



PREPARED FOR

Sika Koudou
Diana Chevere-Flores
Social Security Administration
Annex Building
6401 Security Boulevard
Baltimore, MD 21235

PREPARED BY

Sarah Prenovitz, Laura Peck,
Jason Brinkley, Weston McManus,
Melanie Ward, Michelle Wood,
Sarah Gibson, Tom McCall
Abt Associates
6130 Executive Boulevard
Rockville, MD 20852

August 8, 2023
Final Report

An Investigation on Bias in the Quick Disability Determinations Model

Evaluation, Analysis & Options Report

Contract # 28321318A00040031 Order # 28321322FA0010264

About this Report

The Social Security Administration (SSA) asked Abt Associates to examine the Quick Disability Determinations (QDD) predictive models to determine whether some groups of Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) claimants, specifically those defined by race/ethnicity, sex, age, and medical condition, are systematically excluded from QDD identification. Abt analyzed data from more than 5.5 million SSDI and SSI claimants whose applications were subject to the 2020 QDD prediction algorithm.

Contacts

Project Manager: Sarah Gibson, Principal Associate, sarah_gibson@abtassoc.com

Co-Principal Investigator: Jason Brinkley, Principal Associate, jason_brinkley@abtassoc.com

Co-Principal Investigator: Laura Peck, Principal Scientist, laura_peck@abtassoc.com

Research Team Lead: Sarah Prenovitz, Senior Associate, sarah_prenovitz@abtassoc.com

Project Quality Reviewer: Amanda Parsad, Principal Associate, amanda_parsad@abtassoc.com

Management Reviewer: Michelle Wood, Principal Associate, michelle_wood@abtassoc.com

Report Authors: Sarah Prenovitz; Laura Peck; Jason Brinkley; Weston McManus, Associate, weston_mcmanus@abtassoc.com; Melanie Ward, Associate, melanie_ward@abtassoc.com; Michelle Wood; Sarah Gibson; Tom McCall, Associate, tom_mccall@abtassoc.com

Project Quality Reviewer: Amanda Parsad, Principal Associate, amanda_parsad@abtassoc.com

Management Reviewer: Michelle Wood, Principal Associate, michelle_wood@abtassoc.com

Acknowledgements

This report incorporates information, in places verbatim, from documents the SSA provided Abt Associates on the QDD predictive models, including the Scoring Services for Processing Times and Scoring Services for Allowances business rules; how the fields on the Form SSA-3367, SSA Form 3368, and Form SSA-3820 map to the fields in the QDD predictive models; the order of operations of the QDD predictive models; and trainings the SSA prepared for Abt Associates.

CONTENTS

Acronymns4

Executive Summary5

 Background5

 Key Findings.....6

References.....17

LIST OF EXHIBITS

Exhibit ES-1. Percentage Point Differences in QDD Identification Rate, Adult Modeling Populations	7
Exhibit ES-2. Percentage Point Differences in QDD Identification Rate, Title 16 Children	8

Acronymns

BIFSG	Bayesian Improved First Name Surname Geocoding
DDS	Disability Determination Service
PDD	pervasive developmental disorders
QDD	Quick Disability Determinations
SSA	Social Security Administration
SSAL	Scoring Services for Allowances
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income
SSPT	Scoring Services for Processing Times

Executive Summary

The Social Security Administration (SSA) has enacted an Equity Action Plan as part of its response to Executive Order 13985.¹ As SSA examines its programs and policies to advance equity for all, it is pursuing a wide range of initiatives in several focus areas: Improving Data Collection and Conducting Analyses to Identify Inequities in Programs; Increasing Outreach and Program Participation; Increasing Equitable Service for Underrepresented Claimants in the Disability Application Process; Increasing Gender Equity and Equality; and Increasing Equitable Access to Research Grants and Procurement Opportunities for Historically Black Colleges and Universities, Minority Serving Institutions, and Small and Disadvantaged Businesses. This call order, *An Investigation on Bias in the Quick Disability Determinations Model*, is one example of the specific actions SSA is taking to advance equity in its programs.

Background

Applying for Social Security Disability Insurance (SSDI) or Supplemental Security Income (SSI) benefits can be a lengthy, complex process, in which applicants provide detailed information about their medical condition and other topics and state Disability Determination Service (DDS) examiners assess whether the applicant meets SSA’s definition of disability. Average processing time for initial claims was 223 days in April 2023 (SSA 2023b).

