std.Dev. of Dep. Var.	0.492	0.493	0.437
Panel D: Risk Aversion			
Risk-averse	-0.1331	-0.0938	-0.1182
	(0.091)	(0.091)	(0.086)
Observations	207	207	207
Mean of Dep. Var.	0.595	0.590	0.745
Std.Dev. of Dep. Var.	0.492	0.493	0.437

Notes - Data are drawn from Fitbit data, weekly surveys collected during the weeks of the intervention, and the first-day survey. The sample is restricted to subjects in the treatment group.

Table D.9: Determinants of Bedtime Targets

	(1)	(2)
Variables	Bedtime Target	
Survey Time	0.0128*	
	(0.0065)	
Average Bedtime		-0.00875**
in Previous Week		(0.00339)
Observations	584	584
R-Squared	0.047	0.059
Mean of Dep. Var.	24.40	24.40
Std.Dev. of Dep. Var.	0.743	0.764

Notes - The table above shows regressions of survey time and average bedtime the previous week on bedtime target. Both regressions include controls for week and subject fixed effects. Standard errors clustered at the individual level are reported in parentheses. Column 1 includes observations with survey times after the first surveys were sent (6 am) and before midnight. The sample is restricted to treated subjects in the treatment weeks.

Table D.10: Predicting Achievement Rate

Prediction	In	Week 1	Week 2	Week 3
For	Week 1	2.83		
	Week 2	2.81	2.71	
	Week 3	2.91	2.71	2.60
Achievement		Week 1	Week 2	Week3
In	Week 1	1.81		
	Week 2	1.78	1.84	
	Week 3	1.47	1.50	1.46
Pred-Ach	In	Week 1	Week 2	Week 3
For	Week 1	1.02		
	Week 2	1.03	0.87	
	Week 3	1.44	1.21	1.14

Notes - This table provides averages for the number of nights subjects predict they will meet their target and the number of nights subjects actually meet their target.

Table D.11: Commitment Devices and Target Achievement

	Bedtime target Before 1am	Bedtime target 1am	% Difference	p-value
% Target achieved	53%	46%	7%	0.17
Achieved at least once	93%	84%	9%	0.065
	Sleep target >7hrs	Sleep target 7hrs	% Difference	p-value
% Target achieved	51%	48%	3%	0.48
Achieved at least once	92%	85%	6%	0.16
	Bedtime before 1 am & Sleep >7hrs	Bedtime 1am or Sleep=7hrs	% Difference	p-value
% Target achieved	55%	47%	8%	0.08
Achieved at least once	95%	85%	10%	0.02

Notes - Data is drawn from Fitbit data and weekly surveys collected during the weeks of the intervention.

Table D.12: Demand for Commitment and Sleep

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		7≤Sleep≤9				Sleep less than 7 hours		
Bedtime <1 am	0.0702**				-0.0591*			
	(0.033)				(0.034)			
Sleep Duration>7hrs		0.1266***				-0.1132***		
		(0.025)				(0.023)		
Bedtime<1 am & Sleep Duration>7hrs			0.1531***				-0.1349***	
			(0.027)				(0.026)	
Dominated Contract (Treatment 3)				0.0496				-0.0551
				(0.080)				(0.077)
Observations	1,420	4,566	4,566	710	1,420	4,566	4,566	710
Mean of Dep. Var.	0.511	0.146	0.102	0.0549	0.511	0.146	0.102	0.0549
Std.Dev. of Dep. Var.	0.500	0.353	0.303	0.228	0.500	0.353	0.303	0.228

Notes - All estimates include a control for insufficient sleep at baseline. The sample is restricted to treated subjects during the treatment weeks. Standard errors are clustered at the individual level and are reported in parentheses.

Table D.13: Incentives and Sleep, Other Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep.Var.	Sleep Hours		ln(Sleep Hours)		Sleep<7		Sleep<6		Sleep<5	
Treatment	0.240***	0.124*	0.049***	0.029**	-0.079***	-0.041**	-0.058***	-0.032*	-0.033**	-0.015
	(0.089)	(0.071)	(0.016)	(0.013)	(0.025)	(0.018)	(0.022)	(0.016)	(0.016)	(0.013)
Post-Treatment	0.206	0.104	0.046*	0.026	-0.039	-0.006	-0.059*	-0.042**	-0.031	-0.010
	(0.127)	(0.082)	(0.023)	(0.016)	(0.038)	(0.023)	(0.032)	(0.021)	(0.023)	(0.015)
Individual FE		YES		YES		YES		YES		YES
Observations	7,690	7,690	7,690	7,690	7,690	7,690	7,690	7,690	7,690	7,690
Mean of DV	6.922	6.922	1.891	1.891	0.466	0.466	0.250	0.250	0.131	0.131
SD of DV	1.849	1.849	0.332	0.332	0.499	0.499	0.433	0.433	0.337	0.337
Number of id		280		280		280		280		280

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2, 4, 6, 8 and 10 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

Table D.14: Incentives, Bedtime, and Wake-Up Time

	(1)	(2)	(3)	(4)
VARIABLES	Bedtime		Wake up Time	
Treatment	-0.3222***	-0.2117***	-0.1111	-0.1023
	(0.105)	(0.058)	(0.127)	(0.092)
Post-Treatment	0.0015	0.0368	0.2669	0.2491
	(0.181)	(0.076)	(0.197)	(0.154)
Individual fixed effects		YES		YES
Observations	7,690	7,690	7,690	7,690
Mean of Dep. Var.	00.94	00.94	8.011	8.011
Std.Dev. of Dep. Var.	1.631	1.631	3.046	3.046
Number of id		273		273

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clustered at the individual level are reported in parentheses.

Table D.15: Baseline Characteristics and Sample Attrition in Follow-Up Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Female	Age	White	Black	Asian	Other	Last month sleep
No follow up	0.044	0.685**	-0.041	0.054*	0.030	-0.043	0.123
	(0.053)	(0.326)	(0.053)	(0.030)	(0.045)	(0.034)	(0.104)
Observations	359	359	359	358	359	359	359
Dep. Var.	Sleep during term	Sleep $<$ 6 hrs during term	Ever smoked	Ideal sleep hours	BMI	Overweight	Obese
No follow up	0.190	0.008	0.066	-0.034	1.033	-0.032	0.019
	(0.142)	(0.054)	(0.045)	(0.085)	(1.037)	(0.047)	(0.026)
Observations	359	359	359	359	359	359	359

Notes - Data are drawn by the Day 1 Survey. Each column reports a univariate regression estimate of the dependent variable (baseline characteristics) on a dummy indicating whether the individual did not respond to the follow-up survey.

Table D.16: Follow-up Sleep Quality

Variables	Sleep Quality Z-Score	
Treatment	0.459**	
	(0.229)	
Achievement Rate		0.433**
		(0.211)
R-Squared	0.132	0.164
Mean of Dep. Var.	-0.018	0.164
Std.Dev. of Dep. Var.	0.596	0.573

Notes - The table above shows regressions of treatment and bedtime target achievement rate on a zZ-score of sleep quality formed from the sleep quality questions asked in the follow up survey. All estimates include controls for wave, gender, and a quadratic in age.

Table D.17: Naps

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Nap		Nap duration		7<Sleep<9	
Treatment	-0.0122	-0.0110	-1.0161	-0.8449	0.0489***	0.0490***
	(0.008)	(0.009)	(0.706)	(0.746)	(0.017)	(0.017)
After treatment	-0.0088	-0.0049	-0.7387	-0.3508	0.0168	0.0169
	(0.011)	(0.012)	(0.916)	(1.004)	(0.023)	(0.023)
Nap					-0.098	
					0.020	
Nap duration						-0.0011***
						(0.000)
Individual fixed effects		YES		YES	YES	YES
Observations	8,738	8,738	8,738	8,738	8,738	8,738
Mean of Dep. Var.	0.0570	0.0570	4.638	4.638	0.456	0.456
Std.Dev. of Dep. Var.	0.232	0.232	19.73	19.73	0.498	0.498
Number of id		319		319	319	319

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2, 4, and 6 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table D.18: Incentives and Sleep, Weekends

	(1)	(2)	(3)	(4)
VARIABLES	7<Sleep<9		Sleep<6 hours	
Treatment	0.0707	0.0698	-0.137**	-0.102
	-0.0682	-0.0692	-0.0676	-0.0669
Individual fixed effects		YES		YES
Observations	3342	3342	3342	3342
Individuals	280	280	280	280
Mean of Dep. Var.	0.453	0.453	0.250	0.250
Std.Dev. of Dep. Var.	0.498	0.498	0.433	0.433

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects.Standard errors clutered at the individual level are reported in parentheses.

Table D.19: Baseline Characteristics and Sample Attrition in Time-Use Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Female	Age	White	Black	Asian	Other	Last month sleep
No follow up	0.045	-0.887***	-0.165***	0.102***	0.073	-0.010	-0.200*
	(0.053)	(0.326)	(0.053)	(0.030)	(0.045)	(0.034)	(0.105)
Observations	359	359	359	359	359	359	359
R-squared	0.002	0.020	0.027	0.031	0.007	0.000	0.010
Dep. Var.	Sleep during term	Sleep < 7hrs during term	Ever smoked	Ideal sleep hours	BMI	Overweight	Obese
No follow up	-0.062	0.095*	0.072	-0.168**	-0.527	0.017	0.033
	(0.144)	(0.054)	(0.045)	(0.085)	(1.047)	(0.047)	(0.026)
Observations	359	359	359	359	359	359	359
R-squared	0.001	0.009	0.007	0.011	0.001	0.000	0.005
0.000	0.003	0.001	0.002				

Notes - Data are drawn by the Day 1 Survey. Each column reports a univariate regression estimate of the dependent variable (baseline characteristics) on a dummy indicating whether the individual did not respond to the time-use survey.

