Center for Financial Security
UNIVERSITY OF WISCONSIN-MADISON

Retirement & Disability Research Center

Yang Wang
University of Wisconsin – Madison
Email: ywang26@wisc.edu

Muzhe Yang
Lehigh University
Email: muzheyang@lehigh.edu

Disparities by Race and Gender in SS(D)I Applications and Awards

Center for Financial Security

University of Wisconsin-Madison

1300 Linden Drive
Madison, WI 53706

608-890-0229
cfs@mailplus.wisc.edu
cfs.wisc.edu

The research reported herein was performed pursuant to a grant from the US Social Security Administration (SSA) funded as part of the Retirement and Disability Consortium. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA or any agency of the Federal Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation, or favoring by the United States Government or any agency thereof.

Abstract

This study examines racial and gender disparities in the applications and awards of Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) programs, using administrative records from the Social Security Administration and self-reported data from the Health and Retirement Study. Our research aims to enrich the toolset for analyzing disparities by adopting a more data-driven approach with fewer subjective assumptions. Our study highlights the sensitivity of disparity estimates obtained from the ordinary least squares (OLS) method, commonly used in disparity examinations. We find that OLS estimates are sensitive not only to the control variables in a regression model but also to the interaction terms between the group indicator (race or gender) and these control variables. These sensitivities are tied to the subjective assumptions researchers must make when estimating disparities using OLS.

To address these limitations, we employ a new method known as the double/debiased machine learning estimator, proposed by Chernozhukov et al. (2018). This alternative approach enables us to estimate racial and gender disparities with fewer subjective assumptions, and it flexibly accounts for numerous potential interaction effects arising from various characteristics like education and income. Using this estimator, we find minimal evidence of racial disparities (White vs. Black) in SS(D)I applications and awards. However, suggestive evidence points to a lower prevalence of SSDI applications among women. This contrast suggests the need for increased outreach efforts to facilitate SSDI applications for women. Meanwhile, racial disparities in SS(D)I applications and awards may be attributed to disparities in other socioeconomic dimensions.

Keywords: racial and gender disparities; SSDI and SSI; machine learning

JEL codes: H53; J18; C18

1. Introduction

The Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) programs, administered by the Social Security Administration (SSA), play a vital role in providing financial support to individuals with disabilities. SSI also provides income assistance to the aged who have limited income and assets. According to the Monthly Statistical Snapshot of February 2023, approximately 8.7 million Americans received SSDI, while 7.5 million Americans received SSI benefits during that period.[1]

The impact of SSDI and SSI on beneficiaries' lives, including their labor supply, earnings, income, and health, has been extensively studied (Autor, Duggan, Greenberg, and Lyle 2016; Deshpande 2016; Favreault, Johnson, and Smith 2013). However, there have always been concerns about potential disparities in the application and award processes of SS(D)I, particularly along racial and gender lines. Addressing these potential disparities is crucial to ensure fairness, impartiality, and public trust in these programs. As Godtland et al. (2007) noted, "For the public to perceive that SSA's disability programs are run with the highest degree of integrity, it is of the utmost importance that the agency's decisions to award cash benefits to people with disabilities are accurate and made in a fair and impartial manner, without regard to race, sex, or other factors not related to a person's impairment." Despite ongoing research efforts on the application, appeal, and award processes (e.g., Benítez-Silva et al. 1999; GAO 2003; Kreider and Riphahn 2000), concerns about potential disparities along racial and gender lines persist.

To examine these disparities, it is essential to consider that individuals decide whether to apply for benefits and the Social Security Administration makes decisions regarding benefit awards. Therefore, a thorough analysis should encompass both application rates and award rates to gain a comprehensive understanding of the disparities present. In this study, we contribute to the growing literature on racial and gender disparities in SS(D)I applications and awards by conducting an in-depth empirical analysis using different estimation methods and rich datasets including administrative records from the Social Security Administration (SSA) and self-reported data from the Health and Retirement Study (HRS). Our goal is to shed new light on the detection of these disparities by reducing subjective assumptions and adopting a more data-driven approach.

[1] https://www.ssa.gov/policy/docs/quickfacts/stat_snapshot/2023-02.html (accessed on June 20, 2023).

Understanding and addressing potential racial and gender disparities in the applications and awards of SS(D)I is important. First, addressing racial and gender disparities in the distribution of social welfare benefits is essential for promoting equity and social justice, and achieving social equality is a fundamental goal in society. Understanding the presence and scope of these disparities in the context of SS(D)I applications and awards is crucial to ensure fair access to support for individuals with disabilities, regardless of their race or gender. By investigating and understanding these disparities, we can work towards creating a more inclusive and equitable society where access to disability benefits is not influenced by race or gender.

Second, racial and gender disparities in SS(D)I applications and awards can have far-reaching economic consequences. Disability benefits play a significant role in individuals' economic well-being and financial stability. If certain racial or gender groups face systemic barriers in accessing disability benefits, it can exacerbate existing economic inequalities, perpetuate poverty cycles, and hinder social mobility. Understanding these disparities can help us identify barriers to access, develop strategies to promote equal economic opportunities for all individuals, and inform targeted interventions to mitigate any potential negative economic impacts.

Third, access to disability benefits can have a profound impact on the health and quality of life of individuals with disabilities. If racial or gender disparities exist in accessing SS(D)I benefits, it can have detrimental effects on the physical and mental health of individuals who are disproportionately affected. By uncovering these disparities, we can work towards reducing health disparities and ensuring that all individuals receive the care and assistance they need.

Finally, exploring racial and gender disparities in SS(D)I applications and awards sheds light on the presence of social stigmas and biases that may affect certain groups' willingness to seek support. Understanding these dynamics helps to challenge and dismantle these stigmas, creating a more inclusive and supportive environment for individuals with disabilities.

Previous studies have identified racial and gender disparities in SS(D)I applications and awards, highlighting the need to understand the factors contributing to these disparities (e.g., Godtland et al. 2007). Despite the valuable insights provided by the existing literature on racial and gender disparities within the disability determination process, there are certain limitations that need to be acknowledged. These limitations arise from the methodologies employed and the assumptions made in previous studies.

One common approach to evaluating disparities involves examining the differences between two groups, such as White versus non-White, in terms of an "average treatment effect" (ATE). In this context, the "treatment" could refer to factors like race or gender. Essentially, this approach aims to answer the question of how, overall, an average person would be treated or experience certain outcomes if that person *were* of a different race or gender, while keeping all other factors constant. To ensure that all other variables are held constant, regression models are widely used due to their convenience in explicitly controlling for confounding factors in estimating the ATE. Those regression analyses are often done in two ways.

One way is that researchers often estimate the coefficient on a binary variable indicating a treatment, while simultaneously controlling for many observable factors (e.g., age, education, income, and health) additively in a regression model. Doing so essentially assumes that the treatment effect or disparity is constant (homogeneous) across all individuals, but this assumption may not be true. For instance, if the disparity is more pronounced for individuals with less education or lower for those with more education, it indicates an interaction effect between the treatment and education, meaning that the disparity is not constant, but rather varies (heterogeneous) across different education levels. To account for this heterogeneity, researchers often conduct subsample analyses along various dimensions (e.g., education and income) and report many subsample treatment effect estimates. While these analyses provide valuable insights into the treatment effect evaluation, they do not directly offer an overall assessment of the treatment effect for the entire population. This overall assessment, represented as a single numeric value representing the ATE for the whole population, can be interpreted as a weighted average of those subsample treatment effect estimates.

The other way researchers use a regression model to estimate the ATE is by controlling for as many "treatment × observables" interactions as possible in the regression model. They estimate the coefficients on the treatment variable and these many "treatment × observables" interaction terms, and subsequently compute the "marginal effect" of the treatment. Doing so allows the treatment effect to vary by observable factors, thus accommodating the heterogeneity in the treatment effect due to these observables. Additionally, this method implicitly calculates a weighted average of the subsample treatment effect estimates discussed earlier. However, there are two challenges of using interaction terms. First, specifying interaction terms can be a subjective decision, especially when there is little guidance in theory. Second, the estimation of ATE relies on the assumption of overlap, which assumes that the distribution of characteristics is similar for both treated and untreated groups. Estimating this heterogeneous ATE becomes infeasible when the overlap assumption is violated due to the use of many observable factors as control variables. This violation is often explicitly addressed in the implementation of propensity-score-based estimators (by estimating ATE for a subpopulation where the overlap assumption holds). But when using arguably the most popular estimator for linear regression models—ordinary least squares (OLS)—we may not receive an error message indicating that the overlap assumption is violated.

The limitations discussed above highlight the complexities involved in detecting the presence of racial and gender disparities in SS(D)I applications and awards. The subjectivity in specifying interaction terms and the potential violation of the overlap assumption when controlling for numerous variables underscore the need for caution when interpreting the results of previous studies. In our study, we use a new method, proposed by Chernozhukov et al. (2018) and named "double/debiased machine learning (ML)," to estimate the ATE, which we interpret as the disparity between racial or gender groups. This approach allows for individual-level variation in treatment effects and captures the interactions between treatment indicators and observable characteristics in a flexible, data-driven manner. Our study considers a large number of observable characteristics—830—coming from individual-level characteristics (linear term and quadratic term), state dummy variables, and state dummy variables interacted with individual-level characteristics. To satisfy the overlap assumption, the new method utilizes ML techniques to conduct "dimension reduction," that is, selecting only relevant variables for predicting the dependent variable based on the data used for analysis.

To the best of our knowledge, we are the first to apply this estimator to the analysis of disparities. Notably, this estimator leverages ML techniques in a way that allows for statistical inference (i.e., conducting hypothesis tests). This contrasts with many commonly seen and used ML techniques, which are primarily designed for generating predicted values for a dependent variable and may not be well-suited for conducting hypothesis tests.

Our initial findings highlight some intriguing patterns. We observe different proportions of SSDI and SSI applications and receipts[2] in different samples, indicating potential differences

[2] Throughout our paper, we use the words "awards" and "receipts" interchangeably. One reason for using the word "receipt" is that the study done by Hyde and Harrati (2021), which we follow closely in the construction of variables related to SSDI and SSI, uses the word "receipt."

between self-reported data from the HRS and administrative records from the SSA. Moreover, proportions based on different definitions of applications and receipts reveal nuanced differences, and proportions calculated using SSA administrative records are consistently higher than those based on HRS self-reports, aligning with prior research.

When using the OLS regression models to explore racial and gender disparities in SS(D)I applications and receipts and controlling for a comprehensive set of individual-level demographic variables, we find that racial and gender disparities appear to diminish as additional control variables are introduced. However, it is crucial to note that our OLS estimates have limitations as discussed above. Thus, we further extend our analysis by employing the double/debiased ML estimator. The results provide valuable insights into racial and gender disparities in SS(D)I applications and receipts, when interactions between race or gender and observed variables are accounted for in a flexible, data-driven manner. Here, we find that including interaction terms substantially affects the disparity estimates, particularly for the racial disparity estimates. Furthermore, our analysis reveals a statistically significant gender disparity in SSDI applications, suggesting a decrease among females compared to males, but no statistically significant disparity was found in SSI applications or awards (or SSDI awards in the case of using the HRS data). To ensure the robustness of our findings, we conduct sensitivity analyses, exploring alternative specifications for the least absolute shrinkage and selection operator (LASSO) technique—a critical component of the double/debiased ML estimator. Our results demonstrate the robustness of the estimates, particularly in the case of gender disparities.

We want to clarify certain aspects of our study. First, our focus lies in comparing estimates obtained using the OLS method, a traditional approach widely used in the field, with the double/debiased ML estimator. For this comparison, we employ the same sample for both estimations, without conducting sample-selection-bias corrections. While the sample used may contain sample-selection bias, our main objective is not to obtain an unbiased or consistent ATE estimate. Instead, we aim to examine whether the disparity suggested by the OLS estimates persists when alternative methods, such as the double/debiased ML estimator, are employed.

Second, our study primarily aims to detect the presence of any disparity using the double/debiased ML estimator. While our method accounts for possible interactions between race/gender and observable characteristics in a flexible, data-driven manner, it does not identify the underlying factors causing the disparity. Therefore, the conclusions of our study are limited to whether any disparity is detected or not.

Lastly, to tighten the focus of our study, we concentrate on comparing White vs. Black and Male vs. Female. This choice is driven by the data available in the HRS, which is our main data source. The HRS data have three racial categories: "White/Caucasian," "Black/African American," and "Other," with the White vs. Black consisting of 94 percent of the data. Ethnicity information in the HRS data pertains to "Hispanic or not," with a Hispanic proportion of 10 percent. The gender variable in the HRS data consists of two values: "Male" and "Female," based on self-reported information during HRS interviews. Therefore, we interpret this gender variable as representing gender rather than sex assigned at birth. While our study focuses on examining White vs. Black and Male vs. Female disparities, the methodology we employ can be readily extended to analyze disparities along other dimensions such as ethnicity and sex.

This study contributes to the ongoing discourse on racial and gender disparities in SS(D)I applications and awards by utilizing rich datasets and applying advanced analytical techniques, and it contributes to methodological advancements in the study of disparities in social welfare programs in general. It adds to the literature by employing an advanced econometric technique, the double/debiased ML, to analyze racial and gender disparities in SS(D)I applications and awards. By utilizing advanced methodologies, this study enhances the methodological toolkit for future researchers studying disparities in various social welfare programs and other socioeconomic contexts, contributing to the broader field of economics.

This research also has important policy implications. First, our findings on racial and gender disparities in SS(D)I applications and awards provide insights into the effectiveness of the SS(D)I program in reaching individuals with disabilities across different racial and gender groups. These findings can be useful for evaluating the programs' performance, identifying potential gaps or biases in the program, and implementing targeted policy interventions to address these disparities. Policies could focus on providing additional resources and support to disadvantaged groups, improving outreach efforts to ensure equal access to information about the disability benefits program, and reducing barriers that disproportionately affect those groups. Specifically, in addressing the disparities, it becomes imperative to devise targeted interventions that can foster a more equitable SSDI and SSI application and award landscape. One salient avenue could be the provision of targeted financial support, especially for those unable to secure attorney representation. Furthermore, there is a palpable need to bolster community outreach programs, emphasizing the intricacies of application procedures, common pitfalls leading to denials, and the paramount importance of legal counsel during the application process. In tandem, training modules for decision-makers can be instrumental in negating potential biases, thereby ensuring a fairer evaluation process.

Second, our research findings can be useful for implementing measures aimed at mitigating bias and discrimination in the SS(D)I application and award process. This may involve reviewing and revising the evaluation criteria, ensuring that decision-making processes are fair and unbiased, and providing training and guidance to program administrators to address implicit biases that may impact application outcomes.

Third, our findings emphasize the importance of comprehensive data collection and monitoring systems to track and address potential disparities in SS(D)I applications and awards. This may involve incorporating data on race and gender in program evaluation processes and regularly assessing the programs' performances in reaching diverse populations. Robust data collection enables researchers and policymakers to identify emerging trends, evaluate the impact of policy changes, and ensure accountability in addressing disparities.

In the subsequent sections of this paper, we review the literature in Section 2 and present our data and methods in Section 3 and Section 4. Section 5 discusses our results, followed by concluding remarks in Section 6.