To alleviate some of the burden on claimants that long decision times impose, in 2008 SSA developed the *Quick Disability Determinations (QDD) process*. The QDD algorithm identifies SSDI and SSI claims at the initial level “where a favorable disability determination is highly likely and medical evidence is readily available,” and expedites the review of those claims (SSA, n.d.). The QDD process uses separate computer-based prediction models for each of the four modeling populations—*Title 2, Title 16 Adults, Concurrent Title 2/Title 16, and Title 16 Children*.

Through this call order, Abt Associates examined the QDD identification process to determine whether QDD identification rates differ for claimant groups defined by race/ethnicity, sex, age, and medical condition, and whether any such differences indicate bias. We conducted the analysis in two phases: the first phase examined *whether* there were differences in QDD identification rates, and the second phase examined possible explanations for those differences.

¹ Executive Order 13985, “Advancing Racial Equity and Support for Underserved Communities through the Federal Government,” January 20, 2021, <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>. President Biden signed a second equity-focused Executive Order 14091, “Further Advancing Racial Equity and Support for Underserved Communities” on February 16, 2023, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/02/16/executive-order-on-further-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>.

The analyses focused on differences in QDD identification rates for people with varying identities. We explored differences for claimants who are (1) American Indian/Alaskan Native, Asian/Pacific Islander, Black, Hispanic, or White;² (2) female versus male; (3) of varying ages, for both adults and children; and (4) of varying medical condition categories, distinct for adults and for children. Because applicants for SSDI and SSI have multiple intersecting identities, the Abt team examined these characteristics and groupings concurrently. Examining the characteristics and groupings concurrently means that results can be interpreted as focusing on one set of characteristics while holding all other characteristics constant. For example, if we saw that older claimants were more likely to be identified for QDD, we examined whether that difference was just a factor of older people being more likely to have qualifying medical conditions. Because our estimates of age group differences adjust for differences in medical condition categories we can see the influence of age free from the direct influence of medical condition.

We examined each of the four modeling populations separately. We focused on applicants whose applications were allowed, quickly, at Step 3 of the initial determination process—called here the *QDD-like sample*. By examining differences in QDD identification rates *within* that sample of applicants, we could ascertain how often the QDD algorithm predicts quick allowances for applicants whose applications are otherwise allowed quickly. See the **Summary of Methods** below for more on this sample and analysis methods.

Key Findings

This executive summary (ES) reports the direction and magnitude of the observed differences in QDD identification for each identity characteristic of focus. It also describes possible sources for these differences. In each of the comparisons, the group that is largest in number was selected as the comparison group. For example, for the race/ethnicity comparisons, each racial/ethnic group was compared to White applicants.³

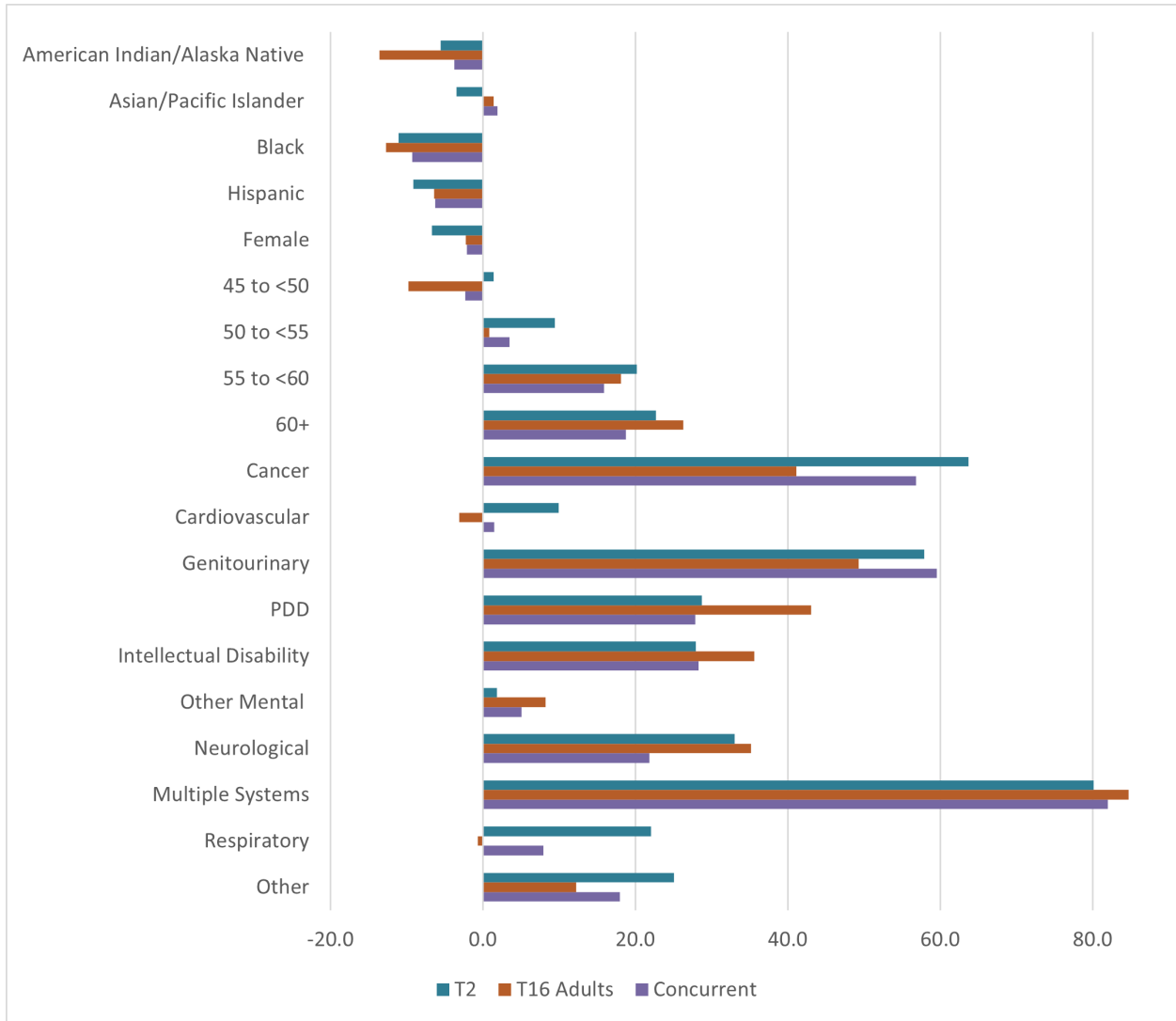
Phase 1 Findings: Investigation of Differences