Table D.20: Incentives to Sleep and Screen Time Near Bedtime

	(1) Screen time (hours) after 8 pm	(2) Any screen time after 8 pm
Night on which treatment was achieved	-0.224***	-0.125***
	(0.050)	(0.040)
Post-treatment	-0.132*	-0.081*
(achieved 50% of nights)	(0.077)	(0.048)
Observations	1,106	1,106
R-squared	0.039	0.030
Mean of Dep. Var.	0.452	0.372
Std.Dev. of Dep. Var.	0.793	0.483

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Standard errors are clustered at the individual level.

Table D.21: Incentives and Sleep Regularity

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Std.Dev. Sleep Hours		Std.Dev. Bedtime		Std.Dev. Wake up time	
Treatment Treatment	-0.1025*	-0.0625	-0.1353***	-0.0881**	-0.0895	-0.0473
	(0.053)	(0.040)	(0.050)	(0.045)	(0.069)	(0.059)
Post-treatment	-0.0254	0.0084	0.0237	0.0606	-0.1718*	-0.1575*
	(0.089)	(0.072)	(0.072)	(0.068)	(0.097)	(0.083)
Individual fixed effects		YES		YES		YES
Observations	7,690	7,690	7,690	7,690		7,690
R-squared	0.048	0.049	0.018	0.017	0.023	0.031
Mean of Dep. Var.	1.289	1.289	1.085	1.085	1.155	1.155
Std.Dev. of Dep. Var.	0.854	0.854	0.762	0.762	1.040	1.040

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2, 4, and 6 include individual fixed effects.Standard errors clutered at the individual level are reported in parentheses.

Table D.22: Incentives and Sleep: Timing of the Incentives

	(1)	(2)	(3)	(4)
	Sleep7≤Sleep≤9		Sleep<6 hours	
Weekly incentive	0.0932***	0.0374	-0.0560**	-0.0143
	(0.031)	(0.024)	(0.028)	(0.021)
Post-weekly incentive	0.0560	0.0085	-0.0718*	-0.0458*
	(0.043)	(0.028)	(0.037)	(0.024)
Bi-weekly incentive	0.0675**	0.0537*	-0.0884***	-0.0711***
	(0.034)	(0.031)	(0.028)	(0.025)
Post-biweekly incentive	0.0116	-0.0073	-0.0449	-0.0283
	(0.037)	(0.032)	(0.028)	(0.023)
Individual fixed effects		YES		YES
Observations	8,738	8,738	8,738	8,738
Mean of Dep. Var.	0.456	0.456	0.245	0.245
Std.Dev. of Dep. Var.	0.498	0.498	0.430	0.430

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table D.23: Incentives and Sleep: the Role of the Size of the Financial Incentive

	(1)	(2)	(3)	(4)
	Sleep7≤Sleep≤9		Sleep<6 hours	
Strong Treatmet	0.0845***	0.0495***	-0.0607***	-0.0351**
	(0.024)	(0.018)	(0.021)	(0.016)
Post Strong Treatment	0.0507	0.0175	-0.0609*	-0.0477**
	(0.036)	(0.024)	(0.031)	(0.021)
Weak Treatment	0.0481	0.0214	-0.0473	-0.0173
	(0.052)	(0.040)	(0.039)	(0.032)
Post Weak Treatment	-0.0066	0.0009	-0.0402	-0.0492
	(0.078)	(0.063)	(0.057)	(0.045)
Individual fixed effects		YES		YES
Observations	8,738	8,738	8,738	8,738
R-squared	0.014	0.005	0.015	0.007
Mean of Dep. Var.	0.456	0.456	0.245	0.245
Std.Dev. of Dep. Var.	0.498	0.498	0.430	0.430
Number of id		319		319

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

Table D.24: Incentives and Sleep

	(1)	(2)	(3)	(4)
VARIABLES	Sleep7<Sleep<9		Sleep<6 hours	
Any Incentive	0.0792***	0.0451***	-0.0588***	-0.0325**
	(0.022)	(0.017)	(0.020)	(0.015)
Post-Treatment (any incentive)	0.0483	0.0147	-0.0600**	-0.0469**
	(0.035)	(0.023)	(0.030)	(0.020)
Individual fixed effects		YES		YES
Observations	8,738	8,738	8,738	8,738
Mean of Dep. Var.	0.456	0.456	0.245	0.245
Std.Dev. of Dep. Var.	0.498	0.498	0.430	0.430
Number of id		319		319

Notes - All estimates include controls for gender, a quadratic in age, week of the semester dummies and day of the week dummies, and a control for the experiment wave and the location of the experiment (Oxford, Pittsburgh). Columns 2 and 4 include individual fixed effects. Standard errors clutered at the individual level are reported in parentheses.

4.5 E: Elicitation of Risk and Time Preferences

We used choice lists to elicit participants' risk and time preferences. The subjects could choose from two columns, representing Option A and Option B. On each list, one of the two options was fixed, and the other option changed from one row to the next. In each row, subjects had to indicate their preferred option: Option A or Option B. To avoid multiple switching points on a single list, the subjects only had to choose in which row they wanted to switch from choosing Option A to choosing Option B. The subjects were given examples and the opportunity to practice before making decisions that counted for payment. When payments involved a future date, the subjects would receive the corresponding amount via email in the form of a gift card.

To elicit the risk preference parameter, we used two lists. On each list, Option A was a fixed lottery: a 50% chance of getting GBP 6 and a 50% chance of getting GBP 0. Option B was always a sure amount. The lists we used are illustrated in Figures E.1 and E.2.

To elicit the time preference parameters, we used four lists. On each list, Option A was associated with a monetary payment at a sooner time and Option B implied some monetary payment at a later time. The amount to be gained at the later time is fixed at GBP 6, and the amount to be gained at the sooner time varied on each list. Among the lists, the sooner time is either today or in 4 weeks, and the delay between the later and the sooner time is either 4 weeks or 8 weeks. The lists we used are illustrated in Figures E.3, E.4, E.5 and E.6.

Figure E.1: Choice List for Risk Preference 1

Option A	Option B
50% Chance of £6 and 50% Chance of £0	£0.00
50% Chance of £6 and 50% Chance of £0	£0.30
50% Chance of £6 and 50% Chance of £0	£0.60
50% Chance of £6 and 50% Chance of £0	£0.90
50% Chance of £6 and 50% Chance of £0	£1.20
50% Chance of £6 and 50% Chance of £0	£1.50
50% Chance of £6 and 50% Chance of £0	£1.80
50% Chance of £6 and 50% Chance of £0	£2.10
50% Chance of £6 and 50% Chance of £0	£2.40
50% Chance of £6 and 50% Chance of £0	£2.70
50% Chance of £6 and 50% Chance of £0	£3.00
50% Chance of £6 and 50% Chance of £0	£3.30
50% Chance of £6 and 50% Chance of £0	£3.60
50% Chance of £6 and 50% Chance of £0	£3.90
50% Chance of £6 and 50% Chance of £0	£4.20
50% Chance of £6 and 50% Chance of £0	£4.50
50% Chance of £6 and 50% Chance of £0	£4.80
50% Chance of £6 and 50% Chance of £0	£5.10
50% Chance of £6 and 50% Chance of £0	£5.40
50% Chance of £6 and 50% Chance of £0	£5.70
50% Chance of £6 and 50% Chance of £0	£6.00

Figure E.2: Choice List for Risk Preference 2

Option A	Option B
50% Chance of £6 and 50% Chance of £0	£0.00
50% Chance of £6 and 50% Chance of £0	£0.30
50% Chance of £6 and 50% Chance of £0	£0.60
50% Chance of £6 and 50% Chance of £0	£0.90
50% Chance of £6 and 50% Chance of £0	£1.20
50% Chance of £6 and 50% Chance of £0	£1.50
50% Chance of £6 and 50% Chance of £0	£1.80
50% Chance of £6 and 50% Chance of £0	£2.10
50% Chance of £6 and 50% Chance of £0	£2.40
50% Chance of £6 and 50% Chance of £0	£2.70
50% Chance of £6 and 50% Chance of £0	£3.00
50% Chance of £6 and 50% Chance of £0	£3.30
50% Chance of £6 and 50% Chance of £0	£3.60
50% Chance of £6 and 50% Chance of £0	£3.90
50% Chance of £6 and 50% Chance of £0	£4.20
50% Chance of £6 and 50% Chance of £0	£4.50
50% Chance of £6 and 50% Chance of £0	£4.80
50% Chance of £6 and 50% Chance of £0	£5.10
50% Chance of £6 and 50% Chance of £0	£5.40
50% Chance of £6 and 50% Chance of £0	£5.70
50% Chance of £6 and 50% Chance of £0	£6.00

Figure E.3: Choice List for Time Preference 1

Option A	Option B
Receive £5.80 today	Receive £6 in 4 weeks
Receive £5.60 today	Receive £6 in 4 weeks
Receive £5.40 today	Receive £6 in 4 weeks
Receive £5.20 today	Receive £6 in 4 weeks
Receive £5.00 today	Receive £6 in 4 weeks
Receive £4.80 today	Receive £6 in 4 weeks
Receive £4.60 today	Receive £6 in 4 weeks
Receive £4.40 today	Receive £6 in 4 weeks
Receive £4.20 today	Receive £6 in 4 weeks
Receive £4.00 today	Receive £6 in 4 weeks
Receive £3.80 today	Receive £6 in 4 weeks
Receive £3.60 today	Receive £6 in 4 weeks
Receive £3.40 today	Receive £6 in 4 weeks
Receive £3.20 today	Receive £6 in 4 weeks
Receive £3.00 today	Receive £6 in 4 weeks
Receive £2.80 today	Receive £6 in 4 weeks
Receive £2.60 today	Receive £6 in 4 weeks
Receive £2.40 today	Receive £6 in 4 weeks
Receive £2.20 today	Receive £6 in 4 weeks
Receive £2.00 today	Receive £6 in 4 weeks
Receive £1.80 today	Receive £6 in 4 weeks
Receive £1.60 today	Receive £6 in 4 weeks
Receive £1.40 today	Receive £6 in 4 weeks
Receive £1.20 today	Receive £6 in 4 weeks
Receive £1.00 today	Receive £6 in 4 weeks
Receive £0.80 today	Receive £6 in 4 weeks
Receive £0.60 today	Receive £6 in 4 weeks
Receive £0.40 today	Receive £6 in 4 weeks
Receive £0.20 today	Receive £6 in 4 weeks

