Identifying Disability with Medicaid Claims Data

Prepared for

The Office of the Assistant Secretary for Planning and Evaluation (ASPE) at the U.S. Department of Health & Human Services

by

Mathematica

September 2025

Office of the Assistant Secretary for Planning and Evaluation

The Assistant Secretary for Planning and Evaluation (ASPE) advises the Secretary of the U.S. Department of Health and Human Services (HHS) on policy development in health, disability, human services, data, and science; and provides advice and analysis on economic policy. ASPE leads special initiatives; coordinates the Department's evaluation, research, and demonstration activities; and manages cross-Department planning activities such as strategic planning, legislative planning, and review of regulations. Integral to this role, ASPE conducts research and evaluation studies; develops policy analyses; and estimates the cost and benefits of policy alternatives under consideration by the Department or Congress.

Office of Behavioral Health, Disability, and Aging Policy

The Office of Behavioral Health, Disability, and Aging Policy (BHDAP) focuses on policies and programs that support the independence, productivity, health and well-being, and long-term care needs of people with disabilities, older adults, and people with mental and substance use disorders. Visit BHDAP at https://aspe.hhs.gov/about/offices/bhdap for all their research activity.

This research was funded by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation under Contract Number #HHSP233201500035I/ 75P00122F37071 and carried out by Mathematica. Please visit the Office of Behavioral Health, Disability, and Aging Policy (BHDAP) page or ASPE Behavioral Health page for additional research in this area.

IDENTIFYING DISABILITY WITH MEDICAID CLAIMS DATA

Authors

Jonathan Gellar, Afroze Chughtai, and Andrea Wysocki

Mathematica

This page has been left blank for double-sided copying.

Contents

Tables

Figures

I. Introduction

Medicaid provides health insurance coverage and services and supports for people with disabilities, as well as groups of children and adults based on financial eligibility. While researchers have an understanding about some of the people with disabilities enrolled in Medicaid, such as those who qualify for Medicaid on the basis of a disability, less is known about those persons who might be disabled but may not have indicators in administrative data that allow for easy identification, such as the eligibility pathway indicators. Such identification is increasingly important, as states implement provisions in the One Big Beautiful Bill Act of 2025 (OBBA,Pub. L. 119-21) including Medicaid work requirements. More broadly, states and researchers would like to utilize more readily available administrative data sources for budgeting and planning.

Researchers have developed algorithms to identify people with disabilities in the Medicare population, but much less work has been done to develop algorithms to identify people with disabilities in the Medicaid population. The objective of this project was to develop a model that incorporates multiple claims-based indicators to identify Medicaid beneficiaries who have a disability using data from the Medicaid Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF). To meet this objective, we first conducted a targeted environmental scan of existing algorithms to identify disabilities in claims; summarized results for the team within the Office of Behavioral Health, Disability, and Aging Policy (BHDAP); and worked with BHDAP to select relevant predictor and service use indicators. Next, we identified beneficiaries with a known disability (defined below) indicated in the TAF. We then fit a model to capture how various claims-based indicators were associated with people with a known disability, and used the model to identify people who had many of these indicators in their claims history but were otherwise not known to be disabled. This report summarizes the findings from our analysis to address the following research questions:

1. How well do claims-based algorithms identify disabled populations in TAF claims data?
2. What is the prevalence of people identified with a disability using various methods of identification in claims data?[1]
3. What are the characteristics of people with known disabilities? How do these characteristics compare to those of people who do not have a known disability, but (a) have a suspected disability based on claims data, and (b) do not have a suspected disability based on claims data?

[1] Based on discussions with BHDAP, the definition of disability was expanded beyond those in a disability pathway to also include those receiving Supplemental Security Income (SSI) or Social Security Disability Insurance (SSDI) or using personal care services.

Study highlights

- The models for adults and children showed strong predictive performance based on area under the receiver operating characteristic curve (AUROC) statistics well above 0.8.
- The potential predictors included a broad list of demographic, diagnostic, and utilization indicators, and many of these different types of predictors were selected in the models. Some of the most important predictors included demographic characteristics that are straightforward to identify from TAF. The diagnostic indicators that were most predictive for both groups included those related to learning disabilities, autism spectrum disorder, and psychiatric conditions.
- Although the AUROC showed strong predictive performance, known disability among the child and adult populations in Medicaid is still a relatively rare outcome and resulted in low model sensitivity. This finding underscores the difficulty in capturing concepts such as functional limitations from claims-based indicators.
- Future research can further explore relevant thresholds for defining disability, identify an external data source that would allow validation for the people with a known disability, and examine the utilization and cost patterns of people with and without known disability based on the predicted probability of having a disability.

II. Approach

A. Overview

Our procedure for identifying Medicaid beneficiaries with disability in TAF consisted of 3 primary steps:

1. First, we defined a subset of beneficiaries with *known disability*, based on observable data. For this analysis, anyone who was eligible for Medicaid based on disability, received Supplemental Security Income (SSI) or Social Security Disability Insurance (SSDI) benefits, or received personal care home and community-based services (HCBS) during our study period is considered to have a known disability.
2. Next, we fit a predictive model to the outcome of known disability, based on demographic characteristics and claims-based indicators.
3. Last, we predicted the outcome (known disability) for all individuals, based on their predictor values. Beneficiaries who do not have a known disability but have a high predicted probability of having one are classified as having a *suspected disability*.

It is important to note that our predictive model does not provide the probability of having a *true disability*, which is the outcome that we are most interested in, because this outcome is unknowable based on claims data. Rather, we are using the model as a tool to identify predictors that are associated with having a known disability, as well as the strength of this association. We then make the assumption that beneficiaries with a true (but unknown) disability will resemble those with a known disability, in terms of their predictors. This assumption implies that beneficiaries with higher predicted probabilities of having a known disability are more likely to have a true disability than those with lower predicted probabilities.

We also note that the use of a *predictive* model does not imply that we are attempting to predict which beneficiaries will become disabled in the future. The goal is to predict (estimate) which beneficiaries have a disability concurrently with the study period. Therefore, we used one annual cross-section of data for our study. Specifically, we used 2019 TAF data to define relevant indicators, including predictors and outcomes of interest for descriptive output, and to develop the prediction model. We included 2018 TAF for condition categories that required a lookback period of more than 12 months to define (described in more detail later in this report). Ultimately, this approach could be used in other studies to conduct analyses among people with a disability or to stratify analyses based on disability status.

B. Study population

The study population consisted of 2019 Medicaid enrollees. We excluded from the population any enrollees who were dually eligible for Medicaid and Medicare, as we did not have access to Medicare claims and therefore would have missed important information about these enrollees' conditions and service use. A small number of Medicaid enrollees aged 65 and older were not dually eligible; we excluded this group as well to avoid small sample issues.

We also restricted the sample to enrollees with full-scope benefits given BHDAP's focus on considering disability in relation to people who need support services and may require Medicaid HCBS. Last, we imposed a restriction for 12 months of continuous enrollment in 2019 to have complete measures for conditions and other variables that informed the prediction model and interpretation.

The final sample consisted of 52,668,685 Medicaid enrollees, 30,276,091 (57 percent) of whom were children aged 18 or younger.

C. Environmental scan

To inform the selection of potential predictors for the analysis and the modeling approach for our analysis, we reviewed literature from 2015 to 2025 that focused on claims-based algorithms to predict disability or frailty.[2] Appendix A includes a brief summary of results from our scan. With input from BHDAP, we selected the set of predictors from algorithms identified in the scan that could most feasibly be constructed within the time frame of our analysis (described below). We also determined that we should develop separate models for children and adults due to their distinct characteristics and different sets of algorithms that were validated and tested for these populations in work identified through the environmental scan.

