HOW MANY MEDICAID RECIPIENTS MIGHT BE ELIGIBLE FOR SSI?

Michael Levere and David Wittenburg

CRR WP 2023-20
November 2023

Center for Retirement Research at Boston College
Haley House
140 Commonwealth Avenue
Chestnut Hill, MA 02467
Tel: 617-552-1762 Fax: 617-552-0191
https://crr.bc.edu

About the Center for Retirement Research

The Center for Retirement Research at Boston College, part of a consortium that includes parallel centers at the National Bureau of Economic Research, the University of Michigan, and the University of Wisconsin-Madison, was established in 1998 through a grant from the Social Security Administration. The Center's mission is to produce first-class research and forge a strong link between the academic community and decision makers in the public and private sectors around an issue of critical importance to the nation's future. To achieve this mission, the Center conducts a wide variety of research projects, transmits new findings to a broad audience, trains new scholars, and broadens access to valuable data sources.

Abstract

Children's participation in the federal Supplemental Security Income (SSI) program has declined substantially over the past decade. Many children with disabilities might be eligible for SSI, yet barriers such as a lack of knowledge of the program or perceived challenges with applying may limit participation. In this paper, we use machine learning models on Medicaid administrative data to estimate the number and characteristics of children who are potentially eligible for SSI but do not currently receive benefits.

The paper found that:

- A substantial number of children are potentially eligible for SSI. Depending on the exact probability used to define potential eligibility, the increase could likely range from 10 percent to 55 percent increase in enrollment (relative to the current number of SSI recipients).
- Children potentially eligible for SSI have intensive health care usage, often more intensive than current child SSI recipients. This is particularly notable given the pattern that SSI recipients have much more intensive usage than non-SSI recipients, consistent with the requirement that SSI recipients must have a disability.

The policy implications of the findings are:

- Local-level estimates presented in this paper, which identify counties and states where many children are potentially eligible for SSI, might facilitate more effective outreach to families who are potentially eligible for SSI. Especially when supplemented with additional data on socioeconomic deprivation, more narrowly targeting outreach to areas where many children are likely to be eligible might allow for an efficient use of limited resources.
- Creating a direct link between Medicaid and the Social Security Administration's (SSA) data might improve the overall administration of the SSI program. For example, SSA could use the data here to identify children who might be eligible based on their pattern of claims and conduct outreach to them. Alternatively, the data might facilitate more streamlined disability determinations for new applications by allowing SSA to review recent patterns of health care claims.

Introduction

Take-up of social benefits programs is often incomplete, with numerous eligible people who do not participate (Currie 2006). A variety of barriers may prevent people from participating, such as stigma (Celhay, Meyer, and Mittag 2022), administrative burden (Herd et al. 2013), or lack of awareness (Chetty, Friedman, and Saez 2013), among others. High take-up rates mean that people eligible for benefits are accessing the supports they need from the program. For programs that are entirely income driven, like the Earned Income Tax Credit, it is straightforward to identify those who are eligible based on income data. Yet it is challenging to identify those who are eligible for Supplemental Security Income (SSI) or Social Security Disability Insurance (SSDI), which require a person to have a significant disability, because this disability criterion cannot be directly measured in readily available data.

Declines in participation in child SSI, which provides cash assistance to low-income families who have a child with a disability, suggest that many who are eligible may not be participating in the program. From 2013 to 2021, the number of child SSI recipients fell by nearly 20 percent. During the COVID-19 pandemic, applications declined dramatically. Yet this latter decline was not uniform across regions (Levere, Hemmeter, and Wittenburg 2023a), consistent with longstanding geographical variation in child SSI participation (Schmidt and Sevak 2017). The Social Security Administration (SSA), which administers SSI, has sought to reach vulnerable populations to support more equitable access, though it is difficult to determine how to target those efforts most efficiently. One critical step to make this determination is more fully understanding who might potentially be eligible for the program.

We use Medicaid data to estimate the number and characteristics of children who are potentially eligible but do not currently receive SSI benefits. We focus on Medicaid recipients for three reasons. First, the program is means-tested, so children who are already receiving benefits come from relatively low-income families and are thus presumably more likely than an average child to meet income and resource limits associated with SSI. Second, though we cannot directly observe whether someone has a disability who meets SSA's criteria, measures of health care utilization available in the data may suggest someone's likely disability status. Third, because Medicaid recipients are already participating in a government benefits program, various aspects related to the fragmentation of the safety net—such as stigma or limited knowledge— may be a relatively smaller barrier to participation for them (Michener 2018).

Using machine learning tools, we identified children who are potentially eligible for SSI based on an array of health care utilization measures in Medicaid's data. We limited our analysis to states where we could reliably infer whether someone is eligible for SSI based on administrative data from both SSA and the Centers for Medicare & Medicaid Services (CMS). We found 32 "high-match" states where the number of child SSI recipients in Medicaid data were sufficiently similar to SSA administrative records. To generate a probability of SSI eligibility for each child who was not receiving SSI, we estimated a separate random forest model in each of these 32 states in 2019. The model includes over 1,000 health care utilization measures. Our results did not change when we used data from earlier years (2017 and 2018).

A substantial number of children are potentially eligible for SSI. If we define potential eligibility as those with a probability of receipt of over 40 percent in our model, our estimates indicate that nearly 115,000 children might be potentially eligible for SSI (a 9.9 percent increase relative to current child SSI recipients). As we lower this probability threshold, we find more children who might be eligible (and as we increase the threshold, we find fewer who might be eligible). For example, with a threshold of 30 percent, the potentially eligible population would be over 260,000 (a 23 percent increase); with a threshold of 20 percent, the potentially eligible population would be nearly 650,000 (a 55 percent increase). The raw numbers reflect both the increase in the 32 high-match states from which we estimated the model and the increase in the remaining 18 states (and the District of Columbia) applying the same percentage increase to the current child SSI recipients from SSA administrative reports.

Children potentially eligible for SSI have intensive health care usage, often more intensive than current child SSI recipients. For example, in some states, those likeliest to be eligible for SSI had more than double the number of prescription drug claims as current SSI recipients, who already have substantially higher claims than the average non-SSI recipient. Such children also commonly have chronic conditions indicating that they have developmental delays. These developmental delays are quite rare among non-SSI recipients, but they are highly prevalent among children who are potentially eligible for SSI (up to nearly 75 percent of those with the highest probability of SSI receipt). Many claims and conditions exhibit a similar pattern, where SSI recipients have more intensive usage of care than non-SSI recipients, with differentially higher intensive usage of care the higher the probability of SSI receipt.

We then use these state-specific models to understand the extent to which cross-state variation in child SSI participation stems from policy factors as opposed to population factors. Using a set of four health care claims profiles, which represent typical claims for an existing SSI recipient with a given condition, we estimate the probability of SSI receipt in each state model. If population factors (like the underlying health conditions of children in a state) contributed to the cross-state variation in child SSI receipt, we would expect to find that people with the same condition would have similar predicted probability regardless of the state. Instead, we find that children with the same condition have notably different probabilities of SSI receipt depending on the state they live in. The cross-state variation is unrelated to the current level of child SSI participation. Together, these suggest that policy-specific factors likely play a critical role in explaining the substantial cross-state variation in current child SSI participation.

The findings contribute to a growing literature on leveraging big data and machine learning approaches to enhance the delivery of social programs. For example, Sansone and Zhu (2023) use administrative data on people who contribute to the Australian social security system to predict whether people will need income support. They find machine learning techniques can effectively identify people in need, allowing the government to potentially reach these at-risk individuals in a timely fashion. Heller et al. (2022) show that using machine learning on arrest and victimization data can accurately predict people's risk of being shot in Chicago. Using these predictions to target social services to prevent ensuing gun violence could have substantial economic benefits (in addition to improving public safety). Numerous other papers show the potential for using machine learning to improve policy related to education (e.g., Chalfin et al. [2016] on identifying effective teachers to promote), health (e.g., Hastings, Howison, and Inman [2020] on flagging opioid prescriptions that might be high risk for subsequent addictions), and tax collection (e.g., Battaglini et al. [2022] on effectively targeting tax audits to detect tax evasion). However, machine learning is not effective in all circumstances – for example, Bazzi et al. (2022) show that it is challenging to accurately predict local outbreaks of violence using detailed data from Colombia and Indonesia.

Our findings might also be especially useful for policymakers in exploring ways to promote higher take-up of programs, such as simplifying user experiences. A wide literature highlights that reducing administrative burden can promote participation in programs. For example, in the context of SSA disability programs, Deshpande and Li (2019) find that the

closure of SSA field offices reduced SSI and SSDI applications, both in the counties where the field office closed and in neighboring counties (because of increased congestion). In the context of Medicaid, Fox, Feng and Reynolds (2023) find that simplifying program rules to reduce the learning and psychological costs promotes Medicaid enrollment. Numerous other papers consistently find that reducing administrative burden in these and other programs (like the Supplemental Nutrition Assistance Program [SNAP]) promotes enrollment, while imposing additional administrative burdens reduces enrollment (e.g., Homonff and Somerville 2021, Heinrich et al. 2022, Herd et al. 2023, Giannella et al. 2024).

Linking programs may be an especially effective way to promote enrollment. Program recipients often receive benefits through multiple programs, with substantial overlap across SSI, Medicaid, SNAP, and others (Schmidt, Shore-Shepard, and Watson 2016). This in turn often requires families to navigate multiple complex processes to maintain eligibility and might lead people to fall through the cracks of these fragmented programs. Linking programs together may thus reduce administrative burden and help families better access the benefits to which they are eligible (Schmidt, Shore-Shepard, and Watson 2024). In the context of programs considered here, people who qualify for SSI automatically qualify for Medicaid in most states. Yet information is not shared in the opposite direction. Our findings highlight that many Medicaid recipients are likely eligible for SSI, showing that strengthening the program connections—such as by linking data across agencies—might help eligible children qualify for benefits.

Institutional Context

The SSI program, administered by SSA, requires recipients to meet specific disability and financial eligibility requirements. The disability criterion for children requires a "marked and severe functional limitation" resulting from a physical or mental impairment that significantly impacts the child's daily activities and is expected to last at least a year or lead to death. The eligibility requirements also include a limit on allowable assets and income. SSA manages a comprehensive eligibility determination process that involves conducting a thorough review of the child's medical history, daily functioning, and financial status. Once financial eligibility is confirmed, the state's Disability Determination Service evaluates the disability criterion by examining health provider information and gathering inputs from those involved in the child's daily life. Children who qualify for SSI are eligible for a cash payment and could qualify for

services from other programs. In 2023, the federal maximum payment from SSI is $914 per month.

The most common disorders for youth receiving SSI are autism spectrum disorders, developmental disorders, and other mental health disorders (which can frequently include attention deficit hyperactivity disorder [ADHD]). About 60 percent of child SSI recipients have one of these three diagnoses (SSA 2022). The potential health care needs of children likely vary depending on diagnosis. For instance, autism spectrum disorders might necessitate speech and language therapy, occupational therapy, behavioral therapy, and sometimes medications for associated symptoms. Developmental disorders may require similar treatments depending on the specific condition and could additionally require physical therapy or specialized educational support. Mental health conditions such as ADHD often involve a combination of medication, behavioral therapy, and ongoing counseling or psychotherapy. Other less common disorders, such as nervous system and sense organ disorders (7.0 percent of current child SSI recipients) and congenital anomalies (5.6 percent), likely require very different types of treatment to effectively manage. Our measures of health care utilization based on Medicaid claims, described below, capture a broad range of metrics that cover the diverse sets of health care needs that children with disabilities are likely to have.

In 2021, about 1 million children received SSI; however, this number has been declining since 2013 (Figure 1). The program generally experienced broad increases since its inception in 1974, with a large expansion in the early 1990s after the *Zebley* decision. As part of welfare reform in 1996, the disability criterion became more stringent, leading to a slight reduction in participation at that time. Although the rules have not changed since 1996, program participation continuously increased from 2000 to 2013, and then decreased since.

Figure 1. *Child SSI Participation, 1974-2021*
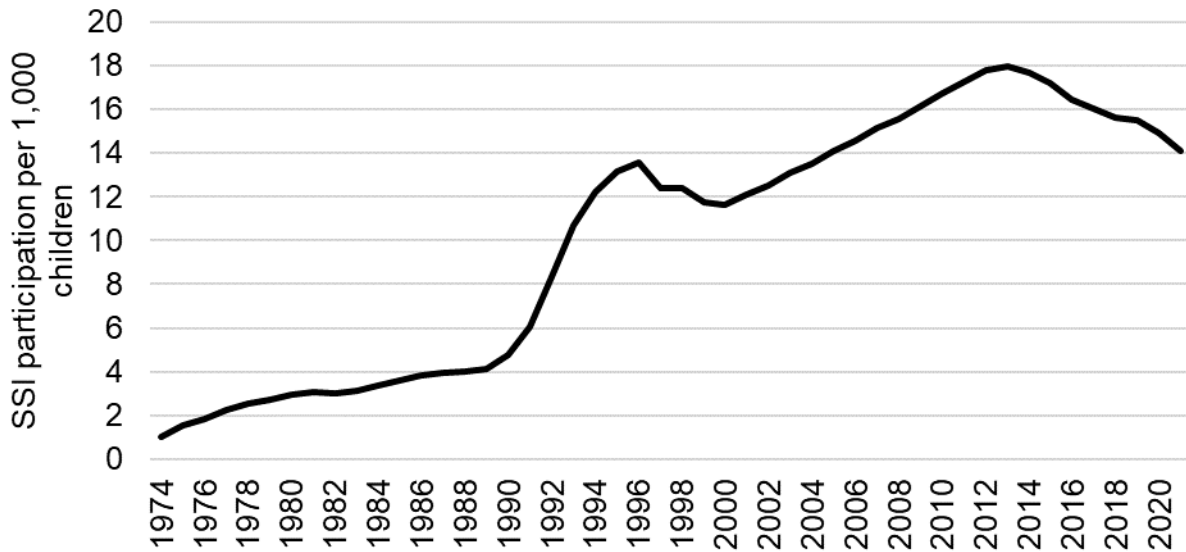
Source: SSA (2022).

The declining caseloads and state variations have prompted policy efforts to identify underserved youth. The Social Security Act authorizes SSA to collaborate with various entities to conduct outreach to potentially eligible populations. In response to the significant decrease in applications during the pandemic, SSA received increased funding in fiscal year 2021 to enhance outreach efforts targeting potential child SSI applicants (SSA 2021). These initiatives aim to address the challenges associated with declining participation.

In most states, children who receive SSI automatically qualify for health insurance coverage through Medicaid, though some states can have separate eligibility requirements. In 34 states and the District of Columbia, a newly awarded SSI recipient will also automatically be enrolled in Medicaid. However, nine states are known as 209(b) states, in which the Medicaid income criteria can be more stringent than the SSI criteria, meaning some children who receive SSI are not eligible for Medicaid. To qualify for Medicaid in these states, children (and families) must file a separate application. An additional seven states are considered SSI states, in which a newly awarded SSI recipient is automatically eligible for Medicaid, but qualifying also requires a separate application. Thus, in these 16 states where SSI receipt does not automatically lead to Medicaid receipt, some children might receive SSI but not Medicaid.

Medicaid is larger in scale than SSI, with eligibility primarily based on income. In December 2021, about 40 million children throughout the United States were enrolled in

Medicaid (Kaiser Family Foundation 2023). In contrast, only about 1 million children receive SSI. Medicaid eligibility for children is relatively generous, in part because of the Children's Health Insurance Program (CHIP). The latter, first established in 1997, led many states to substantially increase the income eligibility level and thus led to many more children qualifying for Medicaid (Cohen-Ross et al. 2009). Children make up nearly half of the total Medicaid and CHIP recipients. Though income criteria are typically more generous for Medicaid than for SSI, a relatively small share of people have incomes that would lead them to qualify for Medicaid but not SSI: a recent study by Levere et al. (2019) indicated substantial overlap in the income eligibility for children who receive Medicaid and SSI.