Figure E.4: Choice List for Time Preference 2

Option A	Option B
Receive £5.80 today	Receive £6 in 8 weeks
Receive £5.60 today	Receive £6 in 8 weeks
Receive £5.40 today	Receive £6 in 8 weeks
Receive £5.20 today	Receive £6 in 8 weeks
Receive £5.00 today	Receive £6 in 8 weeks
Receive £4.80 today	Receive £6 in 8 weeks
Receive £4.60 today	Receive £6 in 8 weeks
Receive £4.40 today	Receive £6 in 8 weeks
Receive £4.20 today	Receive £6 in 8 weeks
Receive £4.00 today	Receive £6 in 8 weeks
Receive £3.80 today	Receive £6 in 8 weeks
Receive £3.60 today	Receive £6 in 8 weeks
Receive £3.40 today	Receive £6 in 8 weeks
Receive £3.20 today	Receive £6 in 8 weeks
Receive £3.00 today	Receive £6 in 8 weeks
Receive £2.80 today	Receive £6 in 8 weeks
Receive £2.60 today	Receive £6 in 8 weeks
Receive £2.40 today	Receive £6 in 8 weeks
Receive £2.20 today	Receive £6 in 8 weeks
Receive £2.00 today	Receive £6 in 8 weeks
Receive £1.80 today	Receive £6 in 8 weeks
Receive £1.60 today	Receive £6 in 8 weeks
Receive £1.40 today	Receive £6 in 8 weeks
Receive £1.20 today	Receive £6 in 8 weeks
Receive £1.00 today	Receive £6 in 8 weeks
Receive £0.80 today	Receive £6 in 8 weeks
Receive £0.60 today	Receive £6 in 8 weeks
Receive £0.40 today	Receive £6 in 8 weeks
Receive £0.20 today	Receive £6 in 8 weeks

Figure E.5: Choice List for Time Preference 3

Option A	Option B
Receive £5.80 in 4 weeks	Receive £6 in 8 weeks
Receive £5.60 in 4 weeks	Receive £6 in 8 weeks
Receive £5.40 in 4 weeks	Receive £6 in 8 weeks
Receive £5.20 in 4 weeks	Receive £6 in 8 weeks
Receive £5.00 in 4 weeks	Receive £6 in 8 weeks
Receive £4.80 in 4 weeks	Receive £6 in 8 weeks
Receive £4.60 in 4 weeks	Receive £6 in 8 weeks
Receive £4.40 in 4 weeks	Receive £6 in 8 weeks
Receive £4.20 in 4 weeks	Receive £6 in 8 weeks
Receive £4.00 in 4 weeks	Receive £6 in 8 weeks
Receive £3.80 in 4 weeks	Receive £6 in 8 weeks
Receive £3.60 in 4 weeks	Receive £6 in 8 weeks
Receive £3.40 in 4 weeks	Receive £6 in 8 weeks
Receive £3.20 in 4 weeks	Receive £6 in 8 weeks
Receive £3.00 in 4 weeks	Receive £6 in 8 weeks
Receive £2.80 in 4 weeks	Receive £6 in 8 weeks
Receive £2.60 in 4 weeks	Receive £6 in 8 weeks
Receive £2.40 in 4 weeks	Receive £6 in 8 weeks
Receive £2.20 in 4 weeks	Receive £6 in 8 weeks
Receive £2.00 in 4 weeks	Receive £6 in 8 weeks
Receive £1.80 in 4 weeks	Receive £6 in 8 weeks
Receive £1.60 in 4 weeks	Receive £6 in 8 weeks
Receive £1.40 in 4 weeks	Receive £6 in 8 weeks
Receive £1.20 in 4 weeks	Receive £6 in 8 weeks
Receive £1.00 in 4 weeks	Receive £6 in 8 weeks
Receive £0.80 in 4 weeks	Receive £6 in 8 weeks
Receive £0.60 in 4 weeks	Receive £6 in 8 weeks
Receive £0.40 in 4 weeks	Receive £6 in 8 weeks
Receive £0.20 in 4 weeks	Receive £6 in 8 weeks

Figure E.6: Choice List for Time Preference 4

Option A	Option B
Receive £5.80 in 4 weeks	Receive £6 in 12 weeks
Receive £5.60 in 4 weeks	Receive £6 in 12 weeks
Receive £5.40 in 4 weeks	Receive £6 in 12 weeks
Receive £5.20 in 4 weeks	Receive £6 in 12 weeks
Receive £5.00 in 4 weeks	Receive £6 in 12 weeks
Receive £4.80 in 4 weeks	Receive £6 in 12 weeks
Receive £4.60 in 4 weeks	Receive £6 in 12 weeks
Receive £4.40 in 4 weeks	Receive £6 in 12 weeks
Receive £4.20 in 4 weeks	Receive £6 in 12 weeks
Receive £4.00 in 4 weeks	Receive £6 in 12 weeks
Receive £3.80 in 4 weeks	Receive £6 in 12 weeks
Receive £3.60 in 4 weeks	Receive £6 in 12 weeks
Receive £3.40 in 4 weeks	Receive £6 in 12 weeks
Receive £3.20 in 4 weeks	Receive £6 in 12 weeks
Receive £3.00 in 4 weeks	Receive £6 in 12 weeks
Receive £2.80 in 4 weeks	Receive £6 in 12 weeks
Receive £2.60 in 4 weeks	Receive £6 in 12 weeks
Receive £2.40 in 4 weeks	Receive £6 in 12 weeks
Receive £2.20 in 4 weeks	Receive £6 in 12 weeks
Receive £2.00 in 4 weeks	Receive £6 in 12 weeks
Receive £1.80 in 4 weeks	Receive £6 in 12 weeks
Receive £1.60 in 4 weeks	Receive £6 in 12 weeks
Receive £1.40 in 4 weeks	Receive £6 in 12 weeks
Receive £1.20 in 4 weeks	Receive £6 in 12 weeks
Receive £1.00 in 4 weeks	Receive £6 in 12 weeks
Receive £0.80 in 4 weeks	Receive £6 in 12 weeks
Receive £0.60 in 4 weeks	Receive £6 in 12 weeks
Receive £0.40 in 4 weeks	Receive £6 in 12 weeks
Receive £0.20 in 4 weeks	Receive £6 in 12 weeks