2. Literature Review

The literature on disability benefit programs, specifically SSDI and SSI, has identified racial and gender disparities in the application and award rates. Several studies have examined the differential treatment and outcomes experienced by different racial and gender groups within the disability determination process.

A notable study by the US Government Accountability Office (GAO) in 1992 reported consistently lower SSDI award rates for African Americans compared to whites between 1961 and 1985, as well as lower SSI award rates for African Americans between 1971 and 1989. This finding highlighted the existence of racial disparities in benefit receipts.

Hu, Lahiri, Vaughan, and Wixon (2001) focused on analyzing SSA administrative data matched with the 1990 Survey of Income and Program Participation. Using multistage logit models, they investigated the determinants of disability decisions and found significant racial differences in denial rates at Step 2 of the determination process. Younger African American claimants, particularly those under the age of thirty-five, were more likely to be denied benefits at this step compared to their white counterparts.

Benítez-Silva et al. (1999) also analyzed HRS data and investigated the likelihood of favorable disability decisions at the initial and appellate levels. Their results revealed no racial disparities, as white claimants were no more likely than minority claimants to receive favorable disability decisions.

Godtland et al. (2007) employed multivariate econometric models and Oaxaca decomposition methods to examine racial/ethnic differences in benefit award rates at the appellate level. They found that when claimants were represented by attorneys, there were no statistically significant differences in benefit award rates between whites and African Americans. However, for claimants without attorney representation, significant disparities existed, with lower benefit award rates for African Americans compared to whites.

Baldwin (1997) focused on the initial level decisions and used a multistage logistic model to estimate factors affecting disability determinations. Although race was not included in her model, she explored gender differences and found that women over the age of fifty-five were more likely to be rejected based on vocational criteria, even after controlling for applicant characteristics and nature of impairment.

Kreider and Riphahn (2000) utilized data from the HRS to examine disability benefit awards at different stages of the process. They aimed to explore whether men and women exhibited different responses to disability policy changes. Their findings indicated that at the initial level, white females and males were no more likely to be awarded benefits than nonwhites. However, at the appellate level, white males were significantly less likely to be awarded benefits compared to nonwhites.

These studies collectively highlight the existence of racial and gender disparities within the disability determination process, with African Americans and women experiencing lower award rates and higher denial rates, particularly at certain stages. However, the literature review also suggests variations in findings depending on factors such as age, attorney representation, and the specific stage of the determination process.

The existing research has certain limitations that need to be acknowledged. One limitation is the assumption of constant disparities. For example, some studies have estimated the coefficient on a binary indicator (d) representing racial groups (e.g., d = 1 for Black and d = 0 for White), controlling for various characteristics (x) such as age, education, income, and health. However, this approach assumes that the disparity is constant for all individuals, which may not be true. Some studies have controlled for interaction terms between d and x, allowing the treatment effect to vary by individual characteristics. This approach aims to estimate the ATE, considering the heterogeneity of treatment effects due to observable factors. However, specifying these interaction terms can be subjective when theoretical guidance is limited. Moreover, estimating the average treatment effect becomes problematic when a large number of control variables are included, potentially leading to biased estimates due to violations of the overlap assumption. These limitations highlight the need for caution in interpreting previous findings.

To address these challenges, as discussed in the Introduction section, we are the first to use an alternative method that uses ML techniques, specifically the double/debiased ML estimator, to investigate the existence of racial and gender disparities in SS(D)I applications and receipts. Our focus lies in comparing estimates obtained by the widely used OLS method with those from the double/debiased ML estimator. Essentially, we examine whether our conclusion about the presence of an overall disparity—that is, the disparity averaged across different characteristics of the population—will change if we examine the disparity characteristic by characteristic, that is on a “case-by-case” basis (which is equivalent to allowing the treatment to interact with observable characteristics) using a data-driven approach. The OLS, although widely popular and commonly used, has limitations in conducting such “case-by-case” analyses (which we discussed in the Introduction section). However, such analyses are critical in forming an overarching conclusion about racial or gender disparities. Given the significance of disparity-related studies, the need to enrich the toolset becomes relevant, especially for economic studies that often aim to disentangle the effects of single factors. But, as noted by Athey and Imbens (2019, p. 686), “the acceptance of ML methods [has] been so much slower in economics compared to the broader statistics community.” Thus, our use of the double/debiased ML estimator represents an effort to bridge the gap in the literature by incorporating alternative tools to assess racial and gender disparities. Furthermore, Athey and Imbens (2019, p. 695) point out that “modern methods are particularly good at detecting severe nonlinearities and high-order interactions.” While ML techniques offer advantages, we must acknowledge that they, by themselves, do not uncover the mechanisms that generate these disparities. Uncovering such mechanisms often requires theories and structural models capable of explaining complex relationships among variables.

3. Data and Variables

Our study utilizes three data sources: the Health and Retirement Study (HRS), the US Social Security Administration (SSA), and the University of Kentucky Center for Poverty Research (UKCPR).

3.1. HRS

The HRS focuses on a nationally representative sample of adults over the age of fifty. It is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and has been conducted by the University of Michigan every other year since 1992. For our study, we utilize the RAND HRS Longitudinal File (henceforth, the RAND HRS), a cross-wave consistent version of the HRS developed by the RAND Center for the Study of Aging. The RAND HRS offers several advantages as it standardizes and processes HRS variables consistently across survey years. This standardized approach greatly facilitated our analysis at an aggregate level, allowing us to calculate averages and examine variables across multiple survey years.

Specifically, in our study we use the "RAND HRS Longitudinal File 2018 (V2)" version, which was the most recent version available at the time of our research. This version encompasses HRS data from 1992 to 2018.[3] Additionally, we obtain the HRS restricted data known as the "Cross-Wave Geographic Information (State) [1992–2018]" file, which contains information regarding the state of residence for each HRS respondent.[4]

3.2. SSA

In our study we use restricted data files that establish links between individual level HRS data on SSDI and SSI with the SSA administrative records. These files include the Form 831 Respondent Disability Records (referred to as Form 831), the Disability Analysis File (referred to as DAF), and Permissions Consent History (referred to as Consent).[5]

[3] For details, see https://hrsdata.isr.umich.edu/data-products/rand-hrs-archived-data-products (accessed in March 2023).
[4] For details, see https://hrs.isr.umich.edu/data-products/restricted-data/available-products/9705 (accessed in March 2023).
[5] For details, see https://hrs.isr.umich.edu/data-products/restricted-data/available-products/9695, https://hrs.isr.umich.edu/data-products/restricted-data/available-products/11489, and https://hrs.isr.umich.edu/data-products/restricted-data/available-products/11516 (accessed in March 2023).

The Form 831 file provides information about SS(D)I applications, specifically the information about the initial application process. The DAF file contains information about SS(D)I receipts. The Consent file provides information about whether HRS respondents consented to having their HRS data linked to the SSA's administrative records. It is important to note that for those who did not provide consent, the SSA will not have any information about their SS(D)I applications or receipts. However, even for those who did consent, the information about their SS(D)I applications and receipts provided by the Form 831 and the DAF is still limited. This means that not all HRS respondents who have applied for or received SS(D)I can be identified in the HRS-SSA linked data, even if they granted permission for the data linkage. Below we list two main reasons, for which Hyde and Harrati (2021) give more detailed explanations.

Firstly, the Form 831 contains information only about the initial applications for SS(D)I that have been deemed eligible by the SSA. This means that it encompasses individuals who have medical conditions that meet the SSA's definition of disability or fulfill the financial criteria of federal disability programs. However, it is important to note that Form 831 does not contain information regarding initial applications that have been denied by the SSA or applications that are currently under appeal. Consequently, Form 831 only represents a subset of the total initial applications.

Secondly, even among HRS respondents who have given consent for their data to be linked to the SSA records, there is a possibility that their SSA records may not be located unless they have provided the necessary information required for the linkage process. This includes, for example, accurate social security numbers, names, and dates of birth.

3.3. UKCPR

We obtain state-level variables from the UKCPR's National Welfare Data,[6] which include population, unemployment rate, poverty rate, minimum wage, and political factors. The UKCPR data are annual and provide longitudinal information for each state.

3.4. Construction of Variables

The HRS data years used in our study are 2006–2018. Henceforth, we refer to this period as the study period.[7] While the HRS provides variables related to self-reported SS(D)I applications and receipts from 1992 onwards, we have chosen 2006 as the starting year of our study period. This is

[6] For details, see https://ukcpr.org/resources/national-welfare-data (accessed in March 2023).
[7] Year 2006 is HRS's wave 8. Year 2018 is HRS's wave 14.

because for those HRS respondents who consented to having their information linked to the earning and benefit records maintained by the SSA, we need to use the SSA's Form 831 and DAF files to make that linkage; according to the study by Hyde and Harrati (2021), the Form 831 and the DAF files are only available for HRS respondents who consented to the linkage in 2006 or later.[8]

The outcome variables of our study are whether an HRS respondent ever applied for and/or received SS(D)I between 2006 and 2018. We construct the measures of SS(D)I applications and receipts by closely following Hyde and Harrati (2021), who give a thorough guide (in Appendix A of their paper) regarding which variables from the RAND HRS, Form 831, DAF, and Consent files to use, as well as how to generate SS(D)I application and receipt variables using these variables.

Hyde and Harrati (2021) categorize these variables into two groups, and we follow their classification. One group is based on self-reported data from the HRS, and the other group relies on administrative data obtained from the SSA. For the former, we follow the detailed instructions provided by Hyde and Harrati (2021) to create two subgroups based on separate definitions. The first subgroup is created using definition 1, which is affirmative self-reporting of confirming applications for or receipts of either SSDI or SSI, or both. The second subgroup is created using definition 2, which includes more self-reports that are affirmative than definition 1. Specifically, the additional self-reports that are affirmative and included by definition 2 (but skipped by definition 1) are from the HRS respondents who are unsure or do not recall exactly which SSA program (SSDI or SSI) they applied to or received benefits from.

The focus of our study is investigating disparities related to race and gender, rather than the dynamics of SS(D)I applications and receipts. Because of this focus, we aggregate the data from multiple waves of the HRS survey, resulting in a cross-sectional dataset. We now give further explanations regarding this data aggregation below.

For the outcome variables of SS(D)I applications and receipts, we follow Hyde and Harrati (2021) to first create a binary indicator for each of those outcomes and for each wave of the HRS data within our study period, spanning from 2006 to 2018. This indicator takes the value of one if the corresponding outcome occurs (such as applying for SSDI) and zero otherwise. Subsequently,

[8] From Hyde and Harrati (2021, p. 42): "Based on correspondence with HRS staff, we learned that the DAF and 831 file were only available for HRS respondents who consented to the linkage in 2006 and later."

we calculate the average value of this indicator over our study period to establish an "ever applied for or received the benefit" variable (e.g., ever applied for SSDI during 2006–2018). This variable is a binary indicator that equals one if the average value is positive, and zero otherwise.

To construct the consent variable based on the Consent file, we follow the approach adopted by Hyde and Harrati (2021). Specifically, we create a binary variable for each wave of the HRS data, indicating whether consent was given or not. A value of one represents consent given, while zero indicates no consent.[9] Subsequently, we compute the average value of this binary variable over our study period, resulting in an "ever consent" variable. This variable is a binary indicator that equals one if the average value is positive, indicating consent given at some point during the period of 2006–2018. Conversely, it equals zero if the average value is zero, implying no consent given. For our analyses that utilize SSA-based measures, we conduct them conditionally on a subsample of HRS respondents who provided consent during our study period. This subsample consists of individuals for whom the "ever consent" variable equals one, indicating consent given.

All HRS respondents included in our study period are aged fifty or above. When considering a demographic variable in the HRS, we handle two cases: 1) if that variable is continuous, such as age, we calculate the average value over our study period; 2) if that variable is binary, with a value of one indicating a "yes" answer and zero representing a "no" answer, we generate an "ever" variable. This "ever" variable takes on a value of one if the binary variable equals one in any year within the study period. Conversely, it is assigned a value of zero if the binary variable remains zero throughout the entire study period.

For our study period (2006–2018), the HRS provides a sampling weight for each wave and for each respondent, whether they reside in a personal home or a nursing home. These weights are referred to as the combined respondent weight and nursing home resident weight. Because our study focuses on a cross-sectional dataset aggregated over 2006–2018, we use the average value of the combined respondent weight and nursing home resident weight as the sampling weight for each HRS respondent in our statistical analysis.[10]

---

[9] Hyde and Harrati (2021) also point out that although there is a variable in the Consent file, indicating whether there is a match between the SSA records and the information of an HRS respondent who gives the consent, that variable is not consistently measured. As a result, we follow Hyde and Harrati (2021) and do not use that variable for selecting a sample which we want to have matched records.

[10] We use the average value of the combined respondent weight and nursing home resident weight as the sampling weight for each HRS respondent in our statistical analysis that either uses the HRS-based measures or the measures

When considering state-level variables obtained from the UKCPR, we handle two cases: 1) if the state-level variable is continuous, such as state's minimum wage, we calculate the average value over our study period; 2) if the state-level variable is binary, such as indicating whether the state's governor is a Democrat, we also compute the average value over our study period. We interpret this average value as the percentage of time the state's governor is a Democrat. As mentioned earlier, our study focuses on a cross-sectional dataset. Consequently, when we use state fixed effects to account for state-level time-invariant confounders, the aforementioned state-level variables become unnecessary for our analysis.

4. Methods

The double/debiased ML estimator used by our study is applied to the following model that includes an outcome-equation and a "treatment"-equation.

The outcome-equation is described as follows:

$$y = g(d, \boldsymbol{x}) + u, \qquad where\ E(u|d, \boldsymbol{x}) = 0$$

Here, y represents the outcome variable, such as the SS(D)I application or receipt; d represents the binary "treatment," equal to one for Black (or Female), and equal to zero for White (or Male); $\boldsymbol{x}$ is a vector of covariates; and u is the error term.

The "treatment"-equation is described as follows:

$$d = m(\boldsymbol{z}) + v, \qquad where\ E(v|\boldsymbol{z}) = 0$$

Here, $m(\boldsymbol{z})$ represents the likelihood of being in the "treatment" group (i.e., $d = 1$), conditional on a vector of covariates denoted by $\boldsymbol{z}$; and v is the error term.

The parameter of interest is defined as follows:

$$\theta = E[g(1, \boldsymbol{x}) - g(0, \boldsymbol{x})] = \text{"Average Treatment Effect (ATE)"}$$

Here, θ represents the average of individual-level heterogeneous treatment effects, where the treatment effect can vary from person to person. Additionally, θ allows for flexible interactions

based on the HRS-SSA linked data. Although the HRS provides a file named "Cross-Wave Social Security Weights" (source: https://hrs.isr.umich.edu/data-products/restricted-data/available-products/9696, accessed in March 2023), that file is not suitable for our study. Here is the reason. The "Summary of SS [social security] weight usage" table reported in that file's "Data Description (2004–2016)" document (source: https://hrs.isr.umich.edu/sites/default/files/restricted_data_docs/SSWgts_JP_DD%20v2.pdf, accessed in March 2023) states that the cross-wave social security weights should not be used for the Form 831 data, but our study uses the Form 831 data. The reason is given in footnote #3 of that "Data Description (2004–2016)" document: "The construction of SS weights does not adjust for the likelihood of Disability Insurance (DI) application."

between d and $\boldsymbol{x}$. This means that there is no specific functional form, such as "$d \cdot \boldsymbol{x}$", required to describe how d interacts with $\boldsymbol{x}$.