Exhibit ES-1 displays the Phase 1 findings for the three Adult modeling populations (Title 2, Title 16 Adults, and Concurrent, and ES-2 displays the Phase 1 findings for Title 16 Children. These exhibits display percentage point differences from the excluded groups. Excluded groups are White and male for all modeling populations, age 18 to <45 and musculoskeletal conditions for Adult modeling populations, and age 6 to <13 and pervasive developmental disorders (PDD) for Title 16 Children.

² Because SSA data contain race and ethnicity on only a subset of claims (about 60 percent of adults and about 17 percent of children), we used an established approach—called Bayesian Improved First Name Surname Geocoding (BIFSG)—to predict claimants’ race and ethnicity as proxies for the missing data, permitting us to use all available claims data for analysis.

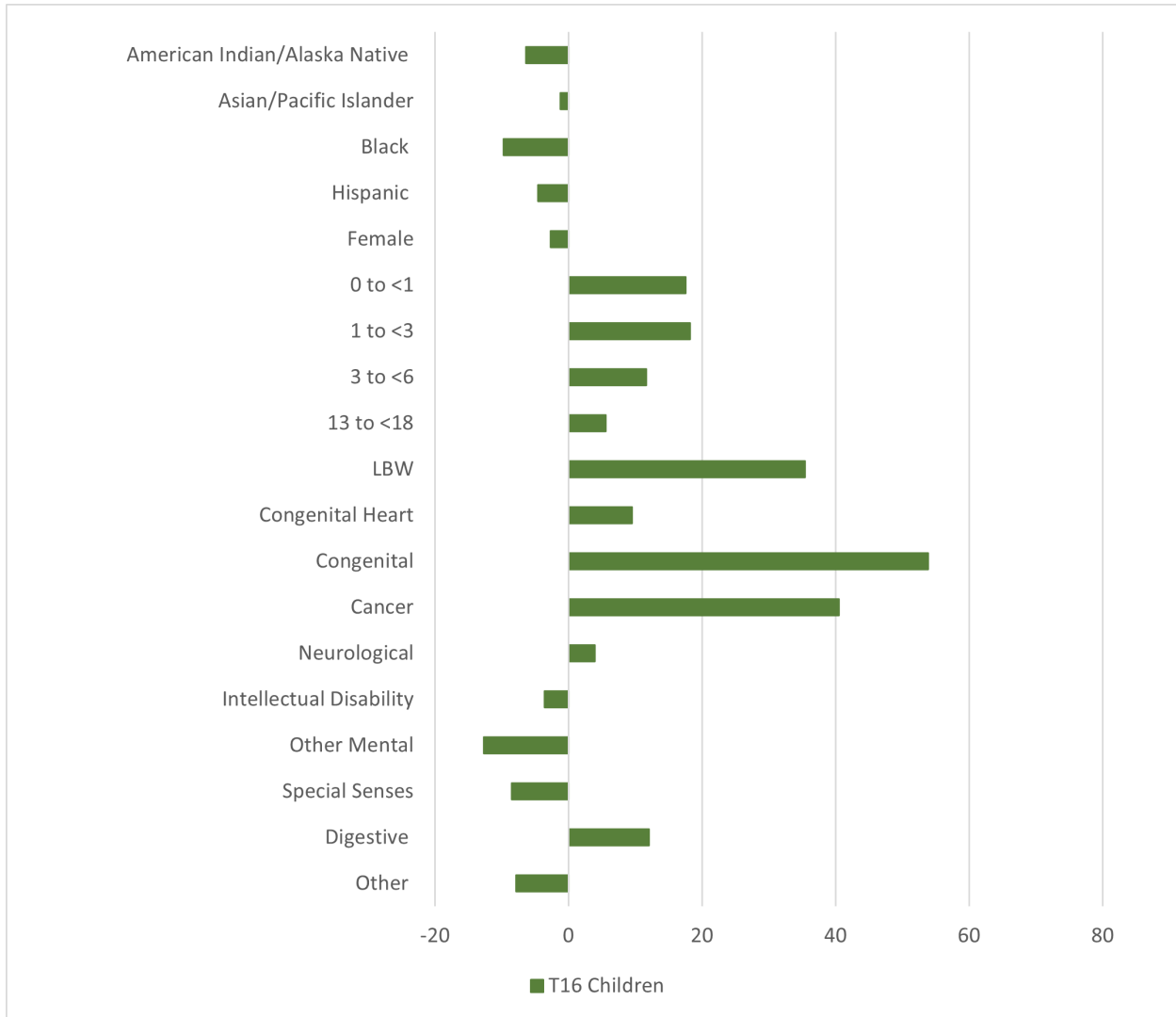
³ Though it is standard analytic practice to compare each group to the largest group, we recognize that comparing non-White racial/ethnic groups to White applicants can appear to imply that the experiences of White applicants is the norm and the standard against which the experiences of other groups are measured. In this report we have followed that standard analytic practice, but we are hopeful that future work will offer viable alternatives to it.

Exhibit ES-1. Percentage Point Differences in QDD Identification Rate, Adult Modeling Populations



Source: Authors' analyses of data from the SSA Structured Data Repository

Exhibit ES-2. Percentage Point Differences in QDD Identification Rate, Title 16 Children



Source: Authors' analyses of data from the SSA Structured Data Repository

Race/ethnicity

Across all modeling populations, compared to White applicants, QDD identification rates are:

