Which Medicaid Recipients Might Be Eligible for SSI?

Michael Levere[^] and David Wittenburg[+]

July 2025

Though safety net programs offer important benefits, take-up is often incomplete. Using machine learning models on Medicaid data, we estimate the take-up rate for children's Supplemental Security Income (SSI), and the characteristics of potentially eligible children with disabilities who do not receive benefits. Using over 1,000 measures of health care utilization, we estimate state-specific models that generate the probability that each child not receiving SSI is eligible. Using the expected number of potentially eligible children, the implied take-up of SSI is approximately 70 percent. Potentially eligible children have intensive health care usage, often more intensive than current child SSI recipients.

[^] Colgate University

[+] Westat

I. INTRODUCTION

Social safety net programs are designed to help the neediest members of society navigate the financial challenges associated with day-to-day living. Yet take-up of these social programs is often incomplete, with numerous eligible people who do not participate (Currie 2006). Various factors can contribute to this limited take-up (Ko and Moffitt 2024). Commonly cited reasons include stigma (Celhay, Meyer, and Mittag 2022), administrative burden (Herd et al. 2013), and lack of awareness (Chetty, Friedman, and Saez 2013), among others. Low take-up rates mean that people eligible for benefits are not accessing the supports they need (and qualify for) from the program.

A key goal of this paper is to estimate the extent to which eligible children with disabilities take up Supplemental Security Income (SSI) benefits. SSI offers cash assistance to low-income families where a child has a disability. To our knowledge, no reliable estimates exist quantifying the SSI take-up rate, likely because it is difficult to measure the disability criterion in readily available data. Ensuring that these families – who face a double disadvantage stemming from both their income status and their child's disability – can access the benefits to which they are eligible is an important goal. Children who maintain access to SSI as they enter adulthood are less likely to commit crimes, particularly those related to income generation like theft or burglary (Deshpande and Mueller-Smith 2022). More generally, when families get access to a new source of income, this leads to improvements in children's long-run mental health (Akee, Copeland, and Simeonova 2024).

Recent declines in participation in child SSI suggest that many who are eligible may not be participating in the program, particularly considering concurrent declines in children's mental health. From 2013 to 2021, the number of child SSI recipients fell by nearly 20 percent. Yet the prevalence of mental health issues like major depressive episodes and suicidal ideation substantially increased over this period (NSDUH 2019; National Center for Health Statistics 2021), though some of that increase may relate to reporting rather than underlying health (Corredor-Waldron and Currie 2024). 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 that 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 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 was 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.[1] 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, with an estimated take-up rate of about 70 percent. By summing the individual probabilities across children not receiving SSI, we can estimate the expected number of potentially eligible recipients. Our preferred estimate is roughly 480,000, which excludes children with probabilities less than 5 percent. This corresponds to a take-up rate of about 70 percent – dividing the 1.16 million child SSI recipients by the current recipients plus potential recipients.[2] These take-up rates implicitly assume that those potentially eligible based on health care claims also meet the income and resource criteria associated with SSI – we present results suggesting that this is indeed the case. It also does not consider non-Medicaid recipients as potentially eligible, so may overestimate the actual take-up rate. This take-up rate is roughly consistent with the estimated take-up rate for other US social benefits programs like the Earned Income Tax Credit, Medicaid, and SNAP (Currie 2006; Kopczuk and Pop-Eleches 2007; Decker, Abdus, and Litman 2022; Ko and Moffitt 2024). It is also broadly in line with the estimated 63 percent take-up rate for SSI among the elderly from Davies et al. (2001/2002), though the elderly need only have low-income and resources but do not need to have a disability to qualify. Take-up rates vary across states, ranging

[1] We also report results from a direct classification approach where we use this probability to create an indicator for whether the child is potentially eligible, with eligibility determined by the probability exceeding some threshold. Yet determining the precise threshold that counts as potentially eligible is inherently arbitrary and can lead to widely varying estimates of the take-up rate. Instead, focusing primarily on the expected value allows us to inherently place more weight on those with high probabilities in predicting the number of children eligible for SSI. Such an approach is consistent with findings on the health care utilization that ultimately use this direct classification approach, showing that utilization of care is typically more intensive as the probability of SSI eligibility increases.
[2] If we include all children, the implied take-up rate is 57 percent.

from 68 percent in Ohio to 86 percent in Hawaii. Interestingly, these take-up rates are not correlated with the existing level of child SSI participation per capita.

Children identified as having a high probability of potential eligibility 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 like developmental delays, intellectual disabilities, learning disabilities, and cerebral palsy. For example, 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—things like general supports that states offer to promote enrollment, varying acceptance rates, local administrative burden, and more—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. (2024) 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. [2024] 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). Foote, Grosz, and Rennane (2019) find that offering online applications to SSDI, which lowered the transaction costs in applying for benefits, led to increased rates of application. 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., Homonoff 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. Papers such as Burns and Dague (2017) and Levere et al. (2019) have shown Medicaid may play an important role in motivating people to enroll in SSI. 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.

II. 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 2024, the federal maximum payment from SSI is $943 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 2024). 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.[3] 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.

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

[3] This decision, from the 1990 Supreme Court case *Sullivan v. Zebley*, eased standards for children (and particularly children with mental disorders) to qualify for SSI (Levere 2021). The decision required that children and adults face similar eligibility criteria so that children with a disability of comparable severity to an adult who qualified might also themselves qualify. Previously, children could only qualify if their disabling condition met a specific listing of impairments, but after the decision, children could qualify if an individualized functional assessment determined that their condition created a "marked and severe functional limitation". SSA also revised its listing of mental impairments to recognize more conditions.

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. Income criteria are typically more generous for Medicaid than for SSI, but only 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.[4] Child SSI participation also varies across counties and states, with much interest in the factors that drive these local differences (e.g., Aizer, Gordon, and Kearney 2013; 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.

III. 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 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

[4] 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.

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 2024). 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.[5] Among the "high-match" states, the most common metric that matched the SSA published statistics was the eligibility group variable (26 of 32 states).[6] One caution with this benchmarking exercise is that it only requires the aggregate number of

[5] 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.
[6] In the other six states, three (Idaho, Kansas, and Minnesota) matched SSA published statistics with the SSI indicator, one (Connecticut) matched with the SSI status, and two (South Dakota and Washington) matched when considering the union of all three potential variables in Medicaid data to identify SSI recipients.

recipients to be similar across the two data sources, though the actual children flagged as SSI recipients in Medicaid data might not be correct.