D. Outcome and predictor definitions

We defined Medicaid beneficiaries as having a known disability if they met one of the following four criteria: (1) were eligible for Medicaid on the basis of disability, (2) received SSI benefits based on the SSI flag in the TAF Annual Demographic and Eligibility File (DE),[3] (3) received SSDI benefits based on the SSDI flag in the TAF DE file, or (4) received HCBS for personal care in 2019.

Based on the environmental scan, we selected a comprehensive set of demographic, geographic, clinical, and utilization-based indicators as potential predictors of disability (Table 1). We integrated multiple frameworks to capture a range of characteristics, diagnoses, and conditions that may be associated with disability prevalence. The predictors varied for the children's and adult's models based on what was identified in the environmental scan for each population.

To characterize the sample, we defined key demographic attributes for each person, including age, sex, race and ethnicity, and urban versus rural residence. We also captured geographic variation by including the state of residence, whether the state had adopted Medicaid expansion, and whether the beneficiary lived in an urban or rural area.

[2] We searched PubMed for relevant research. We prioritized articles that described and/or conducted comparative analysis of multiple indices. We then conducted a targeted snowball search from the citations of the most relevant articles to identify additional indicators of interest. We excluded studies that were redundant with studies already identified for extraction, studies that focused on narrow disease-specific algorithms, non-claims-based studies, descriptive studies that were not focused on predicting disability, and international studies (except for those captured in one included systematic review).

[3] Beneficiaries in an SSI-specific eligibility category are already captured among those who are Medicaid-eligible on the basis of disability, so this separate criteria for the outcome definition captures anybody with an SSI flag in the TAF DE who is not otherwise captured based on their Medicaid eligibility pathway.

We captured service utilization by including indicators of any inpatient stays and 30-day hospital readmissions in the past year, and use of HCBS using the full HCBS taxonomy.[4] Note that we did not include HCBS for personal care use as a predictor, as we used this indicator in our outcome definition. We also included indicators of durable medical equipment (DME) and supplies. These included indicators for ambulance use, walking aids and wheelchairs, hospital beds, and home oxygen.[5] Including DME utilization enabled us to capture evidence of functional support needed that might be missing from a diagnosis-based algorithm due to coding variation or general missingness in data.

Table 1. Categories of predictors included in each model

Predictor category	Number of predictors included adult model	Number of predictors included in child model
Demographic	6	7
Inpatient stays and readmissions	2	2
Use of HCBS[a]	16	16
DME	10	10
Chronic conditions indicators:		
CCW	50	50
CDPS	86	86
CCC[b]	0	18
CWDA[b]	0	5
PMCA[b]	0	36
Total	**170**	**230**

[a] Use of HCBS is defined based on the HCBS taxonomy that relies on claims to identify HCBS use. We did not use indicators of HCBS program enrollment separately for the TAF DE file to identify HCBS use. We also did not include use of personal care HCBS as a predictor, because it is part of the definition of known disability.

[b] The CCC algorithm, CWDA, and PMCA are defined for children only.

CCC = Complex Chronic Conditions; CCW = Chronic Conditions Data Warehouse; CDPS = Chronic Illness and Disability Payment System; CWDA = Children with Disabilities Algorithm; DE = Annual Demographic and Eligibility File; DME = durable medical equipment; HCBS = home- and community-based services; PMCA = Pediatric Medical Complexity Algorithm; TAF = T-MSIS Analytic File; T-MSIS = Transformed Medicaid Statistical Information System.

Finally, we included indicators from several chronic condition and diagnostic classification systems. Two of these systems are designed for the entire population: (1) the Chronic Conditions Data Warehouse (CCW) categories and (2) the Chronic Illness and Disability Payment System (CDPS).[6] The CCW categories incorporate chronic health, mental health, substance abuse, and potentially disabling conditions. Many disabilities stem from chronic conditions, and CCW categorization is used widely because it encompasses an array of condition types. Using CDPS helped us cast a wider net to capture any conditions missed by

[4] For more information on the HCBS taxonomy, see Identifying and Classifying Medicaid Home and Community-Based Services Claims in the Transformed Medicaid Statistical Information System, 2016-2020 Issue Brief.

[5] We adapted existing code from the Claims-Based Frailty Index in Kim et al. (2020) to include relevant predictors related to DME and supplies.

[6] The CDPS uses International Classification of Disease codes to assign CDPS categories that indicate illness burden related to major body systems (for example, cardiovascular) or types of chronic disease (for example, diabetes). Within each major category is a hierarchy reflecting the clinical severity of the condition and its expected effect on future costs. Each hierarchical CDPS category is assigned a CDPS weight. CDPS weights are additive across major categories.

the CCW categorization for the overall population. For more nuanced information for the pediatric population, we included three additional condition algorithms designed for children up to age 18: (1) the Complex Chronic Conditions (CCC) algorithm,[7] (2) the Pediatric Medical Complexity Algorithm (PMCA),[8] and (3) the Children with Disabilities Algorithm (CWDA).[9]

E. Predictive model

To identify how predictors were associated with known disability, we fit a logistic regression model with a Least Absolute Shrinkage and Selection Operator (LASSO) penalty. Logistic regression is the most common and well-understood regression approach for modeling binary outcomes (such as known disability), while the LASSO penalty is a long-established machine-learning approach for handling situations when the model includes a very large number of predictor variables, many of which are potentially correlated. The LASSO penalty builds in an automatic variable selection procedure, effectively dropping variables from the model that are not independently important for classifying individuals into groups based on known disability versus no known disability. Including too many predictors runs the risk of overfitting the model, which is detrimental to prediction. LASSO logistic models are relatively straightforward to implement and interpret compared with other, more complex machine-learning techniques.

We fit separate models for adults and children, which has two advantages. From a conceptual perspective, individual predictors are likely to have different associations with disability among adults compared with children, and fitting separate models implicitly enables us to estimate these effects separately. From a practical perspective, we have a longer list of predictors for children compared with adults (due to the

---

[7] The CCC Classification System (CCC-CS) was developed to identify complex medical conditions within the pediatric population that can be reasonably expected to last at least 12 months and involve either several organ systems or one organ system severely enough to require pediatric care and hospitalization. CCC-CS body systems include cardiovascular, hematologic, malignancy, metabolic, neonatal, neuromuscular, renal, respiratory, transplant, and technology dependence domains. We adapted programming code available at https://www.childrenshospitals.org/content/analytics/toolkit/complex-chronic-conditions.

[8] The PMCA identifies significant chronic conditions in two or more body systems expected to last at least a year and require healthcare resources and treatment to control, or one progressive condition associated with deteriorating health and decreased life expectancy, or continuous dependence on technology for at least six months, or progressive or metastatic malignancy impacting life function. The PMCA stratifies children into three groups: (1) CCC, (2) noncomplex chronic conditions, and (3) no chronic conditions. PMCA domains include cardiac, craniofacial, dermatological, endocrinological, gastrointestinal, genetic, genitourinary, hematological, immunological, malignancy, metabolic, musculoskeletal, neurological, ophthalmological, otologic, otolaryngological, progressive, pulmonary-respiratory, renal, and mental health. We adapted programming code available at https://kpwashingtonresearch.org/index.php/our-research/our-scientists/Mangione-Smith-Rita/measurement-tools-research-dr-rita-mangione-smith.

[9] The original CWDA included 669 International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes classified as having a ≥75 percent likelihood of indicating a child with a disability. Based on the extracted studies from our scan, Chien et al. (2024) updated the CWDA to crosswalk to ICD-10-CM codes and to classify the associated impairment type (physical, sensory, developmental, psychiatric, intellectual) and number of associated impairment types to include in classification, which adds additional dimensions to the algorithm beyond just the individual condition codes.

inclusion of the CCC, CWDA, and PMCA instruments), and separating the modeling between adults and children allows us to incorporate different sets of predictors more easily.