4.6 F: Appendix Figures and Tables

Figure F.1: Age-Adjusted Mortality Summary Statistics

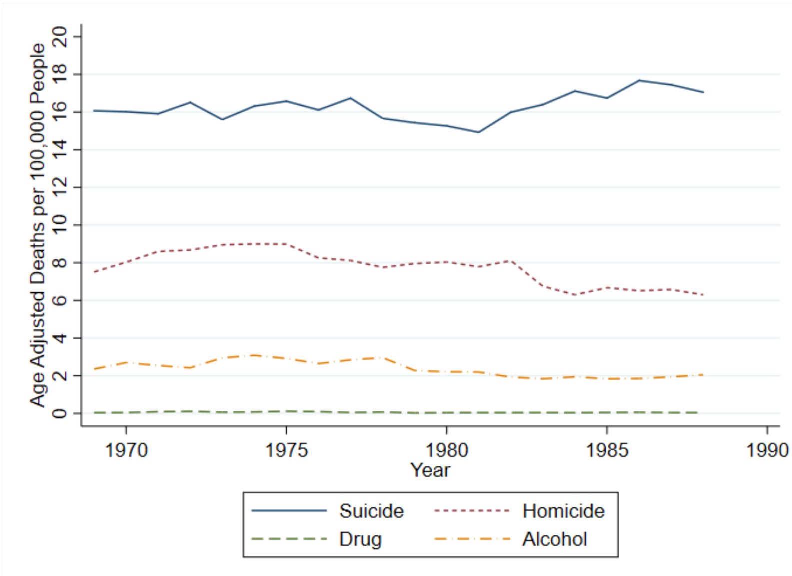


Figure F.2: Age-Adjusted Mortality by Urban/Rural

(a) Urban

(b) Rural

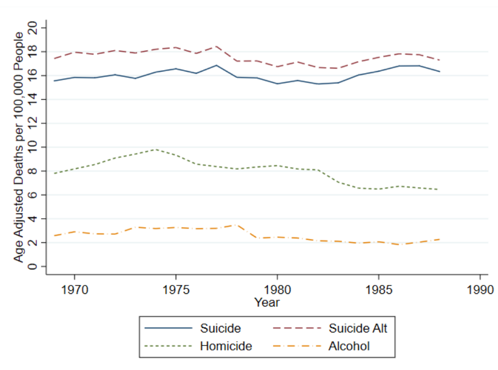


Figure F.3: Event-Study: Suicide among Young Whites

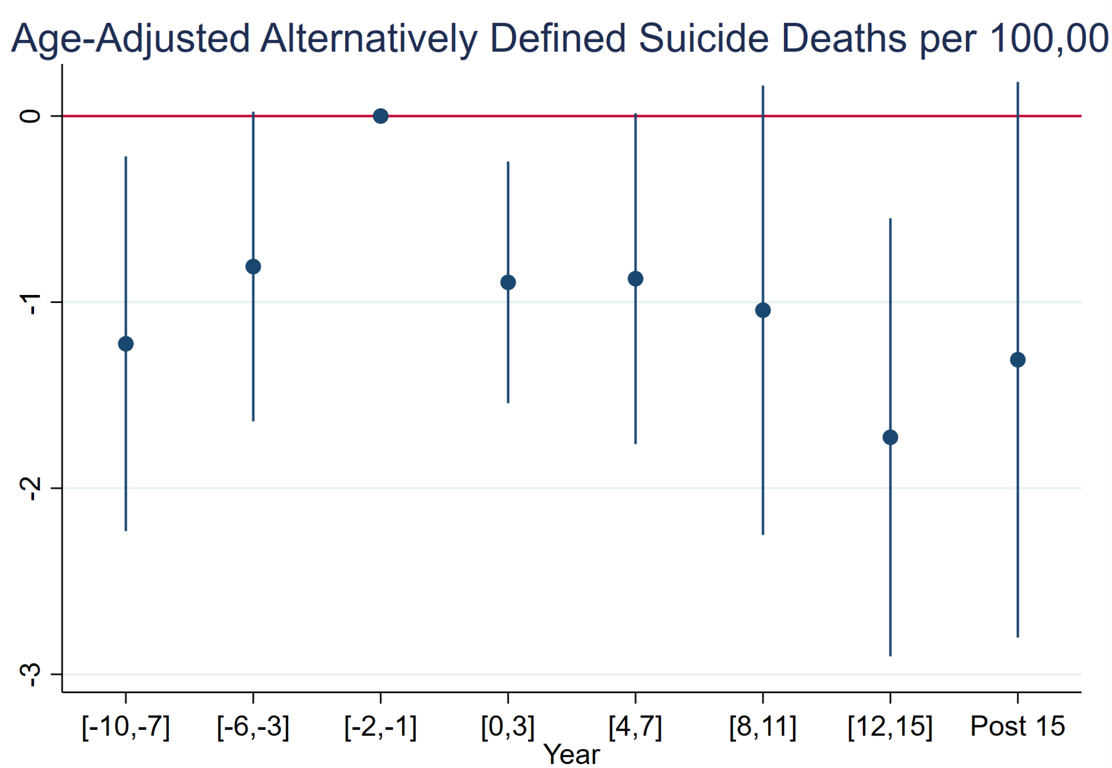


Figure F.4: Event-Study: Homicide among Young Adult Non-Whites

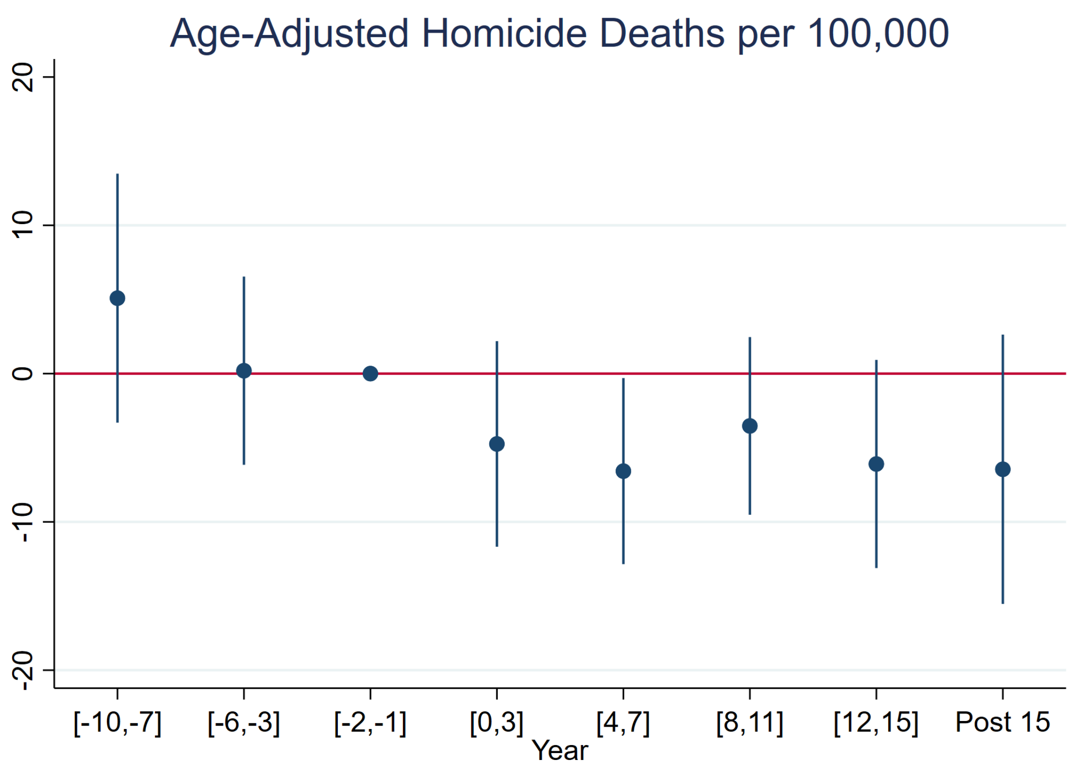


Figure F.5: Event-Study: Homicide among Adult Non-Whites

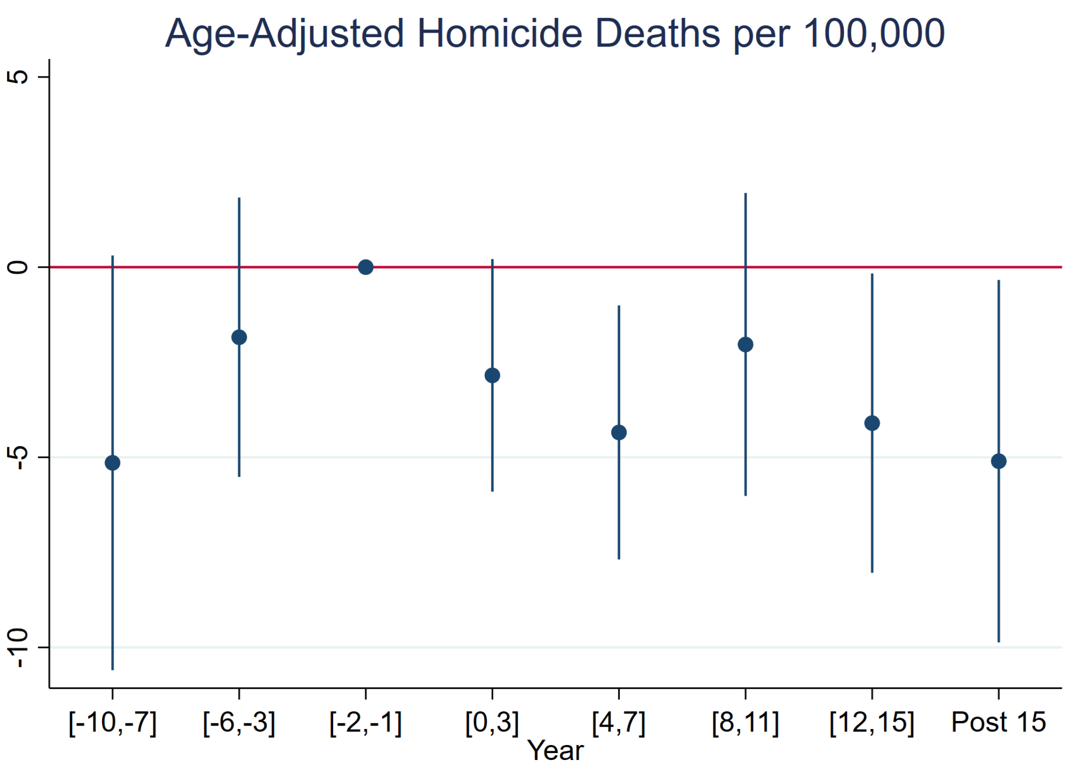


Figure F.6: Event-Study: Homicide among Elderly Non-Whites

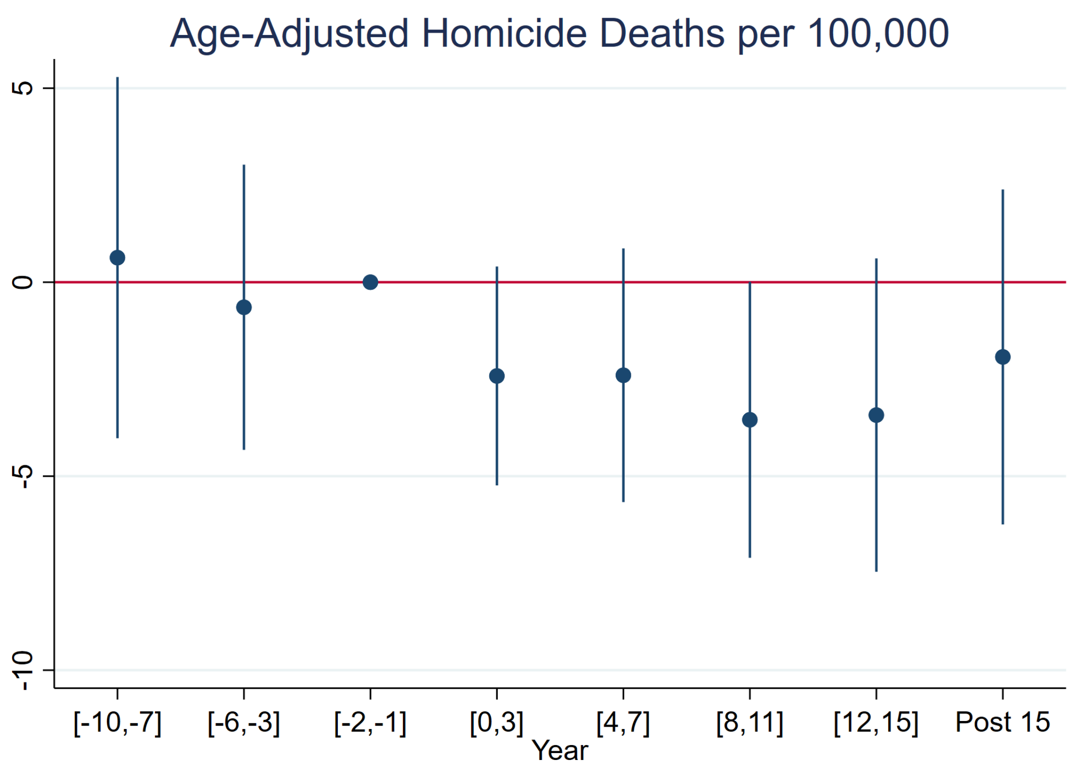


Figure F.7: Event-Study: Alcohol Related Mortality among Adult Non-Whites

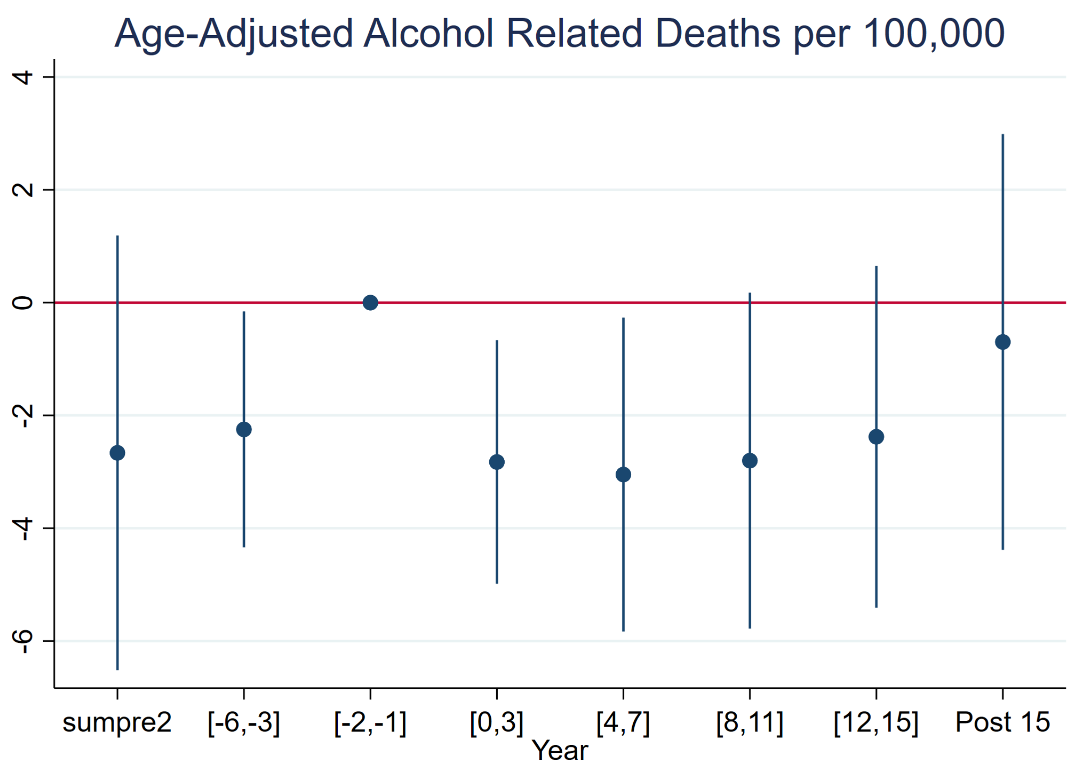


Table F.1: Urban vs. Rural Sub-group Analysis - White Subsample

	(1)	(2)	(3)
	Suicide	Homicide	Alcohol
Panel A: Urban CMHC			
CMHC	-0.008	-0.002	0.021
	(0.010)	(0.020)	(0.054)
Panel B: Rural CMHCs			
CMHC	0.040	-0.034	0.058
	(0.282)	(0.053)	(0.115)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes- Clustered standard errors are in parentheses.*$p < .10$, **$p < .05$, ***$p < .01$. Years included: 1969-1988.

Table F.2: Boundary Analysis: Sub-group Analysis by Age - White Population Only

	(1)	(2)	(3)
	Suicide	Homicide	Alcohol
Panel A: Young			
CMHC	-0.476	0.193	0.005
	(0.296)	(0.241)	(0.027)
Panel B: Young Adult			
CMHC	0.150	0.079	-0.010
	(0.275)	(0.247)	(0.106)
Panel C: Adult			
CMHC	-0.436	0.195	0.093
	(0.316)	(0.178)	(0.296)
Panel D: Elderly			
CMHC	-0.701	-0.105	-0.063
	(0.440)	(0.175)	(0.221)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes- Standard errors clustered at the county level are in parenthesis *$p < .10$, **$p < .05$, ***$p < .01$. Years included: 1969-1988. Limited to boundary analysis sample. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate. Regressions are weighted by county population.

Table F.3: Boundary Analysis: Sub-group Analysis by Age - Non-White Population Only

	(1)	(2)	(3)
	Suicide	Homicide	Alcohol
Panel A: Young			
CMHC	-0.035	-2.777	0.096
	(0.612)	(2.966)	(0.118)
Panel B: Young Adult			
CMHC	-0.919	-5.705**	0.135
	(0.663)	(2.673)	(0.559)
Panel C: Adult			
CMHC	0.313	-2.599**	-1.984*
	(0.673)	(1.306)	(1.141)
Panel D: Elderly			
CMHC	-0.028	-2.107*	0.134
	(0.745)	(1.206)	(0.797)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes- Standard errors clustered at the county level are in parenthesis $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$. Years included: 1969-1988. Limited to boundary analysis sample. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate. Regressions are weighted by county population.

Table F.4: Boundary Analysis: Urban vs. Rural Sub-group Analysis - Non-White Subsample

	(1)	(2)	(3)
	Suicide	Homicide	Alcohol
Panel A: Urban CMHC			
CMHC	-0.202	-3.059	-0.205
	(0.414)	(2.057)	(0.438)
Panel B: Rural CMHCs			
CMHC	-0.081	-9.876***	-3.170**
	(1.259)	(3.045)	(1.290)
County Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
State-Specific Linear Time Trend	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes- Standard errors clustered at the county level are in parenthesis *$p < .10$, **$p < .05$, ***$p < .01$. Years included: 1969-1988. Limited to boundary analysis sample. Regressions include year fixed effects, state-linear time trends, urban-category-by-year linear time trends, and controls for percent less than high school education, percent high school education, unemployment rate, and labor force participation rate. Regressions are weighted by county population.

5.0 Bibliography

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Abuse, S. and Administration, M. H. S. (2018). *Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health*. (HHS Publication No. SMA 17-5044, NSDUH Series H-52). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse, and Mental Health Services Administration.

Acland, D. and Levy, M. R. (2015). Naiveté, projection bias, and habit formation in gym attendance. *Management Science*, 61(1):146–160.

