National Bureau of Economic Research
Retirement and Disability Research Center

Social Security Administration Grant RDR18000003

Final Project Report: NB21-21

RDRC Medicaid Data Initiative:
Understanding Racial Disparities among SSI Recipients

Submitted by
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

January 11, 2023

Center Director: Nicole Maestas

Final Project Report, NB21-21

RDRC Medicaid Data Initiative: Understanding Racial Disparities among SSI Recipients

Nicole Maestas (Harvard Medical School and NBER) and Jessica Van Parys (Hunter College)

The purpose of the RDRC Medicaid Data Initiative is to build a library of administrative Medicaid data for the purpose of studying racial disparities in the health, health care, and program outcomes among SSI recipients. Because virtually all SSI recipients automatically qualify for Medicaid and most state Medicaid programs record recipient race/ethnicity, Medicaid data make it uniquely possible to measure and track progress toward eliminating health disparities among the universe of SSI recipients.

This project report describes the data purchased by the NBER for the RDRC Medicaid Data Initiative and outlines a tentative procedure for making the data available to other RDRC researchers in the future.

Description of the Data: The files, collectively known as the T-MSIS Analytic Files (TAF), consist of all Medicaid claims for all health care service types for 100% of recipients during 2018-2019 (plus a provider information file). In addition, the DUA granted NBER permission to re-use certain 2016-2017 data files that were purchased for a separate RDRC project (NB21-12). A copy of the executive summary from the current DUA is attached to this memorandum. We are in the process of filing a DUA amendment to establish re-use for the rest of the 2016-2017 files, and to purchase all of the 2020 files.

Data Acquisition Process: NBER entered into a Data Use Agreement (DUA) with the Centers for Medicare and Medicaid Services (CMS) on March 8, 2022 after a seven-month process of iteration and review with CMS' contractor, Research Data Assistance Center (ResDAC). After the DUA was approved, CMS contractor Acumen was instructed to prepare and deliver the files to NBER. Because Acumen was behind schedule due to the COVID-19 pandemic, NBER did not receive the files until July 2022.

Data Preparation for Researcher Use: Once the files were received at NBER, our programmers began a systematic process of extracting, testing, and implementing security protocols for the files. This process took several months because the data contain many enormous data files. For example, for each calendar year (2018, 2019), there are five primary file groups based on health care service type – inpatient (IP), other services (OT), long-term care (LT), pharmacy (RX), and demographics/eligibility (DE) – and secondary files. Each yearly file group has two files, a line-item file and a claim file. Each file took 40 hours to read in the absence of any errors and took longer when errors were encountered. A file mix-up by the contractor was discovered which resulted in further delays. To make it easier to access and store the files, NBER programmers split the files by state, year, and service type. NBER researchers were allowed to begin working with the files for their approved RDRC projects in December 2022.

Challenges: Throughout this process we learned that delays are inevitable, but we anticipate the data acquisition process will be somewhat shorter in the future as we gain experience with the DUA/amendment process and the pandemic-induced backlogs at the CMS contractors resolve. We

also learned that even after we receive the data, testing and preparing the data for researchers is a time- and resource-intensive activity for NBER programmers.

Wider Data Access: As we started to explore ways for NBER to make this Medicaid data available to RDRC researchers, it quickly became apparent that CMS reuse rules and NBER data security practices would require that researchers have a tie to the NBER. We can satisfy this requirement for RDRC researchers who propose a project (or project component) through the NBER RDRC. We'd ask for a modest budget to cover costs associated with the DUA itself, IRB review (required by CMS), data preparation, and data management. We note that NBER's RDRC already accepts proposals from and offers a limited appointment to researchers who are not NBER affiliates.

In the beginning stages of the data initiative, we anticipate demand for data access to consist of approximately 1-2 projects per year across the Consortium. If demand increases or as experience reveals new challenges, we will revisit these access procedures in an effort to keep NBER's workload and costs realistic and sustainable.