Modeling approach. We fit and tuned models using standard best practices in machine learning. Due to the very large sample sizes (30.3 million child beneficiaries and 11.2 million adult beneficiaries), fitting the models to the full data required an excessive amount of computational power. To reduce the computational demands of the models and speed up the model-fitting process, we randomly selected a "training set" for each model: a 25 percent sample for the child model, and a 50 percent sample for the adult model. A larger portion was chosen for the adult model because there is a smaller total sample size of adults, so a 50 percent sample was still computationally feasible, while a smaller, 25 percent sample was required for computational feasibility for the larger child model.

Within each training set, we further split the data using a 70/30 ratio, allocating 70 percent of the data to a "tuning" set, and 30 percent of the data to a "validation" set. The tuning set was used to fit the model, and the validation set was used to assess its performance under various tuning parameters. This tuning process adjusted the strength of the LASSO penalty, effectively determining the final number of parameters that were selected into the final model. Though we originally intended to use cross-validation to select model parameters, we did not do so in order to reduce the computational complexity. Additionally, cross-validation is most beneficial when the total sample is small, which is not the case in this analysis. For these large data sets, a single 70/30 split is sufficient for model tuning.

Model testing. After fitting the models, we assessed their performance using standard machine learning validation approaches. Our primary metrics were the following:

- Area under the receiver operating characteristic curve (AUROC): The AUROC measures how well the model distinguishes between beneficiaries with known disability and those without known disability, based on their predictors. This metric is independent of the threshold chosen for classifying beneficiaries as having a high likelihood of disability and is often used as a measure of the overall predictive ability of a model. More specifically, an AUROC of 0.7, for example, means that 70% of the time, a randomly selected beneficiary with a known disability has a higher predicted probability than a randomly selected beneficiary without a known disability.
- Sensitivity: The sensitivity is the proportion of beneficiaries with a known disability who have a model-based predicted probability above a given threshold.
- Specificity: The specificity is the proportion of beneficiaries without a known disability who have a model-based predicted probability below a given threshold.

We calculated the AUROC using the validation set to serve as a measure of overall model predictive ability. Using the validation set ensured a valid assessment of the model, because assessing model performance on the same data used to fit the model typically leads to overestimation. However, for reasons of interpretability, we calculated sensitivity and specificity using the entire set of child and adult beneficiaries. For example, using the entire sample implies that the proportion of beneficiaries with suspected disability, among those with no known disability, is 100 percent minus the specificity (in other words, a specificity of 99 percent implies that 1 percent of beneficiaries without a known disability have a suspected disability).

State adjustment. State-specific regulations, including differences in rules regarding Medicaid eligibility through disability pathways, SSI or SSDI receipt, and allowance for personal care HCBS, will have an impact on whether a person is predicted to have a known disability. Because we wanted beneficiaries' utilization and clinical history to drive our predicted probabilities, not their place of residence, we controlled for state by marginalizing the predictions over the state variable[10]. More specifically, for all Medicaid beneficiaries, we produced a separate predicted probability for the beneficiary under the assumption that the beneficiary lived in each state. We then took a weighted average of the state-specific predictions for each beneficiary, weighting by the proportion of all Medicaid beneficiaries who reside in the state. This procedure ensures that two beneficiaries with the same demographic and clinical history, who differ only by their place of residence, will receive the same predicted probability of having a disability.

[10] For the purposes of this analysis, we consider the "state" to be a place of residence with 53 levels: the 50 United States, as well as Washington, DC; Puerto Rico; and the Virgin Islands. We use the term "state" throughout this report.

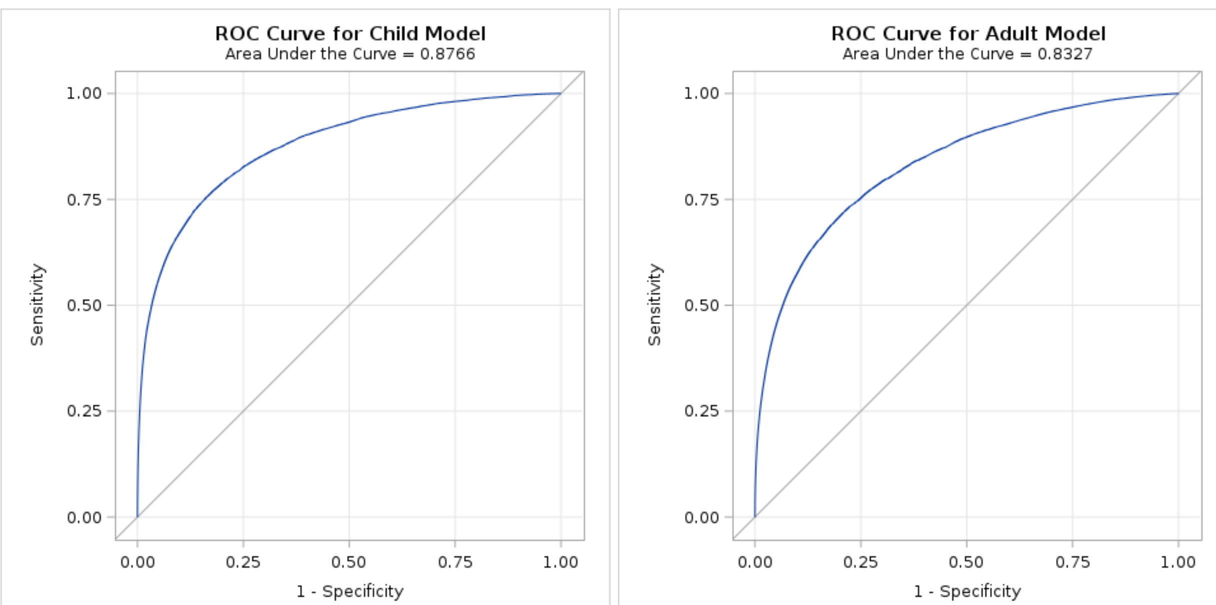
III. Results

Model performance. Table 2 summarizes the performance of the two models, both before and after state adjustment. Both models show strong ability to distinguish between known disability and no known disability, as evidenced by AUROC statistics well above 0.8; these statistics are further reflected in the plots of ROC curves (Figure 1). After state adjustment, a randomly selected child with a known disability has an 88 percent chance of having a higher predicted probability than a randomly selected child with no known disability; this probability is 83 percent among adults.

Table 2. Summary of model performance, both before and after state adjustment

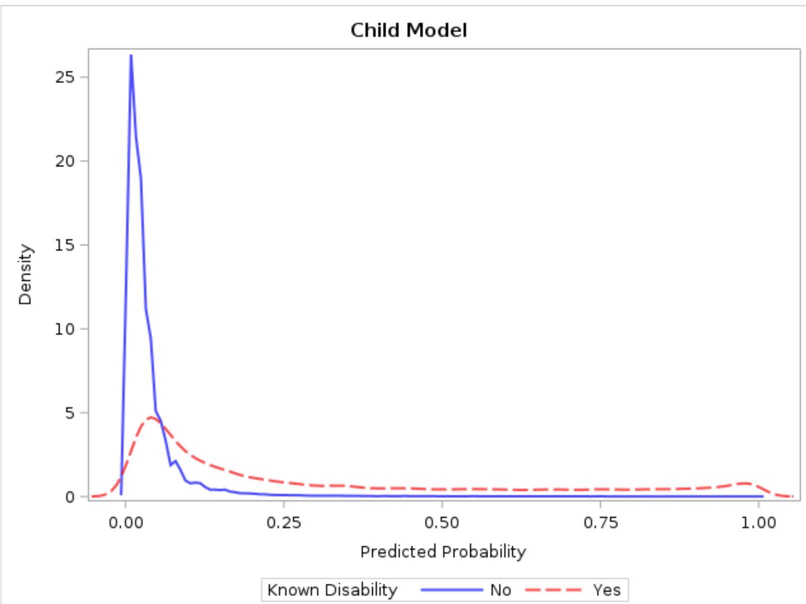
Sample	N	Training set size (%)	Adjustment	AUROC	Sensitivity (80% threshold)	Specificity (80% threshold)
Children (0–18)	30,276,091	7,569,023 (25%)	Unadjusted	0.900	13.4%	99.9%
Children (0–18)	30,276,091	7,569,023 (25%)	State-Adjusted	0.877	11.8%	99.9%
Adults (19–64)	22,392,594	11,196,297 (50%)	Unadjusted	0.874	18.7%	99.7%
Adults (19–64)	22,392,594	11,196,297 (50%)	State-Adjusted	0.833	13.5%	99.7%