Aguiar, M. and Hurst, E. (2007). Measuring trends in leisure: The allocation of time over five decades. *The Quarterly Journal of Economics*, pages 969–1006.

Aguiar, M., Hurst, E., and Karabarbounis, L. (2013). Time use during the great recession. *The American Economic Review*, 103(5):1664–1696.

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–56.

Antonovics, K. L. and Sander, R. H. (2013). Affirmative action bans and the “chilling effect”. *American law and economics review*, 15(1):252–299.

Ashraf, N., Karlan, D., and Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *Quarterly Journal of Economics*, 121(2):635–672.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115.

Ayalew, S., Manian, S., and Sheth, K. (2018). Discrimination from below: Experimental evidence on female leadership in ethiopia. Technical report, mimeo.

Ayres, I. and Brooks, R. (2004). Does affirmative action reduce the number of black lawyers. *Stan. L. Rev.*, 57:1807.

Bachrach, L. L. (1974). *Patients at federally funded rural community mental health centers in 1971*, volume 102. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration.

Bagde, S., Epple, D., and Taylor, L. (2016). Does affirmative action work? caste, gender, college quality, and academic success in india. *American Economic Review*, 106(6):1495–1521.

Bai, L., Handel, B., Miguel, E., and Rao, G. (2017). Self-control and demand for preventive health: Evidence from hypertension in india. Technical report, National Bureau of Economic Research Working Paper No.W23727.

Bailey, M. J. and DiPrete, T. A. (2016). Five decades of remarkable but slowing change in us women's economic and social status and political participation. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(4):1–32.

Bailey, M. J. and Goodman-Bacon, A. (2015). The war on poverty's experiment in public medicine: Community health centers and the mortality of older americans. *American Economic Review*, 105(3):1067–1104.

Balafoutas, L. and Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, 335(6068):579–582.

Baltrunaite, A., Bello, P., Casarico, A., and Profeta, P. (2014). Gender quotas and the quality of politicians. *Journal of Public Economics*, 118:62–74.

Banerjee, R., Gupta, N. D., and Villeval, M. C. (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review*, 101:567–604.

Barber, B. M. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1):261–292.

Bartel, A. and Taubman, P. (1986). Some economic and demographic consequences of mental illness. *Journal of Labor Economics*, 4(2):243–256.

Barth, E., Kerr, S. P., and Olivetti, C. (2017). The dynamics of gender earnings differentials: Evidence from establishment data. Technical report, National Bureau of Economic Research.

Basner, M., Fomberstein, K. M., Razavi, F. M., Banks, S., William, J. H., Rosa, R. R., and Dinges, D. F. (2007). American time use survey: sleep time and its relationship to waking activities. *Sleep*, 30(9):1085–1095.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., and Topalova, P. (2009). Powerful women: does exposure reduce bias? *The Quarterly journal of economics*, 124(4):1497–1540.

Beattie, Z., Pantelopoulos, A., Ghoreyshi, A., Oyang, Y., Statan, A., and Heneghan, C. (2017). Estimation of sleep stages using cardiac and accelerometer data from a wrist-worn device. In *Sleep*, volume 40, pages A26–A26. OXFORD UNIV PRESS INC JOURNALS DEPT, 2001 EVANS RD, CARY, NC 27513 USA.

Beaurain, G. and Masclet, D. (2016). Does affirmative action reduce gender discrimination and enhance efficiency? new experimental evidence. *European Economic Review*, 90:350–362.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, 9(3):226–232.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, pages 493–517.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–64.

Benham, L. and Benham, A. (1982). Employment, earnings, and psychiatric diagnosis. In *Economic aspects of health*, pages 203–220. University of Chicago Press.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–55.

Besley, T., Folke, O., Persson, T., and Rickne, J. (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American economic review*, 107(8):2204–42.