- *not meaningfully different* for Asian/Pacific Islander applicants (3.4 percentage points lower for Title 2, 1.4 percentage points higher for Title 16 adults, 1.9 percentage points higher for Concurrent, and 1.2 percentage points lower for Title 16 children);
- *lower* for American Indian and Alaska Native applicants (5.6 percentage points lower for Title 2, 13.6 percentage points lower for Title 16 adults, 3.7 percentage points lower for Concurrent, and 6.4 percentage points lower for Title 16 children);
- *lower* for Black applicants (11 percentage points lower for Title 2, 13 percentage points lower for Title 16 adults, 9 percentage points lower for Concurrent, and 10 percentage points lower for Title 16 children); and

- *lower* for Hispanic applicants (9 percentage points lower for Title 2, 6 percentage points lower for Title 16 adults, 6 percentage points lower for Concurrent, and 5 percentage points lower for Title 16 children).

Summary of Methods

Research Questions. This investigation addresses two research questions: (1) To what extent does the QDD identification process reveal evidence of bias against certain groups of disability claimants? and (2) If there is bias, what are the likely root causes?

Data Sources. The investigation analyzed data from approximately 5.5 million SSDI and SSI claimants who submitted an initial claim between August 1, 2020, and May 17, 2022. In particular, we examined 557,353 SSDI and SSI claims that were allowed at Step 3 with processing times that meet SSA’s definition of “quick” determination. Because claims that are identified for the QDD process are intentionally different from claims overall, we focus on these claims that are allowed at Step 3 with a quick decision time to identify the claims from the full applicant population QDD aims to identify. We refer to these claims used in our analysis as being “QDD-like.”