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 in the share of states that are high match by whether the state automatically confers Medicaid 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. While a natural concern may be that this leads us to underestimate take-up rates (because the number of actual SSI recipients is lower than observed in the data for some states), we show below that attempts to correct for this underestimate do not change the estimated take-up rate.

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.[7] These other services include

[7] The Data Quality Atlas also assesses reliability of variables related all of these measures of healthcare utilization, such as claim file completeness, service use information, provider information, and more. However, because we use such a wide array of variables across the 32 separate state-specific models—including some constructed measures that draw on multiple facets of the underlying data—we do not restrict which variables are included in the predictive model based on the Data Quality Atlas assessment. In addition to the logistical challenge restricting on a case-by-case basis would pose, our approach produces models that appear to be extremely reliable (based on the measure of AUC discussed below in Section IV). Even if the data are of poor quality for certain variables, there is no reason to expect that quality to differ systematically for SSI recipients and non-SSI recipients. This sort of measurement error

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;[8] the type of provider who treated the patient;[9] the type of services provided;[10] the benefit type code;[11] and the type of medications prescribed.[12] 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, intellectual disabilities, cerebral palsy, and more—as well as an indicator for if the child was continuously enrolled for eleven or more months of the year. Finally, we also included several other measures for intensity of care, such as whether the child had any inpatient stays or emergency department visits and the length of those encounters. In total, we considered approximately 1,300 variables related to health care utilization.

A descriptive comparison indicates that child SSI recipients have much more intensive health care utilization than child non-SSI recipients (Table 1).[13] For example, the claims of child

would make any such variables less predictive, which should lead the model to place lower weight on such unreliable variables.

[8] 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.

[9] This can include providers like a "Physician" or "Speech Language Pathologist." In total, there are 57 unique provider types.

[10] 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.

[11] 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.

[12] 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."

[13] While this comparison of child non-SSI recipients includes some who are potentially eligible, children with high likelihood of eligibility cannot meaningfully change the average for the overall group because so few children have a high likelihood of eligibility. As shown in Table 1, there are over 29 million child Medicaid beneficiaries not receiving SSI. As shown below in Table A5, about 1.4 million child Medicaid beneficiaries not receiving SSI have a predicted probability of SSI eligibility over 10 percent (4.6 percent of all non-SSI recipients). Fewer than 100,000

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), 13 times as likely to have another developmental delay chronic condition (18.4 percent versus 1.4 percent), and more than 20 times as likely to have intellectual disabilities or cerebral palsy. 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.

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.[14] 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 family income data that the DQ Atlas considers as "low concern") to show that our results were similar when we considered several variations to

have a probability over 40 percent (0.3 percent of all non-SSI recipients). This group therefore is too small to meaningfully change the average values in Table 1. This is particularly so for the indicator variables (such as having a specific condition), where outliers cannot skew the average.

[14] 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/

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 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 Table A1.

IV. 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 expected to be potentially eligible by summing up the probabilities across all non-SSI recipients, and in turn estimate the take-up rate for SSI.

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, which occurs when the model too closely hews to the data used in training the model but does not perform well out-of-sample (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. 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 to perform out-of-sample validation.

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.[15] We used the same exact approach in terms of specifying the random forest model, such as hyperparameters (key parameters that guide how the model learns and is estimated) 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.

[15] To ensure the model was computationally feasible, 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 child Medicaid recipients in the state.

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 models from four states for simplicity of presentation: Arkansas, Louisiana, Massachusetts, and Colorado (we also show results for California, New York, Pennsylvania, and Texas—high population states—in the appendix). 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).[16] Additionally, Massachusetts and Colorado have reliable income data (considered as "low concern" in the DQ Atlas), which allow us to explore the sensitivity of our findings to including income in the model, as discussed previously.

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.

[16] 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.

For indicators on types of health care utilization, we excluded characteristics that met both of the following 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[17] between SSI recipients and non-SSI recipients is less than 2 standard deviations.[18] If we exclude the indicator for any utilization, we then also exclude the continuous measure for number of claims of that type.[19] 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." Table A2 lists the variables most frequently ranked among the top 50 most important features across the 32 state-specific models in 2019. The table also reports the number of states for which each variable is in the top 50 for the 2017 and 2018 models. The correlation between each feature's relative rank is around 0.5 between each two-year pair for the 2017, 2018, and 2019 models. Together, these trends indicate relative stability in the models, even when estimating the model separately for different years.

Across all states, the random forest model appears to generate reasonable and reliable estimates of the probability of SSI eligibility. We calculated the AUC, or area under the curve, for each state, both overall and for the untrained observations (Table A3); using the untrained observations validates that our model performs well out-of-sample. The AUC is a useful way to

[17] 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.
[18] This second criterion can actually be thought of as including rare characteristics that differ substantially between child SSI and non-SSI recipients.
[19] 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.

capture the overall diagnostic accuracy of a test in a single number.[20] As established in Mandrekar (2010), an AUC value of 0.7 to 0.8 indicates the model is acceptable, 0.8 to 0.9 indicates the model is excellent, and more than 0.9 indicates the model is outstanding. Out of the 32 states, nine have scores on the untrained data that are outstanding, 22 have scores that are excellent, and only one state has a score that is acceptable. Additionally, Figure 3 shows the percentage within each bucket covering a range of 5 percent probability that are on SSI for the four main states (e.g., the first point on the left of each graph represents Medicaid recipients with a 0 to 5 percent estimated probability). 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. However, it is notable that despite the high rates of diagnostic accuracy indicated by the AUC, the graphs all show similar patterns of deviations tending to be above the 45-degree line, with overpredicting low-likelihood and underpredicting high-likelihood. This may stem from the fact that 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. Figure 3 shows that nobody in the four states has a predicted probability above 90 percent. Nonetheless, 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 consistent when estimating the model out-of-sample: the pattern is nearly identical using

[20] Intuitively, the AUC assesses the accuracy of the model's predictions. It is based on the model's sensitivity and specificity, which depend on the ratios of true and false positives and negatives.

only the testing sample that was left out when estimating the model (Figure A1). Figures A2 and A3 show the analogous graphs for the four high population states.