Bessone, P., Rao, G., Schilbach, F., Schofield, H., and Toma, M. (2018). Sleepless in chennai: The economic effects of sleep deprivation among the poor. Technical report, Working paper. MIT.

Biddle, J. E. and Hamermesh, D. S. (1990). Sleep and the allocation of time. *Journal of Political Economy*, 98(5, Part 1):922–943.

Billari, F. C., Giuntella, O., and Stella, L. (2018). Broadband internet, digital temptations, and sleep. *Journal of Economic Behavior and Organization*.

Bitler, M. and Carpenter, C. (2019). Effects of direct care provision to the uninsured: Evidence from federal breast and cervical cancer programs. Technical report, National Bureau of Economic Research.

Blank, M. B., Fox, J. C., Hargrove, D. S., and Turner, J. T. (1995). Critical issues in reforming rural mental health service delivery. *Community Mental Health Journal*, 31(6):511–524.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Bohren, J. A., Imas, A., and Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review*, 109(10):3395–3436.

Boyd, J. E., Katz, E. P., Link, B. G., and Phelan, J. C. (2010). The relationship of multiple aspects of stigma and personal contact with someone hospitalized for mental illness, in a nationally representative sample. *Social psychiatry and psychiatric epidemiology*, 45(11):1063–1070.

Bracha, A., Cohen, A., and Conell-Price, L. (2019). The heterogeneous effect of affirmative action on performance. *Journal of Economic Behavior & Organization*, 158:173–218.

Breig, Z., Gibson, M., and Shrader, J. (2018). Why do we procrastinate? present bias and optimism. Technical report, Mimeo.

Brown, A. and Chowdhury, S. M. (2017). The hidden perils of affirmative action: Sabotage in handicap contests. *Journal of Economic Behavior & Organization*, 133:273–284.

Brown, S., Inskip, H., and Barraclough, B. (2000). Causes of the excess mortality of schizophrenia. *The British Journal of Psychiatry*, 177(3):212–217.

Bryan, G., Karlan, D., and Nelson, S. (2010). Commitment devices. *Annual Review of Economics*, 2(1):671–698.

Buboltz, W. C., Brown, F., and Soper, B. (2001). Sleep habits and patterns of college students: a preliminary study. *Journal of American College Health*, 50(3):131–135.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447.

Calsamiglia, C., Franke, J., and Rey-Biel, P. (2013). The incentive effects of affirmative action in a real-effort tournament. *Journal of Public Economics*, 98:15–31.

Cappuccio, F. P., D'Elia, L., Strazzullo, P., and Miller, M. A. (2010). Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep*, 33(5):585.

Carrera, M., Royer, H., Stehr, M., and Sydnor, J. (2019a). The structure of health incentives: evidence from a field experiment. *Management Science*.

Carrera, M., Royer, H., Stehr, M., Sydnor, J., and Taubinsky, D. (2019b). How are preferences for commitment revealed? Technical report, National Bureau of Economic Research.

Casaburi, L. and Macchiavello, R. (2019). Demand and supply of infrequent payments as a commitment device: evidence from kenya. *American Economic Review*, 109(2):523–55.

Charness, G. and Gneezy, U. (2009). Incentives to exercise. *Econometrica*, 77(3):909–931.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660.

Coffman, K. B., Collis, M., and Kulkarni, L. (2019a). Stereotypes and belief updating. Technical report, Harvard Business School.

Coffman, K. B., Collis, M., and Kulkarni, L. (2019b). When to apply? Technical report, Harvard Business School.

Cordner, G. (2014). Community policing. *The Oxford handbook of police and policing*, pages 148–171.

Cotton, C., Hickman, B. R., and Price, J. P. (2014). Affirmative action and human capital investment: Evidence from a randomized field experiment. Technical report, National Bureau of Economic Research.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47(2):448–74.

Currie, J. and Stabile, M. (2006). Child mental health and human capital accumulation: the case of adhd. *Journal of health economics*, 25(6):1094–1118.

Czibor, E. and Dominguez Martinez, S. (2019). Never too late: Gender quotas in the final round of a multistage tournament. *The Journal of Law, Economics, and Organization*, 35(2):319–363.

DellaVigna, S. and Malmendier, U. (2006). Paying not to go to the gym. *American Economic Review*, 96(3):694–719.

Dianat, A., Echenique, F., and Yariv, L. (2018). Statistical discrimination and affirmative action in the lab.

Druss, B. G., Zhao, L., Von Esenwein, S., Morrato, E. H., and Marcus, S. C. (2011). Understanding excess mortality in persons with mental illness: 17-year follow up of a nationally representative us survey. *Medical care*, pages 599–604.

Dupas, P. and Robinson, J. (2013). Why don't the poor save more? evidence from health savings experiments. *American Economic Review*, 103(4):1138–71.

Eisenberg, D., Golberstein, E., and Hunt, J. B. (2009). Mental health and academic success in college. *The BE Journal of Economic Analysis & Policy*, 9(1).

Eisenberg, D., Hunt, J., and Speer, N. (2013). Mental health in american colleges and universities: variation across student subgroups and across campuses. *The Journal of Nervous and Mental Disease*, 201(1):60–67.

Erkal, N., Gangadharan, L., and Xiao, E. (2019). Leadership selection: Can changing the default break the glass ceiling? *Available at SSRN 3022386*.

Ettner, S. L., Frank, R. G., and Kessler, R. C. (1997). The impact of psychiatric disorders on labor market outcomes. *ILR Review*, 51(1):64–81.

Ettner, S. L., Maclean, J. C., and French, M. T. (2011). Does having a dysfunctional personality hurt your career? axis ii personality disorders and labor market outcomes. *Industrial Relations: A Journal of Economy and Society*, 50(1):149–173.

Falk, A., Becker, A., Dohmen, T. J., Huffman, D., and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Working Paper*.

Fallucchi, F. and Quercia, S. (2018). Affirmative action and retaliation in experimental contests. *Journal of Economic Behavior & Organization*, 156:23–40.

Fletcher, J. M. (2008). Adolescent depression: diagnosis, treatment, and educational attainment. *Health economics*, 17(11):1215–1235.

Flory, J. A., Leibbrandt, A., and List, J. A. (2014). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies*, 82(1):122–155.

Frank, R. G. and McGuire, T. G. (2000). Economics and mental health. *Handbook of health economics*, 1:893–954.

French, M. T. and Zarkin, G. A. (1998). Mental health, absenteeism and earnings at a large manufacturing worksite. *The Journal of Mental Health Policy and Economics*, 1(4):161–172.

Gang, I. N., Landon-Lane, J. S., and Yun, M.-S. (2003). Does the glass ceiling exist? a cross-national perspective on gender income mobility. Technical report, IZA Discussion Paper.

Garcia, L. T., Erskine, N., Hawn, K., and Casmay, S. R. (1981). The effect of affirmative action on attributions about minority group members. *Journal of Personality*, 49(4):427–437.

Gibson, M. and Shrader, J. (2018). Time use and labor productivity: The returns to sleep. *Review of Economics and Statistics*, 100(5):783–798.

Giné, X., Karlan, D., and Zinman, J. (2010). Put your money where your butt is: a commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2(4):213–35.

Giuntella, O., Han, W., and Mazzonna, F. (2017). Circadian rhythms, sleep, and cognitive skills: Evidence from an unsleeping giant. *Demography*, 54(5):1715–1742.

Giuntella, O. and Mazzonna, F. (2019). Sunset time and the economic effects of social jetlag. evidence from us time zone borders. *Journal of Health Economics*, 65.

Giuntella, O., McManus, S., Mujcic, R., Oswald, A. J., Powdthavee, N., and Tohamy, A. (2019). Why is there so much midlife distress in affluent nations?

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Greiner, B. and Stephanides, M. (2019). Subject pools and recruitment. In *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing.

Grob, G. N. (1994). Government and mental health policy: A structural analysis. *The Milbank Quarterly*, pages 471–500.

Gronfein, W. (1985). Incentives and intentions in mental health policy: A comparison of the medicaid and community mental health programs. *Journal of Health and Social Behavior*, pages 192–206.

Gujral, K. and Basu, A. (2019). Impact of rural and urban hospital closures on inpatient mortality. Technical report, National Bureau of Economic Research.

Guvenen, F., Kaplan, G., and Song, J. (2014). The glass ceiling and the paper floor: Gender differences among top earners, 1981-2012. Technical report, National Bureau of Economic Research.

Hafner, M., Stepanek, M., Taylor, J., Troxel, W. M., and van Stolk, C. (2017). Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis. *Rand Health Quarterly*, 6(4).

Hamermesh, D. S. (2019). *Spending Time: The Most Valuable Resource*. Oxford University Press.

Hamilton, V. H., Merrigan, P., and Dufresne, É. (1997). Down and out: estimating the relationship between mental health and unemployment. *Health economics*, 6(4):397–406.

Handel, B. and Kolstad, J. (2017). Wearable technologies and health behaviors: new data and new methods to understand population health. *American Economic Review: Papers and Proceedings*, 107(5):481–85.

Harcourt, B. E. (2011). An institutionalization effect: The impact of mental hospitalization and imprisonment on homicide in the united states, 1934-2001. *The Journal of Legal Studies*, 40(1):39–83.

Harrison, G. W., Lau, M. I., and Williams, M. B. (2002). Estimating individual discount rates in denmark: A field experiment. *American Economic Review*, 92(5):1606–1617.

Hauser, D. J. and Schwarz, N. (2016). Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants. *Behavior research methods*, 48(1):400–407.

He, J., Kang, S., and Lacetera, N. (2019). Leaning in or not leaning out? opt-out choice framing attenuates gender differences in the decision to compete. Technical report, National Bureau of Economic Research.

Heilman, M. E. (1997). Sex discrimination and the affirmative action remedy: The role of sex stereotypes. In *Women in corporate management*, pages 5–17. Springer.

Heilman, M. E. and Alcott, V. B. (2001). What i think you think of me: Women's reactions to being viewed as beneficiaries of preferential selection. *Journal of Applied Psychology*, 86(4):574.