While the above definition of ATE captures the average of these varying treatment effects, there are two main challenges involved in estimating such an ATE: model specification error and model selection error. The first challenge, model specification error, arises from the need to correctly specify the functional forms of $g(\cdot)$ and $m(\cdot)$. The second challenge, model selection error, pertains to selecting the relevant covariates to include in the functions $g(\cdot)$ and $m(\cdot)$.

To address the model specification error, the double/debiased ML estimator employs the following moment condition (used for estimating θ) that is utilized by the commonly known "doubly-robust estimator":

$$\varphi(\boldsymbol{w};\theta,\eta) = [\hat{g}(1,\boldsymbol{x}) - \hat{g}(0,\boldsymbol{x})] + \frac{d(y-\hat{g}(1,\boldsymbol{x}))}{\widehat{m}(\boldsymbol{z})} - \frac{(1-d)\big(y-\hat{g}(0,\boldsymbol{x})\big)}{1-\widehat{m}(\boldsymbol{z})} - \theta$$

Here, the vector $w = (y,\boldsymbol{x}',\boldsymbol{z}',d)'$ represents the observed variables, and η represents the nuisance parameter comprising $\hat{g}(1,\boldsymbol{x})$, $\hat{g}(0,\boldsymbol{x})$ and $\widehat{m}(\boldsymbol{z})$, which are approximations of $g(1,\boldsymbol{x})$, $g(0,\boldsymbol{x})$ and $m(\boldsymbol{z})$, respectively, obtained using ML techniques. The doubly-robust estimator of θ is defined as the solution to $E\big(\varphi(\boldsymbol{w};\theta,\eta)\big) = 0$. It incorporates correction terms for the treatment group and the control group as $(y-\hat{g}(1,\boldsymbol{x}))/\widehat{m}(\boldsymbol{z})$ and $\big(y-\hat{g}(0,\boldsymbol{x})\big)/(1-\widehat{m}(\boldsymbol{z}))$, respectively. Unlike traditional methods for estimating the ATE, such as regression adjustment (which requires $g(\cdot)$ to be correctly specified) or inverse probability weighting (IPW, which requires $m(\cdot)$ to be correctly specified), the doubly-robust estimator only requires one of the two functional forms to be correctly specified, which helps mitigate, although does not necessarily mitigate, the model specification error.

To address the model selection error, the double/debiased ML estimator leverages the Neyman orthogonality property, a crucial characteristic of the doubly-robust estimator. This property ensures that the estimation of the ATE remains robust to mistakes in model selection made by the ML technique implemented for selecting $\boldsymbol{x}$ and $\boldsymbol{z}$ to predict $g(1,\boldsymbol{x})$, $g(0,\boldsymbol{x})$ and $m(\boldsymbol{z})$. This robustness becomes particularly valuable when estimating a heterogeneous treatment effect, as conventional methods such as regression adjustment, propensity-score matching, and IPW may not be feasible due to high-dimensional $\boldsymbol{x}$ and $\boldsymbol{z}$. The double/debiased ML estimator uses a data-driven process to select a subset of covariates (i.e., dimension reduction) that effectively predict the outcome variable (y) and the treatment variable (d). Despite the possibility of selecting incorrect variables (i.e., model selection error), the Neyman orthogonality property of the doubly-robust estimator allows us to disregard this error in the estimation and hypothesis testing of the ATE.

In summary, the double/debiased ML estimator can be seen as an ML-version of the doubly-robust estimator, with several points that are worth emphasizing here. First, it is important to note that the variables chosen by the ML technique may not necessarily be explanatory variables with causal relationships. In our study, we include the same variables in both the $\boldsymbol{x}$ and $\boldsymbol{z}$ vectors, referring to them as predictor variables. These predictor variables aid in predicting the dependent variable, but they may not have a direct causal effect on it, thus distinguishing them from explanatory variables.

Second, the double/debiased ML estimator incorporates three techniques: LASSO, cross-fitting, and resampling. LASSO aids in dimension reduction, while cross-fitting allows for the selection of a greater number of predictor variables relative to the sample size. In our study, we employ a ten-fold cross-fitting, as using ten folds is a widely adopted practice. However, one drawback of cross-fitting is that it introduces randomness into the estimation of the ATE due to the random division of the original sample into multiple folds. One proposed solution in the ML field is combining cross-fitting with resampling. This involves repeating the cross-fitting procedure multiple times using resamples of the original sample and subsequently averaging the resulting estimates. In our study, we conduct the ten-fold cross-fitting three times by employing three resamples.[11]

Third, an important aspect of implementing the double/debiased ML estimator is determining the appropriate value of LASSO's *tuning parameter*, also known as the *penalty parameter*. This parameter is nonnegative and inversely related to the number of selected

[11] The estimator proposed by Chernozhukov et al. (2018) is called the "double/debiased machine learning" estimator, meaning the use of a machine learning technique in conjunction with cross-fitting and resampling. Here, we give a brief illustration of what cross-fitting does. In a simple case, we randomly partition the original sample into two parts called folds: Fold #1 (the training sample) and Fold #2 (the evaluation sample). In step 1, we apply the LASSO technique using data from Fold #1 to obtain the fitted models of $\hat{g}(\cdot)$ and $\hat{m}(\cdot)$. In step 2, we use those fitted models to compute post-LASSO residuals (e.g., $y - \hat{g}(\cdot)$) for the observations in Fold #2. In step 3, we swap the roles of Fold #1 and Fold #2 (i.e., using Fold #1 as the evaluation sample and using Fold #2 as the training sample), and we repeat step 1 (i.e., modeling fitting) and step 2 (i.e., model evaluation); at the end of step 3, post-LASSO residuals are computed for the full sample. Finally, in step 4, estimation of the ATE is conducted using the moment condition ($\varphi(\boldsymbol{w}; \theta, \eta)$) as well as the post-LASSO residuals obtained for the full sample at the end of step 3. The simple case just described is called a two-fold cross-fitting, which has two rounds of steps 1 through 3. A ten-fold crossing-fitting means doing ten rounds of steps 1 through 3.

predictors. In our study, we employ a plugin approach to determine the optimal value for the penalty parameter. This approach utilizes a formula grounded in theory that ensures the optimal convergence rate for both prediction and statistical inference, such as the estimation and hypothesis testing of the ATE, as detailed by Ahrens, Hansen, and Schaffer (2020). The specific numeric value we assign to the penalty parameter in our implementation of the double/debiased estimator is considered optimal because it is specifically developed for LASSO employed in inference tasks, such as the estimation and hypothesis testing of the ATE. In contrast, other methods used to determine the penalty parameter's numeric value, such as the Bayesian information criterion (BIC), cross-validation (CV), and adaptive CV, are designed only for LASSO employed in prediction tasks, such as generating predicted values for a dependent variable, rather than inference tasks involving hypothesis tests of parameters. Nevertheless, to assess the sensitivity of our ATE estimates to different values of the penalty parameter, our study includes a robustness check using penalty parameters determined by the BIC, CV, and adaptive CV approaches.

Fourth, for the double/debiased ML estimation, we use a linear model for the outcome variable and a logit model for the binary treatment variable. We allow LASSO to choose from a comprehensive set of potential predictor variables—830 variables. These variables encompass individual-level characteristics (both linear and quadratic terms), state dummy variables, and state dummy variables interacted with individual-level characteristics. Having a large choice set enables LASSO to approximate the $g(\cdot)$ and $m(\cdot)$ functions as closely as possible. In one of our robustness checks, we require certain variables, such as demographic variables, to be always included as predictor variables rather than relying solely on the LASSO technique to determine their selection. This robustness check helps us assess whether our ATE estimates are sensitive to the data-driven approach employed by LASSO.

Fifth, the current ML estimation of the ATE does not accommodate the use of sampling weights. To address this limitation, we adopt the procedure described by Cameron and Trivedi (2022, Chapter 3 Section 8): first, we obtain individual-level predicted ATE values (obtained through the double/debiased ML estimator), denoted by $\hat{\theta}_i$; next, we calculate the weighted average prediction to obtain $\widehat{ATE}$, with standard error clustered by state and with the weight (w_i) being the sampling weight provided by the HRS (which we described in the Data section) and normalized to sum to N (the sample size):

$$\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N} w_i\hat{\theta}_i$$

Sixth, to account for the interactions between d and many potential $\boldsymbol{x}$'s (such as education, income, and others), the ML method we use does not directly use the product of d and $\boldsymbol{x}$'s (i.e., $d \cdot \boldsymbol{x}$'s). Instead, the ML method conducts separate analyses for two samples: one where $d = 1$ and the other where $d = 0$. In each sample, the relationship between the outcome variable (y), such as the SSDI application, and the covariates ($\boldsymbol{x}$) is determined in a data-driven way (e.g., through the LASSO variable selection technique). Essentially, the interactions between d and the various potential $\boldsymbol{x}$'s are not explicitly modeled as using $d \cdot \boldsymbol{x}$'s (following a specific functional form) but are determined by the data.[12] For the purpose of comparison, our study also included OLS regression analyses. These regressions were weighted by the sampling weights that were previously explained.

5. Results

5.1. Summary Statistics

Tables 1 and 2 present the summary statistics for the variables in our dataset without and with sampling weights, respectively. Both tables include three groups of variables: the outcome variables of our study, demographic variables obtained from the HRS data, and state-level variables obtained from the UKCPR data.

Table 1: Summary Statistics

	Ever consented during 2006–2018 (1/0)			
	0	1	n/a	Total
Ever applied for SSDI during 2006–2018 (1/0); v1, HRS	0.038	0.049	0.013	0.044
	(0.190)	(0.216)	(0.113)	(0.204)
Ever applied for SSDI during 2006–2018 (1/0); v2, HRS	0.040	0.051	0.017	0.046
	(0.195)	(0.220)	(0.128)	(0.209)
Ever applied for SSDI during 2006–2018 (1/0); SSA records	0.001	0.093	0.000	0.065
	(0.028)	(0.291)	(0.000)	(0.246)
Ever received SSDI during 2006–2018 (1/0); v1, HRS	0.024	0.032	0.006	0.028
	(0.152)	(0.175)	(0.078)	(0.165)

[12] The double/debiased ML estimator has been made available in the software *Stata* since its version 17, and the command is called *telasso*. We use *Stata* for implementing this double/debiased ML estimator, with the following important reason: "Stata is the only statistical software package, commercial or open source, with integrated version control that allows scripts and programs written years ago to continue to work in modern versions of the software. If you wrote a script to perform an analysis in 1985, that same script will still run and still produce the same results today. Any dataset you created in 1985, you can read today. And the same will be true in 2050. Stata will be able to run anything you do today." (Source: https://www.stata.com/features/overview/integrated-version-control/, accessed in May 2023).

Ever received SSDI during 2006–2018 (1/0); v2, HRS	0.024	0.032	0.006	0.028
	(0.152)	(0.175)	(0.078)	(0.165)
Ever received SSDI during 2006–2018 (1/0); SSA records	0.000	0.078	0.000	0.054
	(0.000)	(0.268)	(0.000)	(0.226)
Ever applied for SSI during 2006–2018 (1/0); v1, HRS	0.015	0.022	0.008	0.019
	(0.120)	(0.146)	(0.091)	(0.137)
Ever applied for SSI during 2006–2018 (1/0); v2, HRS	0.017	0.026	0.012	0.023
	(0.130)	(0.158)	(0.110)	(0.149)
Ever applied for SSI during 2006–2018 (1/0); SSA records	0.000	0.045	0.000	0.031
	(0.016)	(0.208)	(0.000)	(0.174)
Ever received SSI during 2006–2018 (1/0); v1, HRS	0.005	0.010	0.004	0.008
	(0.069)	(0.099)	(0.062)	(0.090)
Ever received SSI during 2006–2018 (1/0); v2, HRS	0.005	0.010	0.004	0.008
	(0.069)	(0.099)	(0.062)	(0.091)
Ever received SSI during 2006–2018 (1/0); SSA records	0.000	0.027	0.000	0.019
	(0.000)	(0.163)	(0.000)	(0.136)
Ever consented during 2006–2018 (1/0)	0.000	1.000	n/a	0.750
	(0.000)	(0.000)	n/a	(0.433)
White (1/0)	0.769	0.812	0.835	0.804
	(0.421)	(0.391)	(0.372)	(0.397)
Black (1/0)	0.231	0.188	0.165	0.196
	(0.421)	(0.391)	(0.372)	(0.397)
Female (1/0)	0.588	0.569	0.524	0.570
	(0.492)	(0.495)	(0.500)	(0.495)
Hispanic (1/0)	0.085	0.085	0.056	0.083
	(0.279)	(0.279)	(0.231)	(0.276)
Age, avg. over 2006–2018	71.175	69.887	72.750	70.404
	(10.193)	(9.286)	(10.287)	(9.620)
Years of education	12.492	12.710	12.143	12.616
	(3.208)	(3.016)	(3.091)	(3.071)
Ever married (including partnered) (1/0) during 2006–2018	0.653	0.708	0.598	0.687
	(0.476)	(0.455)	(0.490)	(0.464)
Respondent's income (in $1,000), avg. over 2006–2018	13.015	12.485	7.622	12.234
	(34.917)	(27.567)	(30.564)	(29.678)
Ever in the labor force (1/0) during 2006–2018	0.447	0.512	0.235	0.476
	(0.497)	(0.500)	(0.424)	(0.499)
Ever had health problems that limit work (1/0) during 2006–2018	0.599	0.633	0.558	0.619
	(0.490)	(0.482)	(0.497)	(0.486)
Ever self-reported health being poor (including fair) (1/0) during 2006–2018	0.540	0.538	0.540	0.538
	(0.498)	(0.499)	(0.499)	(0.499)
Sum of ADLs where respondent reported any difficulty (0-6), avg. over 2006–2018	0.597	0.463	0.774	0.518
	(1.097)	(0.911)	(1.354)	(1.001)
Sum of IADLs where respondent reported any difficulty (0-5), avg. over 2006–2018	0.510	0.346	0.602	0.403
	(0.953)	(0.705)	(1.108)	(0.810)
Sum of conditions ever had (doctor diagnosed, 0–8), avg. over 2006–2018	2.283	2.349	2.696	2.360
	(1.442)	(1.408)	(1.568)	(1.432)
CESD (0–8), avg. over 2006–2018	1.488	1.490	1.900	1.521
	(1.639)	(1.622)	(2.045)	(1.666)
Body Mass Index, avg. over 2006–2018	27.826	28.399	26.954	28.156
	(5.560)	(5.862)	(6.166)	(5.833)
Ever drank any alcohol (1/0) during 2006–2018	0.624	0.671	0.474	0.645
	(0.484)	(0.470)	(0.500)	(0.479)
Ever smoked (1/0) during 2006–2018	0.561	0.572	0.628	0.574
	(0.496)	(0.495)	(0.484)	(0.495)
Ever covered by federal government health insurance program (1/0) dur. 2006–2018	0.844	0.870	0.848	0.862
	(0.363)	(0.336)	(0.360)	(0.344)
Ever covered by Medicare (1/0) during 2006–2018	0.815	0.840	0.820	0.833
	(0.388)	(0.366)	(0.384)	(0.373)
Ever covered by Medicaid (1/0) during 2006–2018	0.180	0.202	0.127	0.191
	(0.384)	(0.402)	(0.333)	(0.393)