Both Medicaid and SSI have important local variations that influence program participation, and in turn influence our modeling approach. Though all states must follow certain federal guidelines in developing their Medicaid programs – such as requiring mandatory coverage for children in families with income below 138 percent of the federal poverty limit or covering certain mandatory services like hospital and physician care – each state operates its own program. States therefore differ in the extent to which certain populations or services are covered, and potentially within state if the Medicaid program has waiver approval to do so. Though SSI is a federal program, several states provide an optional supplemental payment to children with disabilities.[1] Child SSI participation also varies across counties and states, with much interest in the factors that drive these local differences (e.g., Schmidt and Sevak 2017; Levere, Wittenburg, and Hemmeter 2022). As discussed below, we therefore estimated a separate model predicting SSI eligibility among Medicaid recipients within each state.

Data

We used administrative data covering all Medicaid claims for children under age 18 among the universe of Medicaid beneficiaries. Specifically, we accessed the Transformed Medicaid Statistical Information System Analytic Files (TAF) through the Research Data Assistance Center (ResDAC). CMS compiles this database to facilitate research using administrative records of Medicaid eligibility and claims. We conducted our analyses at the

---

[1] According to the Policy Surveillance Program, 23 states provided an optional supplement through 2018 (the last date of the project update). Details on state supplemental payments for child and adult SSI recipients are at http://lawatlas.org/datasets/supplemental-security-income-for-children-with-disabilities (accessed June 13, 2023).

annual level focusing primarily on data from 2019, though results were nearly identical for 2017 and 2018.

Medicaid data include several variables intended to capture whether a child is receiving SSI benefits, which is a critical element of our analysis. These include monthly measures of the eligibility group code, an indicator for participation in SSI, and an SSI status code. The eligibility group code indicates the reason the person is eligible for Medicaid. Reasons of "Individuals receiving SSI," "Aged, blind, and disabled individuals in 209(b) states," or "Individuals receiving mandatory state supplements" indicate that the person is receiving SSI. The indicator for SSI participation is a zero or one variable, while if the SSI status code indicates the person is receiving SSI or is an SSI-eligible spouse, we consider them to be receiving SSI. We focused on eligibility in December only (as opposed to any time during the year) to match the way SSA reports on child SSI recipients in its Annual Statistical Report, discussed next (SSA 2022). We considered each of the three SSI variables within the Medicaid data separately, as well as whether any of the three variables indicate the person is receiving SSI. We then calculated the total number of child SSI recipients in each year in each state.

For each state, we compared the number of child SSI recipients from SSA administrative statistics to the number in Medicaid data, classifying states where these numbers were sufficiently close as "high-match" states. We calculated the ratio of Medicaid child SSI recipients to SSA-reported child SSI recipients separately in December 2017, 2018, and 2019 using each of the four separate approaches noted in the previous paragraph (eligibility group, SSI indicator, SSI status, or any of these three). To be considered a high-match state, this ratio for a single measure had to fall between 0.87 and 1.13 in all three years, indicating that the numbers were within 13 percent of each other.[2] For the "high-match" states, the eligibility group variable most commonly matched the SSA published statistics (27 of 32 states). One caution with this benchmarking exercise is that it only requires the aggregate number of recipients to be similar across the two data sources, though the actual children flagged as SSI recipients in Medicaid data might not be correct.

---

[2] CMS maintains a Data Quality Atlas to assess the reliability of certain measures with Medicaid data by comparing statistics from TAF to external data sources. It characterizes the quality as being low concern for a given state if two metrics are within 10 percent of each other. We loosened this criterion to 13 percent because of the requirement that it consistently be close enough for all three years. This indicates that the metric does not just capture the level of child SSI recipients correctly but also captures the evolving trend over time.

Thirty-two states have reliable metrics of child SSI participation and are considered "high-match" states, which we then used in our analysis (Figure 2). The percentage of states that are "high-match" differs substantially between those where new SSI recipients automatically receive Medicaid (solid fill; 74 percent) and those where new SSI recipients do not automatically receive Medicaid (striped fill; 38 percent). In these latter states, which are either SSI criteria states or 209(b) states, there may be SSI recipients who are not Medicaid recipients; some SSI recipients may therefore (correctly) not be in the Medicaid data. Thus, this difference by type of states is unsurprising. Because the "low-match" states do not have a reliable way to tell whether someone is currently receiving SSI, we could not use these states in our modeling procedure and thus omitted them from the analysis.

Next, we created extensive measures of health care utilization based on Medicaid eligibility and claims. We assessed all four primary types of claims available in TAF data: inpatient, long-term care, prescription drug, and other services. These other services include categories like physician services and outpatient hospital utilization. We then considered a host of characteristics about the claim, including: the taxonomy code for the provider who treated the patient;[3] the type of provider who treated the patient;[4] the type of services provided;[5] the benefit type code;[6] and the type of medications prescribed.[7] For each of these characteristics, we created variables for whether the child had each type of claim within the year, as well as the number of such claims to measure the intensity of the condition. Additionally, we identified several other characteristics from the claims and eligibility information, such as whether the pattern of claims indicates that the child has a range of comorbidities or diagnoses, such as learning disabilities, developmental delays, neurological disorders, and more. Finally, we also included several other measures for intensity of care, such as whether the child had any inpatient stays or emergency

---

[3] This can include a grouping like "Behavioral Health and Social Service Providers," or a classification under that grouping like "Clinical Neuropsychologist" or "Psychologist." In total, there are 29 unique groupings and 245 unique classifications.
[4] This can include providers like a "Physician" or "Speech Language Pathologist." In total, there are 57 unique provider types.
[5] This can include categories like "Physicians' services" or "Speech, hearing, and language disorders services (when not provided under home health services)." In total, there are 117 unique types of service.
[6] This can include categories like "Physicians' service" or "Physical Therapy and Related Services - Services for individuals with speech, hearing and language disorders." In total, there are 108 unique benefit type codes.
[7] We mapped each National Drug Code identifier, which is reported in the Medicaid data, to a unique set of 44 medication types, such as "ADHD Medications" or "Antidepressant medications."

department visits and the length of those encounters. In total, we considered approximately 1,300 variables related to health care utilization.

Figure 2. *States with Reliable SSI Indicator*

States with reliable SSI indicator

States with unreliable SSI indicator
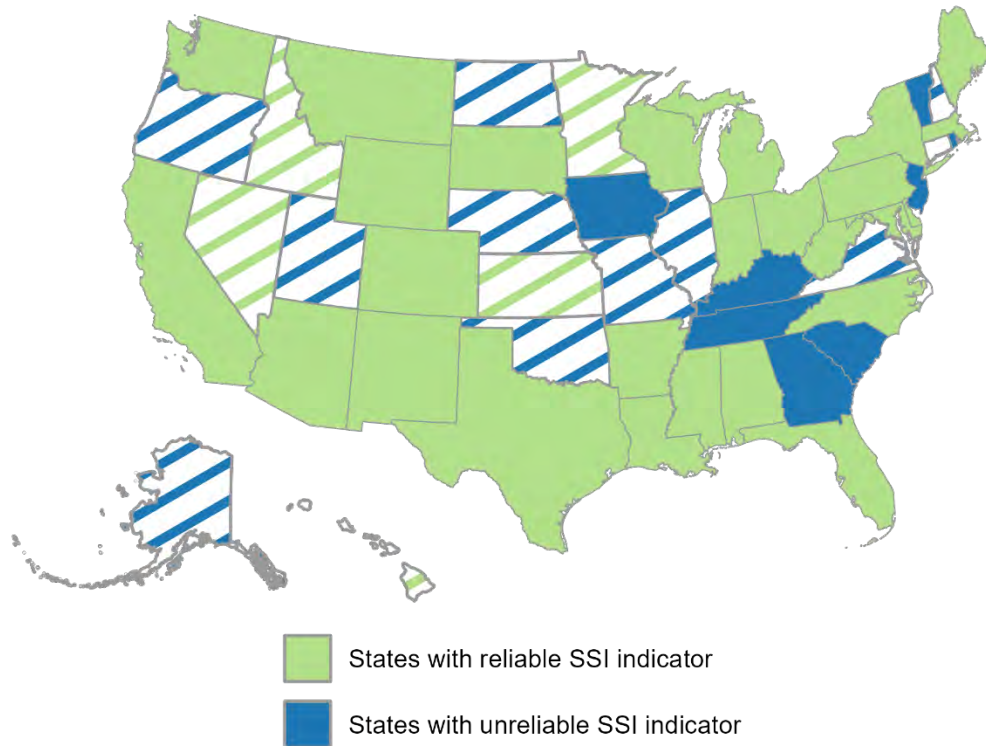
Note: Stripes indicate states in which new SSI awardees do not automatically receive Medicaid, because they are 209(b) states or SSI criteria states.
Source: Authors' calculations using 2017–2019 TAF data and SSI annual statistical report.

A descriptive comparison indicates that child SSI recipients have much more intensive health care utilization than child non-SSI recipients (Table 1). For example, the claims of child SSI recipients indicate that they are on average 8 times as likely as non-SSI recipients to have a learning disability chronic condition (33.2 percent versus 4.0 percent) and 13 times as likely to have another developmental delay chronic condition (18.4 percent versus 1.4 percent). Child SSI recipients are prescribed ADHD medications at a rate 6.5 times as frequently as non-SSI recipients. Though there are small differences between child SSI recipients and non-SSI recipients in having a claim where the type of service is either physician services or prescription drugs, the big difference is in the intensity of usage, as measured by the number of claims: the average child SSI recipient has 3 times as many physician services claims and 4 times as many

prescription drug claims. These differences reflect the key underlying factor that leads children to qualify for SSI, namely that they must have a significant disability. This disability in turn leads to intensive usage of health care.

Table 1. *Characteristics of Child Medicaid Beneficiaries, by Receipt of SSI*

Characteristic	SSI recipients mean (percentage unless otherwise noted)	Non-SSI recipients mean (percentage unless otherwise noted)
Learning disabilities chronic condition	33.2	4.0
Other developmental delays chronic condition	18.4	1.4
Prescribed ADHD medications	28.3	4.3
Has claim with Speech-Language pathologist	7.5	1.1
Has claim with Local Education Agency	16.4	1.6
Has physician services claim	77.2	62.1
Number of physician services claims	9.83	3.32
Has prescription drugs claim	78.1	54.9
Number of prescription drug claims	15.10	3.65
Total population size	894,687	29,141,512

Note: Includes all child Medicaid recipients within the 32 "high-match" states.
Source: 2019 TAF data.

We also included measures of sociodemographic characteristics that are available in the Medicaid data. These include age, race/ethnicity, and sex, as well as income.[8] Income might be especially important given that the income cutoffs are relatively higher for Medicaid than for SSI. Some Medicaid recipients may therefore have health care utilization suggesting they could be eligible for SSI, yet they may not have sufficiently low family income or resources to qualify. Though income data are not available in many states, we present supplemental analyses from Massachusetts and Colorado (which have reliable family income data) to show that our results were similar when we considered several variations to exclude or include income in the model to generate predicted probabilities. In particular, we did two things: (1) assess how many of the same children are above each probability threshold when estimating models that include and exclude income; and (2) assess what share of children flagged as potentially eligible have family

---

[8] Not all of these variables are reliably available for all states; see more information at the Data Quality Atlas at https://www.medicaid.gov/dq-atlas/

income that is above 255 percent of the federal poverty limit, in models that both include and exclude income.

Finally, we supplemented these administrative data with measures of socioeconomic characteristics at the zip code level available from the American Community Survey. In prior work, we found that a measure of socioeconomic deprivation at the local level is highly correlated with child SSI participation (Levere, Wittenburg, Hemmeter 2022). We therefore controlled for all the input characteristics that were included in the calculation of socioeconomic deprivation (which was in turn based on the Area Deprivation Index; see Singh [2003] for more details); these characteristics are all listed in Appendix Table 1.

Methodology

The primary goal of our analysis was to identify the probability that each child Medicaid recipient who is not receiving SSI is, in fact, eligible. With this probability, we can estimate how many children are potentially eligible. We considered a variety of different probability thresholds to determine potential eligibility.

To estimate this probability of SSI eligibility, we used machine learning techniques that algorithmically identify the characteristics most predictive of SSI receipt based on current SSI recipients. Our primary approach is a random forest model (Breiman 2001). The random forest model offers many advantages in terms of flexibly identifying characteristics that are important predictors without overfitting (Mullainathan and Spiess 2017). It creates decision trees by using a random set of input variables to partition the original data into groups that classify the object of interest, which in our setting is whether the child is an SSI recipient. It then creates many such trees, averaging across the various classifications from each tree to estimate a probability that the child is eligible for SSI. This procedure essentially identifies children as being potentially eligible for SSI if they have health care utilization similar to that of children who are currently receiving SSI. To avoid overfitting the actual data, we left out a testing sample of at least 20 percent of child Medicaid recipients in each state who were not used in training the model.

Because data availability and the way that states process health care claims vary across states, we estimated a separate model for each of the 32 "high-match" states. For example, we developed a model using all children in Arkansas to estimate the probability that each child in Arkansas is eligible for SSI (leaving out 20 percent of children as a testing sample). We then

repeated this process for each of the other 31 states.[9] We used the same exact approach in terms of specifying the random forest model, such as hyperparameters and input variables. However, the model may select different characteristics as relatively more or less important in estimating the probability of SSI receipt in each state. This is particularly important, given that (1) Medicaid is inherently a state-specific program and may have different procedures for processing and characterizing claims, and (2) the reliability of certain data characteristics like income may differ across states.

Our predictive models therefore account for the existing interplay of SSI and Medicaid within each state. For example, lower SSI participation in certain states could relate to stringent disability criteria or to general social factors. If low SSI participation relates to stringent disability criteria, leading only children with the most severe disabilities to qualify for benefits, then the potentially eligible population will likely be smaller too as it would only include children with the most severe disabilities. If, instead, low SSI participation is unrelated to the extent of health care utilization, the size of the potentially eligible population might not depend on the current level of SSI participation. A naïve approach that attempted to model these differences across all states could bias results.

Though results are available for all 32 states, the results in this paper cover four states for simplicity of presentation: Arkansas, Louisiana, Massachusetts, and Colorado. These four states cover a range of existing child SSI participation per capita: Arkansas and Louisiana have the two highest rates of child SSI participation among "high-match" states (34.69 and 29.26 per 1,000 children, respectively); Massachusetts has roughly the median rate of child SSI participation (15.72 per 1,000 children); and Colorado has very low SSI participation (6.74 per 1,000 children).[10] Additionally, Massachusetts and Colorado have reliable income data, which allow us to explore the sensitivity of our findings to including income in the model, as discussed previously.

---

[9] To ensure tractability of the model, we could not include more than approximately 1.5 million children in the training sample. So, for states with more than 1.875 million Medicaid recipients (CA, FL, NY, and TX), we randomly sampled 1.5 million children to include in the training sample, leaving the remaining group as the testing sample. We applied the model to calculate the probability of SSI receipt among all Medicaid recipients in the state.
[10] Only Wyoming (6.72 per 1,000 children) and Hawaii (4.05 per 1,000 children) have lower rates of child SSI participation among "high-match" states. Yet because these states are also much lower in population than Colorado, and thus have fewer potential SSI recipients, they are subject to issues related to small sample sizes. We therefore prefer to present results for Colorado as the representative low participation state.