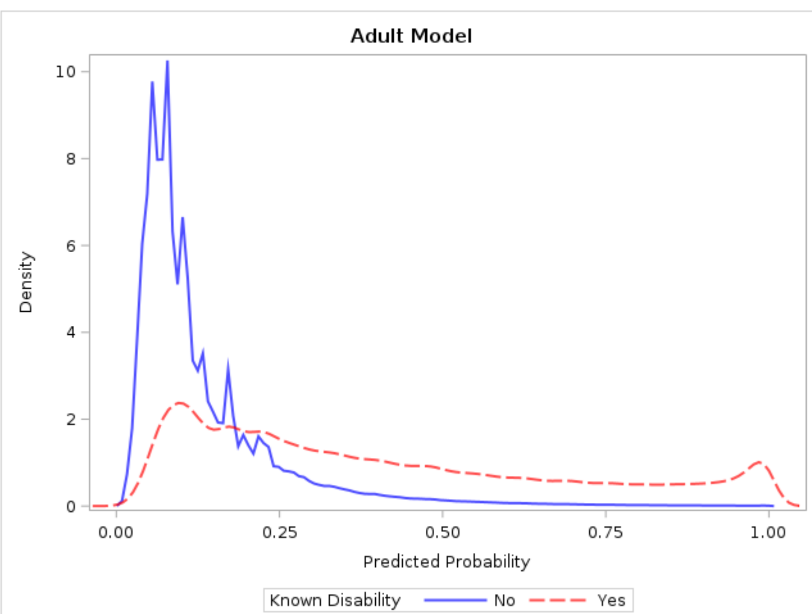
Figure 1. ROC curves for child (left) and adult (right) models, after state adjustment



Using an 80 percent threshold on the predictive probabilities results in a very high specificity, but a low sensitivity. More specifically, only 12 percent of children and 14 percent of adults with known disabilities have predicted probabilities above this 80 percent threshold. Density plots of the predicted probabilities (Figure 2) illustrate that most of the predicted probabilities are quite low, even for the group with known disability: 76 percent of children and 68 percent of adults with known disability have predicted probabilities below 50 percent. While low predicted probabilities are typically a feature of models where the outcome has low prevalence (just 5.6 percent of the child and 16.2 percent of the adult Medicaid population has a known disability), these results imply that there are many beneficiaries with no known disability that have similar predictor profiles as beneficiaries with a known disability. There are two potential explanations for this implication. If we assume that known disability is a good surrogate for true disability, these findings would imply that claims-based predictors are insufficient for determining true disability; other aspects of the beneficiary's health (such as those that can be assessed by a clinician in a face-to-face setting) are needed. On the other hand, these results could also occur if there is a relatively large group of beneficiaries who have a true disability that is not known. The truth is likely a combination of these two phenomena.

Figure 2. Density plots of predicted probabilities by known disability status, for children (top) and adults (bottom)





Choice of threshold. The low sensitivity numbers in Table 2 suggest that an 80 percent threshold for the predicted probabilities may be too restrictive for defining suspected disability; lowering that threshold would increase sensitivity, at the expense of having a lower specificity. This threshold ultimately determines the number of beneficiaries who would be labeled as having a suspected disability based on this methodology. A higher threshold results in a small group labeled as having a suspected disability, while a lower threshold would result in a relatively large suspected disability group. Because we do not know the true disability status of any of the people in the group without known disability, there is no statistical method to select the optimal threshold. However, we can calibrate the choice of threshold based on the resulting sensitivity.

Table 3 illustrates how sensitivity and specificity are affected by the choice of threshold. As expected, lowering the threshold increases sensitivity while lowering specificity. A threshold of 50 percent still results in a sensitivity of 23.9 percent for the child model and 33.0 percent for the adult model, and to achieve 50 percent sensitivity, the threshold would have to be less than 20 percent for children and 30 to 40 percent for adults.

Important predictors of disability. Appendix Tables B.1 and B.2 list the selected model predictors for children and adults, respectively, and rank them in terms of variable importance. Variable importance is measured by the Wald Chi-Squared statistic associated with each predictor, based on a standard logistic regression that only includes the selected predictors. We found that for both children and adults, demographic characteristics, including stage of residence, age, race and ethnicity, and sex, are important predictors. Among children, condition groups representing autism spectrum disorders, learning disabilities, and psychiatric conditions were among the most important diagnostic types predicting disability. Similar conditions groups, such as those for psychiatric conditions and intellectual disabilities and related conditions, were among the most important diagnostic types predicting disability for adults ages 19–64. Several individual categories of HCBS use were selected in both models. A few indicators of DME and the inpatient stay indicator were selected in the adult model, but only one DME indicator was selected in the child model indicating the only select types of service use are predictive of disability among the Medicaid population.

Table 3. Sensitivity and specificity for various choices of the probability threshold

Probability threshold	Child model sensitivity	Child model specificity	Adult model sensitivity	Adult model specificity
10%	60.4%	93.5%	88.7%	52.5%
20%	43.7%	97.9%	68.6%	82.3%
30%	35.1%	98.8%	53.3%	92.1%
40%	29.2%	99.2%	41.7%	96.0%
50%	24.4%	99.5%	32.2%	97.8%
60%	20.0%	99.7%	24.9%	98.8%
70%	15.9%	99.8%	18.8%	99.4%
80%	11.8%	99.9%	13.5%	99.7%
90%	7.3%	100.0%	8.5%	99.9%

Beneficiary characteristics, by known and predicted disability status. Tables 4 and 5 list characteristics of child and adult Medicaid beneficiaries, respectively, stratified by known and predicted disability status (based on the 80 percent threshold). For both children and adults, beneficiaries with known or suspected disabilities tend to be older and are more likely to be male than those with no known or suspected disability. Children with known disability are slightly more likely to live in urban areas, while the opposite is true for adults. For both age groups, those with no known disability are much more likely to live in an expansion state than those with a known disability.

Although those with a known disability are less likely to be non-Hispanic white than those without a known disability, this group has higher representation among those with suspected disability; these seeming contradiction likely reflects confounding between race and other predictors that are associated with known disability. Similarly, other racial groups do not show clear associations between disability status and race.

Seventy-one percent of children and 94 percent of adults with a known disability qualified for Medicaid on the basis of disability, whereas those without a known disability primarily qualified for Medicaid based on financial criteria (through pathways designed for children and adults, respectively). About half (49 percent) of children with a known disability receive SSI benefits, while one-quarter (26 percent) of those with a known disability receive SSDI benefits and 13 percent receive personal care HCBS. Among adults with a known disability, 63 percent receive SSI benefits, 27 percent receive SSDI benefits, and 10 percent receive personal care HCBS.