V. RESULTS

We present three primary sets of results. First, we provide an estimate of the current take-up rate for child SSI, using the expected number of potentially eligible 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. Finally, we also test the sensitivity of our results to the inclusion of income in the model.

A. Child SSI Take-Up Rate

The take-up rate for child SSI is approximately 70 percent. To arrive at this estimate, we first calculated the expected value of the number of potentially eligible non-SSI recipients. This entailed summing the probabilities across non-SSI recipients with a predicted probability of at least 5 percent across the 32 “high-match” states. This leads to an expected increase of 371,662, or a 41.5 percent increase relative to the number of current child SSI recipients identified in the Medicaid data. Applying this percentage to the number of child SSI recipients in “low-match” states that are excluded from our analysis from SSA administrative data in 2019, 108,852 more children could be eligible in the these states. In total, this leads to our preferred estimate of 480,514 children who are potentially eligible but not receiving SSI (Table A4). That in turn corresponds to a take-up rate of 70.7 percent – by dividing the 1,156,721 children receiving SSI by the 1,637,235 children estimated to be eligible (number receiving plus expected number potentially eligible).

If we instead included all children in the estimate, rather than those with a probability exceeding 5 percent, the take-up rate would be 57 percent. As noted above, it is extremely rare for children to have high probabilities of SSI receipt. Yet because there are so many children who are non-SSI recipients, summing even very small probabilities across large swaths of Medicaid beneficiaries leads to a larger estimate. In practice, we prefer the estimate focusing only on those with a probability exceeding 5 percent. From a policy perspective, efforts intended to promote enrollment among many people with such a low probability would be much more inefficient than narrowly targeted efforts among those with higher probabilities. Additionally, as we show below, even those with predicted probabilities as low as 10 percent frequently have health care utilization that is similar to that of the average SSI recipient.[21]

Because some states do not automatically confer Medicaid with SSI enrollment (the striped states in Figure 2), a concern might be that we underestimate take-up rates. In particular, our take-up rate estimate in each state implicitly assumes that the number of child SSI recipients in the state is exactly the number of child SSI recipients in Medicaid data (the numbers must be sufficiently close to be a high-match state in order for us to estimate the state-specific take-up rate). Yet states without automatic enrollment may have more child SSI recipients, in which case the take-up rate would be higher. Indeed, four of the five states with the lowest values for Medicaid SSI recipients as a share of SSI recipients from published statistics are these non-autoenrollment states.

We address this potential underestimation in two steps that confirm its effect on the national take-up rate estimate is minimal. First, we re-calculate the state-specific take-up rate in the five high-match non-autoenrollment states, using the (higher) known number of child SSI

[21] Even if we restricted to only calculate the expected number of potentially eligible children using those with predicted probability exceeding 10 percent, the implied take-up rate would be 76.5 percent (see Table A4).

recipients from published statistics.[22] Second, we re-calculate the national take-up rate using a weighted average of this adjusted take-up rate in non-autoenrollment states (72.9 percent) and in automatic enrollment states (70.6 percent), where the weights are based on the total number of child SSI recipients in each type of state (including all states, not just high-match states). As discussed earlier, the non-autoenrollment states are much more likely to not be high-match states because not all SSI recipients need to have Medicaid, so weighting by the prevalence in the Medicaid data would disproportionately undercount these states. The net result is a national take-up rate of 70.9 percent, essentially identical to the 70.7 percent estimate discussed above.

The estimated take-up rate varies by state (Figure 4). The map in Figure 4 presents the take-up rate based on an estimate of the expected value of child non-SSI recipients that excludes children with less than a 5 percent probability. Maps considering other exclusions (no exclusions, those less than 1 percent) are available in Figure A4, while the implied take-up rate in each state is available in Table A4. The states with the lowest take-up rates are Ohio (68.2 percent), Maryland (68.3 percent), New York (68.5 percent), and Arizona (68.7 percent). Interestingly, there is no relationship between a state's current SSI participation per capita and the take-up rate: when we regress the take-up rate on SSI participation per capita in the state, the coefficient is small and not significant.[23]

Our approach can also generate estimates of the take-up rate at the county level, which might be especially helpful for targeting outreach in specific geographic areas (Figure A5). Within states, the take-up rate often varies across county. These county-level statistics might be

[22] Importantly, both Rennane and Dick (2023) and Rupp and Riley (2016) find rates of Medicaid coverage in non-automatic enrollment rates that are only modestly lower than that in automatic enrollment states, signifying that the vast majority of children in these non-automatic enrollment states still receive Medicaid. This means we are not missing an even larger portion of child SSI recipients, which would lead this modified estimate to still be too low.
[23] The coefficient is also negative, indicating that in states where more children receive benefits, take-up is lower. This is the opposite of what might be expected, where states with higher SSI participation have that higher participation because they capture more of the potentially eligible beneficiaries.

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 2024b). 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.

B. 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.[24] The structure of these figures (such as Figure 5) 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.

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 5). 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

[24] Table A5 gives a sense of the distribution of the underlying predicted probabilities. It shows the number of children nationally not receiving SSI with a probability exceeding each threshold, as well as the implied take-up rate when considering the current number of SSI recipients, for the 2017, 2018, and 2019 models. The similarity across years shows the stability of the model. The implied take-up rate for 2019 depending on a given probability threshold—estimated assuming that everyone above a given probability threshold is in fact eligible, but not placing greater weight on those with higher probabilities as our primary model does—is shown in Figure A6. Table A6 reports state-specific numbers for 2019 only. These in turn correspond to an estimate of the number potentially eligible using a direct classification approach from the model as it treats everyone above the given thresholds as potentially eligible, regardless of the probability.

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. Given that having a disability is a requirement to qualify for SSI, and that having a disability typically involves more intensive usage of care, it makes sense that those where the model is most certain of eligibility are those who (often) have the most intensive usage of care.

Many of the children not currently receiving SSI with the highest probability have claims that signify they have a developmental delays chronic condition (Figure 6). 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. Patterns for the presence of other chronic conditions—intellectual disabilities, learning disabilities, and cerebral palsy—that are consistently found to be important features within the model (Table A2) follow similar general patterns (Figures A7, A8, and A9).