Phase 1 Analysis. To address Research Question 1, we analyzed the differences in QDD identification rates for claimants of varying characteristics among claimants who are “QDD-like” for each of the four modeling populations. Those modeling populations are Title 2, Title 16 Adults, Concurrent Title 2/Title 16 applicants, and Title 16 Children. Using a linear probability model, our multivariable analysis controlled for race/ethnicity, sex, age, and medical condition to estimate the regression-adjusted differences in QDD identification rates.

Phase 2 Analysis. Where we observed differences in Phase 1, we identified possible sources of bias and designed analyses to investigate them further in Phase 2. These analyses used multivariable analysis. Analyses made changes to the Phase 1 linear probability model by adding or changing variables or changing the analysis sample. We then examined how the magnitude of the observed differences changed.

Exploratory Data Analyses. To understand the QDD algorithm itself, we applied additional tools of data science and selected data visualizations. Analyses involved exploring the inputs (data and documentation) and outputs (QDD scores) of the models; examining the treatment of and relationship of text and non-text variables; and visualizing the output scores across modeling populations, identity groups, and some medical conditions.

Expert Input. In addition to our quantitative analyses, we worked closely with SSA staff to inform analytic decisions and interpret our findings. We also interviewed three external experts, two times each, to solicit insights regarding our Phase 1 results, the possible root causes to explore, and implications for Phase 2 results.

Sex

Compared to males, QDD identification rates are *slightly lower or not meaningfully different* for females (5 percentage points lower for Title 2, 2 percentage points lower for Title 16 adults, 1 percentage point lower for Concurrent, and 3 percentage points lower for Title 16 children).

Age

Adult Age Findings. Across all adult modeling populations (Title 2, Title 16 Adults, Concurrent), QDD identification rates increase with age. Specifically, compared to younger adults (age 18 to <45), QDD identification rates are:

- generally *not meaningfully different* for the 45 to <50 and 50 to <55 age categories, with two exceptions: (1) Title 16 Adults in the age 45 to <50 group are about 10 percentage points *less* likely to

be QDD identified than the youngest adult group; and (2) Title 2 Adults in the age 50 to <55 group are about 10 percentage points *more* likely to be QDD identified.

- consistently *higher* for the two oldest age categories: the 55 to <60 age group has rates of about 16 to 20 percentage points greater than the youngest adult age group (20 percentage points higher for Title 2, 18 percentage points higher for Title 16 adults, 16 percentage points higher for Concurrent); and the 60+ age group has rates of about 19 to 26 percentage points greater (23 percentage points higher for Title 2, 26 percentage points higher Title 16 adults, 19 percentage points higher Concurrent). This implies a disadvantage for those in the youngest age group relative to those in the two oldest age categories.

Child Age Findings. Compared to children whose ages run from 6 to <13 years, QDD identification rates are:

- *higher* for all other child age categories: the two youngest groups (age 0 to <1 and age 1 to <3) have a QDD identification rate about 19 points greater; and the oldest group (age 13 to <18) has a rate of about 6 percentage points greater.

Medical Condition

Adult Medical Condition Findings. For the Adult modeling populations, QDD identification rates differ more across medical conditions than across other characteristics, with differences as much as 85 percentage points in magnitude in one instance. Among Adult modeling populations, compared to claimants alleging musculoskeletal conditions, QDD identification rates are:

- *much higher* for claimants with multiple body systems diagnoses (81 to 85 percentage points greater, depending on the modeling population);
- *much higher* for those with cancer diagnoses (41 to 64 percentage points greater, depending on the modeling population);
- *much higher* for those with genitourinary conditions (49 to 60 percentage points greater, depending on the modeling population);
- *higher* for those with “other medical” conditions (12 to 25 percentage points, depending on the modeling population);
- *higher* for those with PDD, intellectual disability, or neurological disorders (between 22 and 43 percentage points, depending on the modeling population);
- *similar or higher* for those with respiratory disorders (-1 to 22 percentage points, depending on the modeling population); and
- *relatively similar* for those with either cardiovascular conditions or “other mental disorders” (with 5 and 10 percentage points, depending on the modeling population).

Child Medical Condition Findings. Among the Title 16 Children, compared to those with PDD, QDD identification rates are:

- *higher* for those with congenital abnormalities or low birth weight (59 and 35 percentage points, respectively);

- *higher* for those with cancer (41 percentage points);
- *slightly higher* for those with digestive disorders or congenital heart disorders (12 and 10 percentage points, respectively);
- *not meaningfully different* for those with neurological disorders or intellectual disability (+/- 4 percentage points); and
- *lower* for those with special senses diagnoses, other mental disorders, or “other medical” conditions (8 to 13 percentage points, depending on the condition).

