Replication Files for
Rules vs. Discretion: Treatment of Mental Illness in U.S. Adolescents

Authors
- Emily Cuddy: *Duke University*
- Janet Currie: *Princeton University*

Description of Data Files and Code

Data Files------
Note: We are not permitted to make the underlying data available (see Restrict Datasets #1 and #2 below); however, researchers should be able to use the data descriptions in these replication files to create similar data using any standard U.S. medical claims data set.

Restricted Dataset #1: BCBS Axis Data
This limited data set is made available through a secure data portal and is drawn from Blue Cross Blue Shield (BCBS) Axis®, the largest source of commercial insurance claims data in the U.S. Accessing insurance claims data often requires extended negotiations with individual insurance carriers, or with government entities. Further information about the BCBS Health of American Initiative, including information about their Axis® data base and contact information is available at: https://www.bcbs.com/the-health-of-america/about.

The full data extract is created using the following procedure:
- Identify all members who are observed before age 11;
- Restrict to members with a universal ID;
- Restrict to members who do not have carved out drug coverage;
- Restrict to members who are covered at least one year between the ages of 10 and 17;
- Restrict to members who have valid geographic and demographic information (i.e., age and sex) *and* have claims dates consistent with their coverage period (N=2,201,566).

These five restrictions yield 2,201,566 members. We extract the diagnosis codes associated with their facility and professional claims into the files:
- c2a_all.txt includes variables FACILITY ID (facility_id), ADMITTING ICD10 DIAGNOSIS CODE (admtng_icd10_diag_cd), PRINCIPAL ICD10 DIAGNOSIS CODE (princ_icd10_diag_cd), ADMITTING ICD9 DIAGNOSIS CODE (admtng_diag_cd), PRINCIPAL ICD9 DIAGNOSIS CODE (princ_icd9_diag_cd), and ID (personid).
- c2b_all.txt includes variables FACILITY ID (facility_id) and SECONDARY ICD10 DIAGNOSIS CODE (sec_icd10_diag_cd).
- c2d_all.txt includes variables FACILITY ID (facility_id) and SECONDARY ICD9 DIAGNOSIS CODES 1-10 (sec_icd9_dag_cd1-10).
- c3a_all_collapse.txt includes variables ID (patientid), PRINCIPAL ICD9 DIAGNOSIS CODE (princ_icd9_diag_cd), and ICD VERSION (icd).
- c3b_all_collapse.txt includes variables ID (patientid), SECONDARY ICD9 DIAGNOSIS CODE (sec_icd9_diag_cd1), and ICD VERSION (icd).

We further extract information on their total health care costs and hospital/ER use into the files:
- Comp_costs.txt includes variables TOTAL SPENDING (costs), ID (personid), YEAR (dyear), and SPENDING TYPE (type). Note: Spending type can refer to one of pharmacy, facility, or professional.

- Comp_hosp.txt includes variables ID (personid), HOSPITAL VISIT (i_hospital), ER VISIT (i_er), and DATE (dates).

Finally, we also extract member-level information on demographics, coverage dates, and location into the files:

- comparison_demog.txt includes variables plain ID (patientid0), BIRTHDAY (memb_brth_dt), and GENDER (memb_gender_cd).
- comparison_carveout_check2.txt includes variables EARLIEST COVERAGE (earliest), LATEST COVERAGE (latest), CARVED OUT PHARMACY (invalid), ID (personid), EARLIEST CLAIM (earliest_claim), LATEST CLAIM (latest_claim), MONTHS UNTIL FIRST CLAIM FROM EARLIEST COVERAGE (months_to_first), and MONTHS UNTIL END OF COVERAGE FROM LAST CLAIM (months_after_last).
- comparison_carveout_3.txt includes variables ID (personid) and MOST FREQUENTLY REPORTED ZIP CODE (member_zip).

The analysis sample imposes an additional restriction:

- Identify all members who had their first insurance claim for depression, anxiety, or adjustment disorder with anxiety or depression between the ages of 10 and 17 (N=202,066).

These six restrictions yield 202,066 members. We extract their facility, professional, and pharmacy claims from the restricted database into files ccc_pull_claims.txt and ccc_pull_mh_drugs.txt. We also extract member-level information on demographics, coverage dates, and location into the files ccc_pull_demographics.txt and ccc_pull_coverage.txt.