As previously noted, the most important predictors of disability among the chronic condition flags, for adults and children, are factors associated with learning and developmental disabilities, including autism. Hearing loss is also associated with disability among children.

Table 4. Characteristics of beneficiaries ages 0 to 18 based on disability group and predicted probability of having a disability, 2019

Characteristic	Known Disability N (%)	Suspected Disability (Predicted probability ≥80%) N (%)	No Known or Suspected Disability (Predicted probability <80%) N (%)
Total	1,691,179	33,975	28,550,937
Age			
0-5 years	273,697 (16.2%)	6,135 (18.1%)	8,662,203 (30.3%)
6-10 years	488,652 (28.9%)	12,681 (37.3%)	8,089,257 (28.3%)
11-18 years	928,830 (54.9%)	15,159 (44.6%)	11,799,477 (41.3%)
Sex			
Female	587,065 (34.7%)	8,548 (25.2%)	14,157,520 (49.6%)
Male	1,104,102 (65.3%)	25,427 (74.8%)	14,390,010 (50.4%)
Unknown	12 (0.0%)	0 (0.0%)	3,407 (0.0%)
Geographic location			
Urban	1,362,600 (80.6%)	27,363 (80.5%)	22,546,779 (79.0%)
Rural	315,079 (18.6%)	6,332 (18.6%)	5,824,718 (20.4%)

Characteristic	Known Disability N (%)	Suspected Disability (Predicted probability ≥80%) N (%)	No Known or Suspected Disability (Predicted probability <80%) N (%)
Unknown	13,500 (0.8%)	280 (0.8%)	179,440 (0.6%)
Race and ethnicity			
American Indian and Alaska Native, Non-Hispanic	13,949 (0.8%)	303 (0.9%)	402,886 (1.4%)
Asian, Non-Hispanic	23,191 (1.4%)	659 (1.9%)	862,483 (3.0%)
Black, Non-Hispanic	367,903 (21.8%)	5,770 (17.0%)	5,148,876 (18.0%)
Hawaiian/Pacific Islander, Non-Hispanic	3,607 (0.2%)	94 (0.3%)	172,668 (0.6%)
Hispanic, All Races	285,505 (16.9%)	3,917 (11.5%)	7,300,107 (25.6%)
Multiracial, Non-Hispanic	4,691 (0.3%)	66 (0.2%)	93,413 (0.3%)
White, Non-Hispanic	497,374 (29.4%)	11,486 (33.8%)	8,855,526 (31.0%)
Unknown	494,959 (29.3%)	11,680 (34.4%)	5,714,978 (20.0%)
Eligibility group			
Disabled	1,208,985 (71.5%)	0 (0.0%)	0 (0.0%)
Pregnant	225 (0.0%)	*	26,719 (0.1%)
Children	276,357 (16.3%)	32,800 (96.5%)	28,087,802 (98.4%)
Adult	0 (0.0%)	0 (0.0%)	0 (0.0%)
Adult expansion	55 (0.0%)	*	35,565 (0.1%)
Unknown	8,059 (0.5%)	1,111 (3.3%)	400,851 (1.4%)
SSI receipt	820,997 (48.5%)	0 (0.0%)	0 (0.0%)
SSDI receipt	436,373 (25.8%)	0 (0.0%)	0 (0.0%)
State of residence			
Expansion state	996,696 (58.9%)	21,891 (64.4%)	17,808,670 (62.4%)
Non-expansion state	694,483 (41.1%)	12,084 (35.6%)	10,742,267 (37.6%)
Service use			
Any personal care service use	213,992 (12.7%)	0 (0.0%)	0 (0.0%)
Any HCBS use	700,498 (41.4%)	22,461 (66.1%)	4,785,944 (16.8%)
Any inpatient stay	216,483 (12.8%)	11,725 (34.5%)	2,116,763 (7.4%)
Top 5 conditions or condition categories			
Extreme low birth weight/preterm conditions (CDPS)	14,750 (0.9%)	481 (1.4%)	10,264 (0.0%)
Autism spectrum disorders (CCW)	354,911 (21.0%)	23,990 (70.6%)	226,489 (0.8%)
Sensory – deafness and hearing impairment (CCW)	52,115 (3.1%)	3,175 (9.3%)	124,116 (0.4%)
Learning disabilities (CCW)	531,055 (31.4%)	26,069 (76.7%)	1,435,474 (5.0%)
Psychiatric, low (CDPS)[a]	768,166 (45.4%)	26,145 (77.0%)	3,237,984 (11.3%)

Source: Mathematica's analysis of 2019 TAF.

Note: An 80% threshold was selected to display characteristics of the sample from the children's model. The sample for the children's model includes Medicaid-only enrollees ages 0 to 18 with full-scope benefits and continuous enrollment in 2019.

[a]The CDPS creates a hierarchy within categories to reflect severity. "Low" designates lower-cost psychiatric disorders.

*Cell is masked due to small sample size.

CCW = Chronic Conditions Data Warehouse; CDPS = Chronic Illness and Disability Payment System; HCBS = home and community-based service; SSI = Supplemental Security Income; SSDI = Social Security Disability Insurance; TAF = T-MSIS Analytic File.

Table 5. Characteristics of beneficiaries ages 19 to 64 based on disability group and predicted probability of having a disability, 2019

Characteristic	Known Disability N (%)	Suspected Disability (Predicted probability ≥80%) N (%)	No Known or Suspected Disability (Predicted probability <80%) N (%)
Total	3,622,524	56,757	18,713,313
Age			
19-30 years	827,795 (22.9%)	11,649 (20.5%)	7,041,445 (37.6%)
31-45 years	811,348 (22.4%)	9,734 (17.2%)	6,624,617 (35.4%)
46-55 years	804,763 (22.2%)	20,864 (36.8%)	2,947,085 (15.7%)
56-64 years	1,178,618 (32.5%)	14,510 (25.6%)	2,100,166 (11.2%)
Sex			
Female	1,848,083 (51.0%)	20,308 (35.8%)	11,425,538 (61.1%)
Male	1,774,426 (49.0%)	36,449 (64.2%)	7,287,706 (38.9%)
Unknown	0 (0.0%)	0 (0.0%)	69 (0.0%)
Geographic location			
Urban	2,790,846 (77.0%)	47,664 (84.0%)	15,290,279 (81.7%)
Rural	786,967 (21.7%)	8,540 (15.0%)	3,293,287 (17.6%)
Unknown	44,711 (1.2%)	553 (1.0%)	129,747 (0.7%)
Race and ethnicity			
American Indian and Alaska Native, Non-Hispanic	43,071 (1.2%)	611 (1.1%)	298,909 (1.6%)
Asian, Non-Hispanic	72,191 (2.0%)	633 (1.1%)	1,108,705 (5.9%)
Black, Non-Hispanic	873,736 (24.1%)	19,112 (33.7%)	3,137,433 (16.8%)
Hawaiian/Pacific Islander, Non-Hispanic	12,006 (0.3%)	188 (0.3%)	110,954 (0.6%)
Hispanic, All Races	396,189 (10.9%)	4,175 (7.4%)	3,604,321 (19.3%)
Multiracial, Non-Hispanic	5,796 (0.2%)	37 (0.1%)	22,478 (0.1%)
White, Non-Hispanic	1,476,852 (40.8%)	23,449 (41.3%)	7,408,741 (39.6%)
Unknown	742,683 (20.5%)	8,552 (15.1%)	3,021,772 (16.1%)
Eligibility group			
Disabled	3,404,785 (94.0%)	0 (0.0%)	0 (0.0%)
Pregnant	2,203 (0.1%)	59 (0.1%)	247,751 (1.3%)
Children	8,371 (0.2%)	2,059 (3.6%)	619,814 (3.3%)
Adult	65,434 (1.8%)	7,608 (13.4%)	5,963,496 (31.9%)
Adult expansion	104,797 (2.9%)	46,625 (82.1%)	11,745,755 (62.8%)