In brief, the differences in QDD identification rates for children with varied medical conditions, relative to children with PDD, vary substantially, with conditions more straightforward to diagnose or document being identified, perhaps not surprisingly, at higher rates.

Phase 2 Findings: Investigation of Sources of Observed Differences

In Phase 2, we consider possible sources of the differences observed. The findings regarding differences in QDD identification based on race/ethnicity, sex, age, and medical condition generally hold even when we control for other variables (e.g., education, geographic location, with the specific magnitudes being qualitatively similar. The investigation offers evidence on several possible explanations—or root causes—of these differences. Some of these explanations reflect differences in application quality resulting from “external” or “upstream” sources of bias—that is, biases that exist in the larger world. Because QDD starts with the completed application, the QDD process itself is not the source for such differences. Though QDD could be designed to offset these biases, this is not the current goal of the process as we understand it. Our analyses explore the extent to which different observable factors might explain the differences in QDD identification rates that we observe, but do not consider all possible factors. We are unable to determine conclusively whether the differences we observed reflect bias within the QDD process, rather than sources of bias outside of the QDD process.

To what extent does medical care quality and resource availability explain the race/ethnicity or age differences?

Medical care quality and medical resource availability vary substantially across the United States (National Academies of Sciences, Engineering, and Medicine, 2017). One dimension along which disparities exist is location, with metropolitan areas having greater quality and resources, in part for their density, compared to micropolitan or rural areas, where care is both sparser and of lower quality (Allred et al., 2015).⁴ Location—used as a proxy for medical care/resource quality—appears to be part of the explanation for race/ethnicity differences in QDD identification rates: Although it reduces differences for some groups, it exacerbates differences for others. Children living in rural areas are 4.5 percentage points less likely to be identified for QDD, holding race/ethnicity, sex, age, and medical condition constant. For adults, location does not appear to meaningfully influence QDD identification rates.

⁴ For example, within Maryland areas within Worcester, Talbot, Queen Anne’s, and Dorchester counties are classified as micropolitan in our analysis.

To what extent do barriers to completing a quality SSDI/SSI application explain the race/ethnicity differences?

We considered five testable barriers related to completing the SSDI/SSI application: (1) the applicant's level of education; (2) whether the applicant can speak and understand English; (3) whether the applicant is experiencing homelessness; (4) whether an appointed representative assisted with the application; and (5) two geographic indicators of the applicant's location (region of the country, metropolitan location versus micropolitan or rural location). Information on education, understanding English, and assistance by an appointed representative are collected on the application. Collectively, this set of indicators reduced the extent of observed difference in QDD identification rates by a relatively small amount (1 to 4 percentage points) for the Black and Hispanic groups and increased differences between the White and Asian/Pacific Islander groups, with Asian/Pacific Islander applicants relatively more likely to be identified for QDD when controlling for barriers. We concluded that application barriers account for a small but non-zero proportion of the difference in QDD identification rates between White and Black applicants. We also concluded that Asian/Pacific Islander applicants can face a lower prevalence of application barriers (or higher prevalence of factors that mitigate such barriers) than do White applicants, otherwise we would have observed lower QDD identification rates for Asian/Pacific Islander applicants.

Because appointed representatives can help address application barriers, we also compared differences in QDD identification by race/ethnicity for applicants who had an appointed representative at the time of the application versus differences in QDD identification for applicants who did not. We found that, relative to White applicants, QDD identification rates for other racial/ethnic groups are higher among applicants with a representative than among those without. We found that QDD identification rates are on average *lower* for those in the QDD-like sample with representatives than those without, after controlling for race/ethnicity, sex, age, and medical condition. That is, we saw less evidence of White applicants having an advantage relative to applicants of other race/ethnicity groups for those with representatives, but also that those with representatives have lower QDD identification rates overall compared with those who do not have representatives.

We also considered how the QDD predictive models handle misspelled words. We noted that various incorrect spellings of a given word appear to receive different weights and that these weights differ from the weight assigned to the correct spelling of the word.

To what extent does the definition of the analysis sample adopted for this investigation explain the adult age differences in QDD identification?

The Phase 1 analysis drew comparisons among applicants whose applications were allowed quickly at Step 3. However, SSA's stated objectives for QDD identification are allowance and speed, without regard to the step in the determination process at which that quick decision is made. It is possible that this more restrictive sample definition explains the differences in QDD identification rates we observed across adult applicants. When we broadened the analysis sample to include quick allowances at any step in the determination process, differences in QDD identification rates between older and younger adults declined

but did not disappear.⁵ As such, the analysis sample definition appears to partially explain adult age differences in QDD identification.