To make the estimation more tractable, we excluded rare health care utilization measures from the model that do not substantively differ between SSI recipients and non-SSI recipients. For indicators on types of health care utilization, we excluded characteristics that met two criteria: (1) fewer than 1 percent of SSI recipients and fewer than 1 percent of non-SSI recipients each have that type of claim, and (2) the standardized difference in the mean[11] between SSI recipients and non-SSI recipients is less than 2 standard deviations. If we exclude the indicator for any utilization, we then also exclude the continuous measure for number of claims of that type.[12] The exclusion is based on pooled data across all 32 "high-match" states. Given that these characteristics are extremely rare and not extensively different between SSI and non-SSI recipients, these characteristics are unlikely to meaningfully affect the estimated probabilities. In total we excluded 760 variables, such as claims where the service provider taxonomy group is either "Podiatric medicine and surgery service providers" or "Dietary and nutritional service providers."

Across all states, the random forest model appears to generate reasonable and reliable estimates of the probability of SSI eligibility (Figure 3). Each bar in Figure 3 represents the group of child Medicaid recipients who fall into a bucket covering a range of 5 percent probability (e.g., the first bar on the left of each graph represents recipients with a 0 to 5 percent estimated probability). The bar then shows the percentage of children in that range who are on SSI (in light gray) and not on SSI (in dark gray). In all states, approximately 0 percent of those with the lowest probability are receiving SSI, as expected. About 90 percent of all non-SSI recipients have predicted probability less than 5 percent in Colorado, Louisiana, and Massachusetts (in Arkansas, only about 80 percent do). As the predicted probability of SSI receipt increases, so does the share of children who are actually receiving SSI. For example, about 20 percent of those with predicted probability of SSI receipt in the range of 20 to 25 percent receive SSI. Though there are very few (if any) children with very high probabilities, the general contours of this figure suggest that the model picks up important predictive information based on the health care utilization characteristics. This pattern is also not the result of the model

---

[11] To calculate the standardized difference, we calculate an effect size that divides the log odds ratio between SSI recipients and non-SSI recipients by 1.65.

[12] There are a few exceptions to this: emergency department visits, emergency department visits leading to inpatient stays, inpatient stays, nursing facility stays, and behavioral health treatment services. In these instances, we could drop the indicator if it did not meet the criteria, but its continuous counterpart would remain.

overfitting: the pattern is nearly identical using only the testing sample that was left out when estimating the model (Appendix Figure 1).

Figure 3. *Distribution of SSI Receipt by Predicted Probability of SSI Receipt*
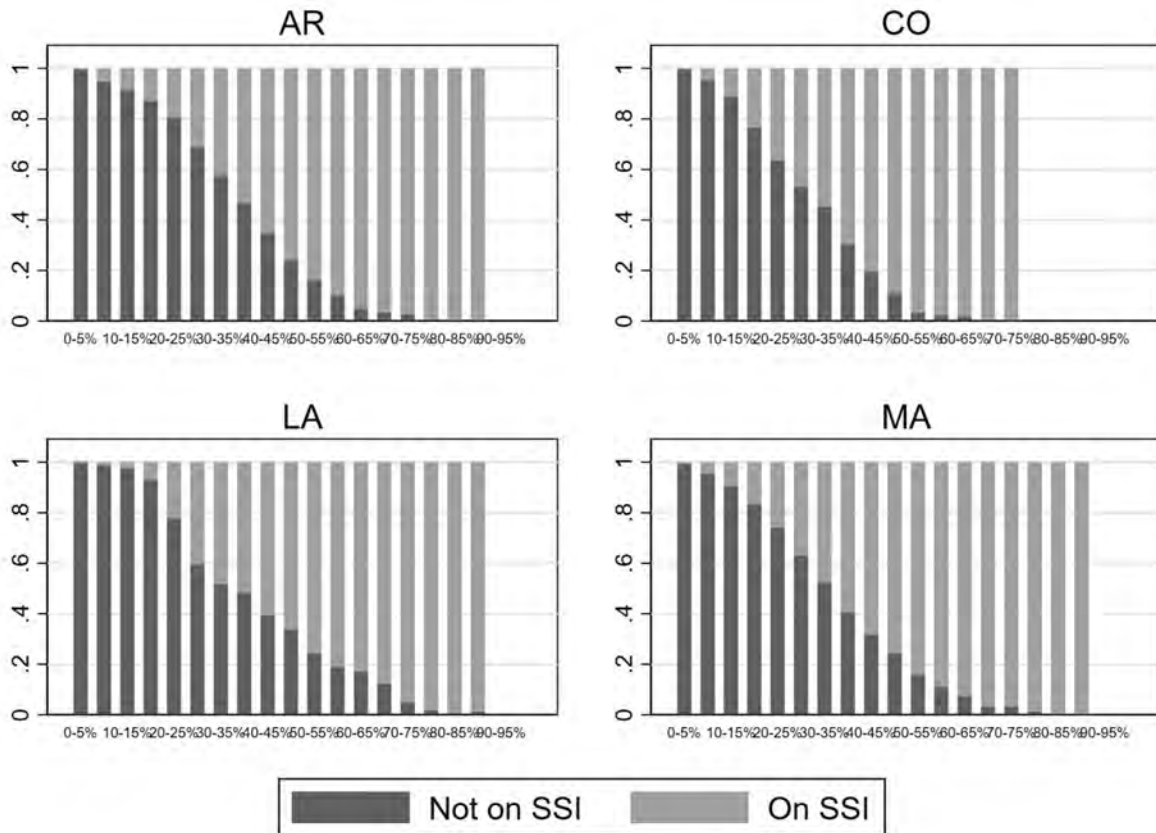
Notes: Indicates the share of children receiving SSI for each ventile of predicted SSI probability. If no children in a state have a probability sufficiently high, that bar is excluded from the figure (e.g., nobody in Colorado has a predicted probability above 80 percent).
Source: Authors' calculations using 2019 TAF data.

Results

We present three primary sets of results. First, we provide an estimate of the number of potentially eligible child SSI recipients, both as a raw number and as a percentage of current child SSI recipients. Second, we describe the characteristics of those who are potentially eligible, comparing these characteristics to child SSI recipients and to child non-SSI recipients. In each of these analyses, we show how results differ when we change the probability threshold

for what constitutes a potential child SSI recipient. Finally, we also test the sensitivity of our results to the inclusion of income in the model.[13]

Number of Potentially Eligible Child SSI Recipients

A substantial number of children might be eligible for SSI based on their health care utilization but do not currently receive benefits (Table 2). For example, if everyone in the 32 "high-match" states with a predicted probability above 40 percent were to qualify for SSI, this would increase SSI participation by about 90,000, or 9.9 percent relative to the number of current child SSI recipients identified in the Medicaid data. Applying this percentage to the number of child SSI recipients from SSA administrative data in 2019, about 25,000 more children could be eligible in the "low-match" states that are excluded from our analysis. In total, with a probability threshold of 40 percent, there might be nearly 115,000 children eligible for SSI. Instead using a more lenient probability threshold of 20 percent, the total number of potentially eligible children would be nearly 650,000. The numbers and shares are roughly similar if we use data from 2017 or 2018 (Appendix Tables 2 and 3).

We focus on probabilities within this range based on the similarity between health care claims of child non-SSI recipients with such probabilities and those of child SSI recipients. As shown below, even those with predicted probabilities of at least 20 percent tend to have health care utilization that is similar to or slightly more intensive than that of the average SSI recipient. For those with predicted probability over 40 percent, the utilization is that much higher. Empirically, because SSI participation is so infrequent (overall, about 3 percent of child Medicaid recipients receive SSI; see Table 1), the model rarely produces high probabilities – for SSI recipients or non-SSI recipients. For example, Figure 3 shows that nobody in the four states has a predicted probability above 90 percent. Thus, even a "low" probability in the range of 30 percent might be thought of as corresponding to a high likelihood of SSI eligibility.

---

[13] We also considered a fourth analysis that would have extrapolated results from the state-specific models to a national estimate of the potentially eligible population given each state model. Ultimately, the assumptions and results from the model were not sufficiently reliable to include. For more detail, see Appendix A.

Table 2. *Potentially Eligible Child SSI Recipients Above Each Probability Threshold*

Predicted probability threshold	Number of children in "high-match" states	As percentage of current child SSI recipients	Number of children in "low-match" states applying same percentage	Total number of potentially eligible children in United States
10 percent	1,366,028	152.7%	400,079	1,766,107
15 percent	790,356	88.3%	231,478	1,021,834
20 percent	493,640	55.2%	144,576	638,216
25 percent	315,934	35.3%	92,530	408,464
30 percent	204,334	22.8%	59,845	264,179
35 percent	134,299	15.0%	39,333	173,632
40 percent	88,849	9.9%	26,022	114,871
45 percent	57,702	6.4%	16,900	74,602
50 percent	35,516	4.0%	10,402	45,918

Notes: The first column reports the total number of children across high-match states who are not receiving SSI but who have a predicted probability above the threshold. The second column expresses this as a percentage of the 894,687 total child SSI recipients across these 32 states within the Medicaid data. The third column multiplies this percentage by the 262,034 child SSI recipients in 2019 in the 18 low-match states and the District of Columbia from the SSI recipients by state and county report. Finally, the fourth column sums the first and the third columns to indicate the total number of potentially eligible children across the United States.
Source: Authors' calculations using 2019 TAF data and 2019 SSI recipients by state and county.

The share of children likely to be eligible as a percentage of current child SSI recipients varies by state, with more potentially eligible children in the southern states that have high SSI participation (Figure 4). The map in Figure 4 calculates the number of child non-SSI recipients exceeding the 40 percent probability threshold in each state as a percentage of current child SSI recipients (referred to subsequently as the "percentage increase"). Maps considering other thresholds (10 percent, 20 percent, 30 percent, and 50 percent) are available in Appendix Figure 2, while the number of children potentially eligible at each threshold in each state is available in Appendix Table 4. The states with the highest percentage increases given the 40 percent threshold are California (16.2 percent), Louisiana (13.6 percent), Texas (11.6 percent), and Florida (11.4 percent). States with higher SSI participation per capita also tend to have larger percentage increases: a regression of the percentage increase on SSI participation per capita in the state is significant and positive at the 5 percent level for every probability threshold except for 10 percent.

Figure 4. *Child Non-SSI Recipients with Probability of Receipt at Least 40 Percent, as Percentage of Current SSI Recipients*
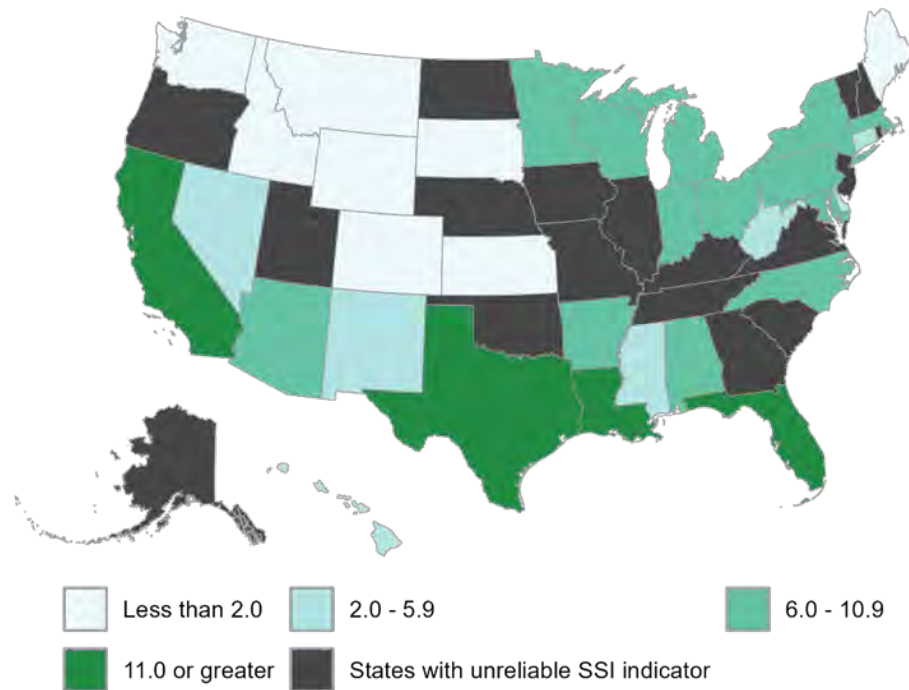
Source: Authors' calculations using 2019 TAF data.

Our approach can also generate estimates of the share of children potentially eligible at the county level, which might be especially helpful for targeting outreach in specific geographic areas (Figure 5). Within states, the percentage increase often varies across county. These county-level statistics might be helpful to policymakers in considering where to target outreach efforts, potentially leveraging local networks. For example, local networks such as schools are important ways that children and families learn about SSI (Levere, Hemmeter, Wittenburg 2023b). SSA has recently deployed Vulnerable Population Liaisons who seek to help potentially eligible people within highly localized areas to apply for SSI. These sorts of local-level statistics can help ensure that resources are targeted to achieve the greatest impact, given a larger potentially eligible population.

Figure 5. *Child Non-SSI Recipients with Probability of Receipt at Least 40 Percent, as Percentage of Current SSI Recipients, County-Level*
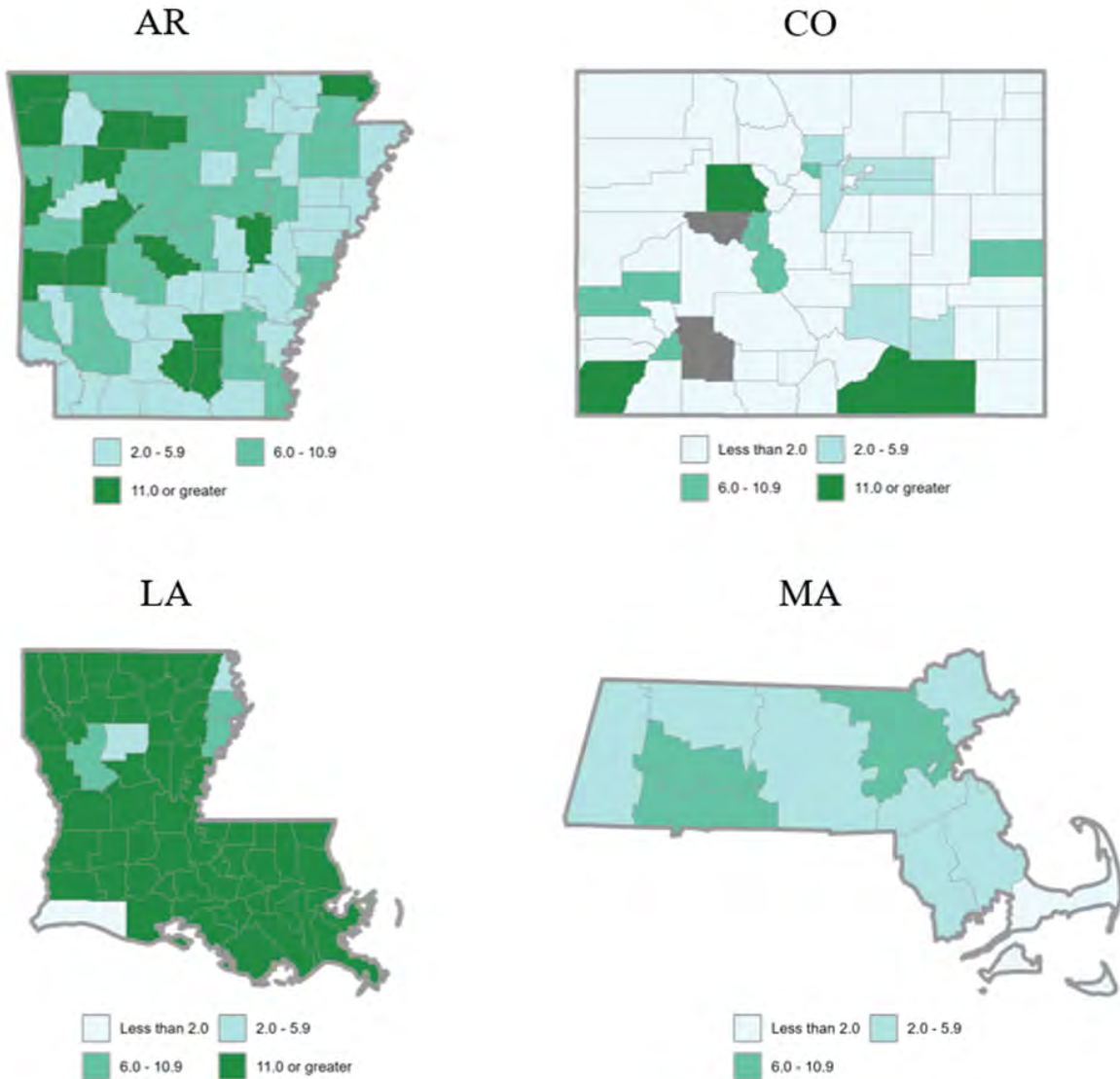
Source: Authors' calculations using 2019 TAF data.