Characteristic	Known Disability N (%)	Suspected Disability (Predicted probability ≥80%) N (%)	No Known or Suspected Disability (Predicted probability <80%) N (%)
Unknown	36,934 (1.0%)	406 (0.7%)	136,497 (0.7%)
SSI receipt	2,269,273 (62.6%)	0 (0.0%)	0 (0.0%)
SSDI receipt	983,267 (27.1%)	0 (0.0%)	0 (0.0%)
State of residence			
Expansion state	2,295,661 (63.4%)	53,860 (94.9%)	16,393,275 (87.6%)
Non-expansion state	1,326,863 (36.6%)	2,897 (5.1%)	2,320,038 (12.4%)
Service use			
Any personal care service use	369,613 (10.2%)	0 (0.0%)	0 (0.0%)
Any HCBS use	1,767,895 (48.8%)	45,259 (79.7%)	4,584,092 (24.5%)
Any inpatient stay	950,987 (26.3%)	35,739 (63.0%)	2,806,199 (15.0%)
Top 5 conditions or condition categories			
Autism spectrum disorders (CCW)	133,494 (3.7%)	5,467 (9.6%)	18,145 (0.1%)
Intellectual disabilities and related conditions (CCW)	291,540 (8.0%)	14,284 (25.2%)	12,017 (0.1%)
Learning disabilities (CCW)	59,343 (1.6%)	3,051 (5.4%)	16,122 (0.1%)
Psychiatric, high (CDPS)[a]	363,216 (10.0%)	28,562 (50.3%)	133,641 (0.7%)
Psychiatric, low (CDPS)[a]	988,835 (27.3%)	18,011 (31.7%)	2,629,500 (14.1%)

Source: Mathematica's analysis of 2019 TAF.

Note: An 80% threshold was selected to display characteristics of the sample for the adult model. The sample for the adult model includes Medicaid-only enrollees ages 19 to 64 with full-scope benefits and continuous enrollment in 2019.

[a]The CDPS creates a hierarchy within categories to reflect severity. "High" designates high-cost psychiatric disorders and "low" designates lower-cost psychiatric disorders.

CCW = Chronic Conditions Data Warehouse; CDPS = Chronic Illness and Disability Payment System; HCBS = home and community-based service; SSI = Supplemental Security Income; SSDI = Social Security Disability Insurance; TAF = T-MSIS Analytic File.

This page has been left blank for double-sided copying.

IV. Conclusions

The models for adults and children showed strong predictive performance based on AUROC statistics well above 0.8, which suggests that the models include relevant predictors that help distinguish known disability among the sample. The predictors included a broad list of demographic, diagnostic, and utilization indicators, and many of these different types of predictors were selected in the models. This suggests that including a variety of characteristics and diagnostic factors was an important feature of the model. Some of the most important predictors included demographic characteristics that are straightforward to identify from TAF. For example, for both adults and children, beneficiaries with a known disability are more likely to be male and in an older age bracket than those with no known disability. The diagnostic indicators that were most predictive for both groups included those related to intellectual and learning disabilities, autism spectrum disorders, and psychiatric conditions. Only a narrow set of utilization indicators were selected in the models, mostly reflecting specific types of HCBS use.

Although the AUROC showed strong predictive performance, known disability among the child and adult populations in Medicaid is still a relatively rare outcome and resulted in low model sensitivity. Specifically, most of the predictions from both models were less than 25 percent for both groups, highlighting that the presence of other factors we were unable to capture in the models to predict disability. This finding underscores the difficulty in capturing concepts such as functional limitations from claims-based indicators, and likely makes claims insufficient for estimating exact numbers of persons with disabilities in the program.

Using an 80 percent threshold, the models identified 33,975 children and 56,757 adults as potentially disabled but without a known disability, representing 0.12 and 0.30 percent, respectively, of the populations with no known disability. Although lowering the threshold would increase the number of beneficiaries identified as disabled, the trade-off would be a lower specificity, or less confidence that the identified beneficiaries have a true disability. Although no threshold will be able to distinguish perfectly between people with and without a true disability, it may be worth carefully considering whether a threshold should be more or less inclusive.

This work provided a strong foundation for further exploring claims-based models that predict disability among the Medicaid population. Next steps could include identifying an external data source that would allow validation for the people with a known disability and examining the utilization and cost patterns of each of the groups.

This page has been left blank for double-sided copying.

Appendix A

Summary of Environmental Scan Findings

This page has been left blank for double-sided copying.

Extracted studies

- Ben-Shalom, Y., and Stapleton, D.C. "Predicting Disability among Community-Dwelling Medicare Beneficiaries Using Claims-Based Indicators." *Health Services Research* 51.1 (2016): 262–281.
- Leyenaar, J.K., Schaefer, A.P., Freyleue, S.D., Austin, A.M., Simon, T.D., Van Cleave, J., Moen, E.L., O'Malley, A.J., and Goodman, D.C. "Prevalence of Children With Medical Complexity and Associations With Health Care Utilization and In-Hospital Mortality," *JAMA pediatrics* 176.6 (2022), e220687-e220687.
- Kim, D.H., Patorno, E., Pawar, A., Lee, H., Schneeweiss, S., and Glynn, R.J. "Measuring Frailty in Administrative Claims Data: Comparative Performance of Four Claims-Based Frailty Measures in the U.S. Medicare Data." *The Journals of Gerontology. Series A* 75.6 (2020): 1120–1125.
  - Compared: Disability Index: Davidoff (2013); ADL dependency Index: Faurot (2015); Frailty Index: Segal (2017); Frailty (CFI) Index: Kim (2018)
- Heins, S.E., Agniel, D., Mann, J., and Sorbero, M.E., "Comparative Performance of Three Claims-Based Frailty Measures Among Medicare Beneficiaries." *Journal of Applied Gerontology* 43.6 (2024): 765–774.
- Shashikumar S.A., Huang K, Konetzka R.T., and Joynt Maddox K.E. "Claims-based Frailty Indices: A Systematic Review." *Medical care* 58.9 (2020): 815-825.
  - Reviewed: Shashikumar (2020); De la Garza Ramos (2016); Gilbert (2018); Hope (2018); Hope (2015); Joynt (2017); Kim (2020); Kim (2015); Kim (2019); Lunney (2002); Moldovan (2020); Olsen (2018); Orkaby (2019); Segal (2017); Segal (2017); Soong (2015); Soong (2018); Wu (2019)
- Chien A.T., Spence S.J., Okumura M.J., Lu S., Chan C.H., Houtrow A.J., Kuo D.Z., Van Cleave J.M., Shanske S.A., Schuster M.A., Kuhlthau K.A., and Toomey S.L., "Impairment Types and Combinations Among Adolescents and Young Adults with Disabilities: Colorado 2014-2018." *Academic pediatrics* 24.4 (2024): 587-595.
- Straub L., Bateman B.T., Hernandez-Diaz S., York C., Zhu Y., Suarez E.A., Lester B., Gonzalez L., Hanson R., Hildebrandt C., Homsi J., Kang D., Lee K.W.K., Lee Z., Li L., Longacre M., Shah N., Tukan N., Wallace F., Williams C., Zerriny S., Mogun H., and Huybrechts K.F. "Validity of claims-based algorithms to identify neurodevelopmental disorders in children." *Pharmacoepidemiology and drug safety* 30.12 (2021): 1635-1642.