To what extent does the age group definition explain adult age differences observed for Title 16 Adults?

The youngest adult age category includes claimants who range from 18 to <45 years old. For Title 16 Adults, a substantial proportion of applicants in this age group are persons with long-standing conditions who are applying as they age into adulthood; such young adults are far less prevalent in the Title 2 and Concurrent applicant populations. Applicants with lifelong disabilities who are aging into adulthood could have different conditions and medical histories from those who experience a new disability onset in young to middle adulthood in ways that make allowance and speed more predictable.

We split this large group into two smaller age groups, claimants age 18 to <23 and claimants age 23 to <45. With that split in age groups, we found that among Title 16 Adult applicants, the latter group were the least likely to be QDD identified, and the former group were likely to be identified for QDD at similar rates as the older age group categories (45 to <50, 50 to <55, 55 to <60, and 60+). Our analysis concludes that the large number of persons with lifelong disabilities aging into adulthood among the Title 16 Adult population explains the different pattern in QDD identification by age for Title 16 Adults applicants in contrast to Title 2 and Concurrent applicants.

To what extent does the model show signs of overfitting related to a cancer diagnosis, and could this explain the medical condition differences in QDD identification?

Overfitting happens when a computer model trains (or “fits”) the training data very well but does not predict future observations reliably because it is too closely aligned to the specific training data used. When an algorithm is trained on datasets where one group dominates the sample or dominates the “targets” (those who have the outcome being predicted), the algorithm can overemphasize the factors that are predictive for that group, or even those that identify that group. Such overfitting can result in an algorithm that makes excellent predictions for members of that dominant group but that does not perform as well for others.

In our Phase 1 investigations we noted that QDD identification rates for cases with cancer as the primary diagnosis were larger than those for most other conditions, suggesting that the algorithm is particularly good at identifying quick allowances among these cases. Overfitting is one of the reasons that such a pattern can arise, so we examined differences between applicants with cancer and with other conditions more closely. We found that although applicants with a primary medical condition of cancer make up only 5 percent of all applicants, they account for 38 percent of those allowed quickly at Step 3. That is, they make up a relatively large share of the targets despite being a reasonably small share of the sample, which can make models prone to overfitting. We also investigated the information value of terms related to cancer to examine how powerful these terms are at predicting QDD score. These scores indicate possible overfitting for Title 2 and Concurrent modeling populations.

One possible result of overfitting is that an algorithm may have relatively little information to use to predict outcomes among those not in the population to which the model is overfit and may overemphasize

⁵ Compared to the main analysis sample, the distribution of primary medical conditions differs for this sample. Most notably, the proportion of applicants alleging a musculoskeletal condition is higher for this sample. The sample is also older. Distributions of sex and predicted race/ethnicity are similar in the two samples.

information from the largest groups as a result. We examined QDD identification rates by sex for applicants with and without cancer, to investigate whether this might be occurring. We found that among applicants with cancer, male and female applicants are identified for QDD at similar rates; but among applicants with non-cancer conditions, male applicants are identified for QDD 29.1 percentage points more frequently than females are. For Title 2 and Concurrent modeling populations, the difference is 36.5 and 27.8, respectively. This pattern is consistent with the model overemphasizing information from the largest group, though we are unable to rule out the possibility that it reflects other differences between the male and female applicants with and without cancer. Investigations of differences by age group and self-reported race-ethnicity show different patterns for applicants with and without cancer as a primary medical condition, but are not clearly consistent with the model overemphasizing information from the largest categories of those characteristics more for non-cancer applications than for cancer applications.

We conclude that there are patterns in the data that are consistent with overfitting, as well as reasons to be wary that overfitting could occur. In order to determine conclusively whether overfitting is occurring, and if so to measure its implications, additional analyses would be required that were not possible with the data available to our investigation.

Options for SSA Follow-Up

This investigation on bias in the QDD model revealed differences in QDD identification rates for groups of applicants defined by race/ethnicity, sex, age, and medical condition. The investigation found the largest differences in QDD identification based on medical condition. Despite the smaller magnitude, differences observed in QDD identification by race/ethnicity and age also raise concerns for equity.