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 (Figure A10). 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. These patterns are 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.[25]

C. Importance of Income as a Predictive Variable

Our results are similar when including or excluding family income in the predictive model (Table 2). Because almost half of states have income data considered "unusable" by the DQ Atlas, 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 income data considered "low concern", 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 2 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.

[25] For the six patterns of claims presented in the paper and appendix, we also present analogous graphs in Figures A11 through A16 for four high population states: California, New York, Pennsylvania, and Texas.

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 (Table A7). 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.

VI. UNDERSTANDIG CROSS-STATE VARIATION IN CHILD SSI PARTICIPATION

Numerous papers highlight that child SSI participation varies substantially across local areas (e.g., Aizer, Gordon, Kearney 2013; Schmidt and Sevak 2017; Levere, Hemmeter, and Wittenburg 2024a). 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, we first identified children who had one of the following four chronic conditions—ADHD, behavioral disorders, other developmental disorders, and cerebral palsy. These were mostly non-overlapping (i.e., the behavioral disorders also conditioned on not having ADHD) but allowed for the presence of other chronic conditions. Next, we found the typical claims behavior among child SSI recipients at the 25th percentile of cost among those with each condition. We focused specifically on children in North Carolina because it was the largest state that had a high percentage of children who were not covered by managed care plans, and thus can calculate the cost of claims directly.[26] Finally, we feed each of these four profiles through each of the 32 state-specific models to estimate the probability of SSI receipt associated with each condition in each state.[27] The primary goal of this analysis is to show the variation across states given a fixed pattern of health claims, not to draw conclusions about the specific conditions (the claims patterns are not intended to be nationally representative of a particular condition).

Children with the same condition and health claims experience very different probability of SSI receipt across states (Figure 7). 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

[26] North Carolina has similar patterns of utilization among SSI recipients, non-SSI recipients, and potentially eligible SSI recipients as the other states presented throughout the paper and appendix – see Figure A17. Figure A18 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.

[27] 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.

panels. However, the likelihood of receipt is typically correlated within state across conditions – 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 7). 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 current level of child SSI participation cannot explain the predicted probability of SSI receipt.

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.

VII. CONCLUSION

We used machine learning tools to identify children potentially eligible for SSI based on their patterns of health care utilization. The estimates imply that the take-up rate for child SSI is about 70 percent. Children most likely to be potentially eligible often have very intensive usage of health care, frequently exceeding those of current SSI recipients. Importantly, these might underestimate the number of potentially eligible SSI recipients: the estimates are based on training a model on data where SSI take-up is incomplete, because of the reasons cited at the outset (e.g., administrative burden, limited awareness, stigma). In other words, the model predicts a probability of SSI eligibility, holding fixed the amount of administrative burden which currently exists in the program. Additionally, the model only considers current Medicaid beneficiaries, whereas some children not receiving Medicaid may also be eligible for SSI. To the extent more children are eligible, our estimate of the take-up rate is an upper bound.

Given this, the critical question is how these findings can ultimately inform policy. Indeed, much of the literature on program take-up focuses on identifying the reasons that take-up is incomplete or policies that might increase take-up (e.g., Aizer 2007; Bhargava and Manoli 2015; Finkelstein and Notowidigdo 2019; Goldin et al. 2022; Linos et al. 2022). 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. Such efforts may be particularly important given the magnitude of the potential increase in SSI recipients were a large share of these potentially eligible children to apply.

Increasing child SSI take-up would be costly, though may also lead to savings in other forms. In 2023, the average monthly payment for children receiving SSI was $793.21, resulting in approximately $9.4 billion in total annual payments (SSA 2024). If the number of child SSI recipients increased by 41.5 percent (the equivalent of going from a current take-up rate of 70.7 percent to 100 percent), that would entail a roughly $3.9 billion annual increase in child SSI payments. Such a calculation assumes that the new child recipients would receive similar average payments as current child SSI recipients. Yet it is also worth noting that receiving SSI may generate savings in other areas: Deshpande and Mueller-Smith (2022) show that the savings from lower SSI payments when people exit SSI are nearly entirely offset by increases in the cost associated with increased criminal activity. We are unaware of reliable estimates as to the impacts of SSI receipt on health, while SSI receipt may also lead to lower labor market activity (Deshpande 2016, Levere 2021). A full cost-benefit analysis of the potential increased take-up, however, goes beyond the scope of this paper.

More generally, our results suggest that leveraging big data and machine learning 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.

Acknowledgments and Disclaimers

The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) as part of the Retirement and Disability Research Consortium. The findings and conclusions 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, expressed 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. We are grateful to Carolina Castilla, Laura Dague, Jeffrey Hemmeter, Kathleen Mullen, Jody Schimmel Hyde, Lucie Schmidt, and participants at an SSA work-in-progress seminar; the 2024 ASHEcon Annual Conference; the 2024 APPAM Annual Conference; and the Colgate University economics workshop for feedback on early findings. Claire Erba and Addison Larson provided excellent research assistance.

References

Aizer, Anna. 2007. "Public health insurance, program take-up, and child health." *Review of Economics and Statistics* 89(3): 400-415.

Aizer, Anna, Nora Gordon, and Melissa Kearney. 2013. "Exploring the Growth of the Child SSI Caseload." NBER Disability Research Center Paper no. NB 13-02, Cambridge, MA: National Bureau of Economic Research.

Akee, Randall, William Copeland, and Emilia Simeonova. 2024. "Child Mental Health, Family Circumstance, and Long-Term Success: The Effect of Household Income." *Journal of Human Resources* 59(S): S77-S107.

Battaglini, Marco, Luigi Guiso, Chiara Lacava, Douglas L. Miller, and Eleonora Patacchini. 2024. "Refining Public Policies with Machine Learning: The Case of Tax Auditing." *Journal of Econometrics*, forthcoming.

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.

Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment." *American Economic Review* 105(11): 3489-3529.

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

Burns, Marguerite, and Laura Dague. 2017. "The effect of expanding Medicaid eligibility on Supplemental Security Income program participation." *Journal of Public Economics* 149: 20-34.

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.

Corredor-Waldron, Adriana, and Janet Currie. 2024. "To what extent are trends in teen mental health driven by changes in reporting? The example of suicide-related hospital visits." *Journal of Human Resources* 59(S): S14-S40.

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.