Health Care Utilization of Potentially Eligible Child SSI Recipients

We next summarize characteristics among the group of children who exceed a given probability of SSI receipt, comparing them to children currently receiving SSI and those not receiving SSI. The structure of these figures (such as Figure 6) is as follows: the black solid line represents the average value for all child SSI recipients. The red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all child

non-SSI recipients in the state with probability at least that high. Note that at higher

probabilities, the size of the non-SSI group with sufficiently high probability gets smaller.

Figure 6. *Number of Prescription Drug Claims, by State, Receipt of SSI, and Estimated Probability of SSI Receipt*
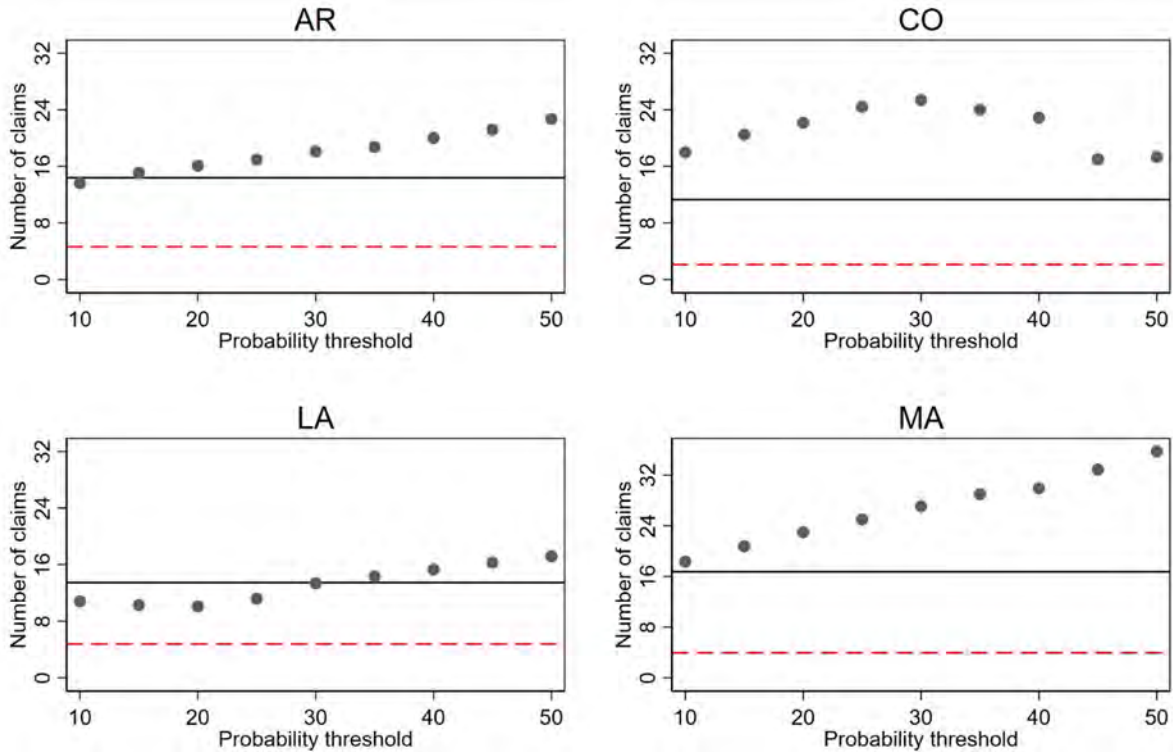
Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
*Source:* Authors' calculations using 2019 TAF data.

Health care utilization for children with very high predicted probability of SSI receipt is

very intensive, and often more intensive than that for average child SSI recipients (Figure 6).

For example, in Arkansas, the average child SSI recipient had 14.4 prescription drug claims in

2019, while the average non-SSI recipient had 4.6 such claims. As children's probability of SSI

receipt increases, so do their average prescription drug claims. For those with at least a 10

percent probability of SSI receipt, the average number of prescription drug claims is 13.6 (95

percent of the SSI recipient mean). Meanwhile, for those with at least a 50 percent probability,

the average number of prescription drug claims is 22.7 (67 percent higher, or 157 percent of the

SSI recipient mean). Patterns for prescription drug claims in other states are mostly similar, with

more such claims among SSI recipients than non-SSI recipients and differentially more intensive prescription drug claims with higher probability of SSI receipt. In Massachusetts, children with probability over 50 percent have more than double the average prescription drug claims of child SSI recipients in the state.

Many of the children not currently receiving SSI with the highest probability have claims that signify they have a developmental delays chronic condition (Figure 7). This condition is quite rare among child non-SSI recipients, with fewer than 5 percent having it across the four states. Yet it is highly prevalent among those potentially eligible for SSI – for children with the highest probability of SSI receipt, more than half in Arkansas and two-thirds to three-quarters in Colorado and Massachusetts have a developmental delays chronic condition. Louisiana follows a different pattern, with this condition not appearing to be especially predictive of SSI receipt.

Figure 7. *Presence of Other Developmental Delays Chronic Condition, by State, Receipt of SSI, and Estimated Probability of SSI Receipt*
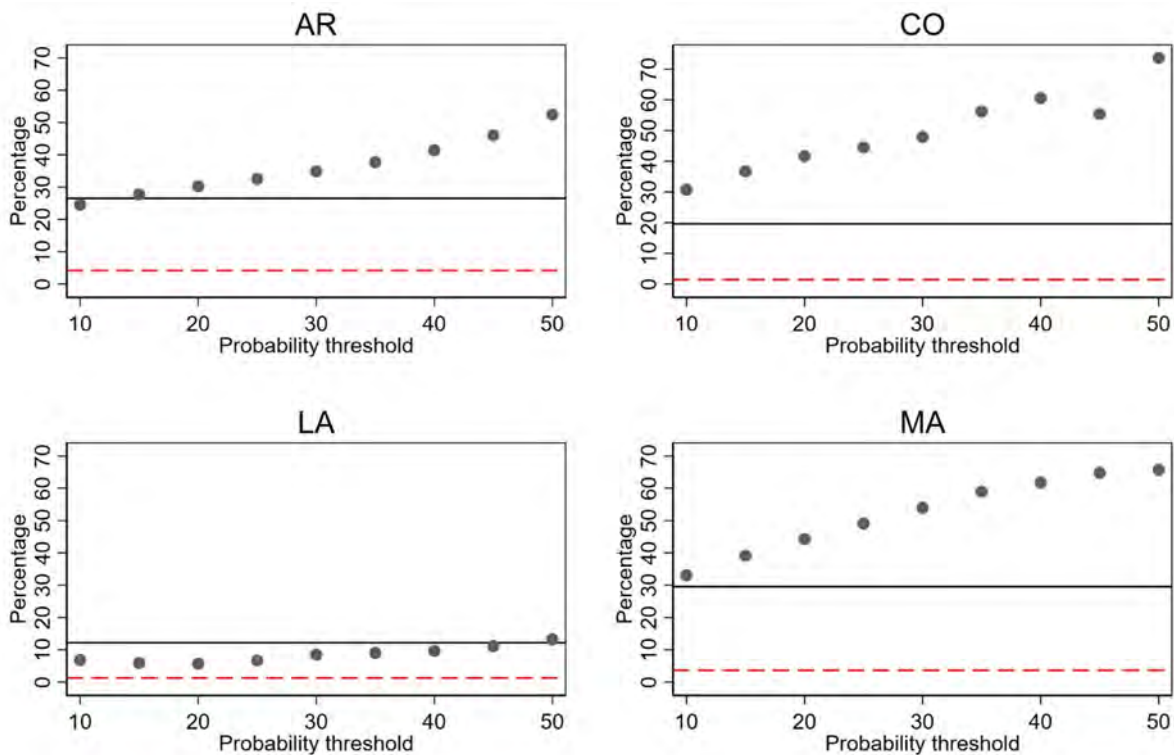
Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
Source: Authors' calculations using 2019 TAF data.

The pattern of results for a given characteristic often differs across states, reinforcing findings related to heterogeneous local patterns in SSI participation. For example, the pattern of being prescribed medication for ADHD in Arkansas and Louisiana mostly follows that of prescription drug claims: children with the highest probability of SSI receipt have higher rates of ADHD prescriptions, with the prevalence even higher than that of current child SSI recipients (Appendix Figure 3). Yet in Colorado, the pattern is the opposite, as the higher the probability of SSI receipt, the lower the rate of being prescribed ADHD medications. Claims with a speech-language pathologist appear to be very predictive of SSI receipt in Arkansas: about 66 percent of those likeliest to be eligible for SSI had such a claim, nearly double the rate among current SSI recipients (Appendix Figure 4). Yet in Massachusetts and Louisiana, almost nobody has such a claim, meaning it is not an effective way to identify potentially eligible SSI recipients. These patterns reiterate the importance of taking a state-specific modeling approach, given that states can differ extensively in how they characterize certain types of claims. They are also broadly consistent with findings showing local variation in SSI eligibility and participation (e.g., Levere, Wittenburg, Hemmeter 2022). Patterns for other types of claims and conditions are available upon request, but are omitted for space constraints.[14]

Importance of Income as a Predictive Variable

Our results are similar when including or excluding family income in the predictive model (Table 3). Because many states do not have reliable income data, a concern is that if income were included in the model our findings would differ, or that the model may be identifying children who are in fact ineligible for SSI because of high family income. However, when we estimate models both including and excluding income in Colorado and Massachusetts, both of which have reliable income data, we find substantial overlap in the Medicaid beneficiaries identified as potentially eligible for SSI. For example, our main model including income yields 1,507 children in Massachusetts with a predicted probability exceeding 30 percent. Of this group, Table 3 shows that 75.8 percent of them (998) would still have predicted probability exceeding 30 percent if we re-ran the model with all the same characteristics but excluded income.

---

[14] For the four patterns of claims presented in the paper and appendix, we also present analogous graphs in Appendix Figures 5 through 8 for four high population states: California, New York, Pennsylvania, and Texas.

22

The models also do not identify children with high family income as potentially eligible for SSI, even when income is not included in the predictive model (Appendix Table 5). In Colorado, at least 98 percent of the Medicaid beneficiaries identified as potentially eligible at all probability thresholds between 10 and 50 percent had family income under 255 percent of the federal poverty limit. This is sufficiently low that a family might be likely to qualify for SSI: the income threshold to receive any SSI benefits is 235 percent of the federal poverty limit for families with one child and two parents and only earned income (Levere et al. 2019). In the model that excluded income, the corresponding number was at least 96 percent. In Massachusetts, over 90 percent of those identified as potentially eligible were on the lower end of the income distribution.

Table 3. *Overlap in Potentially Eligible Child SSI Beneficiaries between Models that Include and Exclude Income*

Predicted probability threshold	Colorado	Massachusetts
20 percent	79.4	84.1
25 percent	78.6	79.7
30 percent	75.1	75.8
35 percent	68.4	70.3
40 percent	64.1	66.2
45 percent	49.2	60.2
50 percent	47.4	60.0

Notes: The numbers in the table show the percentage of child Medicaid beneficiaries who have a probability that exceeds the predicted probability threshold in the random forest model that includes income as a predictor who also have a probability exceeding the same threshold in the random forest model that does not include income as a predictor.
Source: Authors' calculations using 2019 TAF data.

Understanding Cross-State Variation in Child SSI Participation

Numerous papers highlight that child SSI participation varies substantially across local areas (e.g., Schmidt and Sevak 2017; Levere, Hemmeter, and Wittenburg 2023a). Variation may stem from policy, program, and administrative factors. Such factors include the availability of other cash supports or the generosity of state supplemental payments or administrative differences across local Disability Determination Service (e.g., differences in acceptance rates in initial applications). Variation may also stem from population factors, such as if characteristics

of the population differ, especially in the prevalence of conditions that would lead children to meet the definition of disability.

To isolate the role of population factors from the role of policy, program and administrative factors, we used our state-specific models to generate the predicted probability of SSI receipt for several fixed health care claims profiles. Specifically, given a set of health care claims behavior associated with each of four conditions—ADHD, behavioral disorders, other developmental disorders, and cerebral palsy—we estimate the probability of SSI receipt in each of the 32 states with a state-specific model.[15] To create these health care claims profiles, we used data from North Carolina. We chose North Carolina because it was the largest state that had a high percentage of children who were not covered by managed care plans.[16] A critical part of the analysis was identifying the cost of claims to distinguish high-cost and low-cost Medicaid beneficiaries.[17] Because it is challenging to identify the cost of claims with managed care plans, for the purposes of estimating cost we dropped people who were enrolled in managed care plans. We then found the typical claims behavior among child SSI recipients at the 25th percentile of cost among those with the given condition.[18]

Children with the same condition and health claims experience very different probability of SSI receipt across states (Figure 8). For example, the predicted probability of SSI receipt for a child with other developmental disabilities ranges from 2.5 percent in New Mexico to 42.7 percent in Montana. For children with cerebral palsy, predicted probabilities range from 10.5 percent in Washington to 56.4 percent in West Virginia. The probability of receipt also varies within states across conditions, as can be seen by comparing the same state across the four panels. However, the likelihood of receipt is typically correlated within state across conditions –

---

[15] Given the differing availability of income data, we only used models that did not include the income variables as predictors. As discussed earlier, the inclusion of these income variables does not meaningfully impact the results.
[16] North Carolina has similar patterns of utilization among SSI recipients, non-SSI recipients, and potentially eligible SSI recipients as the other states presented in Figure 6, Figure 7, Appendix Figure 3, and Appendix Figure 4 – see Appendix Figure 9.
[17] Appendix Figure 10 shows the total claim payment amounts during 2019 for each Medicaid beneficiary in North Carolina. Similar to the findings presented in earlier parts of the paper, health care utilization is more intensive for SSI recipients than non-SSI recipients, with even more intensive use of care among those with high predicted probabilities of receipt. For example, claims amounts for SSI recipients were nearly seven times higher for SSI recipients ($14,488) than for non-SSI recipients ($2,107). For children with predicted probability over 50 percent, the average claims amount was $66,894, or more than 4.5 times as high as the average SSI recipient.
[18] We also created high-cost profiles using the claims behavior among child SSI recipients at the 75th percentile of cost among those with the given condition. The patterns contrasting the high and low-cost claims for ADHD, other developmental disabilities, and cerebral palsy are available in Appendix Figure 11. There was not enough data to identify the 75th percentile of claims for children with behavioral disorders.