Ever covered by other health insurance (1/0) during 2006–2018	0.336	0.353	0.226	0.339
	(0.472)	(0.478)	(0.419)	(0.473)
Ever covered by long-term care insurance (1/0) during 2006–2018	0.197	0.230	0.104	0.213
	(0.398)	(0.421)	(0.306)	(0.409)
Number of private insurance plans, avg. over 2006–2018	0.592	0.572	0.578	0.577
	(0.460)	(0.455)	(0.528)	(0.462)
Number of people living in the household, avg. over 2006–2018	2.074	2.176	2.046	2.142
	(0.959)	(0.992)	(0.959)	(0.983)
Number of living children, avg. over 2006–2018	3.059	3.217	3.149	3.176
	(1.998)	(2.068)	(2.139)	(2.059)
Household (respondent and spouse) income (in $1,000), avg. over 2006–2018	63.302	64.140	71.479	64.509
	(115.134)	(134.449)	(718.313)	(234.791)
Household total wealth (in $1,000), avg. over 2006–2018	532.699	475.075	593.568	497.441
	(1307.354)	(1004.726)	(4409.138)	(1607.622)
State's population (in 1,000,000), avg. over 2006–2018	12.852	13.133	13.343	13.085
	(9.902)	(10.308)	(10.120)	(10.202)
State's unemployment rate (0–100), avg. over 2006–2018	6.422	6.436	6.411	6.431
	(0.897)	(0.915)	(0.884)	(0.908)
State's poverty rate (0–100), avg. over 2006–2018	13.711	13.700	13.727	13.704
	(2.589)	(2.620)	(2.472)	(2.601)
State minimum wage, avg. over 2006–2018	7.336	7.367	7.342	7.358
	(0.793)	(0.810)	(0.793)	(0.805)
Percentage of time State's governor is Democrat (0–100), avg. over 2006–2018	43.151	42.660	41.932	42.717
	(32.690)	(32.726)	(31.987)	(32.661)
Percentage of State House that is Democrat (0–100), avg. over 2006–2018	49.249	49.139	48.040	49.080
	(12.477)	(12.279)	(12.146)	(12.318)
Percentage of State Senate that is Democrat (0–100), avg. over 2006–2018	46.277	46.324	45.330	46.237
	(13.370)	(13.115)	(13.019)	(13.169)
Sample size	3,942	11,853	1,312	17,107

Notes: The mean and standard deviation (in parenthesis) are reported for each variable.

Our analysis is based on two different samples: the full sample, which relies on the HRS's self-reported SS(D)I applications/receipts data and consists of 17,107 observations; and the consent sample, which utilizes the SSA's administrative records and includes only those HRS respondents who had consented to linking their information with the SSA's records during our study period. The consent sample consists of 11,853 observations. Our findings reveal that during our study period, a significant majority of HRS respondents (75.0 percent in Table 1 and 76.1 percent in Table 2) had given their consent, aligning with the results reported by Hyde and Harrati (2021).[13]

Table 2: Summary Statistics (weighted)

	Ever consented during 2006-2018 (1/0)			
	0	1	n/a	Total
Ever applied for SSDI during 2006–2018 (1/0); v1, HRS	0.041	0.056	0.014	0.050

[13] In their study, they also find that the rates of consent differ by survey cohorts and over time, stating (in the abstract of their paper) that: "Older cohorts in the HRS are more likely than younger ones to have consented to having their HRS data linked to SSA administrative records. Younger cohorts, however, are more likely to have consented in recent years." In contrast, our study does not conduct the wave-by-wave analysis. Instead, we use one cross-sectional dataset with aggregate information for the period of 2006–2018. We gave detailed explanations about the construction of this cross-sectional dataset in Section 3.

	(0.198)	(0.230)	(0.117)	(0.217)
Ever applied for SSDI during 2006–2018 (1/0); v2, HRS	0.043	0.057	0.017	0.052
	(0.203)	(0.233)	(0.128)	(0.221)
Ever applied for SSDI during 2006–2018 (1/0); SSA records	0.001	0.109	0.000	0.078
	(0.025)	(0.311)	(0.000)	(0.267)
Ever received SSDI during 2006–2018 (1/0); v1, HRS	0.027	0.036	0.008	0.032
	(0.162)	(0.187)	(0.087)	(0.177)
Ever received SSDI during 2006–2018 (1/0); v2, HRS	0.027	0.036	0.008	0.032
	(0.162)	(0.188)	(0.087)	(0.177)
Ever received SSDI during 2006–2018 (1/0); SSA records	0.000	0.092	0.000	0.066
	(0.000)	(0.290)	(0.000)	(0.248)
Ever applied for SSI during 2006–2018 (1/0); v1, HRS	0.014	0.023	0.008	0.020
	(0.119)	(0.150)	(0.090)	(0.141)
Ever applied for SSI during 2006–2018 (1/0); v2, HRS	0.017	0.027	0.011	0.024
	(0.129)	(0.163)	(0.105)	(0.152)
Ever applied for SSI during 2006–2018 (1/0); SSA records	0.000	0.041	0.000	0.029
	(0.009)	(0.197)	(0.000)	(0.167)
Ever received SSI during 2006–2018 (1/0); v1, HRS	0.004	0.010	0.004	0.008
	(0.065)	(0.098)	(0.064)	(0.090)
Ever received SSI during 2006–2018 (1/0); v2, HRS	0.004	0.010	0.004	0.008
	(0.065)	(0.098)	(0.064)	(0.090)
Ever received SSI during 2006–2018 (1/0); SSA records	0.000	0.023	0.000	0.017
	(0.000)	(0.151)	(0.000)	(0.128)
Ever consented during 2006–2018 (1/0)	0.000	1.000	n/a	0.761
	(0.000)	(0.000)	n/a	(0.426)
White (1/0)	0.859	0.894	0.893	0.886
	(0.348)	(0.308)	(0.309)	(0.318)
Black (1/0)	0.141	0.106	0.107	0.114
	(0.348)	(0.308)	(0.309)	(0.318)
Female (1/0)	0.554	0.537	0.514	0.539
	(0.497)	(0.499)	(0.500)	(0.498)
Hispanic (1/0)	0.062	0.060	0.046	0.060
	(0.242)	(0.238)	(0.209)	(0.237)
Age, avg. over 2006–2018	68.869	67.792	70.525	68.211
	(10.120)	(9.060)	(10.896)	(9.464)
Years of education	12.964	13.152	12.513	13.068
	(3.033)	(2.871)	(2.940)	(2.917)
Ever married (including partnered) (1/0) during 2006–2018	0.678	0.723	0.612	0.706
	(0.467)	(0.448)	(0.488)	(0.456)
Respondent’s income (in $1,000), avg. over 2006–2018	18.589	18.099	12.249	17.826
	(44.863)	(35.621)	(46.343)	(38.654)
Ever in the labor force (1/0) during 2006–2018	0.526	0.594	0.307	0.560
	(0.499)	(0.491)	(0.461)	(0.496)
Ever had health problems that limit work (1/0) during 2006–2018	0.552	0.591	0.519	0.578
	(0.497)	(0.492)	(0.500)	(0.494)
Ever self-reported health being poor (including fair) (1/0) during 2006–2018	0.480	0.483	0.505	0.484
	(0.500)	(0.500)	(0.500)	(0.500)
Sum of ADLs where respondent reported any difficulty (0–6), avg. over 2006–2018	0.509	0.394	0.702	0.440
	(1.054)	(0.850)	(1.294)	(0.937)
Sum of IADLs where respondent reported any difficulty (0–5), avg. over 2006–2018	0.430	0.286	0.539	0.334
	(0.899)	(0.643)	(1.074)	(0.746)
Sum of conditions ever had (doctor diagnosed, 0–8), avg. over 2006–2018	2.096	2.197	2.533	2.197
	(1.439)	(1.409)	(1.597)	(1.432)
CESD (0–8), avg. over 2006–2018	1.399	1.396	1.909	1.430
	(1.635)	(1.585)	(2.081)	(1.638)
Body Mass Index, avg. over 2006–2018	27.715	28.493	27.140	28.231
	(5.457)	(5.880)	(6.214)	(5.827)
Ever drank any alcohol (1/0) during 2006–2018	0.669	0.706	0.513	0.685
	(0.471)	(0.456)	(0.500)	(0.465)

Ever smoked (1/0) during 2006–2018	0.564	0.567	0.607	0.569
	(0.496)	(0.496)	(0.489)	(0.495)
Ever covered by federal government health insurance program (1/0) dur. 2006–2018	0.790	0.825	0.755	0.813
	(0.408)	(0.380)	(0.430)	(0.390)
Ever covered by Medicare (1/0) during 2006–2018	0.757	0.795	0.720	0.781
	(0.429)	(0.404)	(0.449)	(0.413)
Ever covered by Medicaid (1/0) during 2006–2018	0.142	0.164	0.108	0.155
	(0.349)	(0.370)	(0.311)	(0.362)
Ever covered by other health insurance (1/0) during 2006–2018	0.335	0.356	0.211	0.342
	(0.472)	(0.479)	(0.408)	(0.474)
Ever covered by long-term care insurance (1/0) during 2006–2018	0.198	0.227	0.106	0.212
	(0.399)	(0.419)	(0.308)	(0.409)
Number of private insurance plans, avg. over 2006–2018	0.656	0.643	0.642	0.646
	(0.456)	(0.451)	(0.531)	(0.458)
Number of people living in the household, avg. over 2006–2018	2.071	2.171	2.057	2.141
	(0.920)	(0.956)	(0.969)	(0.950)
Number of living children, avg. over 2006–2018	2.780	2.941	2.987	2.908
	(1.859)	(1.948)	(2.008)	(1.933)
Household (respondent and spouse) income (in $1,000), avg. over 2006–2018	79.467	78.569	112.907	81.014
	(147.787)	(151.035)	(1153.823)	(328.842)
Household total wealth (in $1,000), avg. over 2006–2018	667.371	555.523	764.398	594.110
	(1605.222)	(1179.947)	(5952.431)	(1970.722)
State's population (in 1,000,000), avg. over 2006–2018	12.507	12.583	13.281	12.611
	(9.882)	(10.139)	(10.452)	(10.104)
State's unemployment rate (0–100), avg. over 2006–2018	6.386	6.397	6.379	6.393
	(0.919)	(0.934)	(0.918)	(0.929)
State's poverty rate (0–100), avg. over 2006–2018	13.503	13.489	13.507	13.493
	(2.624)	(2.612)	(2.525)	(2.609)
State minimum wage, avg. over 2006–2018	7.375	7.403	7.407	7.397
	(0.792)	(0.801)	(0.777)	(0.797)
Percentage of time State's governor is Democrat (0–100), avg. over 2006–2018	44.894	44.973	45.069	44.962
	(32.276)	(32.542)	(31.589)	(32.420)
Percentage of State House that is Democrat (0–100), avg. over 2006–2018	49.817	49.522	49.214	49.568
	(12.722)	(12.324)	(12.476)	(12.424)
Percentage of State Senate that is Democrat (0–100), avg. over 2006–2018	47.067	46.731	46.521	46.792
	(13.840)	(13.394)	(13.446)	(13.498)
Sample size	3,942	11,853	1,312	17,107

Notes: The mean and standard deviation (in parenthesis) are reported for each variable. The means are weighted by the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018.

In Tables 1 and 2 we made the following observations. First, the proportions of SSDI and SSI applications/receipts are higher in the consent sample than in the full sample. Second, the proportions of SSDI and SSI applications/receipts based on definition 2 are slightly higher than those based on definition 1, which is reasonable since definition 2 is more inclusive than definition 1, as explained in Section 3.4. Third, the proportions calculated using the HRS data (self-reports) are lower than those calculated using the SSA data (administrative records) for both the full sample and the consent sample. This finding is consistent with the results reported by Hyde and Harrati (2021). Fourth, unweighted and weighted summary statistics are similar. Fifth, in our study that uses a cross-sectional dataset covering the period from 2006 to 2018 (explained in Section 3), the demographic characteristics between consenters and non-consenters are similar, and this similarity is even more pronounced in the weighted summary statistics. However, there are some differences in ADL (activities of daily living), IADL (instrumental activities of daily living), and household total wealth. The residential state-level characteristics also exhibit similarities, which are further emphasized in the weighted summary statistics.

To aid in visualizing our estimation results and facilitating comparisons, the subsequent subsections employ "rope ladder" plots to present point estimates and the associated confidence intervals. Further elaborated findings are documented in tables, which can be found in the appendix.

5.2. OLS Estimates

Figure 1 provides an overview of the OLS estimates for racial and gender disparities. For more comprehensive results on racial disparity, please refer to Appendix A[14], and for gender disparity, see Appendix B[15]. The regression models used in Figure 1 include a large number of control variables and state-fixed effects. Additional information on these control variables can be found in column 4 of Appendix A and column 4 of Appendix B.

An important observation from Figure 1 is the limited evidence of racial and gender disparities in most SSDI/SSI application/receipt outcomes, once a comprehensive set of observable characteristics is taken into account. However, a few exceptions stand out, such as gender disparities in SSDI applications and receipts (based on SSA administrative data) and racial disparities in SSI applications and receipts (based on SSA administrative data). Notably, the detailed estimates in Appendix A demonstrate a diminishing racial disparity as more control variables are introduced.

---

[14] Appendix A presents detailed information about the OLS estimates for the analysis of racial disparity, considering various control variables. This analysis incorporates individual-level demographic variables, including female (1/0), continuous covariates, and discrete covariates. The continuous covariates include age, years of education, number of people living in the household, number of living children, sum of conditions ever had (doctor diagnosed), the Center for Epidemiological Studies Depression (CESD) score, sum of ADLs where respondent reported any difficulty, sum of IADLs where respondent reported any difficulty, Body Mass Index (BMI), number of private insurance plans, HRS respondent's income, household (HRS respondent and spouse) income, and household total wealth. The discrete covariates include Hispanic (1/0), ever married (including partnered) (1/0), ever had health problems that limit work (1/0), ever self-reported health being poor (including fair) (1/0), ever drank any alcohol (1/0), ever smoked (1/0), ever in the labor force (1/0), ever covered by federal government health insurance program (1/0), ever covered by Medicare (1/0), ever covered by Medicaid (1/0), ever covered by other health insurance (1/0), and ever covered by long-term care insurance (1/0). Additionally, state-level variables include population, unemployment rate, poverty rate, minimum wage, percentage of time the state's governor is a Democrat, percentage of the State House that is Democrat, and percentage of the State Senate that is Democrat.

[15] Appendix B presents detailed information about the OLS estimates for the analysis of gender disparity, using the same control variables as Appendix A, with the exception that the female (1/0) variable in Appendix A is replaced by Black (1/0).