Davies, Paul S., Minh Huynh, Chad Newcomb, and Paul O'Leary. 2001/2002. "Modeling SSI financial eligibility and simulating the effect of policy options." *Social Security Bulletiun* 64(2): 16-45.

Decker, Sandra L., Salam Abdus, and Brandy J. Lipton. 2022. "Eligibility for and enrollment in Medicaid among nonelderly adults after implementation of the Affordable Care Act." *Medical Care Research and Review*, 79(1): 125-132.

Deshpande, Manasi. 2016. "Does welfare inhibit success? The long-term effects of removing low-income youth from the disability rolls." *American Economic Review*, 106(11): 3300-3330.

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.

Deshpande, Manasi, and Michael Mueller-Smith. 2022. "Does welfare prevent crime? The criminal justice outcomes of youth removed from SSI." *The Quarterly Journal of Economics* 137(4): 2263-2307.

Finkelstein, Amy, and Matthew J. Notowidigdo. 2019. "Take-up and targeting: Experimental evidence from SNAP." *Quarterly Journal of Economics* 134(3): 1505-1556.

Foote, Andrew, Michel Grosz, and Stephanie Rennane. 2019. "The effect of lower transaction costs on social security disability insurance application rates and participation." *Journal of Policy Analysis and Management* 38(1): 99-123.

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. 2024. "Administrative Burden and Procedural Denials: Experimental Evidence from SNAP." *American Economic Journal: Economic Policy* 16(4): 316-340.

Goldin, Jacob, Tatiana Homonoff, Rizwan Javaid, and Brenda Schafer. 2022. "Tax filing and take-up: Experimental evidence on tax preparation outreach and benefit claiming." *Journal of Public Economics* 206: 104550.

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. 2024. "Machine Learning Can Predict Shooting Victimization Well Enough to Help Prevent It." *The Review of Economics and Statistics*, forthcoming.

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/

Ko, Wonsik, and Robert A. Moffitt. 2024. "Take-up of social benefits." *Handbook of Labor, Human Resources and Population Economics*: 1-42.

Kopczuk, Wojciech and Cristian Pop-Eleches. 2007. "Electronic filing, tax preparers and participation in the Earned Income Tax Credit." *Journal of Public Economics*, 91(7-8): pp.1351-1367.

Levere, Michael. 2021. "The labor market consequences of receiving disability benefits during childhood." *Journal of Human Resources* 56(3): 850-877.

Levere, Michael, Jeffrey Hemmeter, and David Wittenburg. 2024a. "The Local Declines in Child SSI Applications and Awards During COVID" *Journal of Disability Policy Studies*, forthcoming.

Levere, Michael, Jeffrey Hemmeter, and David Wittenburg. 2024b. “The Importance of Schools in Driving Children’s Applications for Disability Benefits.” *Journal of Public Economics* 239: 105239.

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.

Linos, Elizabeth, Allen Prohofsky, Aparna Ramesh, Jesse Rothstein, and Matthew Unrath. 2022. “Can nudges increase take-up of the EITC? Evidence from multiple field experiments.” *American Economic Journal: Economic Policy* 14(4): 432-452.

Mandrekar, Jayawant N. 2010. “Receiver operating characteristic curve in diagnostic test assessment.” *Journal of Thoracic Oncology* 5(9): 1315-1316.

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

Mullainathan, Sendhil and Jann Spiess. 2017. “Machine Learning: An Applied Econometric Approach.” *Journal of Economic Perspectives* 31(2): 87-106.

National Center for Health Statistics (NCHS). 2021. “Health, United States, 2019: Table 007.” National Center for Health Statistics. https://www.cdc.gov/nchs/hus/contents2019.htm.

National Survey on Drug Use and Health (NSDUH). 2019. “2002–2019 National Survey on Drug Use and Health Final Analytic File (Codebook).” National Survey on Drug Use and Health. https://bit. ly/3e1d8Sg

Rennane, Stephanie, and Andrew Dick. 2023. “Effects of Medicaid automatic enrollment on disparities in insurance coverage and caregiver burden for children with special health care needs.” *Medical Care Research and Review* 80(1): 65-78.

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

Rupp, Kalman, and Gerald F. Riley. 2016. “State Medicaid eligibility and enrollment policies and rates of Medicaid participation among disabled supplemental social security income recipients.” *Social Security Bulletin* 76(3): 17-40.

Sansone, Dario and Anna Zhu. 2023. "Using Machine Learning to Create an Early Warning System for Welfare Recipients." *Oxford Bulletin of Economics and Statistics* 85(5): 959-992.

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. "Safety Net Program Spillovers and Administrative Burden." *National Tax Journal* 77(1): 199-221.

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. 2024. "SSI Annual Statistical Report, 2023." Baltimore, MD. Available at: https://www.ssa.gov/policy/docs/statcomps/ssi_asr/2023/

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
Intellectual disabilities chronic condition	10.5	0.4
Cerebral palsy chronic condition	4.3	0.1
Prescribed ADHD medications	28.3	4.3
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

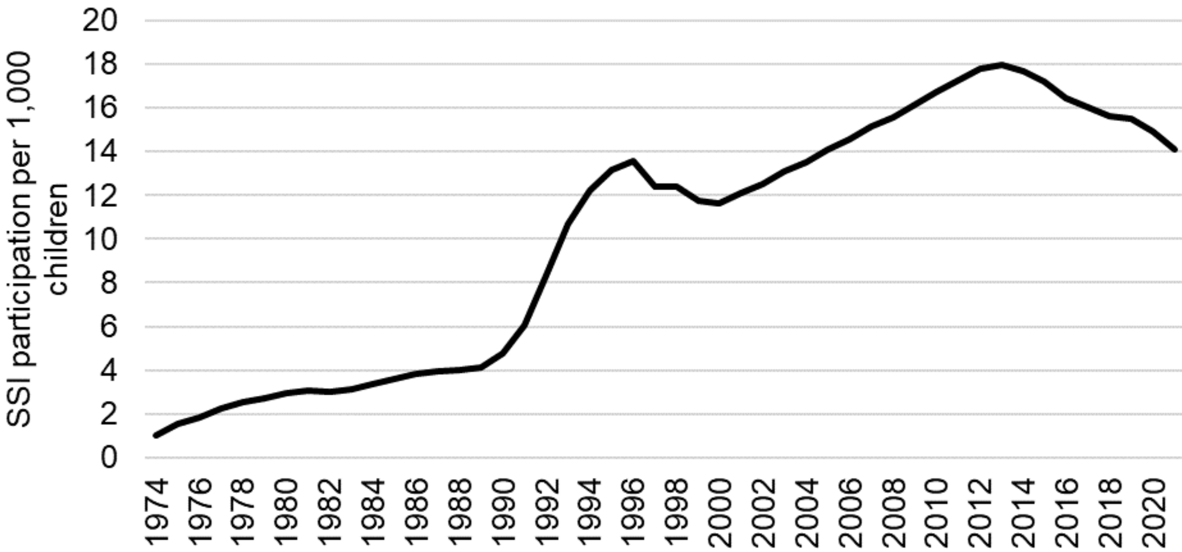
Note: Includes all child Medicaid recipients within the 32 "high-match" states.
Source: 2019 TAF data.