24

the six possible two-way correlation coefficients across these claims profiles ranges from 0.69 to 0.82.

The predicted probability of SSI receipt does not systematically vary with the current level of SSI participation (Figure 8). Louisiana and Arkansas, which have high current SSI receipt, have low predicted probabilities across all conditions. For the four fitted regression lines shown in the figure, the estimated slopes are all close to zero and have *p*-values exceeding 0.50. The highest R^2 value for any of these four relationships is 0.011, signifying that the predicted the current level of child SSI participation cannot explain the predicted probability of SSI receipt.

Figure 8. *Predicted Probability of SSI Receipt, by State and Condition*
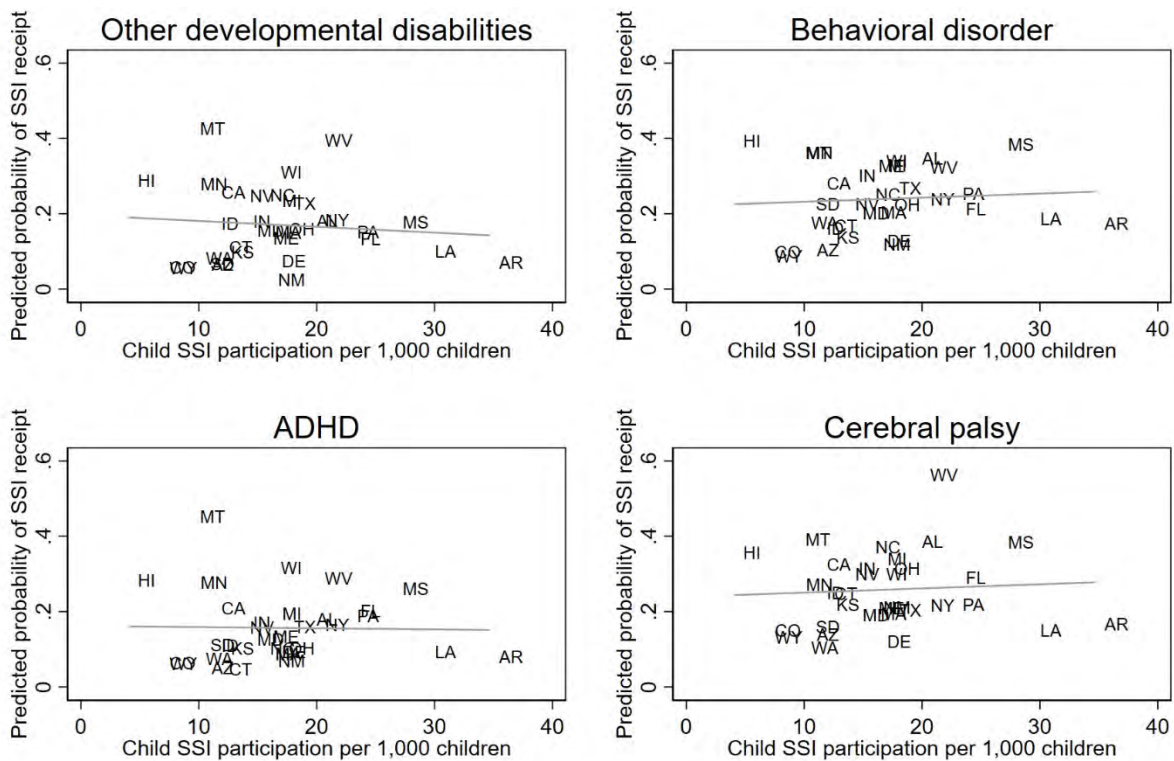
Notes: Estimates the probability of SSI receipt using the state specific model, holding fixed the health care claims associated with each condition. The gray line represents a line of best fit based on a linear regression. See the text for a description of how we identified the health care claims associated with each condition.
Source: Authors' calculations using 2019 TAF data.

Taken together, these findings indicate that variation in current child SSI participation is likely driven by a combination of local program, policy, and administrative factors rather than population factors. Children with the same condition have very different probabilities of SSI

receipt depending on the state they live in. This probability is not necessarily higher in states where many children already receive SSI. Other state and local-level policy factors in turn must contribute to the varying receipt of SSI across localities. However, identifying those specific policy-level factors is beyond the scope of this paper.

Conclusion

We used machine learning tools to identify children potentially eligible for SSI based on their patterns of health care utilization. Many children not currently receiving SSI might be eligible. Depending on the probability threshold that is considered to represent potential eligibility, the increase relative to current SSI recipients could range from about 10 percent (40 percent threshold) to 55 percent (20 percent threshold). Those potentially eligible SSI recipients often have very intensive usage of health care, frequently exceeding those of current SSI recipients.

Given this, the critical question is how these findings can ultimately inform policy. An important first step might be data linkages across organizations, such as CMS and SSA. For example, if SSA could observe health care claims and essentially replicate the analysis done here with a consistently reliable measure of SSI participation, it might be able to conduct outreach to those likely to be eligible. Such an outreach effort would need to be mindful of privacy concerns. Outreach might be most effective if it occurs through existing relationships, such as the child's and family's health care providers. A direct data linkage could also be used in making disability determinations: the local disability determination services could use recent health care claims in assessing whether someone meets SSA's definition of disability. This might help streamline the application process, making things simpler for the applicant (who would need to gather fewer medical records) and for the doctor (who would need to complete less paperwork). A smoother information flow throughout the application process might also make it easier for SSA to administer SSI more effectively. For example, recent budget cuts for SSA staff have led to long delays in processing applications (Romig 2023). Streamlining processes might help combat this trend.

More generally, our results suggest that leveraging big data and machine can complement targeted outreach campaigns to boost program participation. Simplifying user experiences and reducing administrative burdens have been shown to significantly increase enrollment (e.g., Herd

et al. 2023). Research highlights that streamlined program rules and integrated services—where beneficiaries often qualify for multiple programs like SSI, Medicaid, and SNAP—can simplify processes and ensure eligible individuals do not miss out (Schmidt, Shore-Shepard, and Watson 2024). For example, easing Medicaid rules increases enrollments, and linking program data helps identify those eligible for multiple benefits, enhancing access. These approaches, alongside predictive capabilities of machine learning for identifying at-risk individuals or optimizing service delivery, represent a comprehensive strategy for improving program participation.

References

Battaglini, Marco, Luigi Guiso, Chiara Lacava, Douglas L. Miller, and Eleonora Patacchini. 2022. "Refining Public Policies with Machine Learning: The Case of Tax Auditing." Working Paper w30777. Cambridge, MA: National Bureau of Economic Research.

Bazzi, Samuel, Robert A. Blair, Christopher Blattman, Oeindrila Dube, Matthew Gudgeon, and Richard Peck. 2022. "The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia." *Review of Economics and Statistics* 104(4): 764-779.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45: 5-32.

Celhay, Pablo A., Bruce D. Meyer, and Nikolas Mittag. 2022. "Stigma in Welfare Programs." Working Paper w30307. Cambridge, MA: National Bureau of Economic Research.

Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. 2016. "Productivity and Selection of Human Capital with Machine Learning." *American Economic Review* 106(5): 124-127.

Chetty, Raj, John N. Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103(7): 2683-2721.

Cohen-Ross, Donna, Marian Jarlenski, Samantha Artiga, and Caryn Marks. 2009. "A Foundation for Health Reform: Findings of a 50 State Survey of Eligibility Rules, Enrollment and Renewal Procedures, and Cost-Sharing Practices in Medicaid and CHIP for Children and Parents During 2009." Washington, DC: Kaiser Commission on Medicaid and the Uninsured.

Currie, Janet. 2006. "The Take-Up of Social Benefits." In *Public Policy and the Income Distribution*, edited by Alan J. Auerbach, David Card, and John M. Quigley, 80-148. New York, NY: Russell Sage Foundation.

Deshpande, Manasi and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy* 11(4): 213-248.

Fox, Ashley, Wenhui Feng, and Megan Reynolds. 2023. "The Effect of Administrative Burden on State Safety-Net Participation: Evidence from Food Assistance, Cash Assistance, and Medicaid." *Public Administration Review* 83(2): 367-384.

Giannella, Eric, Tatiana Homonoff, Gwen Rino, and Jason Somerville. 2023. "Administrative Burden and Procedural Denials: Experimental Evidence from SNAP." Working Paper w31239. Cambridge, MA: National Bureau of Economic Research.

Hastings, Justine S., Mark Howison, and Sarah E. Inman. 2020. "Predicting High-Risk Opioid Prescriptions Before They Are Given." *Proceedings of the National Academy of Sciences* 117(4): 1917-1923.

Heinrich, Carolyn J., Sayil Camacho, Sarah Clark Henderson, Mónica Hernández, and Ela Joshi. 2022. "Consequences of Administrative Burden for Social Safety Nets That Support the Healthy Development of Children." *Journal of Policy Analysis and Management* 41(1): 11-44.

Heller, Sara B., Benjamin Jakubowski, Zubin Jelveh, and Max Kapustin. 2022. "Machine Learning Can Predict Shooting Victimization Well Enough to Help Prevent It." Working Paper w30170. Cambridge, MA: National Bureau of Economic Research.

Herd, Pamela, Thomas DeLeire, Hope Harvey, and Donald P. Moynihan. 2013. "Shifting Administrative Burden to the State: The Case of Medicaid Take-Up." *Public Administration Review* 73(S1): S69-S81.

Herd, Pamela, Hilary Hoynes, Jamila Michener, and Donald Moynihan. 2023. "Introduction: Administrative Burden as a Mechanism of Inequality in Policy Implementation." *Russell Sage Foundation Journal of the Social Sciences* 9(5): 1-30.

Homonoff, Tatiana, and Jason Somerville. 2021. "Program Recertification Costs: Evidence from SNAP." *American Economic Journal: Economic Policy* 13(4): 271-298.

Kaiser Family Foundation. 2023. "Monthly Child Enrollment in Medicaid and CHIP." State Health Facts. San Francisco, CA. Available at: https://www.kff.org/medicaid/state-indicator/total-medicaid-and-chip-child-enrollment/

Levere, Michael, Jeffrey Hemmeter, and David Wittenburg. 2023a. "Does the Drop in Child SSI Applications and Awards During the COVID-19 Pandemic Vary by Locality?" Working Paper. Princeton, NJ: Mathematica.

Levere, Michael, Jeffrey Hemmeter, and David Wittenburg. 2023b. "The Importance of Schools in Driving Children's Applications for Disability Benefits." Working Paper. Princeton, NJ: Mathematica.

Levere, Michael, David Wittenburg, and Jeffrey Hemmeter. 2022. "What Is the Relationship Between Socioeconomic Deprivation and Child Supplemental Security Income Participation?" *Social Security Bulletin* 82(2): 1-20.

Levere, Michael, Sean Orzol, Lindsey Leininger, and Nancy Early. 2019. "Contemporaneous and Long-Term Effects of Children's Public Health Insurance Expansions on Supplemental Security Income Participation." *Journal of Health Economics* 64: 80-92.

Michener, Jamila. 2018. *Fragmented Democracy: Medicaid, Federalism, and Unequal Politics*. Cambridge University Press, Cambridge

Mullainathan, Sendhil and Jann Spiess. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31(2): 87-106.

Romig, Kathleen. 2023. "Social Security Administration Needs Additional Funding to Avoid Exacerbating Customer Service Crisis." Center on Budget and Policy Priorities. Washington, DC. Available at: https://www.cbpp.org/blog/social-security-administration-needs-additional-funding-to-avoid-exacerbating-customer-service

Sansone, Dario and Anna Zhu. 2023 (forthcoming). "Using Machine Learning to Create an Early Warning System for Welfare Recipients." *Oxford Bulletin of Economics and Statistics*.

Schmidt, Lucie and Purvi Sevak. 2017. "Child Participation in Supplemental Security Income: Cross- and Within-State Determinants of Caseload Growth." *Journal of Disability Policy Studies* 28(3): 131-140.

Schmidt, Lucie, Lara Shore-Sheppard, and Tara Watson. 2016. "The Effect of Safety-Net Programs on Food Insecurity." *Journal of Human Resources* 51(3): 589-614.

Schmidt, Lucie, Lara Shore-Sheppard, and Tara Watson. 2024 (forthcoming). "Safety Net Program Spillovers and Administrative Burden." *National Tax Journal*.

Singh, Gopal K. 2003. "Area Deprivation and Widening Inequalities in US Mortality, 1969–1998." *American Journal of Public Health* 93(7): 1137-1143.

U.S. Social Security Administration. 2021. "SSA Budget Information: FY 2022 Budget Request." FY 2022 Congressional Justification. Baltimore, MD. Available at: https://www.ssa.gov/budget/assets/materials/2022/2022BO.pdf

U.S. Social Security Administration. 2022. "SSI Annual Statistical Report, 2021." Baltimore, MD. Available at: https://www.ssa.gov/policy/docs/statcomps/ssi_asr/2021/

Appendix Table 1. *Zip-Code Level Socioeconomic Controls*

Measure
Population aged 25 and older with less than 9 years of education
Population aged 25 and older who completed at least a high school education
Employed persons aged 16 and older in white collar occupations (management, business, science and arts occupations)
Population aged 16 and older who are unemployed
Owner-occupied housing units (home ownership rate)
Households with more than one person per room
Median monthly mortgage ($)
Median gross rent ($)
Median home value ($)
Median family income ($)
Income disparity (ratio of people with income under $15,000 to people with income over $75,000)
Families below poverty level
Population earning less than 150 percent of the federal poverty limit
Single parent households with children under 18 years old
Households without a motor vehicle
Households without a telephone
Occupied housing units without complete plumbing

Note: Unless otherwise indicated, all measures are percentages.

Appendix Table 2. *Potentially Eligible Child SSI Recipients Above Each Probability Threshold, 2018 Data*

Predicted probability threshold	Number of children in "high-match" states	As percentage of current child SSI recipients	Number of children in "low-match" states applying same percentage	Total number of potentially eligible children in United States
10 percent	1,398,576	151.9%	401,261	1,799,837
15 percent	818,236	88.9%	234,757	1,052,993
20 percent	513,623	55.8%	147,362	660,985
25 percent	326,365	35.4%	93,636	420,001
30 percent	207,464	22.5%	59,523	266,987
35 percent	132,831	14.4%	38,110	170,941
40 percent	84,808	9.2%	24,332	109,140
45 percent	53,629	5.8%	15,387	69,016
50 percent	32,357	3.5%	9,283	41,640

Notes: The first column reports the total number of children across high-match states who are not receiving SSI but who have a predicted probability above the threshold. The second column expresses this as a percentage of the 920,753 total child SSI recipients across these 32 states within the Medicaid data. The third column multiplies this percentage by the 264,170 child SSI recipients in 2018 in the 18 low-match states and the District of Columbia from the SSI recipients by state and county report. Finally, the fourth column sums the first and the third columns to indicate the total number of potentially eligible children across the United States.
Source: Authors' calculations using 2018 TAF data and 2018 SSI recipients by state and county.