Identified algorithms and indicators

Disability

- Chronic Illness and Disability Payment System
- Access Risk Classification System
- SSA Health Information Technology business rules
- Psychiatric, cognitive, and intellectual disorder-specific disability indicators
- Complex Chronic Condition Classification System
- Pediatric Medical Complexity Algorithm

- Children with Disabilities Algorithm and Diagnosis-to-Impairment Types Algorithm
- Claims-based disability indexes (Davidoff, Faurot, Olsen, Wu)
- Straub Claims-Based Neurodevelopmental Disorder Algorithm

Frailty

- Claims-based frailty indexes (Segal, Kim)
- Pre-Hospital Frailty Model
- RAND Activity and Mobility Index
- RAND Memory Index
- Frailty indexes for specific or international populations: Metastatic Spinal Tumor Frailty Index, Hospital Frailty Risk Score, Medically Complex Frailty subgroup, Frail Decedent Subgroup, Statistical Learning-Based Frailty Index, Veteran Frailty Index, Soong Frailty Model

Key takeaways

- **Measurement period:** Among the studies we considered in this targeted environmental scan, the most common measurement period used to capture predictors was 12 months. Indicators used lookback periods ranging from six months to four years but usually 12–24 months. Indicators designed to predict health outcomes often used an index date, such as hospital admission, to define the end of the lookback period.
- **Validation:** Studies most commonly used linked self-reported health status data from surveys or medical record data to validate index performance, sometimes coupled with parent or physician perspectives.
  - Among studies that only used claims data, they used these data to create the index and either (1) identified people in the data set who did not have the condition of interest (frailty or disability) as a comparison group, or (2) randomly generated training and test sets from the same source to calculate predictive probabilities.
- **Data sources:** Studies most commonly used Medicare data, either alone or linked with other data sources for validation. Although less common, a few other studies used all-payer claims data, Medicaid data, other hospital or national survey data, or international data sets (identified in a systematic review).
- **Populations and age groups.** Studies covered a range of age groups, from children to older adults. The studies that considered only older populations typically focused on predicting frailty or limitations on activity of daily living, while those that considered only younger populations typically focused on medical complexity and specific disorders.
- **Algorithm definitions.** All study algorithms used a combination of diagnoses and conditions (whether grouped or included individually), with the total number and type of conditions or diagnoses varying across algorithms.
  - Few algorithms used utilization of services as predictors, but among those that did, it was typically related to use of durable medical equipment (DME). Some also included nursing facility care,

rehabilitation services, hospital-acquired problems, or other variables considered proxies for cognitive or physical impairment.

- – Some algorithms grouped conditions by body system (for example, cardiovascular, skeletal) or condition type (for example, physical, mental) and created an algorithm score based on the totals within each category.
- – Some algorithms simply looked for the presence of certain codes in claims (presence of 1+ or 2+ codes = 1, absence = 0), while others were more complex and assigned probabilistic values or coefficients to different conditions based on their relative importance to the variable of interest.
- – Some algorithms used demographic characteristics along with diagnoses or conditions as predictors, but some of the findings from comparative studies suggest that indexes that included demographic variables are unlikely to provide a large improvement in risk prediction and case-mix adjustment beyond demographic variables (at least based on findings for those measuring frailty among the older population).

- **Conditions.** Some algorithms used existing condition lists to identify relevant conditions, including the Charlson Comorbidity Index, the Social Security Administration Listing of Impairments, or the Centers for Medicare & Medicaid Services Chronic Conditions Data Warehouse. Studies often included hundreds of individual International Classification of Disease, Current Procedural Terminology, and Healthcare Common Procedure Coding System codes in their list of relevant conditions.
  - – Frailty indicators usually included conditions related to aging and functional impairment such as dementia, senility, incontinence, pressure ulcers, malnutrition, difficulty walking, falls, and indicators for DME use.
  - – Indicators of medical complexity included more comprehensive lists of codes related to serious physical health conditions across multiple body systems, such as cardiovascular disease, diabetes, mental health conditions, infections, and renal failure.
  - – General disability indicators included intellectual disabilities, neurological or developmental conditions, psychiatric disorders, and serious physical health conditions, such as cancer, paralysis, Parkinson's disease, and stroke or brain injury, in addition to other indicators such use of an ambulance, life support, or DME.
  - – Indicators focused on specific subsets of disability, such as neurodevelopmental disorders (Straub) or the RAND indexes for activity and for mobility and memory include a narrower list of conditions related to the type of limitations of interest.
- **Index development.** Development methods for the study algorithms varied and included clinical review, Least Absolute Shrinkage and Selection Operator (LASSO) regression, and logistic regression. After predictors were selected, the approaches for defining the algorithms included regression methods (such as logistic regression), index creation by combining various characteristics into a summary score, and identification of a threshold number of factors (such as conditions) to group people by disability, medical complexity, or frailty status.

  - At least some predictors were usually initially selected a priori based on previous literature, or by consensus discussions among clinical experts.

- **Model performance.** The models' ability to predict disability or frailty status or health outcomes was usually evaluated using logistic regression, c-statistic, or area under the receiver operating characteristic curve.

  - Development, validation, and comparison studies calculated sensitivity and specificity to describe each model's discriminative ability (for example, positive predictive value). During validation, some studies identified false positives and false negatives to examine algorithm assignment.

  Comparison studies used odds ratios, descriptive statistics, chi square tests, Spearman correlation coefficients, root mean squared error, bootstrap random sampling, and c-statistics to compare model performance. Studies also used other tests to assess and compare model performance. Some studies assessed model performance within subpopulations defined by demographics, social conditions, or health status.

Appendix B

Important Predictors of Known Disability

This page has been left blank for double-sided copying.

Tables B.1 and B.2 list all predictors that were selected by the two LASSO models. They are listed in order of variable importance, determined by first fitting a standard logistic regression model for known disability that included only the selected predictors, and then ranking predictors based on the Wald Chi-Squared statistic.

Table B.1. Selected predictors of known disability for beneficiaries ages 0 to 18, ranked by variable importance (Scaled Wald Chi-Squared statistic)

Predictor	Type	Degrees of Freedom	Relative importance
State	Demographic	52	100.00
Age	Demographic	3	47.06
Autism spectrum disorders	Chronic condition indicator (CCW)	1	45.73
Medicaid and/or CHIP enrollment	Demographic	4	42.12
Race and ethnicity	Demographic	7	29.62
Learning disabilities	Chronic condition indicator (CCW)	1	18.87
Psychiatric, low	Chronic condition indicator (CDPS)	1	10.96
Sex	Demographic	2	7.24
Round-the-clock services	HCBS	1	5.82
Extreme low birth weight/preterm conditions	Chronic condition indicator (CDPS)	1	5.14
Sensory – deafness and hearing impairment	Chronic condition indicator (CCW)	1	4.27
Caregiver support services	HCBS	1	4.06
Cerebral palsy	Chronic condition indicator (CCW)	1	2.79
Intellectual disabilities and related conditions	Chronic condition indicator (CCW)	1	2.77
Genetic conditions, any	Chronic condition indicator (PMCA)	1	2.71
Other developmental delays	Chronic condition indicator (CCW)	1	2.62
Developmental disability, low	Chronic condition indicator (CDPS)	1	2.02
Renal, low	Chronic condition indicator (CDPS)	1	1.75
Case management services	HCBS	1	1.49
Cardiovascular, low	Chronic condition indicator (CDPS)	1	1.47
Endocrinological, any	Chronic condition indicator (PMCA)	1	1.43
Non-medical transportation services	HCBS	1	1.14
Nursing services	HCBS	1	1.06
Central nervous system, low	Chronic condition indicator (CDPS)	1	1.01
Mental health, any	Chronic condition indicator (PMCA)	1	0.97
Skeletal, medium	Chronic condition indicator (CDPS)	1	0.90
Neuromuscular	Chronic condition indicator (CCC)	1	0.82
Epilepsy	Chronic condition indicator (CCW)	1	0.57
Chronic lung disease	Chronic condition indicator (CCW)	1	0.55
Malignancy	Chronic condition indicator (CCC)	1	0.54
Spina bifida and other congenital anomalies of the nervous system	Chronic condition indicator (CCW)	1	0.53
Metabolic, any	Chronic condition indicator (PMCA)	1	0.42