Table 2. *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

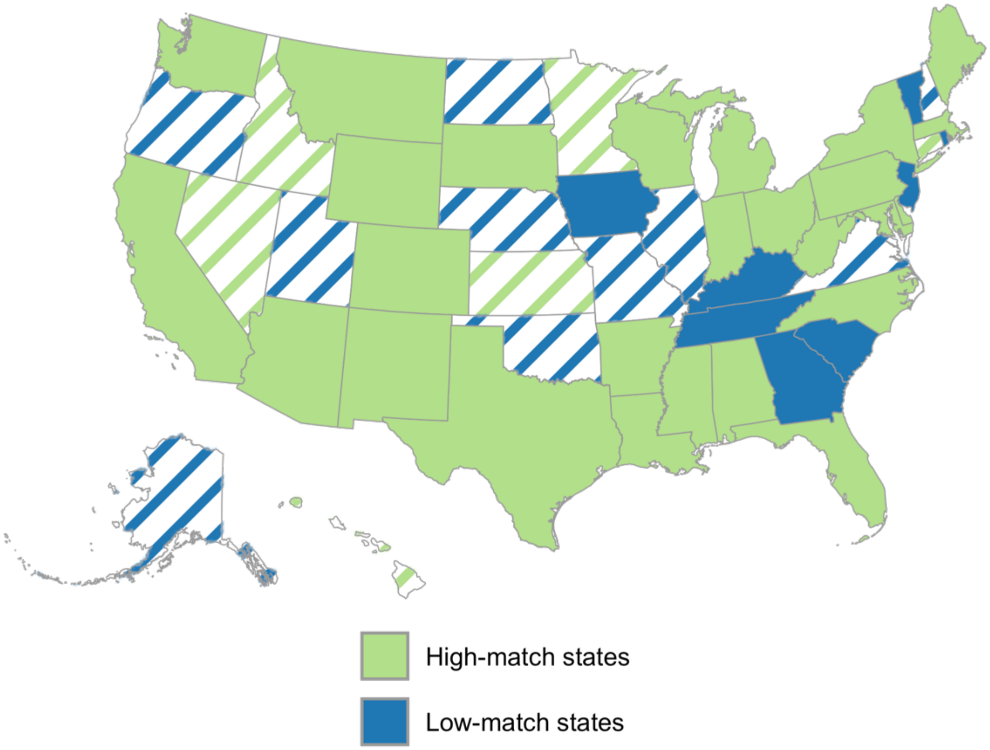
Notes: The numbers in the table show the percentage of child Medicaid beneficiaries who have a probability that exceeds the predicted probability threshold in the random forest model that includes income as a predictor who also have a probability exceeding the same threshold in the random forest model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

Figure 1. *Child SSI Participation, 1974-2021*



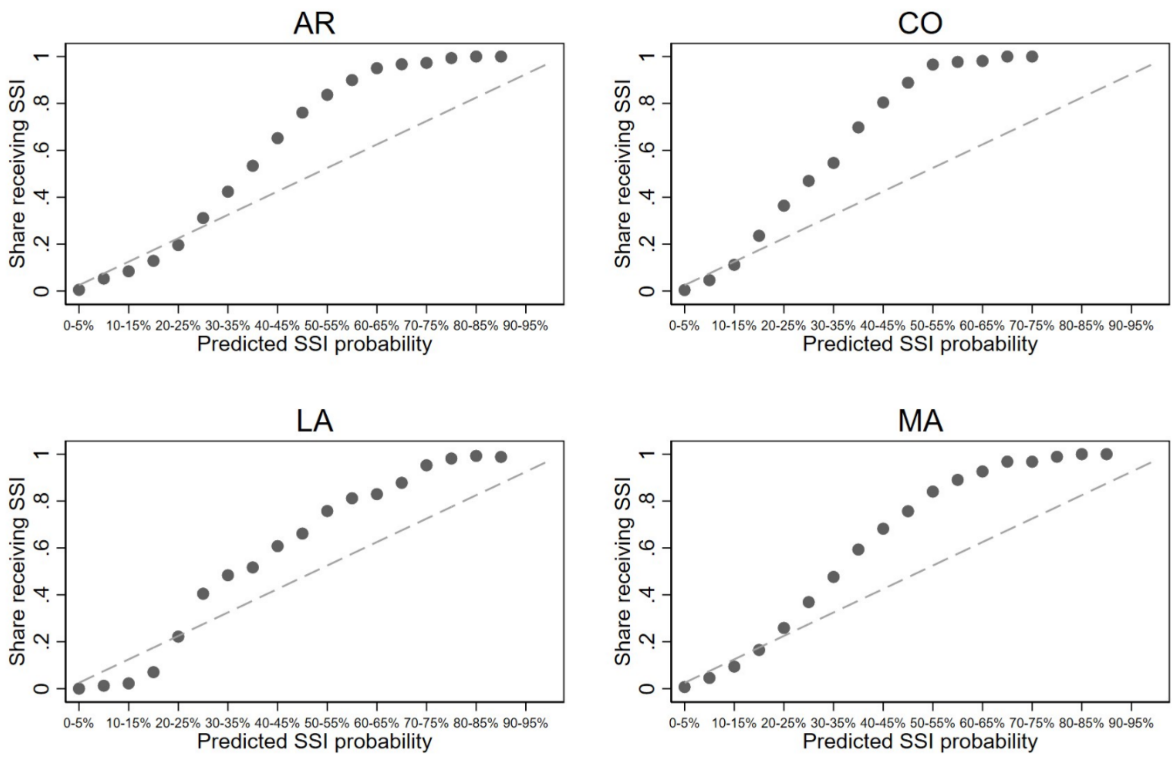
Source: SSA (2022).