Appendix Table 3. *Potentially Eligible Child SSI Recipients Above Each Probability Threshold, 2017 Data*

Predicted probability threshold	Number of children in "high-match" states	As percentage of current child SSI recipients	Number of children in "low-match" states applying same percentage	Total number of potentially eligible children in United States
10 percent	1,420,519	150.0%	405,768	1,826,287
15 percent	833,954	88.0%	238,217	1,072,171
20 percent	522,082	55.1%	149,132	671,214
25 percent	330,327	34.9%	94,357	424,684
30 percent	208,427	22.0%	59,537	267,964
35 percent	131,354	13.9%	37,521	168,875
40 percent	82,479	8.7%	23,560	106,039
45 percent	50,740	5.4%	14,494	65,234
50 percent	30,004	3.2%	8,571	38,575

Notes: The first column reports the total number of children across high-match states who are not receiving SSI but who have a predicted probability above the threshold. The second column expresses this as a percentage of the 947,240 total child SSI recipients across these 32 states within the Medicaid data. The third column multiplies this percentage by the 270,577 child SSI recipients in 2017 in the 18 low-match states and the District of Columbia from the SSI recipients by state and county report. Finally, the fourth column sums the first and the third columns to indicate the total number of potentially eligible children across the United States.
Source: Authors' calculations using 2017 TAF data and 2017 SSI recipients by state and county.

Appendix Table 4. *State-level Estimates of Potentially Eligible Population*

State	Number of current SSI recipients (in Medicaid data)	Potentially Eligible Child SSI Recipients Above Each Probability Threshold								
		10	15	20	25	30	35	40	45	50
AL	23,167	28,186	16,655	10,948	7,482	5,086	3,420	2,247	1,412	868
AR	26,374	38,589	23,439	14,934	9,581	6,206	3,919	2,381	1,385	753
AZ	17,857	30,718	17,639	11,033	7,039	4,400	2,769	1,659	899	442
CA	104,448	153,150	88,189	60,982	44,441	32,637	23,634	16,918	11,552	7,175
CO	8,766	11,938	5,267	2,582	1,366	747	361	170	65	19
CT	8,081	15,131	6,544	2,503	1,183	572	296	174	80	35
DE	3,282	3,373	1,549	823	458	254	154	91	48	31
FL	103,450	147,460	83,858	53,086	35,479	24,400	17,637	12,836	8,836	5,697
HI	1,309	635	336	208	110	65	47	36	23	12
ID	4,529	6,199	3,192	1,438	652	334	179	88	40	21
IN	21,413	29,898	17,349	10,749	6,940	4,487	2,745	1,630	859	398
KS	7,901	12,769	5,782	2,822	1,315	550	235	94	40	17
LA	35,606	33,736	23,759	18,301	12,855	8,594	6,665	5,329	4,234	3,193
MA	23,253	37,255	21,961	13,077	7,701	4,516	2,666	1,507	821	420
MD	19,849	36,219	21,655	13,148	7,818	4,541	2,517	1,316	658	349
ME	3,924	5,305	2,423	1,159	533	262	119	49	15	1-10
MI	35,134	48,417	27,098	16,131	9,806	6,371	4,394	3,049	2,126	1,347
MN	10,922	17,120	10,726	6,800	4,124	2,410	1,372	702	342	142
MS	20,446	32,926	18,687	10,914	6,147	3,273	1,675	887	447	213
MT	2,036	1,706	732	358	168	86	55	29	13	1-10
NC	37,434	60,323	36,085	22,533	13,312	7,673	4,724	3,112	2,081	1,373
NM	7,842	10,990	5,799	3,321	2,012	1,174	675	346	177	93
NV	9,129	11,247	5,867	3,310	1,973	1,191	712	401	216	112
NY	80,231	137,964	89,666	60,745	40,263	25,033	14,471	8,000	4,074	1,808
OH	45,296	81,840	47,760	28,164	16,712	10,287	6,560	4,078	2,449	1,392
PA	57,460	95,661	52,433	29,548	17,011	10,587	7,389	5,745	4,641	3,239
SD	2,104	2,154	725	268	95	41	21	1-10	1-10	1-10
TX	128,218	207,632	119,613	74,370	48,495	32,449	21,587	14,252	9,252	5,877
WA	15,431	25,103	11,877	5,892	3,104	1,535	581	132	16	0
WI	21,710	33,532	19,481	11,382	6,690	3,988	2,390	1,394	776	395
WV	7,245	7,943	3,758	1,918	980	540	307	187	115	75
WY	840	909	452	193	89	45	23	1-10	1-10	1-10

Notes: The first column reports the total number of children flagged as eligible for SSI within Medicaid data. The subsequent columns report the number who have a predicted probability at least as high as the number in the column, who thus might be thought of as potentially eligible child SSI beneficiaries given the threshold. A value of 1-10 means that the number was too low to disclose the exact value. Simplifying by treating a value of 1-10 as 5, the total across all states within each column matches the number of children in "high-match" states column (Column 2) in Table 2.
*Source:* Authors' calculations using 2019 TAF data.

Appendix Table 5. *Percentage of Potentially Eligible Beneficiaries with Family Incomes Below 255 Percent of Federal Poverty Limit*

Predicted probability threshold	Colorado		Massachusetts	
	Model with income	Model without income	Model with income	Model without income
10 percent	98.6	97.6	94.5	93.2
15 percent	98.6	97.5	95.2	93.4
20 percent	98.6	97.0	96.1	93.7
25 percent	99.1	97.0	96.8	94.4
30 percent	99.6	97.0	97.1	94.4
35 percent	99.7	96.6	97.3	94.2
40 percent	99.4	96.3	97.6	94.9
45 percent	98.5	97.7	98.4	94.2
50 percent	100.0	96.4	98.8	94.4

Notes: The numbers in the table show the percentage of potentially eligible child SSI beneficiaries who have a probability that exceeds the predicted probability threshold whose family income is below 255 percent of the federal poverty limit, which indicates that their income might likely be sufficiently low for them to qualify for SSI. We report these findings from random forest models that do and do not include income variables as predictive features. *Source:* Authors' calculations using 2019 TAF data.

Appendix Figure 1. *Distribution of SSI Receipt by Predicted Probability of SSI Receipt, Testing Sample Only*
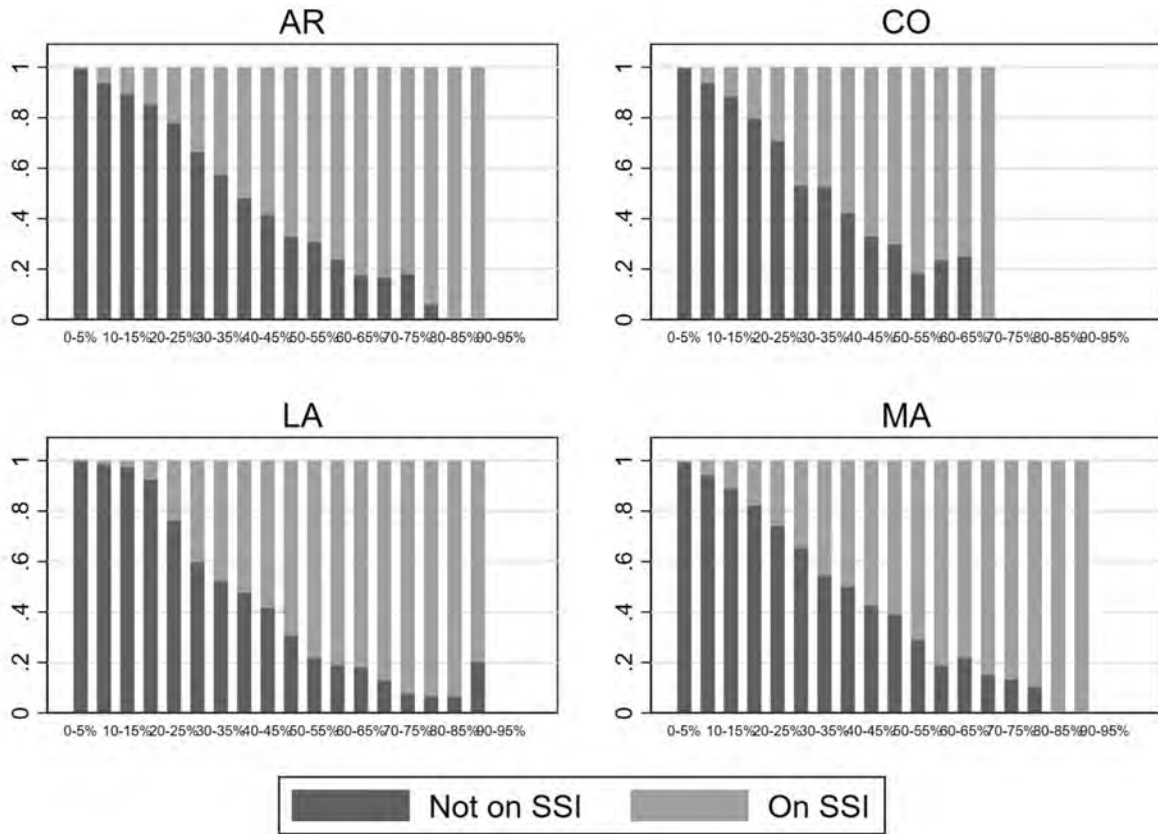
Notes: Indicates the share of children receiving SSI for each ventile of predicted SSI probability, using only those children in the testing sample (not included in estimating the random forest model). If no children in a state have a probability sufficiently high, that bar is excluded from the figure (e.g., nobody in Colorado in the testing sample has a predicted probability above 75 percent).
Source: Authors' calculations using 2019 TAF data.

Appendix Figure 2. *Child Non-SSI Recipients with Probability of Receipt Exceeding Threshold, as Percentage of Current SSI Recipients*
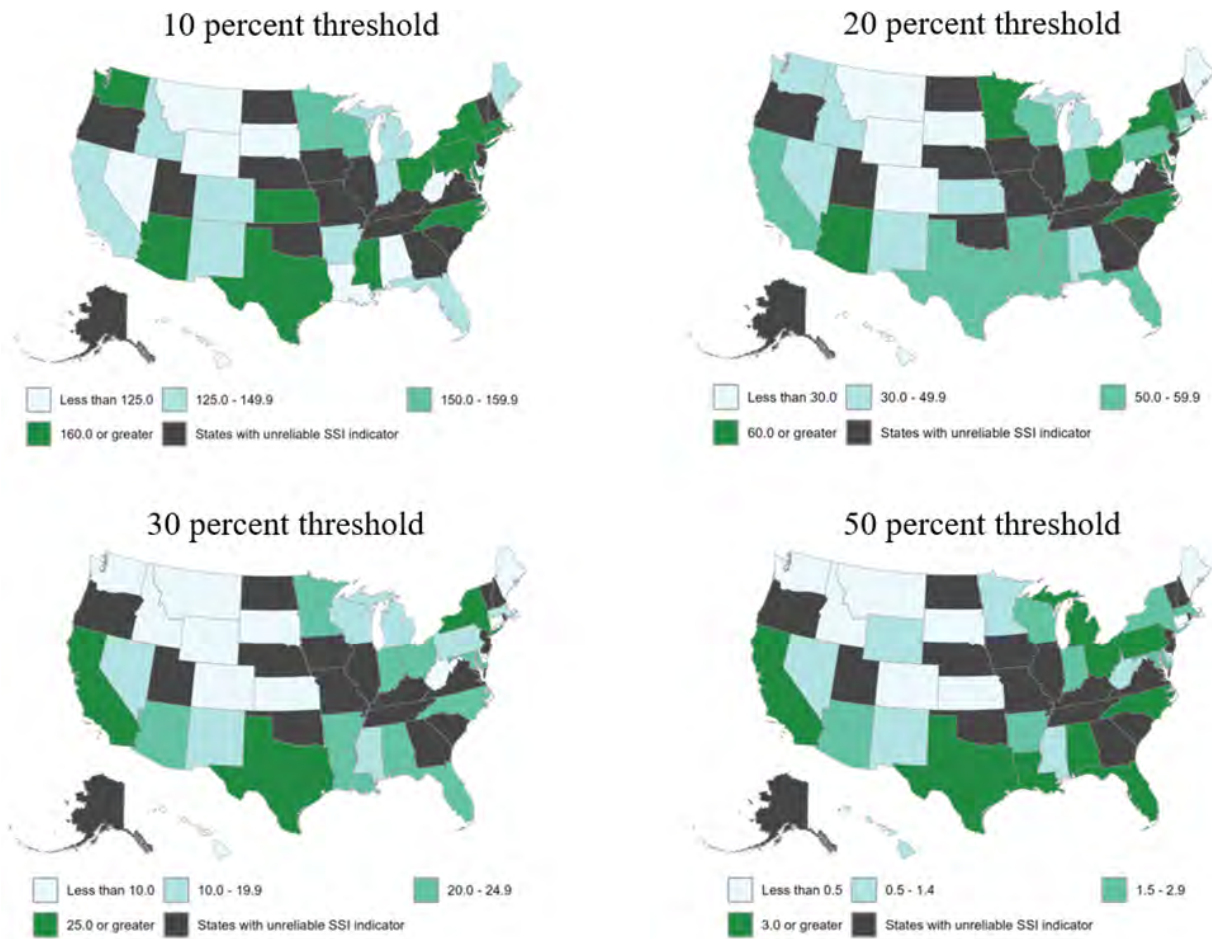
Notes: Each panel shows the potential increase in child SSI recipients if all children with probability exceeding a given threshold were eligible for SSI. Each panel considers a different probability threshold. The 40 percent threshold is presented in Figure 3.
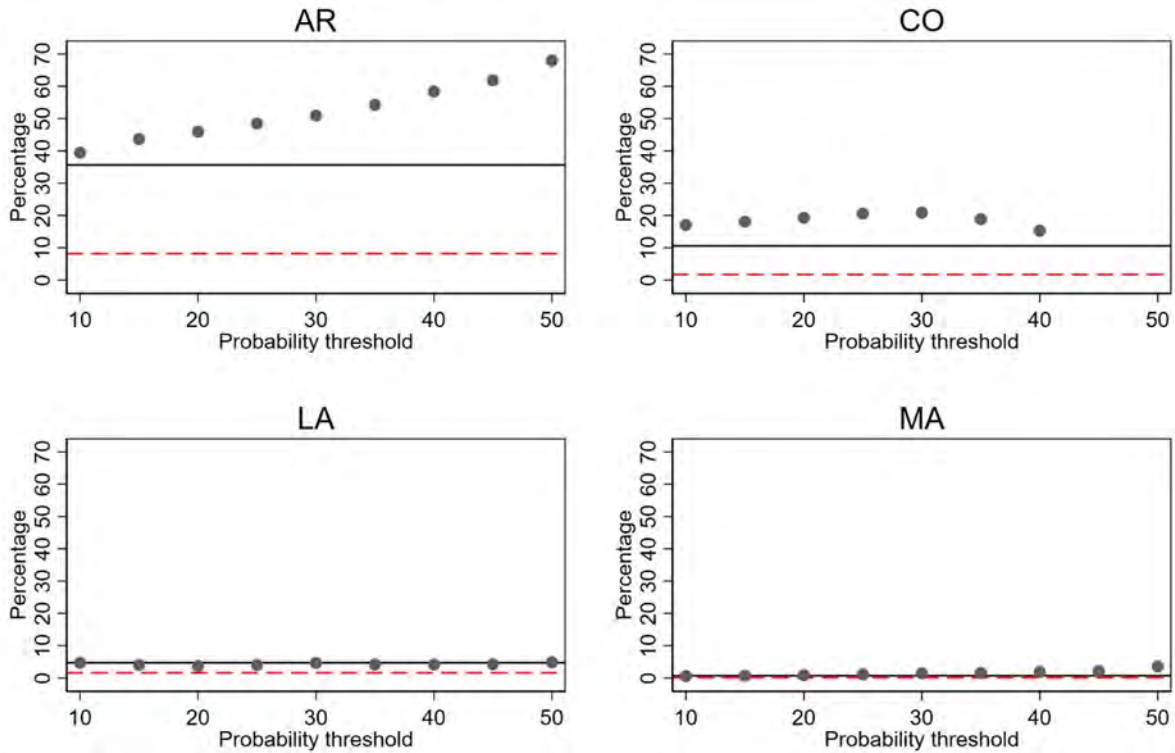*Source:* Authors' calculations using 2019 TAF data.