Heilman, M. E., Block, C. J., and Stathatos, P. (1997). The affirmative action stigma of incompetence: Effects of performance information ambiguity. *Academy of Management Journal*, 40(3):603–625.

Heilman, M. E., Kaplow, S. R., Amato, M. A., and Stathatos, P. (1993). When similarity is a liability: Effects of sex-based preferential selection on reactions to like-sex and different-sex others. *Journal of Applied Psychology*, 78(6):917.

Heilman, M. E., Lucas, J. A., and Kaplow, S. R. (1990). Self-derogating consequences of sex-based preferential selection: The moderating role of initial self-confidence. *Organizational Behavior and Human Decision Processes*, 46(2):202–216.

Heilman, M. E., McCullough, W. F., and Gilbert, D. (1996). The other side of affirmative action: Reactions of nonbeneficiaries to sex-based preferential selection. *Journal of Applied Psychology*, 81(4):346.

Heilman, M. E., Simon, M. C., and Repper, D. P. (1987). Intentionally favored, unintentionally harmed? impact of sex-based preferential selection on self-perceptions and self-evaluations. *Journal of Applied Psychology*, 72(1):62.

Heissel, J. A. and Norris, S. (2018). Rise and shine the effect of school start times on academic performance from childhood through puberty. *Journal of Human Resources*, 53(4):957–992.

Hershner, S. D. and Chervin, R. D. (2014). Causes and consequences of sleepiness among college students. *Nature and Science of Sleep*, 6:73.

Hiday, V. A., Swartz, M. S., Swanson, J. W., Borum, R., and Wagner, H. R. (1999). Criminal victimization of persons with severe mental illness. *Psychiatric services*, 50(1):62–68.

Hiroeh, U., Appleby, L., Mortensen, P. B., and Dunn, G. (2001). Death by homicide, suicide, and other unnatural causes in people with mental illness: a population-based study. *The Lancet*, 358(9299):2110–2112.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental economics*, 14(3):399–425.

Hossain, T. and Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, 80(3):984–1001.

Huang, J. and Kisgen, D. J. (2013). Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics*, 108(3):822–839.

Huffman, D., Raymond, C., and Shvets, J. (2018). Persistent overconfidence and biased memory: Evidence from managers.

Ibañez, M. and Riener, G. (2018). Sorting through affirmative action: Three field experiments in colombia. *Journal of Labor Economics*, 36(2):437–478.

Insel, T. R. (2008). Assessing the economic costs of serious mental illness.

Jagnani, M. (2018). Poor sleep: Sunset time and human capital production. *Mimeo*.

Jakicic, J. M., Davis, K. K., Rogers, R. J., King, W. C., Marcus, M. D., Helsel, D., Rickman, A. D., Wahed, A. S., and Belle, S. H. (2016). Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the idea randomized clinical trial. *JAMA*, 316(11):1161–1171.

Jemelka, R., Trupin, E., and Chiles, J. (1989). The mentally ill in prisons: A review. *Hospital and Community Psychiatry*, 40:481–485.

Jin, L., Ziebarth, N. R., et al. (2015). Sleep and human capital: Evidence from daylight saving time. Technical report, HEDG, c/o Department of Economics, University of York.

Just, D. R. and Price, J. (2013). Using incentives to encourage healthy eating in children. *Journal of Human Resources*, 48(4):855–872.

Kahn-Greene, E. T., Killgore, D. B., Kamimori, G. H., Balkin, T. J., and Killgore, W. D. (2007). The effects of sleep deprivation on symptoms of psychopathology in healthy adults. *Sleep Medicine*, 8(3):215–221.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123(6):1227–1277.

Keeler, E. B., Manning, W. G., and Wells, K. B. (1988). The demand for episodes of mental health services. *Journal of Health Economics*, 7(4):369–392.

Kelleher, K., Chaffin, M., Hollenberg, J., and Fischer, E. (1994). Alcohol and drug disorders among physically abusive and neglectful parents in a community-based sample. *American Journal of Public Health*, 84(10):1586–1590.

Kenig, S. (1992). *Who plays? Who pays? Who cares?: A case study in applied sociology, political economy, and the community mental health centers movement*. Baywood Publishing Company.

Kessler, R. C., Walters, E. E., and Forthofer, M. S. (1998). The social consequences of psychiatric disorders, iii: probability of marital stability. *American Journal of Psychiatry*, 155(8):1092–1096.

Kirk, S. A. (1974). The impact of labeling on rejection of the mentally ill: An experimental study. *Journal of Health and Social Behavior*, pages 108–117.

Kleinjans, K. J. (2009). Do gender differences in preferences for competition matter for occupational expectations? *Journal of Economic Psychology*, 30(5):701–710.

Kremer, M., Rao, G., and Schilbach, F. (2019). Behavioral development economics. *Handbook of Behavioral Economics*, 2.

Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121.

Kurlaender, M. and Grodsky, E. (2013). Mismatch and the paternalistic justification for selective college admissions. *Sociology of Education*, 86(4):294–310.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Laibson, D. (2015). Why don't present-biased agents make commitments? *American Economic Review*, 105(5):267–72.

Laibson, D. (2018). Private paternalism, the commitment puzzle, and model-free equilibrium. In *AEA Papers and Proceedings*, volume 108, pages 1–21.

Lally, P., Van Jaarsveld, C. H., Potts, H. W., and Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European journal of social psychology*, 40(6):998–1009.

Landsman, R. (2018). Gender differences in executive departure. Technical report.

Lang, M. (2013). The impact of mental health insurance laws on state suicide rates. *Health economics*, 22(1):73–88.

Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K., and Rathouz, P. J. (2008a). Self-reported and measured sleep duration: how similar are they? *Epidemiology (Cambridge, Mass.)*, 19(6):838–845.

Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K., and Rathouz, P. J. (2008b). Sleep duration: How well do self-reports reflect objective measures? The cardia sleep study. *Epidemiology (Cambridge, Mass.)*, 19(6):838.

Lee, H.-A., Lee, H.-J., Moon, J.-H., Lee, T., Kim, M.-G., In, H., Cho, C.-H., and Kim, L. (2017). Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. *Psychiatry Investigation*, 14(2):179.

Leibbrandt, A., Wang, L. C., and Foo, C. (2017). Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women. *Management Science*, 64(8):3501–3516.

Leonard, J. S. (1984). The impact of affirmative action on employment. *Journal of Labor Economics*, 2(4):439–463.

Lerner, D. and Henke, R. M. (2008). What does research tell us about depression, job performance, and work productivity? *Journal of Occupational and Environmental Medicine*, 50(4):401–410.

Lewis, G. B. (1986). Gender and promotions: Promotion chances of white men and women in federal white-collar employment. *Journal of Human Resources*, pages 406–419.

Libertella, A. F., Sora, S. A., and Natale, S. M. (2007). Affirmative action policy and changing views. *Journal of business ethics*, 74(1):65–71.

Link, B. G., Andrews, H., and Cullen, F. T. (1992). The violent and illegal behavior of mental patients reconsidered. *American Sociological Review*, pages 275–292.

Link, B. G. and Cullen, F. T. (1986). Contact with the mentally ill and perceptions of how dangerous they are. *Journal of Health and Social Behavior*, pages 289–302.

Liu, C. H., Stevens, C., Wong, S. H., Yasui, M., and Chen, J. A. (2019). The prevalence and predictors of mental health diagnoses and suicide among us college students: Implications for addressing disparities in service use. *Depression and anxiety*, 36(1):8–17.

Liu, Y. (2016). Prevalence of healthy sleep duration among adults—united states, 2014. *MMWR. Morbidity and mortality weekly report*, 65.

Luyster, F. S., Strollo Jr, P. J., Zee, P. C., Walsh, J. K., et al. (2012). Sleep: a health imperative. *Sleep*, 35(6):727–734.

Maggian, V., Montinari, N., and Nicolò, A. (2017). Do quotas help women to climb the career ladder? a laboratory experiment. Technical report, GATE-Lyon Saint-Etienne, WP.

Maniglio, R. (2009). Severe mental illness and criminal victimization: a systematic review. *Acta Psychiatrica Scandinavica*, 119(3):180–191.

Marcotte, D. E. and Wilcox-Gok, V. (2003). Estimating earning losses due to mental illness: a quantile regression approach. *Journal of Mental Health Policy and Economics*, 6(3):123–134.

Matsa, D. A. and Miller, A. R. (2013). A female style in corporate leadership? evidence from quotas. *American Economic Journal: Applied Economics*, 5(3):136–69.

McEwen, B. S. (2006). Sleep deprivation as a neurobiologic and physiologic stressor: allostasis and allostatic load. *Metabolism*, 55:S20–S23.

McGrath, J., Saha, S., Chant, D., and Welham, J. (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiologic reviews*, 30(1):67–76.

McGuire, T. G. (1981). Financing and demand for mental health services. *The Journal of Human Resources*, 16(4):501–522.

McKenna, B. S., Dickinson, D. L., Orff, H. J., and Drummond, S. P. (2007). The effects of one night of sleep deprivation on known-risk and ambiguous-risk decisions. *Journal of sleep research*, 16(3):245–252.

Mechanic, D. (2002). Removing barriers to care among persons with psychiatric symptoms. *Health Affairs*, 21(3):137–147.

Mechanic, D. (2014). More people than ever before are receiving behavioral health care in the united states, but gaps and challenges remain. *Health Affairs*, 33(8):1416–1424.

Milkman, K. L., Minson, J. A., and Volpp, K. G. (2013). Holding the hunger games hostage at the gym: An evaluation of temptation bundling. *Management Science*, 60(2):283–299.

Miller, A. R. and Segal, C. (2012). Does temporary affirmative action produce persistent effects? a study of black and female employment in law enforcement. *Review of Economics and Statistics*, 94(4):1107–1125.

Miller, C. (2017). The persistent effect of temporary affirmative action. *American Economic Journal: Applied Economics*, 9(3):152–90.