It is crucial to highlight that the OLS estimates presented in these analyses have limitations due to the absence of interaction terms between the treatment variable (d) and observables ($\boldsymbol{x}$'s). The use of these interaction terms essentially allows a researcher to estimate the treatment effect on a "case-by-case" basis, that is, taking into account the possibility that the treatment effect varies by observable characteristics. In contrast to the observed pattern in Appendix A, Appendix B reveals a statistically significant gender disparity in SSDI applications in this analysis. However, it is important to acknowledge that the OLS estimations here also lack incorporation of interaction terms between the treatment variable (d) and observables ($\boldsymbol{x}$'s). Consequently, the insights derived from these estimations are limited.

In summary, while Figure 1 indicates minimal racial and gender disparities in most SSDI/SSI outcomes after considering observable characteristics, the use of OLS estimates without interaction terms restricts the depth of understanding we can derive from these results. Furthermore, this figure presents two sets of estimates along with their respective standard errors, computed as either cluster-robust (clustered on state) or heteroskedasticity-robust standard errors. Overall, inference (indicated by the width of a confidence interval) does not seem to be significantly impacted by whether standard errors are clustered on state or adjusted to be robust only to heteroskedasticity. If there is concern about the small number of clusters (i.e., states), we could just conduct inferences by using heteroskedasticity-robust standard errors. Nevertheless, exercising caution, we have opted for cluster-robust standard errors, which may potentially be larger than the actual ones.

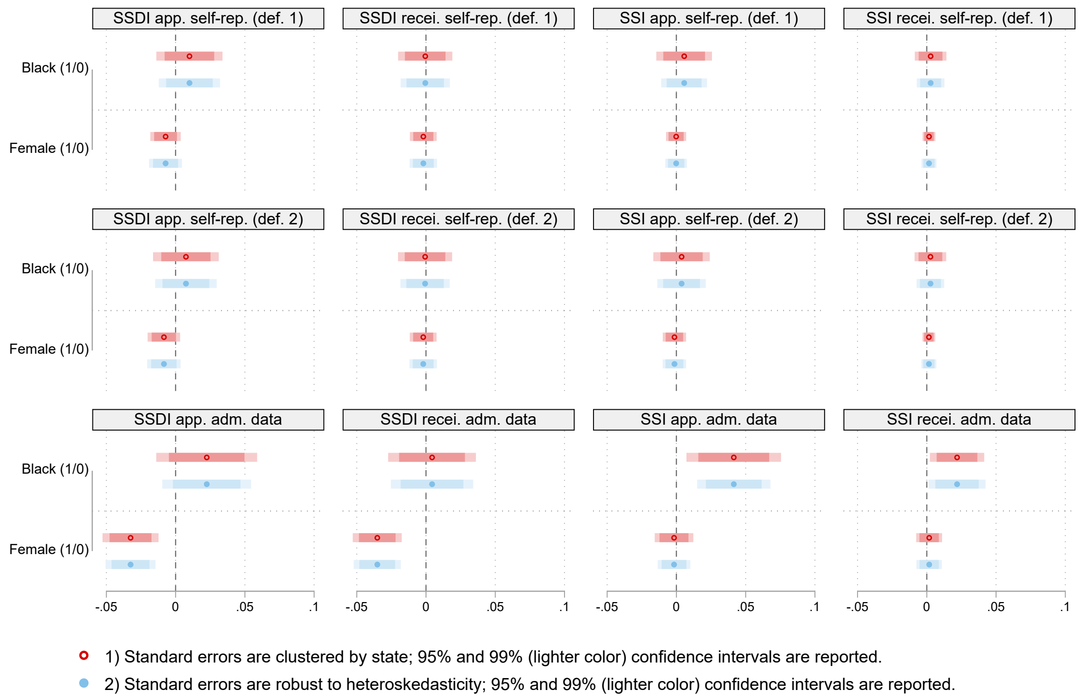


Figure 1: Estimates of Racial Disparity and Gender Disparity Obtained by the Ordinary Least Squares Estimator

Notes: The number of observations for the outcome variables that use the self-reported data (HRS data) is 17,107. The number of observations for outcome variables that use the administrative data (SSA records) is 11,853. Regressions are weighted. The weight variable is the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018. Reported in the figure are the point estimates and the associated standard errors. The regression model controls for individual-level demographic variables and state fixed effects (i.e., the specification used for column 4 of Appendices A and B).[16]

Figures 2 and 3 present the OLS estimates for analyses on racial disparity and gender disparity, respectively. In both figures, the case where no interaction terms are included means that

[16] Appendix C presents a comparison between the estimates obtained from the OLS and logit models. Specifically, for the OLS a linear model is used for the outcome variable; the point estimates and the associated confidence intervals are reported. Notably, in this case, the point estimates are equal to the marginal effect estimates. For the logit model used for the outcome variable, the marginal effect estimates, and the associated confidence intervals are reported. Both the OLS and the logit regressions are weighted, using the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018. The standard errors are clustered on a state level. Additionally, both models control for individual-level demographic variables and state fixed effects, following the specifications outlined in column 4 of Appendix A and Appendix B. Overall, the results from both the OLS and logit models are similar.

the control variables include female (1/0) for the racial disparity analysis and Black (1/0) for the gender disparity analysis, along with the continuous covariates (explained in Appendix A), the discrete covariates (explained in Appendix A), and the state-level variables (explained in Appendix A). The case where interaction terms are included means that: (1) in Figure 2, Black (1/0) is interacted with female (1/0), the continuous covariates, the discrete covariates, and the state-level variables; (2) in Figure 3, female (1/0) is interacted with Black (1/0), the continuous covariates, the discrete covariates (explained earlier), and the state-level variables. Overall, the results depicted in Figure 2 show that the racial disparity estimates (i.e., the point estimates and the associated confidence intervals) change substantially, especially in the case of SSDI and SSI applications/receipts using the SSA administrative data, when interaction terms are included. In contrast, the results displayed in Figure 3 indicate that the gender disparity estimates (i.e., the point estimates and the associated confidence intervals) remain very similar regardless of the inclusion of interaction terms. These patterns suggest that racial disparity could vary substantially with observable characteristics, whereas gender disparity appears to be similar across those observables.

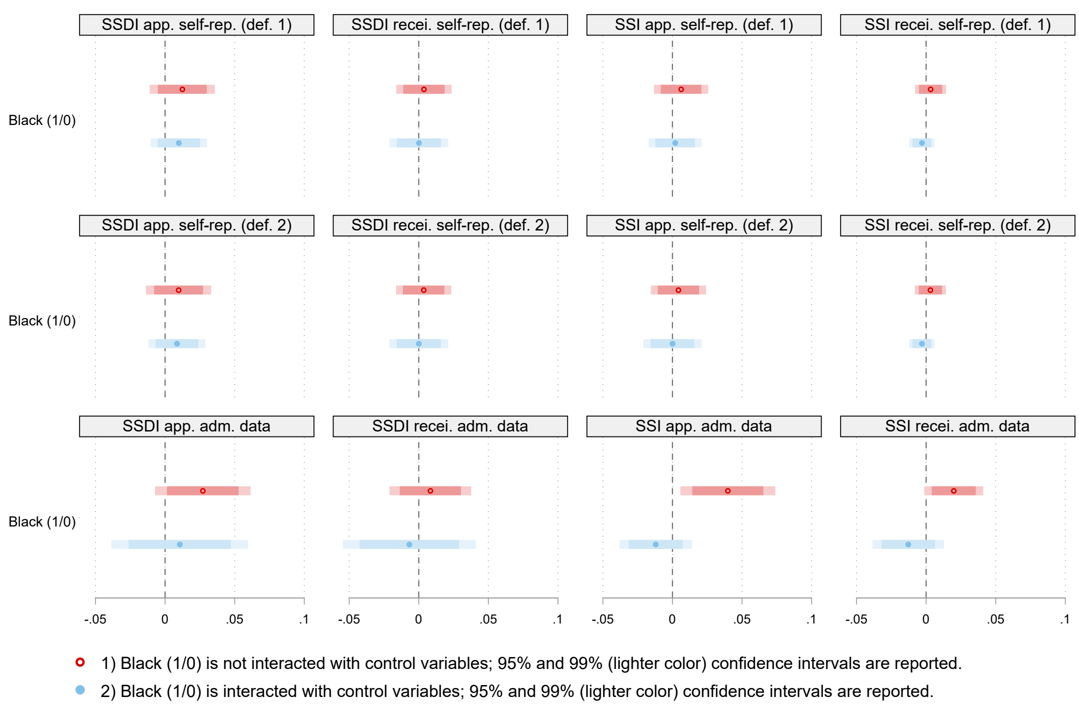


Figure 2: Analysis of the Ordinary Least Squares Estimates of Racial Disparity

Notes: See the notes section of Appendix D.

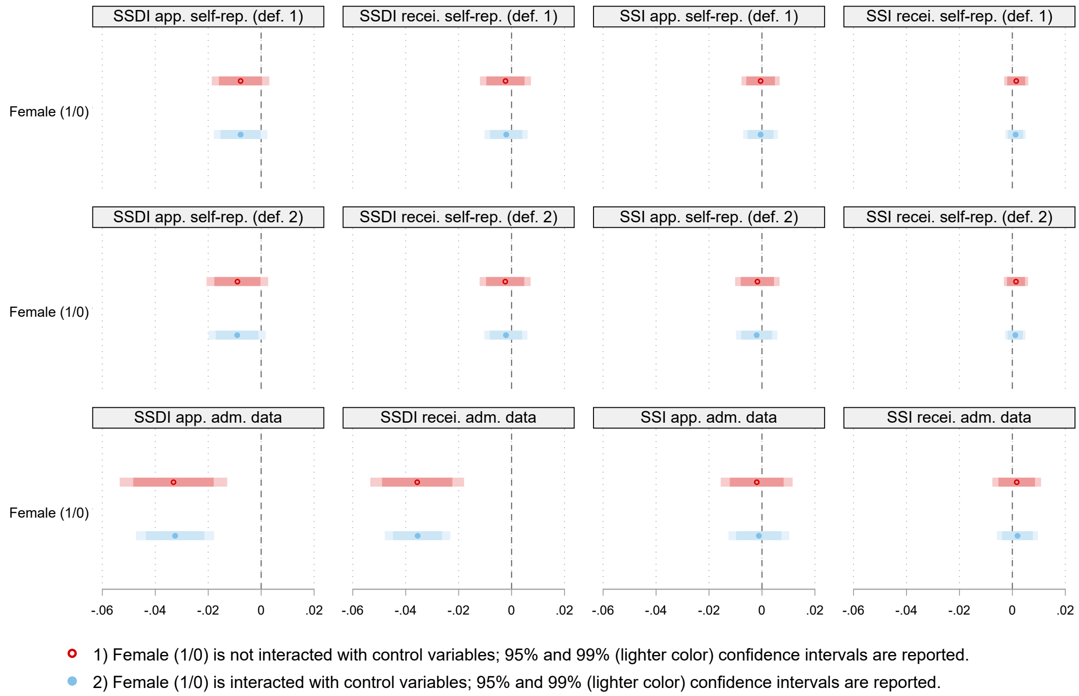


Figure 3: Analysis of the Ordinary Least Squares Estimates of Gender Disparity

Notes: See the notes section of Appendix E.

Appendix D provides more detailed results corresponding to Figure 2. In the racial disparity analysis (Appendix D), it appears that there is no statistically significant disparity when controlling for the interaction terms. Appendix E provides more detailed results corresponding to Figure 3. In the gender disparity analysis (Appendix E), it appears that in the case of SSDI applications using the SSA administrative data there is a robust finding of a decrease in applications among females compared with males.

In Appendix D and Appendix E, we also demonstrate the full replication of the disparity estimates by setting $d = 1$ for all observations, followed by setting $d = 0$ for all observations, and then calculating the average difference. We conduct this replication exercise to address the following rationale: Strictly speaking, there is no "effect" of race or gender because race or gender is an inherent attribute, not a manipulable construct. However, in widely used OLS regression analyses, the disparity estimate obtained is equivalent to evaluating the outcome for the same individual twice, once with $d = 1$ and once with $d = 0$, treating d (representing race or gender) as a manipulable construct. This means that, by using OLS regression analyses, what we are doing is not treating race or gender as an attribute; instead, we are treating race or gender as a manipulable construct, such as a perceived characteristic, which allows us to conduct a "treatment-effect" evaluation.

5.3. Double/Debiased ML Estimates

Figure 4 presents the racial disparity estimates obtained using the double/debiased ML estimator, with more detailed results reported in Appendix F.

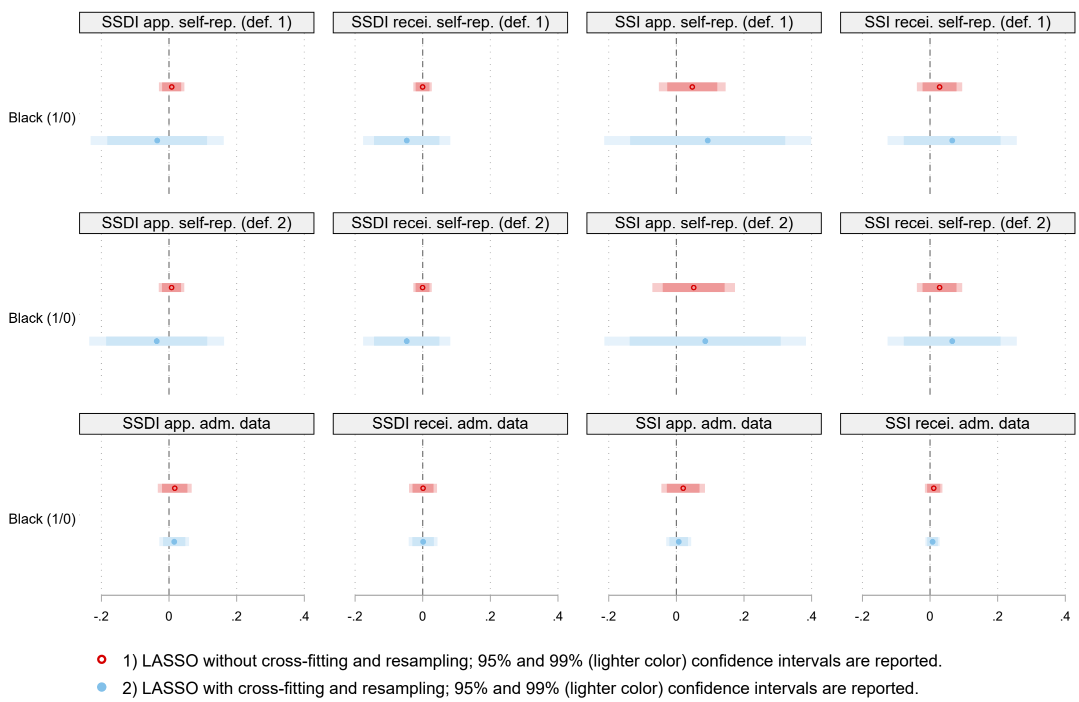


Figure 4: Estimates of Racial Disparity Obtained by the Double/Debiased Machine Learning Estimator

Notes: See the notes section of Appendix F.

Overall, the results are largely similar (except for the confidence intervals in cases where HRS self-reported SSDI and SSI applications/receipts data are used), comparing the case of cross-fitting used together with resampling and the case of no cross-fitting and resampling: There appears to be no racial disparity in SSDI and SSI applications/receipts once we control for as many possible interactions between Black (1/0) and the observed variables in a flexible, data-driven way using the double/debiased estimator.