Appendix Figure 3. *Has ADHD Medication Prescription, by State, Receipt of SSI, and Estimated Probability of SSI Receipt*
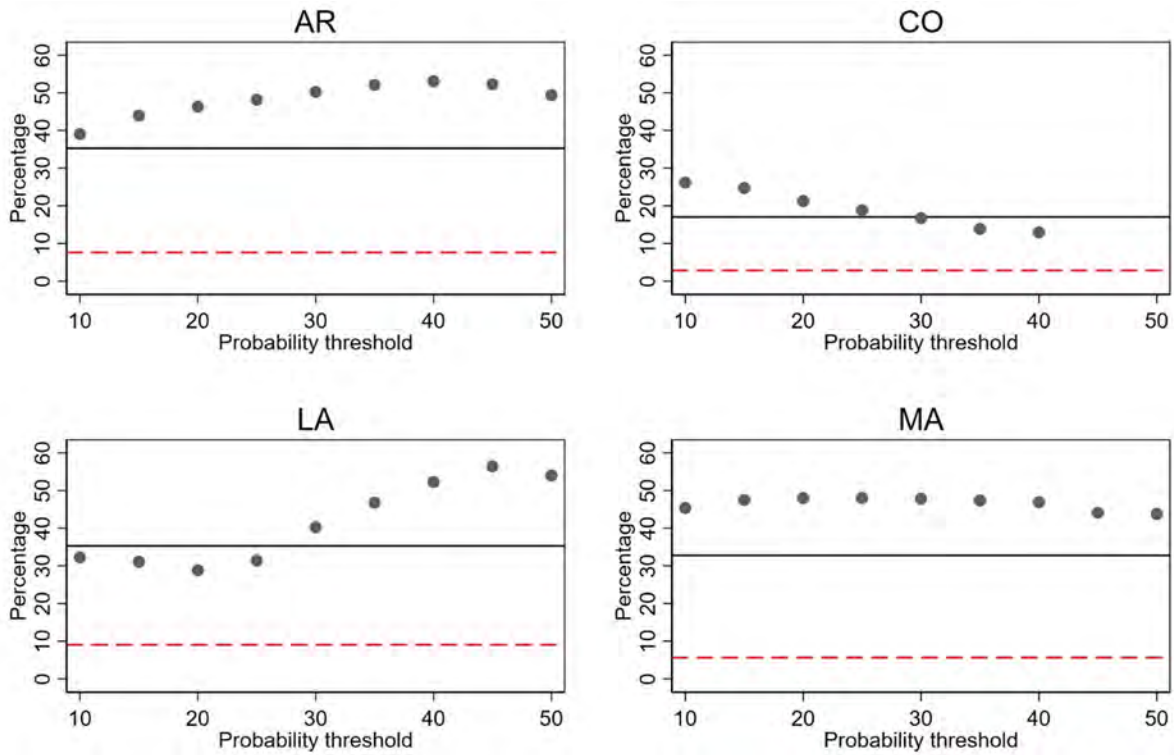
Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
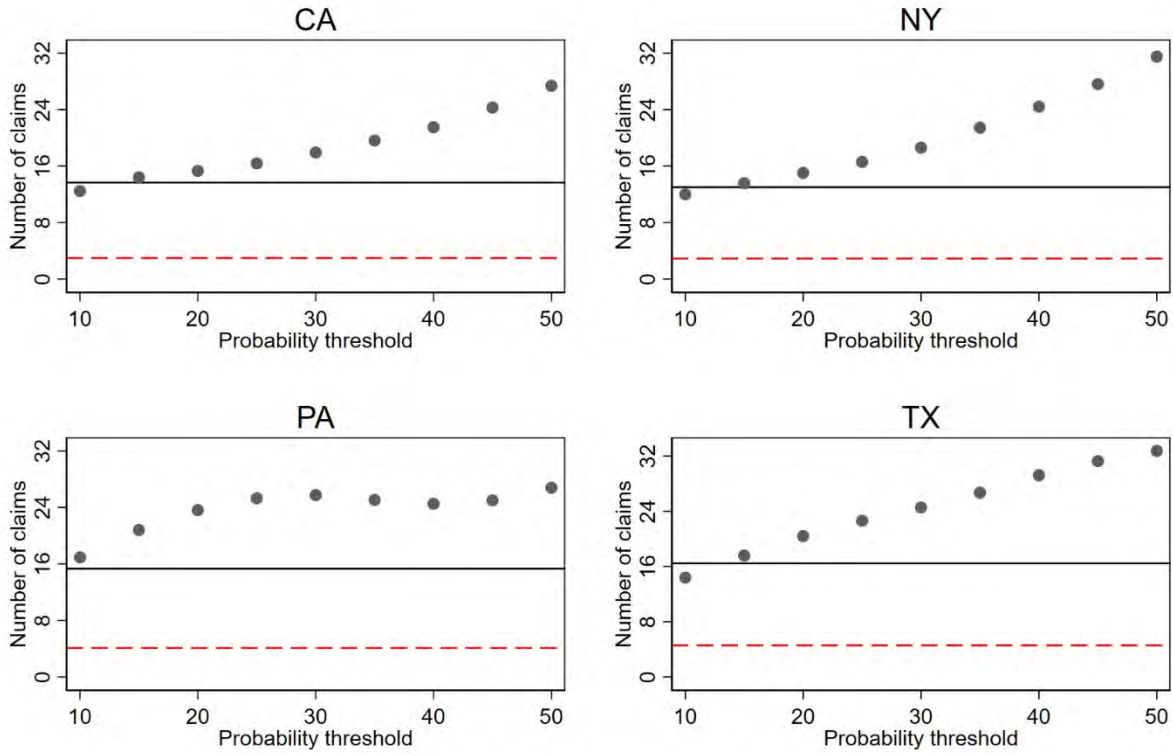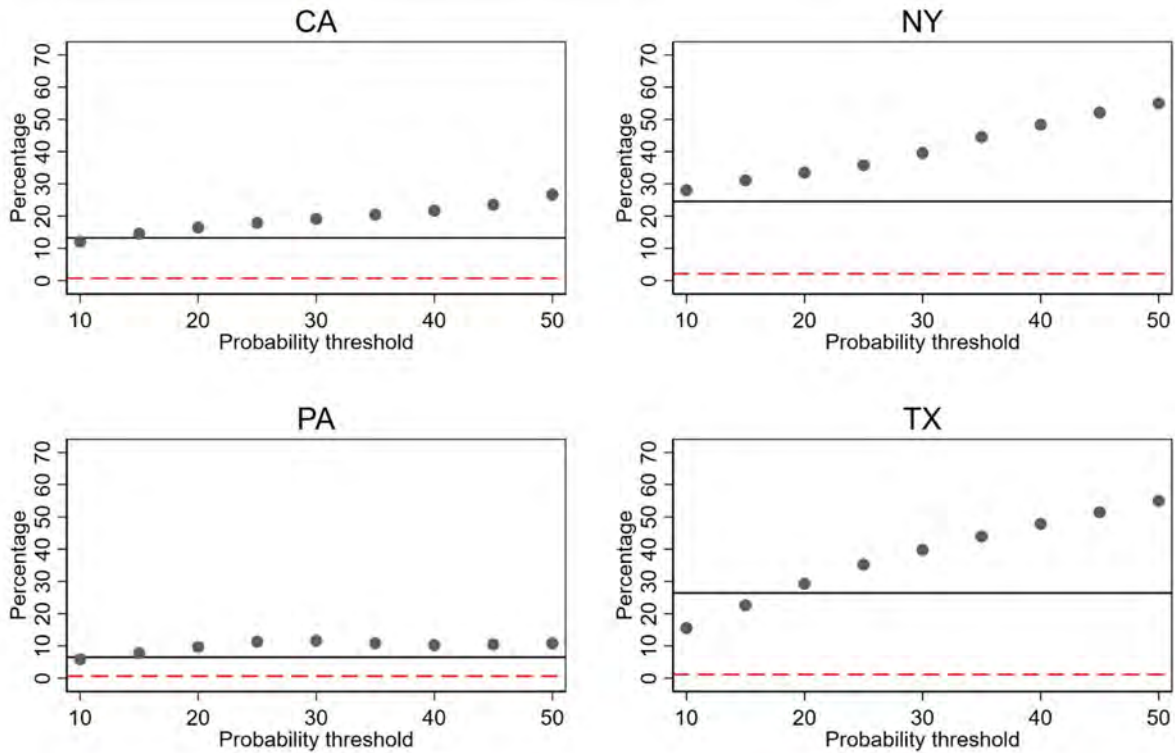*Source:* Authors' calculations using 2019 TAF data.

Appendix Figure 4. *Has Claim with Speech-Language Pathologist, by State, Receipt of SSI, and Estimated Probability of SSI Receipt*

Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
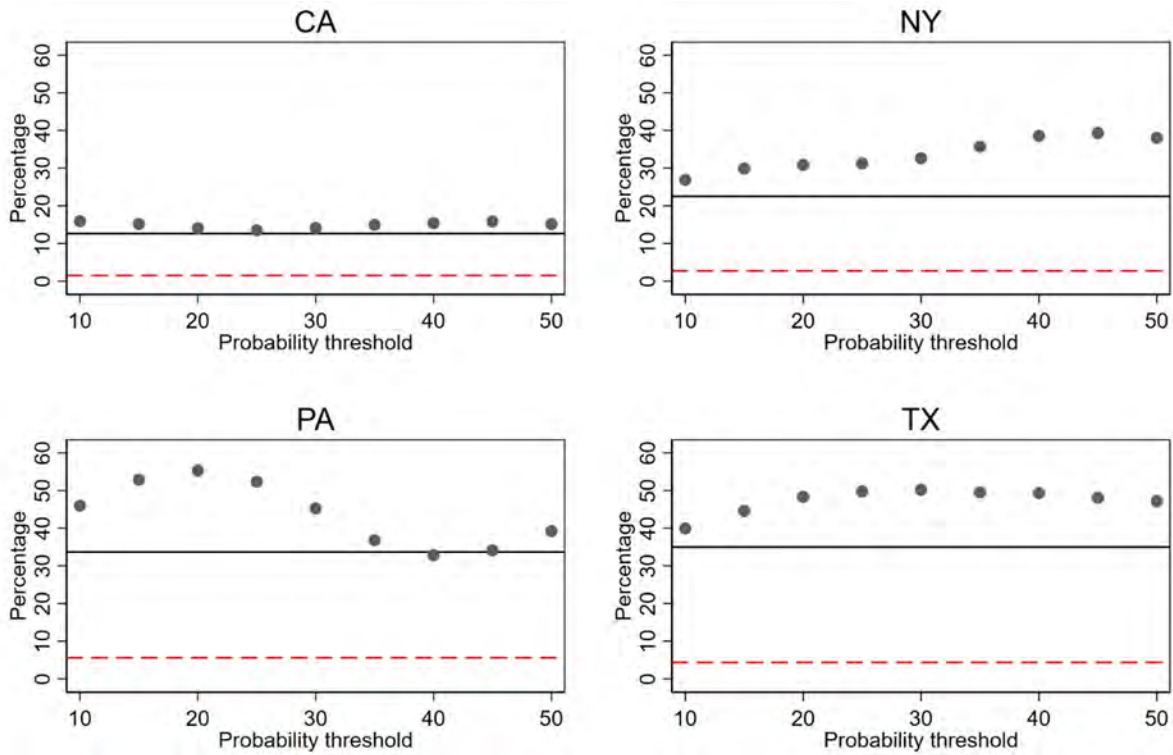*Source:* Authors' calculations using 2019 TAF data.

Appendix Figure 5. *Number of Prescription Drug Claims, by State, Receipt of SSI, and Estimated Probability of SSI Receipt (High Population States)*

Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
Source: Authors' calculations using 2019 TAF data.

Appendix Figure 6. *Presence of Other Developmental Delays Chronic Condition, by State, Receipt of SSI, and Estimated Probability of SSI Receipt (High Population States)*

Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
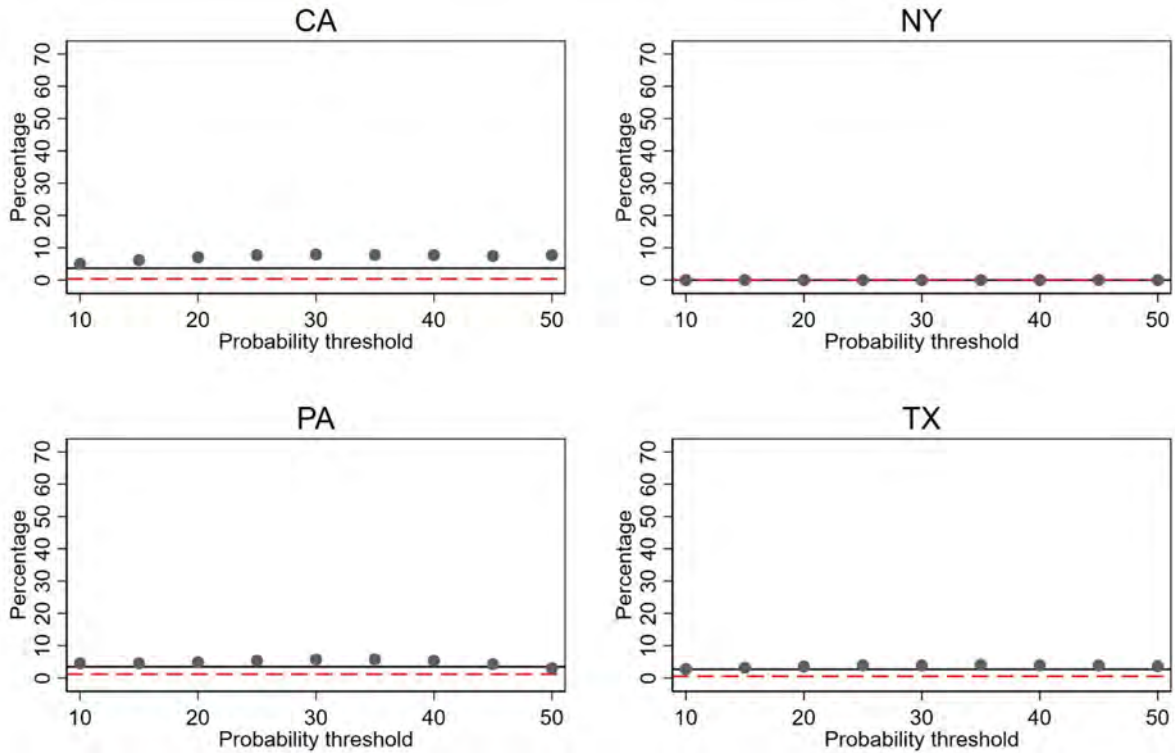*Source:* Authors' calculations using 2019 TAF data.