Concrete steps that SSA could take to examine the sources of bias further or to address these findings include the following:

- SSA could further investigate whether models for the Title 2 and Concurrent populations are overfit for cancer. These investigations require full access to the QDD model, so must be carried out by the contractor who has this access, in collaboration with SSA. If overfitting is found, SSA could investigate the most appropriate and effective methods of addressing this overfitting. One option would be to predict medical condition based on the application contents and estimate separate models for those predicted to have cancer and those predicted to have another medical condition. SSA could also consider subsampling applicants with a primary alleged medical condition of cancer when estimating the model, or summarizing information related to a cancer diagnosis in a small number of variables. Steps like these could improve the models' ability to identify applicants with conditions other than cancer who are likely to be allowed quickly, but further analysis is needed to identify whether changes are needed and which changes would be most effective.
- Regarding the influence of access to quality medical care and resources on QDD identification, SSA could consider reviewing a random sample of applications to better understand the ways in which documented disparities in medical care quality and system resources manifest in applications. With this information, SSA could decide whether any application fields should be excluded from the predictive model or deprioritized based on the extent to which they reflect external disparities.
- Because the use of appointed representatives appears to reduce differences in QDD identification by race/ethnicity, SSA could continue and potentially augment efforts to inform applicants about how to enlist the help of an appointed representative.

- Because relationships between who is allowed and who is allowed quickly vary by characteristics explored here, SSA could consider potential options to revise its current approach of estimating SSAL and SSPT separately and then calculating the combined score as a linear combination of the SSAL and SSPT scores. One option would be to define a single modeling goal, combining information on allowance and speed, rather than modeling allowance separately from speed. SSA could consider the relative performance of different options in terms of their predictive power and bias, and identify the model that best achieves SSA's goals.
- Given that more than 99 percent of the models' input variables are text-based terms, SSA could consider which elements from the application are most important for the QDD process. The application uses free-response text fields to collect information on alleged condition, which is clearly of high importance. However, including a large number of text-based terms in the model creates biases. SSA could consider options for how to counteract the potential bias that can be caused by the inclusion of a large number of text terms, such as making greater use of composite features, where a single feature indicates that one more of a set of related words were found in the application and independent features for those words are not included.
- Because the QDD algorithm is estimated using data from past applications, it can reflect any biases or inequities that existed at the time those applications were processed. As SSA improves its processes, QDD will only reflect these improvements once sufficient data to support estimation has accumulated, estimation has occurred, and the new model parameters put into place. SSA could consider these factors when determining the frequency of model updates or the reference data to be used in model updates.

In closing, we note that SSA's QDD process achieves its goals of fast-tracking a set of applications for which allowances are likely and decisions can be made quickly. This investigation into potential bias reveals that some groups of applicants are underrepresented among claims identified for QDD. The largest disparities arise for claims with non-cancer diagnoses, and responding to this important difference may have the ripple effects of reducing some of the differences for the other groups (defined by race/ethnicity, age, and sex) as well.

We also note that some of the factors discussed here are specific to the QDD process, while others may have implications beyond the QDD process. Barriers to medical care and disparities in the quality of other systems could depress allowance rates, or increase decision speeds, if applicants are less likely to have a clear medical record that documents their condition. Similarly, clinician biases or lack of cultural competency could have similar effects, as could barriers to application quality like limited understanding of English. Further, while we find that these factors explain only a small proportion of variation in QDD identification rates, they may still be quite important for allowances and decision time. That is because our analysis here focuses on additional effects of these factors on QDD identification, beyond any effects they have on having a claim allowed quickly, but we expect that they would also have substantial effects on allowance probability and decision speed. For example, a lack of documentation that results in slower decision times would primarily affect decision speed, with any additional effect on QDD identification being secondary. In pursuing its equity action plan, SSA could consider investigating the effects of these factors on decisions, decision speed, and other program outcomes.

References

Allred J, Duffrin C, Brinkley J, Jones J. High-risk travel distance and number of primary care visits in a North Carolina Medicaid population. *Frontier Public Health Services & Systems Research* 2015;4(3):55–60. <https://uknowledge.uky.edu/frontiersinphssr/vol4/iss3/4>

Social Security Administration. (n.d.). *Quick disability determinations (QDD)*. <https://www.ssa.gov/disabilityresearch/qdd.htm>

Social Security Administration. 2023b. Social Security Administration (SSA) Monthly Data for Combined Title II Disability and Title XVI Blind and Disabled Average Processing Time (excludes technical denials). <https://www.ssa.gov/open/data/Combined-Disability-Processing-Time.html>

Weinstein, J. N., Geller, A., Negussie, Y., & Baciu, A. (2017). The state of health disparities in the United States. In A. Baciu, Y. Negussie, A. Geller et al. (Eds.), *Communities in Action: Pathways to Health in Action - NCBI Bookshelf* (nih.gov)