Figure 2. *States with Reliable SSI Indicator*



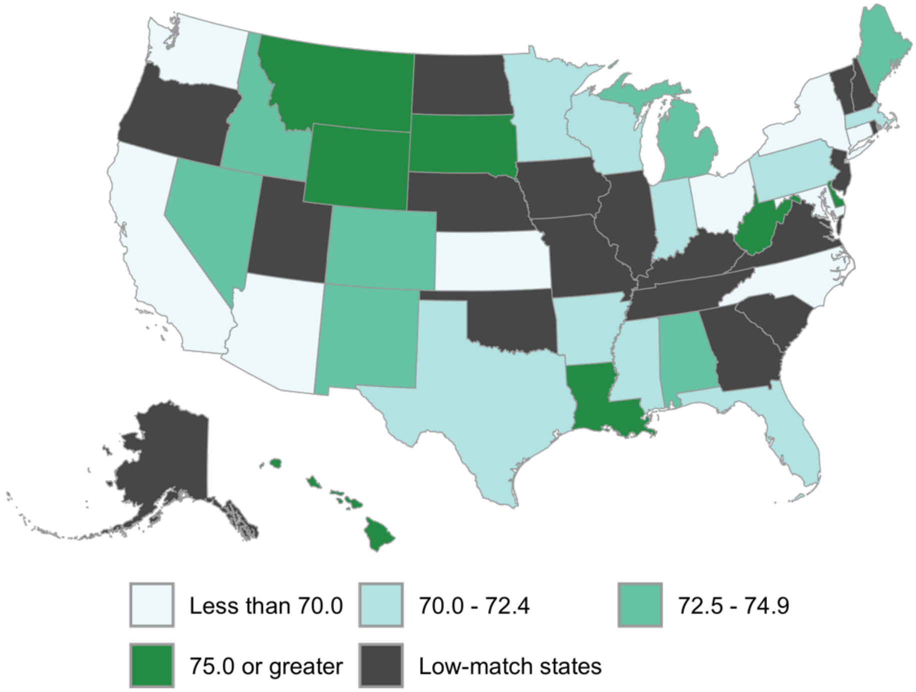
Note: High-match states are those where the SSI indicator available in Medicaid data is reliable: specifically, the number of children receiving SSI in the state according to Medicaid data closely matches the number of children receiving SSI in the state according to official SSA statistics. As discussed in the text, this means the two estimates of the child SSI population are within 13 percent of each other in 2017, 2018, and 2019. Low-match states are those where the SSI indicator available in Medicaid data is not reliable and the two estimates are not within 13 percent of each other in at least one year. High-match states are included in our analysis, whereas low-match states are not. 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.

Figure 3. *Distribution of SSI Receipt by Predicted Probability of SSI Receipt*



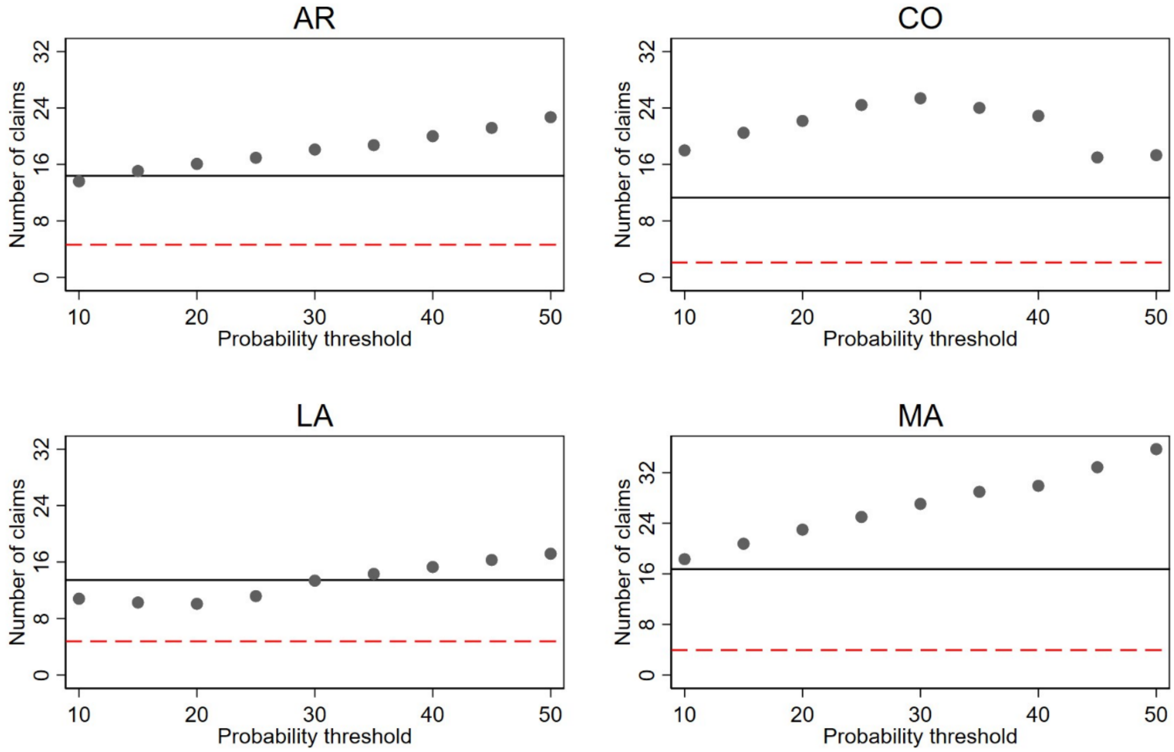
Notes: Each point indicates the empirical share of children receiving SSI (on the y-axis) among each ventile of predicted SSI probability (shown on the x-axis). The 45-degree line thus represents the line of perfect fit: where we might expect between 20-25% of people with a predicted probability of 20-25% to be actually receiving SSI. If no children in a state have a probability sufficiently high, that point is excluded from the figure (e.g., nobody in Colorado has a predicted probability above 80 percent). The difference between the observed values and the 45-degree line is also roughly captured by the model's AUC, which is shown in Table A3.
Source: Authors' calculations using 2019 TAF data.