Appendix Figure 7. *Has ADHD Medication Prescription, by State, Receipt of SSI, and Estimated Probability of SSI Receipt (High Population States)*

Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
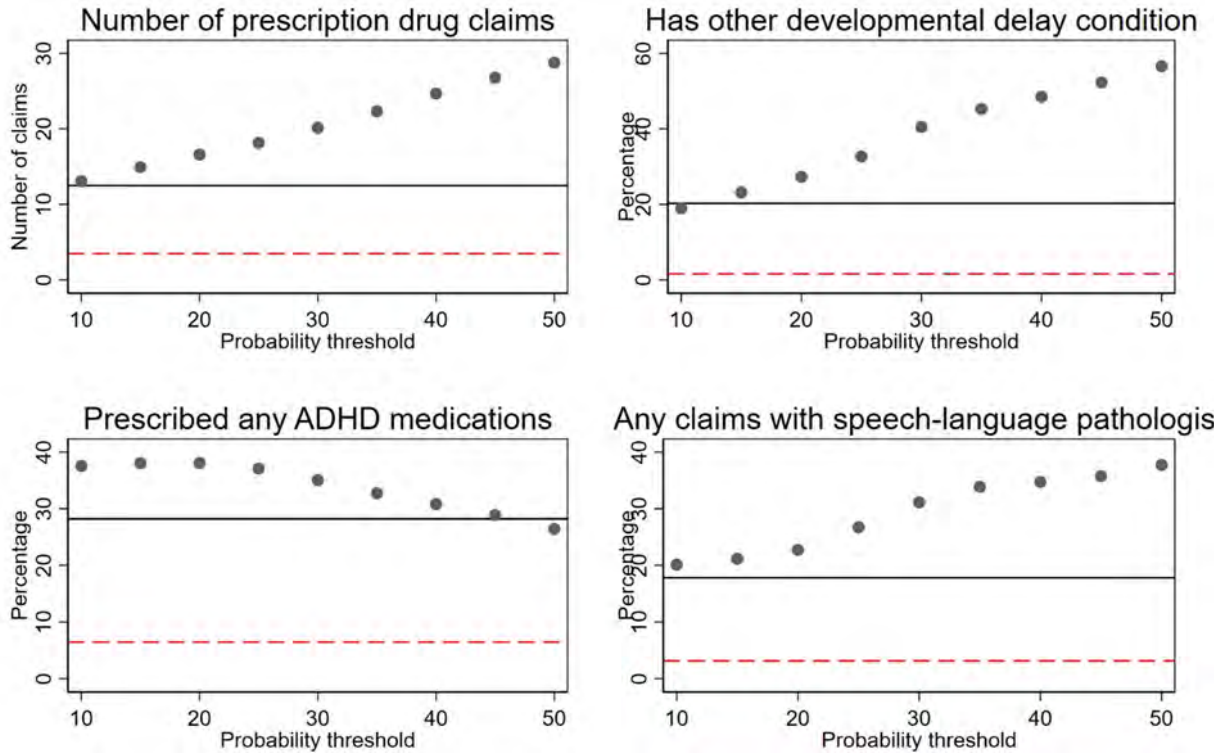*Source:* Authors' calculations using 2019 TAF data.

Appendix Figure 8. *Has Claim with Speech-Language Pathologist, by State, Receipt of SSI, and Estimated Probability of SSI Receipt (High Population States)*

Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
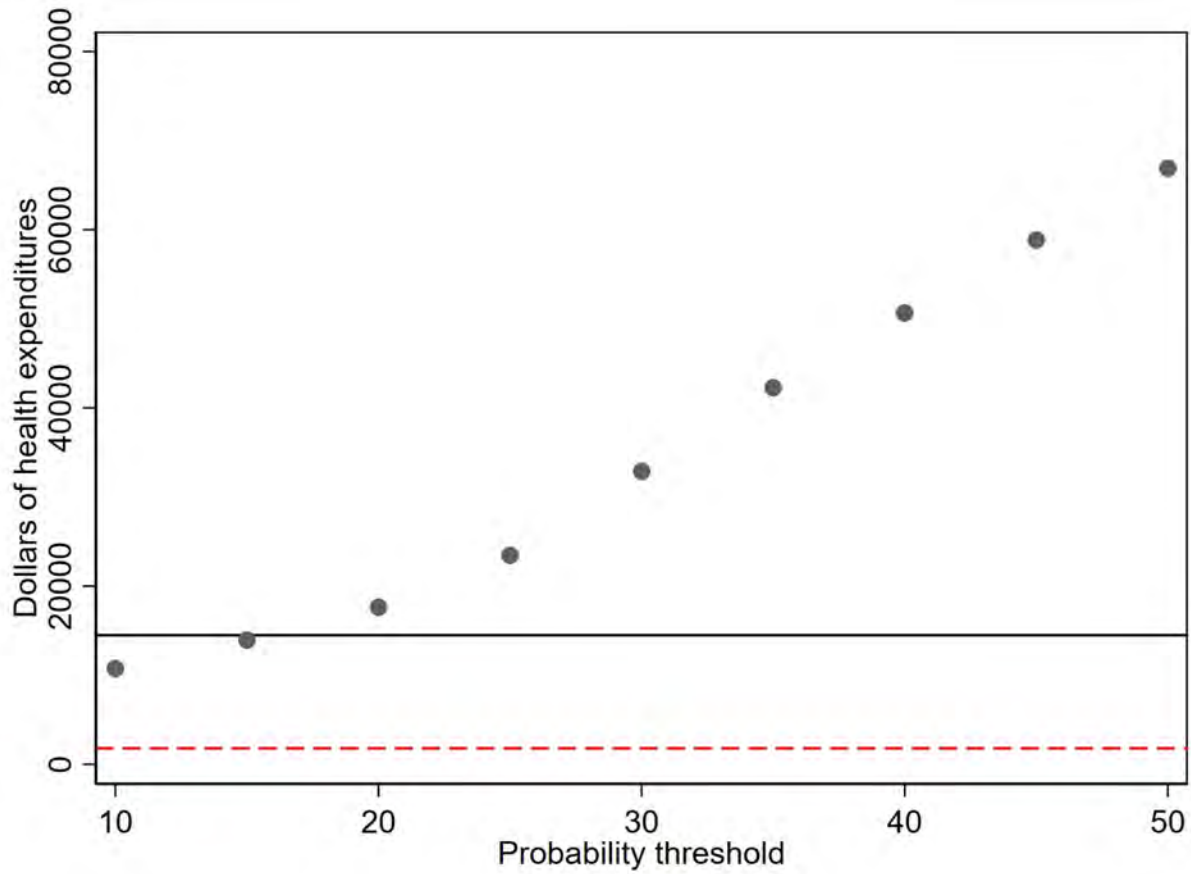*Source:* Authors' calculations using 2019 TAF data.

Appendix Figure 9. *Health Care Utilization in North Carolina, by Receipt of SSI, and Estimated Probability of SSI Receipt*

Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
Source: Authors' calculations using 2019 TAF data.

Appendix Figure 10. *Medicaid Claim Payment Amounts in North Carolina, by Receipt of SSI, and Estimated Probability of SSI Receipt*

Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
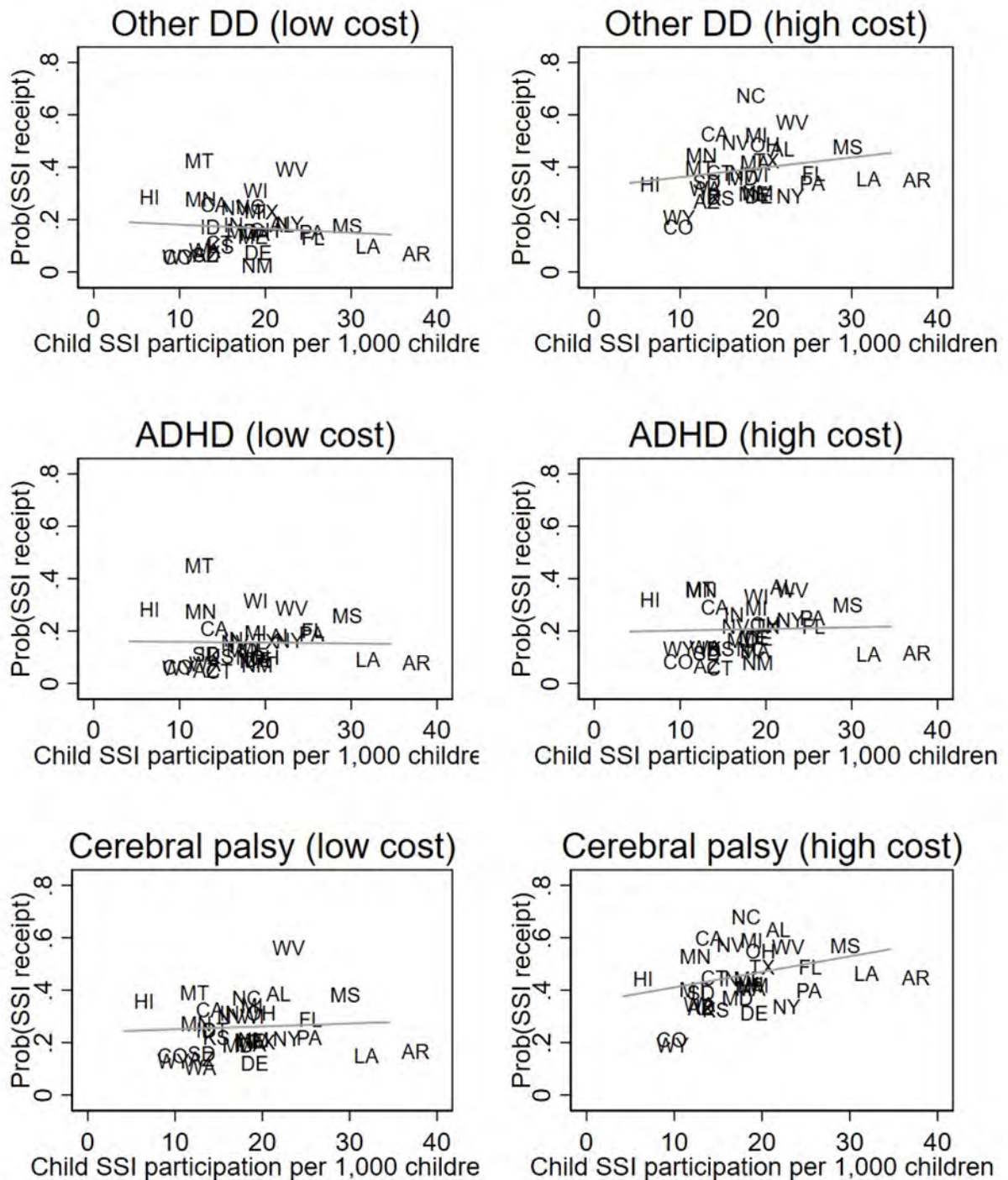*Source:* Authors' calculations using 2019 TAF data.

Appendix Figure 11. *Predicted Probability of SSI Receipt, by State, Condition, and Cost*

Notes: Estimates the probability of SSI receipt using the state specific model, holding fixed the health care claims associated with each condition. The gray line represents a line of best fit based on a linear regression. See the text for a description of how we identified the health care claims associated with each condition. Low cost refers to the claims profile of a typical SSI recipient at the 25th percentile of the cost distribution for a child with that condition. High cost refers to the claims profile of a typical SSI recipient at the 75th percentile of the cost distribution for a child with that condition.
Source: Authors' calculations using 2019 TAF data.

Appendix A

We considered an analysis that would extrapolate the results from each state-specific model to estimate the size of the potentially eligible population if national SSI patterns matched a given state's participation. In our main analysis, we use each state-specific model to estimate the probability of SSI receipt for each child in that state. This national extrapolation would use each state-specific model to estimate the probability of SSI receipt for each child in the entire country, including those states that do not have a reliable indicator of SSI receipt. The extrapolation exercise could be useful from a policy perspective in helping to understand some of the reasons why SSI participation is lower in certain states: for example, SSI participation may be relatively low in Colorado both because it has fewer children with marked and severe functional limitations, or because children with the same limitations do not qualify as frequently as they do in other states.

Though we conducted this analysis, we did not include it in this paper because it did not produce reliable results. Two primary factors render the results of this national extrapolation exercise to be unreliable.

1. Data availability and characterization of claims differ across states. For example, in Colorado, which has reliable income data, income variables can be highly predictive of potentially eligibility. However, many states have missing income data for everyone. If the Colorado model only identifies people as potentially eligible who have low income, then applying the Colorado model to other states will lead few people to be identified as potentially eligible based solely on the difference in data quality. See Section V.C for results that test the sensitivity of our findings to including and excluding income in the model. A similar pattern emerges with characterization of claims – for example, in Arkansas having a claim with a speech language pathologist is predictive of SSI, but some states have no claims with speech language pathologists (perhaps because of billing processes, see Appendix Figure 4).

2. Part of the national extrapolation exercise relies on applying the predictive model in states where we do not know who currently receives SSI. Because SSI receipt is so rare, the predicted probability of SSI eligibility can be very low even for those receiving SSI:

for example, in Massachusetts more than half of SSI recipients have a predicted probability less than 20 percent (98.4 percent of non-SSI recipients have a predicted probability less than 20 percent). It is therefore hard to reliably estimate how many are potentially eligible in states where we cannot distinguish current SSI recipients from children who do not currently qualify.

Though the national extrapolation exercise would have been an interesting analysis, it was not sufficiently informative to warrant inclusion in the paper. Beyond the two main factors noted above, it is also important to consider that Medicaid is ultimately a state-specific program. As emphasized in Section II, though all states must follow certain federal guidelines in developing their Medicaid programs, each state operates its own program. Caution therefore should always be warranted when considering national patterns regarding Medicaid recipients. Here, the two specific factors together with the local nature of Medicaid means that any results from the national extrapolation exercise are too unreliable to be included in the paper.

RECENT WORKING PAPERS FROM THE CENTER FOR RETIREMENT RESEARCH AT BOSTON COLLEGE

Are Older Workers Good for Business?
Laura D. Quinby, Gal Wettstein, and James Giles, November 2023

How Much Do People Value Annuities and Their Added Features?
Karolos Arapakis and Gal Wettstein, November 2023

Experiences, Behaviors, and Attitudes About COVID-19 for People with Disabilities Over Time
Amal Harrati, Marisa Shenk, and Bernadette Hicks, October 2023

Shared Households as a Safety Net for Older Adults
Hope Harvey and Kristin L. Perkins, October 2023

The Impact of Losing Childhood Supplemental Security Income Benefits on Long-Term Education and Health Outcomes
Priyanka Anand and Hansoo Ko, October 2023

What Is the Insurance Value of Social Security by Race and Socioeconomic Status?
Karolos Arapakis, Gal Wettstein, and Yimeng Yin, September 2023

Does Temporary Disability Insurance Reduce Older Workers' Reliance on Social Security Disability Insurance?
Siyan Liu, Laura D. Quinby, and James Giles, September 2023

How Will Employer Health Insurance Affect Wages and Social Security Finances?
Anqi Chen, Alicia H. Munnell, and Diana Horvath, September 2023

What Are the Implications of Rising Debt for Older Americans?
Anqi Chen, Siyan Liu, and Alicia H. Munnell, September 2023

Wills, Wealth, and Race
Jean-Pierre Aubry, Alicia H. Munnell, and Gal Wettstein, August 2023

How Can Social Security Children's Benefits Help Grandparents Raise Grandchildren?
Siyan Liu and Laura D. Quinby, June 2023

Forward-Looking Labor Supply Responses to Changes in Pension Wealth: Evidence from Germany
Elisabeth Artmann, Nicola Fuchs-Schündeln, and Giulia Giupponi, June 2023

All working papers are available on the Center for Retirement Research website (https://crr.bc.edu) and can be requested by e-mail (crr@bc.edu) or phone (617-552-1762).