Predictor	Type	Degrees of Freedom	Relative importance
Equipment, technology, and modifications services	HCBS	1	0.36
Cardiac, any	Chronic condition indicator (PMCA)	1	0.35
Other health and therapeutic services	HCBS	1	0.25
Anemia	Chronic condition indicator (CCW)	1	0.24
Neurological, any	Chronic condition indicator (PMCA)	1	0.24
Ambulance	DME	1	0.20
Pulmonary-Respiratory, multiple	Chronic condition indicator (PMCA)	1	0.09
Diabetes	Chronic condition indicator (CCW)	1	0.08
Pulmonary-Respiratory, any	Chronic condition indicator (PMCA)	1	0.08
Skeletal, low	Chronic condition indicator (CDPS)	1	0.07
Geography (urban or rural)	Demographic	2	0.06
Chronic kidney disease	Chronic condition indicator (CCW)	1	0.06
Central nervous system, high	Chronic condition indicator (CDPS)	1	0.05
Mental health, multiple	Chronic condition indicator (PMCA)	1	0.02
Other mental health and behavioral services	HCBS	1	0.01
Any progressive condition (PMCA classification)	Chronic condition indicator (PMCA	1	0.00
Childhood central nervous system, high	Chronic condition indicator (CDPS)	1	0.00
Asthma	Chronic condition indicator (CCW)	1	0.00
Childhood psychiatric, medium	Chronic condition indicator (CDPS)	1	0.00
Spinal cord injury	Chronic condition indicator (CCW)	1	0.00

Source: Mathematica's analysis of 2019 TAF.

Note: The sample for the children's model includes Medicaid-only enrollees ages 0 to 18 with full-scope benefits and continuous enrollment in 2019. Relative importance is the Wald Chi-Squared statistic, rescaled linearly so that the largest value is 100.

CCC = Complex Chronic Conditions index; CCW = Chronic Conditions Data Warehouse; CDPS = Chronic Illness and Disability Payment System; CHIP = Children's Health Insurance Program; HCBS = home and community-based services; PMCA = Pediatric Medical Complexity Algorithm.

Table B.2. Selected predictors of known disability for beneficiaries ages 19 to 64, ranked by variable importance (scaled Wald Chi-Squared statistic)

Predictor	Type	Degrees of Freedom	Relative importance
State	Demographic	52	100.00
Age	Demographic	2	28.38
Psychiatric, high	Chronic condition indicator (CDPS)	1	25.84
Race and ethnicity	Demographic	7	6.71
Psychiatric, low	Chronic condition indicator (CDPS)	1	6.57
Autism spectrum disorders	Chronic condition indicator (CCW)	1	5.33
Sex	Demographic	1	4.76
Round-the-clock services	HCBS	1	4.08
Intellectual disabilities and related conditions	Chronic condition indicator (CCW)	1	3.15
Case management services	HCBS	1	2.35
Learning disabilities	Chronic condition indicator (CCW)	1	2.15
Cerebral palsy	Chronic condition indicator (CCW)	1	2.11
Psychiatric, medium	Chronic condition indicator (CDPS)	1	1.58
Pregnancy with complications	Chronic condition indicator (CDPS)	1	1.55
Pulmonary, low	Chronic condition indicator (CDPS)	1	1.41
Developmental disability, low	Chronic condition indicator (CDPS)	1	1.33
Ambulance	DME	1	1.33
Chronic Obstructive Pulmonary Disease	Chronic condition indicator (CCW)	1	1.28
Fibromyalgia, chronic pain and fatigue	Chronic condition indicator (CCW)	1	1.26
Central nervous system, high	Chronic condition indicator (CDPS)	1	1.23
Infectious, high	Chronic condition indicator (CDPS)	1	1.12
Rheumatoid arthritis/osteoarthritis	Chronic condition indicator (CCW)	1	1.01
Epilepsy	Chronic condition indicator (CCW)	1	0.99
Genital, extra low	Chronic condition indicator (CDPS)	1	0.91
Day services	HCBS	1	0.84
Oxygen	DME	1	0.82
Non-medical transportation services	HCBS	1	0.81
Caregiver support services	HCBS	1	0.78
Central nervous system, very high	Chronic condition indicator (CDPS)	1	0.74
Renal, low	Chronic condition indicator (CDPS)	1	0.73
Wheelchair	DME	1	0.67
Substance abuse, low	Chronic condition indicator (CDPS)	1	0.61
Central nervous system, low	Chronic condition indicator (CDPS)	1	0.61
Cardiovascular, very high	Chronic condition indicator (CDPS)	1	0.50
Peripheral vascular disease	Chronic condition indicator (CCW)	1	0.48
Cataract	Chronic condition indicator (CCW)	1	0.40
Cerebrovascular, medium	Chronic condition indicator (CDPS)	1	0.39

Predictor	Type	Degrees of Freedom	Relative importance
Viral Hepatitis	Chronic condition indicator (CCW)	1	0.37
Cardiovascular, low	Chronic condition indicator (CDPS)	1	0.35
Sensory – deafness and hearing impairment	Chronic condition indicator (CCW)	1	0.35
Any inpatient stay	Inpatient stays and readmissions	1	0.34
Pregnancy (routine)	Chronic condition indicator (CDPS)	1	0.33
Hypertension	Chronic condition indicator (CCW)	1	0.33
Skeletal, medium	Chronic condition indicator (CDPS)	1	0.30
Substance abuse, very low	Chronic condition indicator (CDPS)	1	0.30
Diabetes, type 2	Chronic condition indicator (CDPS)	1	0.25
Geography (urban or rural)	Demographic	2	0.25
Pulmonary, medium	Chronic condition indicator (CDPS)	1	0.24
Equipment, technology, and modifications	HCBS	1	0.15
Anemia	Chronic condition indicator (CCW)	1	0.14
Diabetes	Chronic condition indicator (CCW)	1	0.12
Central nervous system, medium	Chronic condition indicator (CDPS)	1	0.11
Glaucoma	Chronic condition indicator (CCW)	1	0.10
Cardiovascular, extra low	Chronic condition indicator (CDPS)	1	0.08
Other mental health and behavioral services	HCBS	1	0.07
Heart failure and non-ischemic heart disease	Chronic condition indicator (CCW)	1	0.07
Hyperlipidemia	Chronic condition indicator (CCW)	1	0.06
Skeletal, low	Chronic condition indicator (CDPS)	1	0.04
Lung cancer	Chronic condition indicator (CCW)	1	0.03
Tobacco use disorders	Chronic condition indicator (CCW	1	0.02
Gastro, low	Chronic condition indicator (CDPS)	1	0.02
Other health and therapeutic services	HCBS	1	0.01

Source: Mathematica's analysis of 2019 TAF.

Note: The sample for the children's model includes Medicaid-only enrollees ages 19 to 64 with full-scope benefits and continuous enrollment in 2019. Relative importance is the Wald Chi-Squared statistic, rescaled linearly so that the largest value is 100.

CCW = Chronic Conditions Data Warehouse; CDPS = Chronic Illness and Disability Payment System; CHIP = Children's Health Insurance Program; HCBS = home and community-based services.

This page has been left blank for double-sided copying.