Milojevich, H. M. and Lukowski, A. F. (2016). Sleep and mental health in undergraduate students with generally healthy sleep habits. *PloS one*, 11(6):e0156372.

Mintz, J., Mintz, L. I., Arruda, M. J., and Hwang, S. S. (1992). Treatments of depression and the functional capacity to work. *Archives of General Psychiatry*, 49(10):761–768.

Misra, R. and McKean, M. (2000). College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. *American journal of Health studies*, 16(1):41.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence. Technical report, National Bureau of Economic Research.

Monk, T. H., Buysse, D. J., Carrier, J., Billy, B. D., and Rose, L. R. (2001). Effects of afternoon "siesta" naps on sleep, alertness, performance, and circadian rhythms in the elderly. *Sleep*, 24(6):680–687.

Morin, L.-P. (2015). Do men and women respond differently to competition? evidence from a major education reform. *Journal of Labor Economics*, 33(2):443–491.

Morrissey, J. P. and Goldman, H. H. (1986). Care and treatment of the mentally ill in the united states: Historical developments and reforms. *The Annals of the American Academy of Political and Social Science*, 484(1):12–27.

Mortier, P., Auerbach, R. P., Alonso, J., Bantjes, J., Benjet, C., Cuijpers, P., Ebert, D. D., Green, J. G., Hasking, P., Nock, M. K., et al. (2018). Suicidal thoughts and behaviors among first-year college students: Results from the wmh-ics project. *Journal of the American Academy of Child & Adolescent Psychiatry*, 57(4):263–273.

Mullahy, J. and Sindelar, J. L. (1993). Alcoholism, work, and income. *Journal of Labor Economics*, 11(3):494–520.

Mullainathan, S. (2014). Get some sleep, and wake up the gdp the new york times. *The New York Times*.

Naierman, N., Haskins, B., and Robinson, G. (1978). *Community Mental Health Centers - A decade later*. Abt Books.

Nie, N. H. and Hillygus, D. S. (2002). The impact of internet use on sociability: Time-diary findings. *It & Society*, 1(1):1–20.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How costly is diversity? affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1):1–16.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101.

Niels-Jakob Harbo, H., Karl, H., Erik, Ö., and Hans-Henrik, S. (2019). On using pareto distributions for measuring top-income gender disparities. Technical report.

Noon, M. (2010). The shackled runner: time to rethink positive discrimination? *Work, Employment and Society*, 24(4):728–739.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

O'Donoghue, T. and Rabin, M. (2001). Choice and procrastination. *The Quarterly Journal of Economics*, 116(1):121–160.

O'Donoghue, T. and Rabin, M. (2003). Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review*, 93(2):186–191.

O'Donoghue, T. and Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, 90(10-11):1825–1849.

O'Donoghue, T. and Rabin, M. (2015). Present bias: Lessons learned and to be learned. *American Economic Review*, 105(5):273–79.

O'Donoghue, T., Rabin, M., et al. (2006). Incentives and self-control. *Econometric Society Monographs*, 42:215.

on the Homeless, U. S. I. C. (1990). *The 1990 annual report of the Interagency Council on the Homeless.* Washington D.C.: The Council.

Organization, W. H. (2001). *The World Health Report 2001: Mental Health New Understanding, New Hope*. Geneva: World Health Organization Office of Publication.

Patel, M. S., Asch, D. A., and Volpp, K. G. (2015). Wearable devices as facilitators, not drivers, of health behavior change. *JAMA*, 313(5):459–460.

Petters, L., Schroeder, M., et al. (2017). Justification as a key determinant of the success of affirmative action. Technical report, Cologne Graduate School in Management, Economics and Social Sciences.

Rabin, M., O'Donoghue, T., et al. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

Raphael, S. and Stoll, M. A. (2013). Assessing the contribution of the deinstitutionalization of the mentally ill to growth in the us incarceration rate. *The Journal of Legal Studies*, 42(1):187–222.

Reetz, D. R., Krylowicz, B., and Mistler, B. (2014). The association for university and college counseling center directors annual survey. *Aurora*, 51:60506.

Roan Gresenz, C. and Sturm, R. (2004). Mental health and employment transitions. In *The economics of gender and mental illness*, pages 95–108. Emerald Group Publishing Limited.

Rose, S. M. (1979). Deciphering deinstitutionalization: Complexities in policy and program analysis. *The Milbank Memorial Fund Quarterly. Health and Society*, pages 429–460.

Rothstein, J. and Yoon, A. H. (2008). Affirmative action in law school admissions: What do racial preferences do? Technical report, National Bureau of Economic Research.

Rowan, K., McAlpine, D. D., and Blewett, L. A. (2013). Access and cost barriers to mental health care, by insurance status, 1999–2010. *Health affairs*, 32(10):1723–1730.

Royer, H., Stehr, M., and Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, 7(3):51–84.

Saarni, S. I., Suvisaari, J., Sintonen, H., Pirkola, S., Koskinen, S., Aromaa, A., and Lönnqvist, J. (2007). Impact of psychiatric disorders on health-related quality of life: general population survey. *The British journal of psychiatry*, 190(4):326–332.

Saha, S., Chant, D., and McGrath, J. (2007). A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time? *Archives of general psychiatry*, 64(10):1123–1131.

Sarsons, H. and Xu, G. (2015). Confidence men? gender and confidence: Evidence among top economists. Technical report, Harvard University, Department of Economics, Littauer Center.

Schilbach, F. (2019). Alcohol and self-control: A field experiment in india. *American Economic Review*, 109(4):1290–1322.

Schlesinger, M. and Dorwart, R. (1984). Ownership and mental-health services: A reappraisal of the shift toward privately owned facilities. *New England Journal of Medicine*, 311(15):959–965.

Seierstad, C. and Opsahl, T. (2011). For the few not the many? the effects of affirmative action on presence, prominence, and social capital of women directors in norway. *Scandinavian Journal of Management*, 27(1):44–54.

Sipe, T. A., Finnie, R. K., Knopf, J. A., Qu, S., Reynolds, J. A., Thota, A. B., Hahn, R. A., Goetzel, R. Z., Hennessy, K. D., McKnight-Eily, L. R., et al. (2015). Effects of mental health benefits legislation: A community guide systematic review. *American journal of preventive medicine*, 48(6):755–766.

Smith, N. (2018). Gender quotas on boards of directors. Technical report, IZA World of Labor.

Steadman, H. J., Mulvey, E. P., Monahan, J., Robbins, P. C., Appelbaum, P. S., Grisso, T., Roth, L. H., and Silver, E. (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Archives of general psychiatry*, 55(5):393–401.

Stein, L. I. and Test, M. A. (1980). Alternative to mental hospital treatment: I. conceptual model, treatment program, and clinical evaluation. *Archives of general psychiatry*, 37(4):392–397.

Sutter, M., Glätzle-Rützler, D., Balafoutas, L., and Czermak, S. (2016). Cancelling out early age gender differences in competition: an analysis of policy interventions. *Experimental Economics*, 19(2):412–432.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from vietnam. *American Economic Review*, 100(1):557–71.

Teitler, J. O. and Reichman, N. E. (2008). Mental illness as a barrier to marriage among unmarried mothers. *Journal of Marriage and Family*, 70(3):772–782.

Teplin, L. A., McClelland, G. M., Abram, K. M., and Weiner, D. A. (2005). Crime victimization in adults with severe mental illness: comparison with the national crime victimization survey. *Archives of general psychiatry*, 62(8):911–921.

Test, M. A. and Stein, L. I. (1980). Alternative to mental hospital treatment: Iii. social cost. *Archives of General Psychiatry*, 37(4):409–412.

Thomas, K. A. and Clifford, S. (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, 77:184–197.

Thornicroft, G. (2011). Physical health disparities and mental illness: the scandal of premature mortality. *The British Journal of Psychiatry*, 199(6):441–442.

Torrey, E. F. (1994). Violent behavior by individuals with serious mental illness. *Psychiatric services*, 45(7):653–662.

Torrey, E. F. (1997). *Out of the Shadows: Confronting America's Mental Illness Crisis*. John Wiley and Sons, Inc.

Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, 86(3):859–889.

Trockel, M. T., Barnes, M. D., and Egget, D. L. (2000). Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors. *Journal of American college health*, 49(3):125–131.

Twenge, J. M., Krizan, Z., and Hisler, G. (2017). Decreases in self-reported sleep duration among us adolescents 2009–2015 and association with new media screen time. *Sleep Medicine*, 39:47–53.

Unzueta, M. M., Gutiérrez, A. S., and Ghavami, N. (2010). How believing in affirmative action quotas affects white women's self-image. *Journal of Experimental Social Psychology*, 46(1):120–126.

Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., Galvin, R., Zhu, J., Wan, F., DeGuzman, J., et al. (2009). A randomized, controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*, 360(7):699–709.

Weber, E. U., Blais, A.-R., and Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4):263–290.

Weisbrod, B. A. (1983). A guide to benefit-cost analysis, as seen through a controlled experiment in treating the mentally ill. *Journal of Health Politics, Policy and Law*, 7(4):808–845.

Weisbrod, B. A., Test, M. A., and Stein, L. I. (1980). Alternative to mental hospital treatment: Ii. economic benefit-cost analysis. *Archives of General Psychiatry*, 37(4):400–405.

Williams, W. M. and Ceci, S. J. (2015). National hiring experiments reveal 2: 1 faculty preference for women on stem tenure track. *Proceedings of the National Academy of Sciences*, 112(17):5360–5365.

Wood, W. and Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological review*, 114(4):843.

Zhang, M., Rost, K. M., and Fortney, J. C. (1999a). Earnings changes for depressed individuals treated by mental health specialists. *American Journal of Psychiatry*, 156(1):108–114.

Zhang, M., Rost, K. M., Fortney, J. C., and Smith, G. R. (1999b). A community study of depression treatment and employment earnings. *Psychiatric Services*, 50(9):1209–1213.

Zimmermann, F. (2019). The dynamics of motivated beliefs. *American Economic Review*.