Figure 4. *Child SSI Take-up Rates*



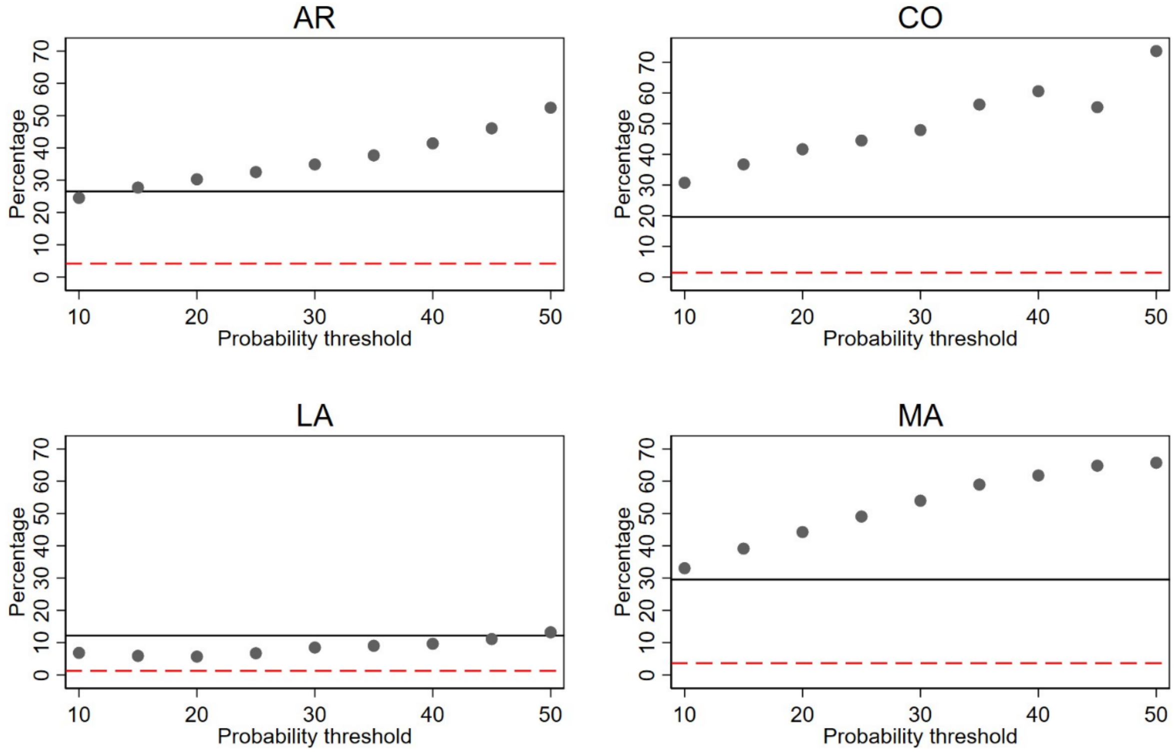
Source: Authors' calculations using 2019 TAF data.

Figure 5. *Number of Prescription Drug Claims, by State, Receipt of SSI, and Estimated Probability of SSI Receipt*



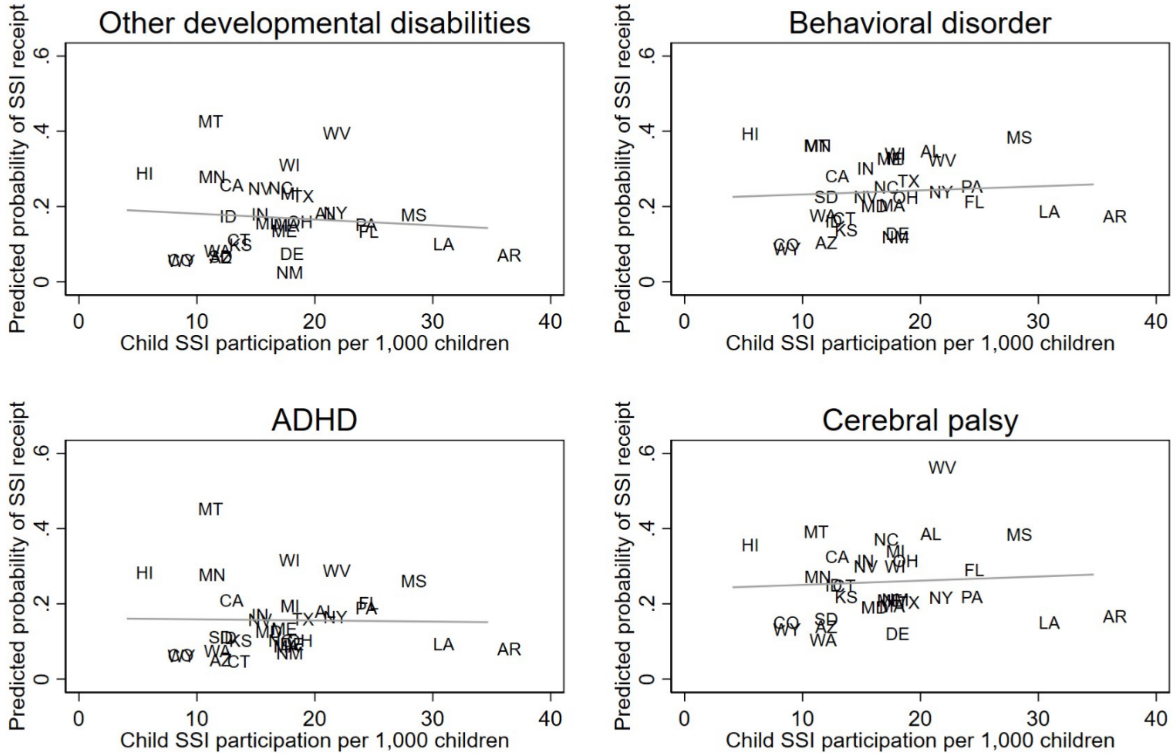
Notes: The black solid line represents the average value for all child SSI recipients, while the red dashed line represents the average value for all child non-SSI recipients. Each circle indicates the average among all non-SSI recipients in the state with probability at least that high.
Source: Authors'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

Figure 6. *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.

Figure 7. *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.