By using this data-driven approach and relying less on subjective assumptions (about exactly what interaction terms should be included), here we essentially examine whether our conclusion about the presence of an overall racial disparity—that is, the disparity averaged across different characteristics of the population—will change if we examine the disparity characteristic by characteristic, that is on a "case-by-case" basis (which is equivalent to allowing the treatment to interact with observable characteristics). The OLS, although widely popular and commonly used, has limitations in conducting such "case-by-case" analyses. However, such analyses are critical in forming an overarching conclusion about the presence of a disparity.

Figure 5 presents the gender disparity estimates obtained using the double/debiased ML estimator, with more detailed results reported in Appendix G.

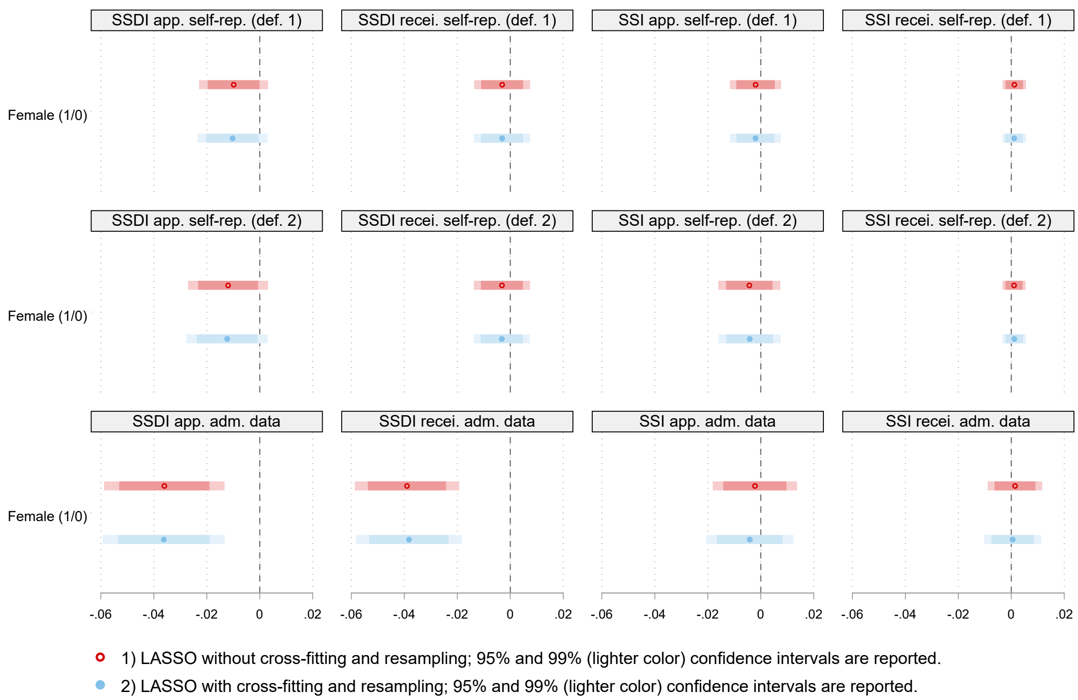


Figure 5: Estimates of Gender Disparity Obtained by the Double/Debiased Machine Learning Estimator

Notes: See the notes section of Appendix G.

In contrast to the findings regarding racial disparity estimates, here we find a statistically significant decrease in the case of SSDI applications among females compared with males, both for the HRS data (where the decrease is significant at the 5 percent level, but insignificant at the 1 percent level) and for the SSA data. Regarding SSDI receipt, we find a statistically significant decrease among females compared with males only in the SSA data. Furthermore, unlike the pattern observed in Figure 4, the results in Figure 5 remain very similar regardless of whether cross-fitting and resampling are used.[17]

In Appendix H and Appendix J we present the results on racial disparity and gender disparity obtained by employing alternative specifications for the LASSO technique, respectively. More detailed results can be found in Appendix I (for Appendix H) and Appendix K (for Appendix J). Overall, our estimates (point estimates and the associated confidence intervals) are robust to alternative specifications for the LASSO technique, which is a crucial component of the double/debiased ML estimator. This robustness is particularly notable in the estimation of gender disparity.

In conclusion, when comparing the estimates obtained from the double/debiased estimator, which uses a data-driven approach, with the estimates obtained from OLS, which involves subjectively modeling interaction terms, we observe a distinct pattern. Racial disparity (White vs. Black) seems to vary significantly with observable characteristics, leading to the potential for substantial changes in conclusions about its presence based on how these interaction terms are incorporated. In this context, a data-driven approach emerges as a valuable alternative to the model-dependent approach employed by OLS. On the other hand, gender disparity (Male vs. Female) appears to remain constant across these observable characteristics. Although the advantage of flexibly controlling for interaction terms may not be apparent in this case, a data-driven approach still holds significance for conducting robustness checks.

6. Discussion and Conclusion

[17] The estimates obtained from cross-fitting and resampling have wider confidence intervals in the case of racial disparity. This may be explained by the "imbalance," that is 23 percent (d=1) versus 77 percent (d=0). In this case, the random partition used by cross-fitting (and resampling) may exacerbate that imbalance, therefore making the estimates more variables (i.e., resulting in wider confidence intervals). In contrast, in the case of gender disparity, there is no imbalance (male vs. female is about 45 percent vs. 55 percent). As a result, the estimates with or without cross-fitting and resampling have similar confidence intervals. Although it is conceivable to randomly select a subset of the larger group to make the two groups similar in size, which has been used in the field of data science, this solution is only for prediction tasks, not for inference tasks; that is, only for the tasks in which the predicted values of the dependent variable are of interest. This solution does not work for inference tasks such as conducting hypothesis tests for a model's parameters. A simple intuition for why this solution does not work for inference tasks is this: Randomly selecting a subset of the larger group to make two groups similar in size is essentially conducting a one-to-one matching between the two groups. In this case, decisions as to which observation in the treatment group is matched to which observation in the control group will affect what is being estimated (i.e., the estimand). As a result, using a randomly selected subset and using the full sample (i.e., without using the randomly selected subset) will have different estimands.

In this study, we have undertaken a comprehensive empirical analysis to investigate racial and gender disparities in SSDI and SSI applications and awards. By utilizing rich datasets and applying advanced analytical techniques, we have contributed to the growing literature on disparities in social welfare programs, specifically in the context of disability benefits.

However, it is important to approach the conclusions drawn from our study with caution. As we have highlighted throughout the paper, the presence of disparities cannot be definitively established due to the lack of theoretical guidance in modeling the outcome of interest. In such cases, a data-driven approach can be a valuable alternative (or complement) to the theoretical modeling driven approach.

The diminishing gaps when estimated by controlling for more factors, particularly with respect to the gender disparity in SSDI applications for women, could be attributed to various socioeconomic and historical reasons. Women are more likely to have lower income than men, which may limit their ability to afford regular medical treatment and obtain sufficient evidence for their disability claims. Furthermore, past generations of women may have been less informed about SSDI, preferring to rely on family or needs-tested programs during times of severe medical impairment. This aligns with observations from the 1970s, where researchers found that despite women reporting higher rates of disability, they were less likely to even apply for SSDI.[18] Historically, women with disabilities leaned more towards relying on their spouses' earnings or collecting public assistance, compared with their male counterparts.

Our study has utilized the double/debiased ML estimator, which offers several advantages in estimating an ATE that is heterogenous in observables, overcoming the problem of violating the overlap assumption (that conventional methods such as propensity-score matching rely on) through the dimension reduction technique such as LASSO. By implementing dimension reduction in the estimation of a propensity score, which is commonly used to handle heterogeneity, our approach allows for dimension reduction using ML techniques such as LASSO.

The decision to employ the double/debiased ML estimator in our study was driven by its potential to handle high-dimensional data and select salient features, allowing for a more nuanced analysis than might be possible with traditional methods. While not all future studies necessarily need to adopt the double/debiased ML estimator, it serves as a valuable alternative when the

[18] See https://www.cbpp.org/research/social-security/women-and-disability-insurance-five-facts-you-should-know (accessed on August 6, 2023).

research goal is to uncover subtle patterns using a more data-driven, rather than theory-driven, approach.

The application of the double/debiased ML estimator to the analysis of racial and gender disparities in SSDI/SSI applications and awards has yielded valuable insights. Our findings suggest that the inclusion of interaction terms between race or gender and observed variables significantly affects the disparity estimates, particularly in the case of racial disparity. Notably, our analysis reveals a statistically significant gender disparity in SSDI applications, indicating a decrease among females compared to males. However, no statistically significant disparities were found in other situations examined (except the case of SSDI awards using the SSA data, instead of the HRS data).

The implications of our research extend beyond understanding the presence of disparities. Policymakers can utilize our findings to evaluate the performance of the SSDI and SSI programs in reaching individuals with disabilities from different racial and gender groups. Our study highlights the need for targeted policy interventions to address potential gaps and biases in the program, improve outreach efforts, and reduce barriers that disproportionately affect certain groups. Moreover, our research underscores the importance of comprehensive data collection and monitoring systems to track and address disparities in disability benefits. Incorporating data on race and gender in program evaluation processes and regularly assessing program performance will enable researchers and policymakers to identify emerging trends, evaluate the impact of policy changes, and ensure accountability in addressing disparities.

In conclusion, this study contributes to the ongoing discourse on racial and gender disparities in SSDI/SSI applications and awards. By employing advanced analytical techniques and utilizing rich datasets, we have shed new light on the extent and nature of these disparities. While we acknowledge the limitations of our study and the caution required in interpreting the results, our research provides valuable insights for policymakers, researchers, and practitioners seeking to promote equity, social justice, and equal access to disability benefits for individuals with disabilities, regardless of their race or gender.

References

Ahrens, Achim, Christian B. Hansen, and Mark E. Schaffer. 2020. “lassopack: Model Selection and Prediction with Regularized Regression in Stata.” *The Stata Journal* 20, no. 1: 176–235.

Athey, Susan, and Guido W. Imbens. 2019. “Machine Learning Methods That Economists Should Know About.” *Annual Review of Economics* 11, no. 1: 685–725.

Autor, David H., Mark Duggan, Kyle Greenberg, and David S. Lyle. 2016. “The Impact of Disability Benefits on Labor Supply: Evidence from the VA’s Disability Compensation Program.” *American Economic Journal: Applied Economics* 8, no. 3: 31–68.

Baldwin, Marjorie L. 1997. “Gender Differences in Social Security Disability Decisions.” *Journal of Disability Policy Studies* 8, no. 1–2: 25–49.

Benítez-Silva, Hugo, Moshe Buchinsky, Hiu Man Chan, John Rust, and Sofia Sheidvasser. 1999. “An Empirical Analysis of the Social Security Disability Application, Appeal, and Award Process.” *Labour Economics* 6, no. 2: 147–178.

Cameron, A. Colin and Pravin K. Trivedi. 2022. *Microeconometrics Using Stata, Second Edition*. College Station, Texas: Stata Press.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. “Double/Debiased Machine Learning for Treatment and Structural Parameters.” *The Econometrics Journal* 21, no. 1: C1–C68.

Deshpande, Manasi. 2016. “The Effect of Disability Payments on Household Earnings and Income: Evidence from the SSI Children's Program.” *The Review of Economics and Statistics* 98, no. 4: 638–654.

Favreault, Melissa M., Richard W. Johnson, and Karen E. Smith. 2013. “How Important Is social security Disability Insurance to Us Workers?.” *Urban Institute*.

Godtland, Erin M., Michele Grgich, Carol Dawn Petersen, Douglas M. Sloane, and Ann T. Walker. 2007. “Racial Disparities in Federal Disability Benefits.” *Contemporary Economic Policy* 25, no. 1: 27–45.

Health and Retirement Study, (RAND HRS Longitudinal File 2018 (V2)) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (July 2022). Accessed September 2023.
https://hrsdata.isr.umich.edu/sites/default/files/documentation/other/1658948491/randhrs1992_2018v2.pdf.

Hu, Jianting, Kajal Lahiri, Denton R. Vaughan, and Bernard Wixon. 2001. "A Structural Model of Social Security's Disability Determination Process." *The Review of Economics and Statistics* 83, no. 2: 348–361.

Hyde, Jody S., and Amal Harrati 2021. "The Alignment between Self-Reported and Administrative Measures of Application to and Receipt of Federal Disability Benefits in the Health and Retirement Study." Working Paper 2021–26. Chestnut Hill, MA: Center for Retirement Research at Boston College.

Kreider, Brent, and Regina T. Riphahn. 2000. "Explaining Applications to the U.S. Disability System: A Semiparametric Approach." *The Journal of Human Resources* 35, no. 1: 82–115.

RAND HRS Longitudinal File 2018 (V2). Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (July 2022). Accessed September 2023. https://hrsdata.isr.umich.edu/sites/default/files/documentation/other/1658948491/randhrs1992_2018v2.pdf.

University of Kentucky Center for Poverty Research. (2023, Feb.). UKCPR National Welfare Data, 1980–2021. Lexington, KY. Accessed in March 2023. http://ukcpr.org/resources/national-welfare-data.

U.S. General Accounting Office (GAO). Social Security: Racial Difference in Disability Decisions Warrants Further Investigation. GAO/HRD-92-56. Washington, DC, April 21, 1992. Accessed September 2023. https://www.gao.gov/assets/hrd-92-56.pdf.

U.S. General Accounting Office (GAO). SSA Disability Decision Making: Additional Steps Needed to Ensure Accuracy and Fairness at the Hearings Level. GAO-04-14. Washington, DC, November 2003. Accessed September 2023. https://www.gao.gov/assets/gao-04-14.pdf.

Appendix A: Estimates of Racial Disparity Obtained by the Ordinary Least Squares Estimator

	(1)	(2)	(3)	(4)
(1) Outcome variable: ever applied for SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Black (1) vs. White (0)	0.0573 ***	0.0150 *	0.0124	0.0100
	(0.0101)	(0.0088)	(0.0087)	(0.0089)
(2) Outcome variable: ever received SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Black (1) vs. White (0)	0.0342 ***	0.0048	0.0037	-0.0006
	(0.0080)	(0.0073)	(0.0074)	(0.0073)
(3) Outcome variable: ever applied for SSI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Black (1) vs. White (0)	0.0325 ***	0.0077	0.0062	0.0056
	(0.0067)	(0.0069)	(0.0073)	(0.0075)
(4) Outcome variable: ever received SSI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Black (1) vs. White (0)	0.0157 ***	0.0038	0.0032	0.0027
	(0.0039)	(0.0040)	(0.0042)	(0.0043)
(5) Outcome variable: ever applied for SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Black (1) vs. White (0)	0.0562 ***	0.0123	0.0097	0.0074
	(0.0101)	(0.0089)	(0.0087)	(0.0088)
(6) Outcome variable: ever received SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Black (1) vs. White (0)	0.0341 ***	0.0047	0.0035	-0.0007
	(0.0080)	(0.0073)	(0.0074)	(0.0073)
(7) Outcome variable: ever applied for SSI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Black (1) vs. White (0)	0.0345 ***	0.0060	0.0043	0.0037
	(0.0071)	(0.0072)	(0.0074)	(0.0076)
(8) Outcome variable: ever received SSI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Black (1) vs. White (0)	0.0157 ***	0.0036	0.0030	0.0026
	(0.0039)	(0.0040)	(0.0042)	(0.0043)
(9) Outcome variable: ever applied for SSDI during 2006–2018 (1/0), administrative data (SSA records)				
Black (1) vs. White (0)	0.1238 ***	0.0295 **	0.0271 **	0.0224
	(0.0144)	(0.0125)	(0.0128)	(0.0136)
(10) Outcome variable: ever received SSDI during 2006–2018 (1/0), administrative data (SSA records)				
Black (1) vs. White (0)	0.0935 ***	0.0092	0.0083	0.0043
	(0.0116)	(0.0110)	(0.0109)	(0.0118)
(11) Outcome variable: ever applied for SSI during 2006–2018 (1/0), administrative data (SSA records)				
Black (1) vs. White (0)	0.1073 ***	0.0389 ***	0.0398 ***	0.0414 ***
	(0.0153)	(0.0124)	(0.0127)	(0.0127)
(12) Outcome variable: ever received SSI during 2006–2018 (1/0), administrative data (SSA records)				
Black (1) vs. White (0)	0.0660 ***	0.0194 **	0.0198 **	0.0218 ***
	(0.0095)	(0.0077)	(0.0079)	(0.0073)
Control variables				
Individual-level demographic variables	No	Yes	Yes	Yes
State-level variables	No	No	Yes	No
State fixed effects	No	No	No	Yes

Notes: The number of observations for the outcome variables that use the self-reported data (HRS data) is 17,107. The number of observations for outcome variables that use the administrative data (SSA records) is 11,853. Regressions are weighted. The weight variable is the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018. Standard errors are clustered on state. * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01.

Individual-level demographic variables include female (1/0), the continuous covariates, and the discrete covariates.

The continuous covariates include: age, years of education, number of people living in the household, number of living children, sum of conditions ever had (doctor diagnosed), the Center for Epidemiological Studies Depression (CESD) score, sum of ADLs where respondent reported any difficulty, sum of IADLs where respondent reported any difficulty, Body Mass Index (BMI), number of private insurance plans, HRS respondent's income, household (HRS respondent and spouse) income, and household total wealth.

The discrete covariates include: Hispanic (1/0), ever married (including partnered) (1/0), ever had health problems that limit work (1/0), ever self-reported health being poor (including fair) (1/0), ever drank any alcohol (1/0), ever smoked (1/0), ever in the labor force (1/0), ever covered by federal government health insurance program (1/0), ever covered by Medicare (1/0), ever covered by Medicaid (1/0), ever covered by other health insurance (1/0), and ever covered by long-term care insurance (1/0).

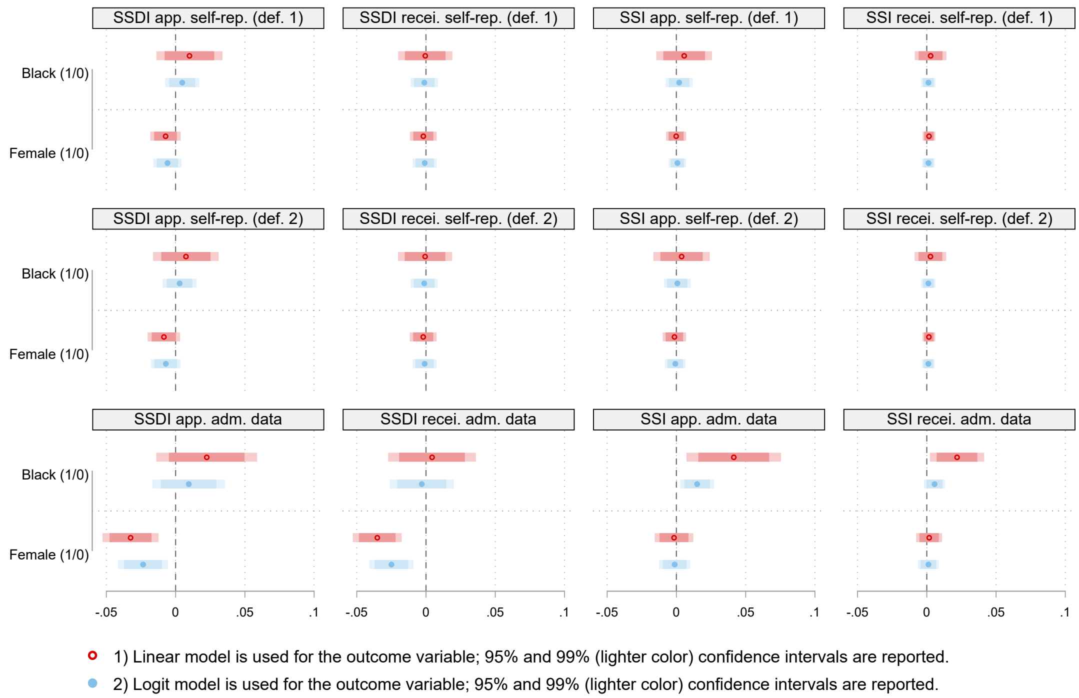
The state-level variables include: population, unemployment rate, poverty rate, minimum wage, percentage of time the state's governor is a Democrat, percentage of the State House that is Democrat, and percentage of the State Senate that is Democrat.

Appendix B: Estimates of Gender Disparity Obtained by the Ordinary Least Squares Estimator

	(1)	(2)	(3)	(4)
(1) Outcome variable: ever applied for SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Female (1) vs. Male (0)	-0.0083 *	-0.0080 *	-0.0078 *	-0.0073 *
	(0.0047)	(0.0041)	(0.0041)	(0.0041)
(2) Outcome variable: ever received SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Female (1) vs. Male (0)	-0.0035	-0.0024	-0.0023	-0.0020
	(0.0031)	(0.0036)	(0.0036)	(0.0036)
(3) Outcome variable: ever applied for SSI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Female (1) vs. Male (0)	0.0004	-0.0006	-0.0005	-0.0002
	(0.0026)	(0.0027)	(0.0027)	(0.0027)
(4) Outcome variable: ever received SSI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Female (1) vs. Male (0)	0.0021	0.0014	0.0014	0.0016
	(0.0014)	(0.0017)	(0.0017)	(0.0017)
(5) Outcome variable: ever applied for SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Female (1) vs. Male (0)	-0.0088 *	-0.0092 **	-0.0090 **	-0.0085 *
	(0.0049)	(0.0044)	(0.0043)	(0.0044)
(6) Outcome variable: ever received SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Female (1) vs. Male (0)	-0.0036	-0.0025	-0.0024	-0.0020
	(0.0032)	(0.0036)	(0.0036)	(0.0036)
(7) Outcome variable: ever applied for SSI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Female (1) vs. Male (0)	-0.0003	-0.0019	-0.0017	-0.0014
	(0.0030)	(0.0031)	(0.0031)	(0.0031)
(8) Outcome variable: ever received SSI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Female (1) vs. Male (0)	0.0021	0.0013	0.0014	0.0015
	(0.0014)	(0.0017)	(0.0017)	(0.0017)
(9) Outcome variable: ever applied for SSDI during 2006–2018 (1/0), administrative data (SSA records)				
Female (1) vs. Male (0)	-0.0170 **	-0.0337 ***	-0.0331 ***	-0.0325 ***
	(0.0070)	(0.0076)	(0.0076)	(0.0075)
(10) Outcome variable: ever received SSDI during 2006–2018 (1/0), administrative data (SSA records)				
Female (1) vs. Male (0)	-0.0186 ***	-0.0363 ***	-0.0357 ***	-0.0352 ***
	(0.0064)	(0.0066)	(0.0066)	(0.0066)
(11) Outcome variable: ever applied for SSI during 2006–2018 (1/0), administrative data (SSA records)				
Female (1) vs. Male (0)	0.0082	-0.0020	-0.0020	-0.0017
	(0.0054)	(0.0051)	(0.0051)	(0.0052)
(12) Outcome variable: ever received SSI during 2006–2018 (1/0), administrative data (SSA records)				
Female (1) vs. Male (0)	0.0100 ***	0.0017	0.0016	0.0017
	(0.0035)	(0.0034)	(0.0034)	(0.0035)
Control variables				
Individual-level demographic variables	No	Yes	Yes	Yes
State-level variables	No	No	Yes	No
State fixed effects	No	No	No	Yes

Notes: The same as Appendix A, except that it is Black (1/0), not female (1/0), that is included as an individual-level demographic variable.

Appendix C: Estimates of Racial Disparity and Gender Disparity Obtained by the Ordinary Least Squares vs. Obtained by the Logit



Notes: The number of observations for the outcome variables that use the self-reported data (HRS data) is 17,107. The number of observations for outcome variables that use the administrative data (SSA records) is 11,853. For the ordinary least squares (OLS), a linear model is used for the outcome variable; the point estimates and the associated standard errors are reported; and in this case the point estimates are equal to the marginal effect estimates. For the logit, a logit model is used for the outcome variable; the marginal effect estimates and the associated standard errors are reported. Both the OLS and the logit regressions are weighted. The weight variable is the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018. Standard errors are clustered on state. Both the OLS and the logit models control for individual-level demographic variables and state fixed effects (i.e., the specification used for column 4 of Appendices A and B).

Appendix D: Analysis of the Ordinary Least Squares Estimates of Racial Disparity

	(1) Ever applied X	(2) Ever applied X	(3) Ever received X	(4) Ever received X
(1) Outcome variable X = SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Black (1) vs. White (0), taking into account interactions terms when they are included	0.0124	0.0098	0.0037	0.0001
	(0.0087)	(0.0076)	(0.0074)	(0.0079)
Average predicted outcome, setting Black = 1 for all observations	0.0607	0.0583	0.0357	0.0321
Average predicted outcome, setting Black = 0 for all observations	0.0483	0.0485	0.0320	0.0320
Difference in average predicted outcome	0.0124	0.0098	0.0037	0.0001
(2) Outcome variable X = SSI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Black (1) vs. White (0), taking into account interactions terms when they are included	0.0062	0.0019	0.0032	-0.0030
	(0.0073)	(0.0071)	(0.0042)	(0.0034)
Average predicted outcome, setting Black = 1 for all observations	0.0257	0.0214	0.0109	0.0046
Average predicted outcome, setting Black = 0 for all observations	0.0195	0.0195	0.0077	0.0076
Difference in average predicted outcome	0.0062	0.0019	0.0032	-0.0030
(3) Outcome variable X = SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Black (1) vs. White (0), taking into account interactions terms when they are included	0.0097	0.0085	0.0035	0.0000
	(0.0087)	(0.0076)	(0.0074)	(0.0079)
Average predicted outcome, setting Black = 1 for all observations	0.0601	0.0591	0.0356	0.0321
Average predicted outcome, setting Black = 0 for all observations	0.0504	0.0506	0.0321	0.0321
Difference in average predicted outcome	0.0097	0.0085	0.0035	0.0000
(4) Outcome variable X = SSI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Black (1) vs. White (0), taking into account interactions terms when they are included	0.0043	0.0000	0.0030	-0.0031
	(0.0074)	(0.0078)	(0.0042)	(0.0034)
Average predicted outcome, setting Black = 1 for all observations	0.0276	0.0233	0.0108	0.0046
Average predicted outcome, setting Black = 0 for all observations	0.0233	0.0233	0.0078	0.0076
Difference in average predicted outcome	0.0043	0.0000	0.0030	-0.0031
(5) Outcome variable X = SSDI during 2006–2018 (1/0), administrative data (SSA records)				
Black (1) vs. White (0), taking into account interactions terms when they are included	0.0271 **	0.0105	0.0083	-0.0068
	(0.0128)	(0.0183)	(0.0109)	(0.0178)
Average predicted outcome, setting Black = 1 for all observations	0.1330	0.1163	0.0999	0.0847

Average predicted outcome, setting Black = 0 for all observations	0.1059	0.1058	0.0916	0.0915
Difference in average predicted outcome	0.0271	0.0105	0.0083	-0.0068
(6) Outcome variable X = SSI during 2006–2018 (1/0), administrative data (SSA records)				
Black (1) vs. White (0), taking into account interactions terms when they are included	0.0398 ***	-0.0121	0.0198 **	-0.0129
	(0.0127)	(0.0097)	(0.0079)	(0.0096)
Average predicted outcome, setting Black = 1 for all observations	0.0760	0.0233	0.0411	0.0078
Average predicted outcome, setting Black = 0 for all observations	0.0363	0.0354	0.0213	0.0206
Difference in average predicted outcome	0.0398	-0.0121	0.0198	-0.0129
Control variables				
Individual-level demographic variables	Yes	Yes	Yes	Yes
State-level variables	Yes	Yes	Yes	Yes
Interaction terms between Black (1/0) and individual-level demographic variables and state-level variables	No	Yes	No	Yes

Notes: The number of observations for the outcome variables that use the self-reported data (HRS data) is 17,107. The number of observations for the outcome variables that use the administrative data (SSA records) is 11,853. Regressions are weighted. The weight variable is the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018. Standard errors are clustered on state. * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01.

In the case where no interaction terms are included, control variables include female (1/0), the continuous covariates (explained in the notes section of Appendix A), the discrete covariates (explained in the notes section of Appendix A), and the state-level variables (explained in the notes section of Appendix A).

In the case where interaction terms are included, control variables include: female (1/0); the continuous covariates (explained in the notes section of Appendix A); the discrete covariates (explained in the notes section of Appendix A); the state-level variables (explained in the notes section of Appendix A); the interactions between Black (1/0) and female (1/0), the continuous covariates, the discrete covariates, and the state-level covariates.

Appendix E: Analysis of the Ordinary Least Squares Estimates of Gender Disparity

	(1) Ever applied X	(2) Ever applied X	(3) Ever received X	(4) Ever received X
(1) Outcome variable X = SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Female (1) vs. Male (0), taking into account interactions terms when they are included	-0.0078 *	-0.0078 **	-0.0023	-0.0020
	(0.0041)	(0.0038)	(0.0036)	(0.0030)
Average predicted outcome, setting Female = 1 for all observations	0.0462	0.0442	0.0314	0.0295
Average predicted outcome, setting Female = 0 for all observations	0.0540	0.0520	0.0337	0.0315
Difference in average predicted outcome	-0.0078	-0.0078	-0.0023	-0.0020
(2) Outcome variable X = SSI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Female (1) vs. Male (0), taking into account interactions terms when they are included	-0.0005	-0.0005	0.0014	0.0012
	(0.0027)	(0.0024)	(0.0017)	(0.0014)
Average predicted outcome, setting Female = 1 for all observations	0.0200	0.0194	0.0088	0.0086
Average predicted outcome, setting Female = 0 for all observations	0.0205	0.0200	0.0073	0.0074
Difference in average predicted outcome	-0.0005	-0.0005	0.0014	0.0012
(3) Outcome variable X = SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Female (1) vs. Male (0), taking into account interactions terms when they are included	-0.0090 **	-0.0091 **	-0.0024	-0.0021
	(0.0043)	(0.0040)	(0.0036)	(0.0030)
Average predicted outcome, setting Female = 1 for all observations	0.0474	0.0457	0.0314	0.0295
Average predicted outcome, setting Female = 0 for all observations	0.0564	0.0548	0.0338	0.0316
Difference in average predicted outcome	-0.0090	-0.0091	-0.0024	-0.0021
(4) Outcome variable X = SSI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Female (1) vs. Male (0), taking into account interactions terms when they are included	-0.0017	-0.0020	0.0014	0.0011
	(0.0031)	(0.0029)	(0.0017)	(0.0014)
Average predicted outcome, setting Female = 1 for all observations	0.0230	0.0226	0.0088	0.0086
Average predicted outcome, setting Female = 0 for all observations	0.0247	0.0246	0.0074	0.0075
Difference in average predicted outcome	-0.0017	-0.0020	0.0014	0.0011
(5) Outcome variable X = SSDI during 2006–2018 (1/0), administrative data (SSA records)				
Female (1) vs. Male (0), taking into account interactions terms when they are included	-0.0331 ***	-0.0325 ***	-0.0357 ***	-0.0355 ***
	(0.0076)	(0.0055)	(0.0066)	(0.0046)
Average predicted outcome, setting Female = 1 for all observations	0.0934	0.0917	0.0759	0.0740

Average predicted outcome, setting Female = 0 for all observations	0.1266	0.1242	0.1116	0.1095
Difference in average predicted outcome	-0.0331	-0.0325	-0.0357	-0.0355
(6) Outcome variable X = SSI during 2006–2018 (1/0), administrative data (SSA records)				
Female (1) vs. Male (0), taking into account interactions terms when they are included	-0.0020	-0.0012	0.0016	0.0019
	(0.0051)	(0.0043)	(0.0034)	(0.0029)
Average predicted outcome, setting Female = 1 for all observations	0.0396	0.0397	0.0242	0.0243
Average predicted outcome, setting Female = 0 for all observations	0.0416	0.0409	0.0225	0.0223
Difference in average predicted outcome	-0.0020	-0.0012	0.0016	0.0019
Control variables				
Individual-level demographic variables	Yes	Yes	Yes	Yes
State-level variables	Yes	Yes	Yes	Yes
Interaction terms between Black (1/0) and individual-level demographic variables and state-level variables	No	Yes	No	Yes

Notes: The number of observations for the outcome variables that use the self-reported data (HRS data) is 17,107. The number of observations for the outcome variables that use the administrative data (SSA records) is 11,853. Regressions are weighted. The weight variable is the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018. Standard errors are clustered on state. * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01.

In the case where no interaction terms are included, control variables include Black (1/0), the continuous covariates (explained in the notes section of Appendix A), the discrete covariates (explained in the notes section of Appendix A), and the state-level variables (explained in the notes section of Appendix A).

In the case where interaction terms are included, control variables include: Black (1/0); the continuous covariates (explained in the notes section of Appendix A); the discrete covariates (explained in the notes section of Appendix A); the state-level variables (explained in the notes section of Appendix A); the interactions between female (1/0) and Black (1/0), the continuous covariates, the discrete covariates, and the state-level covariates.

Appendix F: Estimates of Racial Disparity Obtained by the Double/Debiased Machine Learning Estimator

	(1) Ever applied X	(2) Ever applied X	(3) Ever received X	(4) Ever received X
(1) Outcome variable X = SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Black (1) vs. White (0)	0.0078	-0.0350	-0.0004	-0.0473
	(0.0140)	(0.0734)	(0.0102)	(0.0480)
	[0.5822]	[0.6351]	[0.9651]	[0.3299]
	[-0.0204, 0.0359]	[-0.1827, 0.1126]	[-0.0209, 0.0200]	[-0.1439, 0.0494]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	52	98	47	105
Cross-fitting and resampling used	No	Yes	No	Yes
(2) Outcome variable X = SSI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Black (1) vs. White (0)	0.0469	0.0926	0.0281	0.0655
	(0.0368)	(0.1140)	(0.0250)	(0.0712)
	[0.2085]	[0.4209]	[0.2664]	[0.3623]
	[-0.0271, 0.1210]	[-0.1368, 0.3219]	[-0.0222, 0.0783]	[-0.0777, 0.2087]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	46	157	62	165
Cross-fitting and resampling used	No	Yes	No	Yes
(3) Outcome variable X = SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Black (1) vs. White (0)	0.0073	-0.0365	-0.0005	-0.0473
	(0.0141)	(0.0743)	(0.0102)	(0.0480)
	[0.6057]	[0.6252]	[0.9613]	[0.3294]
	[-0.0210, 0.0356]	[-0.1860, 0.1129]	[-0.0210, 0.0200]	[-0.1440, 0.0493]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	42	97	47	106
Cross-fitting and resampling used	No	Yes	No	Yes
(4) Outcome variable X = SSI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Black (1) vs. White (0)	0.0512	0.0852	0.0280	0.0655
	(0.0455)	(0.1110)	(0.0250)	(0.0712)
	[0.2666]	[0.4467]	[0.2672]	[0.3626]
	[-0.0404, 0.1427]	[-0.1382, 0.3086]	[-0.0222, 0.0783]	[-0.0778, 0.2087]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	44	140	63	161

Cross-fitting and resampling used	No	Yes	No	Yes
(5) Outcome variable X = SSDI during 2006–2018 (1/0), administrative data (SSA records)				
Black (1) vs. White (0)	0.0168	0.0152	0.0009	0.0012
	(0.0187)	(0.0164)	(0.0154)	(0.0159)
	[0.3720]	[0.3575]	[0.9534]	[0.9424]
	[-0.0207, 0.0544]	[-0.0177, 0.0481]	[-0.0302, 0.0320]	[-0.0308, 0.0331]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	46	86	41	76
Cross-fitting and resampling used	No	Yes	No	Yes
(6) Outcome variable X = SSI during 2006–2018 (1/0), administrative data (SSA records)				
Black (1) vs. White (0)	0.0201	0.0068	0.0107	0.0073
	(0.0240)	(0.0137)	(0.0097)	(0.0079)
	[0.4069]	[0.6206]	[0.2727]	[0.3552]
	[-0.0282, 0.0684]	[-0.0208, 0.0345]	[-0.0087, 0.0301]	[-0.0085, 0.0231]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	40	84	38	75
Cross-fitting and resampling used	No	Yes	No	Yes

Notes: The number of observations for the outcome variables that use the self-reported data (HRS data) is 17,107. The number of observations for the outcome variables that use the administrative data (SSA records) is 11,853. The estimate of disparity is the weighted mean of the predicted disparity for each observation generated by the double/debiased ML estimator. The weight variable is the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018. Standard errors for the weighted means are clustered on state. For cross-fitting, the number of folds used is ten. For resampling, the number of resamples used is three. Reported in the table are weighted means, standard errors (in parentheses), p-values (in brackets), and the 95 percent confidence intervals. * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01.

The predictor variables include female (1/0) and the following control variables: linear terms of the continuous covariates, the discrete covariates, and state fixed effects; the quadratic terms of the continuous covariates; and the interaction terms between each state dummy variable and the continuous covariates. The continuous covariates and the discrete covariates are explained in the notes section of Appendix A.

Appendix G: Estimates of Gender Disparity Obtained by the Double/Debiased Machine Learning Estimator

	(1) Ever applied X	(2) Ever applied X	(3) Ever received X	(4) Ever received X
(1) Outcome variable X = SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Female (1) vs. Male (0)	-0.0099 **	-0.0103 **	-0.0031	-0.0031
	(0.0049)	(0.0049)	(0.0039)	(0.0040)
	[0.0481]	[0.0419]	[0.4390]	[0.4360]
	[-0.0196,-0.0001]	[-0.0202,-0.0004]	[-0.0110, 0.0048]	[-0.0111, 0.0049]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	28	91	27	101
Cross-fitting and resampling used	No	Yes	No	Yes
(2) Outcome variable X = SSI during 2006–2018 (1/0), self-reported (HRS data), definition 1				
Female (1) vs. Male (0)	-0.0020	-0.0020	0.0011	0.0011
	(0.0036)	(0.0036)	(0.0016)	(0.0017)
	[0.5855]	[0.5751]	[0.4875]	[0.5204]
	[-0.0092, 0.0053]	[-0.0092, 0.0052]	[-0.0022, 0.0044]	[-0.0023, 0.0044]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	35	145	56	173
Cross-fitting and resampling used	No	Yes	No	Yes
(3) Outcome variable X = SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Female (1) vs. Male (0)	-0.0120 **	-0.0123 **	-0.0032	-0.0032
	(0.0056)	(0.0057)	(0.0039)	(0.0040)
	[0.0385]	[0.0365]	[0.4288]	[0.4205]
	[-0.0233,-0.0007]	[-0.0238,-0.0008]	[-0.0111, 0.0048]	[-0.0112, 0.0048]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	27	94	27	101
Cross-fitting and resampling used	No	Yes	No	Yes
(4) Outcome variable X = SSI during 2006–2018 (1/0), self-reported (HRS data), definition 2				
Female (1) vs. Male (0)	-0.0043	-0.0042	0.0010	0.0011
	(0.0044)	(0.0044)	(0.0016)	(0.0017)
	[0.3255]	[0.3446]	[0.5309]	[0.5066]
	[-0.0131, 0.0044]	[-0.0130, 0.0046]	[-0.0023, 0.0043]	[-0.0022, 0.0045]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	42	149	55	164

Cross-fitting and resampling used	No	Yes	No	Yes
(5) Outcome variable X = SSDI during 2006–2018 (1/0), administrative data (SSA records)				
Female (1) vs. Male (0)	-0.0361 ***	-0.0363 ***	-0.0390 ***	-0.0383 ***
	(0.0085)	(0.0086)	(0.0073)	(0.0075)
	[0.0001]	[0.0001]	[0.0000]	[0.0000]
	[-0.0531,-0.0190]	[-0.0535,-0.0190]	[-0.0538,-0.0243]	[-0.0533,-0.0233]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	34	70	30	56
Cross-fitting and resampling used	No	Yes	No	Yes
(6) Outcome variable X = SSI during 2006–2018 (1/0), administrative data (SSA records)				
Female (1) vs. Male (0)	-0.0022	-0.0042	0.0014	0.0005
	(0.0059)	(0.0062)	(0.0038)	(0.0040)
	[0.7094]	[0.5018]	[0.7177]	[0.9013]
	[-0.0141, 0.0097]	[-0.0166, 0.0082]	[-0.0063, 0.0091]	[-0.0076, 0.0086]
Number of potential predictor variables	830	830	830	830
Number of selected predictor variables	28	58	23	41
Cross-fitting and resampling used	No	Yes	No	Yes

Notes: The same as Appendix F except that it is Black (1/0), not female (1/0), that is included as a predictor variable.

Appendix H: Estimates of Racial Disparity Obtained by the Machine Learning Estimator Using Alternative Specifications

Notes: See the notes section of Appendix I.

Appendix I: Estimates of Racial Disparity Obtained by the Machine Learning Estimator Using Alternative Specifications

Alternative specifications are applied to the least absolute shrinkage and selection operator (LASSO) used by the double/debiased machine learning estimator.
Outcome variable: ever applied for SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1

	(1)	(2)	(3)	(4)	(5)
Black (1) vs. White (0)	0.0078	0.0177	0.0566	0.0148	0.0589
	(0.0140)	(0.0175)	(0.0338)	(0.0138)	(0.0527)
	[0.5822]	[0.3167]	[0.1008]	[0.2890]	[0.2696]
	[-0.0204, 0.0359]	[-0.0175, 0.0530]	[-0.0114, 0.1246]	[-0.0130, 0.0427]	[-0.0471, 0.1648]
Number of observations	17,107	17,107	17,107	17,107	17,107
Number of potential predictor variables	830	830	830	830	830
Number of selected predictor variables	52	47	60	43	52
Method for determining the value of the LASSO penalty parameter (called lambda)	Plugin	BIC	CV	Adaptive CV	Plugin
Lambda for the Black (1/0) equation	0.0204	0.0204	0.0204	0.0204	0.0204
Lambda for the outcome equation for Black = 1	0.0810	0.0176	0.0122	0.0248	0.0809
Lambda for the outcome equation for Black = 0	0.0408	0.0046	0.0042	0.0133	0.0408
Specify variables that should always be included in the outcome equation	No	No	No	No	Yes

Notes: The BIC stands for Bayesian information criterion. The CV stands for cross-validation. The estimate of disparity is the weighted mean of the predicted disparity for each observation generated by the double/debiased ML estimator. The weight variable is the HRS's combined respondent weight and nursing home resident weight, averaged over 2006–2018. Standard errors for the weighted means are clustered on state. Cross-fitting and resampling are not used in these specification checks. Reported in the table are weighted means, standard errors (in parentheses), p-values (in brackets), and the 95 percent confidence intervals. * p-value < 0.10, ** p-value < 0.05, *** p-value < 0.01.

The list of predictor variables is explained in the notes section of Appendix F. In the case where we specify certain variables that should always be included in the outcome equation, these variables are: age, female (1/0), years of education, ever married (including partnered) (1/0), number of people living in the household, and household (HRS respondent and spouse) income.

Appendix J: Estimates of Gender Disparity Obtained by the Machine Learning Estimator Using Alternative Specifications

Notes: See the notes section of Appendix K.

Appendix K: Estimates of Gender Disparity Obtained by the Machine Learning Estimator Using Alternative Specifications

Alternative specifications are applied to the least absolute shrinkage and selection operator (LASSO) used by the double/debiased machine learning estimator.
Outcome variable: ever applied for SSDI during 2006–2018 (1/0), self-reported (HRS data), definition 1

	(1)	(2)	(3)	(4)	(5)
Female (1) vs. Male (0)	-0.0099 **	-0.0099 **	-0.0100 **	-0.0100 **	-0.0094 **
	(0.0049)	(0.0048)	(0.0049)	(0.0048)	(0.0044)
	[0.0481]	[0.0438]	[0.0464]	[0.0436]	[0.0388]
	[-0.0196,-0.0001]	[-0.0196,-0.0003]	[-0.0199,-0.0002]	[-0.0197,-0.0003]	[-0.0182,-0.0005]
Number of observations	17,107	17,107	17,107	17,107	17,107
Number of potential predictor variables	830	830	830	830	830
Number of selected predictor variables	28	36	55	45	29
Method for determining the value of the LASSO penalty parameter (called lambda)	Plugin	BIC	CV	Adaptive CV	Plugin
Lambda for the Female (1/0) equation	0.0204	0.0204	0.0204	0.0204	0.0204
Lambda for the outcome equation for Female = 1	0.0483	0.0046	0.0035	0.0237	0.0483
Lambda for the outcome equation for Female = 0	0.0555	0.0072	0.0050	0.0033	0.0555
Specify variables that should always be included in the outcome equation	No	No	No	No	Yes

Notes: The same as Appendix I except that: (a) it is Black (1/0), not female (1/0), that is included as a predictor variable; (b) in the list of variables that should always be included in the outcome equation, it is Black (1/0), not female (1/0).

Center for Financial Security

School of Human Ecology
University of Wisconsin-Madison

1300 Linden Drive
Madison, WI 53706

608-890-0229
cfs@mailplus.wisc.edu
cfs.wisc.edu