- Ccc_pull_claims.txt includes variables MONTH (dm), YEAR (dy), ID (patientid), MH SPENDING (mh_spend), NON-MH SPENDING (non_mh_spend), TOTAL SPENDING (total_spend), ANY EVALUATION (i_evaluation), ANY THERAPY (i_therapy), ANY OTHER MH TREATMENT (i_otreat), HOSPITAL VISIT (i_hospital), and ER VISIT (i_er).
- Ccc_pull_mhdrugs.txt includes variables ID (patientid), DAYS SUPPLY (days_supply_cnt), DATE (serv_dt), and NDC (prod_serv_id_cd).

  NOTE: Externally, we identify the active ingredient associated with each pharmacy claim using the FDA's National Drug Code Directory and Wolters Kluwer Medi-Span Data file using the first nine digits of the 11-digit NDC code; this yields the variable *active*. From these, we flag if the drug associated with each pharmacy claim is a mental health drug, benzodiazepine, opioid, or tricyclic antidepressant. The resulting information is saved in the file child_index_mini.dta which includes the variables NDC9 (ndc9), MENTAL HEALTH DRUG (i_mh), BENZODIAZEPINE (i_benzo), OPIOID (i_opioid), and TRICYCLIC ANTIDEPRESSANT (i_tca).

- Ccc_pull_demographics.txt includes variables plain ID (patientid0), BIRTHDAY (memb_brth_dt), GENDER (memb_gender_cd), and ID (patientid).
- Ccc_pull_coverage.txt includes variables START COVERAGE (begin_dt), CURRENT ZIP CODE (mbrs_crnt_prim_zip_cd), ID (patientid), EARLIEST MH (earliest), and MONTHS SINCE MH DIAGNOSIS (months_since_diagnosis).

Among these patients, we also extract the IDs of patients with a first mental health event in a hospital, ER, or evaluation setting into files ccc_pull_hospital.txt, ccc_pull_er.txt, and ccc_pull_evaluation.txt, respectively.

- Ccc_pull_hospital.txt includes variables ID (patientid) and FIRST MH IN HOSPITAL (first_hospital).
- Ccc_pull_er.txt includes variables ID (patientid) and FIRST MH IN ER (first_er).
- Ccc_pull_evaluation.txt includes variables ID (patientid).

We also extract detailed information about any facility visits for these members:

- Ccc2_monthly.txt includes variables MONTH (dm), YEAR (dy), ID (patientid), MH SPENDING (mh_spend), NON-MH SPENDING (non_mh_spend), TOTAL SPENDING (total_spend), ANY EVALUATION (i_evaluation), ANY THERAPY (i_therapy), ANY OTHER MH TREATMENT (i_otreat), HOSPITAL VISIT (i_hospital), and ER VISIT (i_er).
- Ccc2_temp2_length.txt includes variables ID (patientid), FIRST DATE IN HOSPITAL (first_serv_dt), and LAST DATE IN HOSPITAL (last_serv_dt.

Finally, we extract information on the diagnoses received by these members from their facility and professional claims into the files:

- facility_a.txt includes variables FACILITY CLAIM ID (facility_id), PLAIN ID (patientid0), ADMITTING ICD10 DIAGNOSIS CODE (admtng_icd10_diag_cd), PRINCIPAL ICD10 DIAGNOSIS CODE (princ_icd10_diag_cd), ADMITTING ICD9 DIAGNOSIS CODE (admtng_diag_cd), PRINCIPAL ICD9 DIAGNOSIS CODE (princ_icd9_diag_cd), ID (personid), and DATE (first_serv_dt).
- facility_b.txt includes variables FACILITY ID (facility_id) and SECONDARY ICD10 DIAGNOSIS CODE (sec_icd10_diag_cd).
- facility_d.txt includes variables FACILITY ID (facility_id) and SECONDARY ICD9 DIAGNOSIS CODES 1-10 (sec_icd9_dag_cd1-10).
- prof_a.txt includes variables PROFESSIONAL ID (professional_id), PRINCIPAL ICD9 DIAGNOSIS CODE (princ_icd9_diag_cd), PRINCIPAL ICD10 DIAGNOSIS CODE (princ_icd10_diag_cd), PLAIN ID (patientid0), ID (patientid), and DATE (serv_end_dt).
- prof_b.txt includes variables PROFESSIONAL ID (professional_id), SECONDARY ICD9 DIAGNOSIS CODES 1-3 (sec_icd9_diag_cd1-3), and SECONDARY ICD10 DIAGNOSIS CODES 1-3 (sec_icd10_diag_cd1-3).

In order to construct our instruments, we compile several additional extracts using a different sample. First, we extract all pharmacy claims for members between the ages of 10 and 18 into file ccc_physician_measure1e.txt. Second, we collapse these records at the physician-year-mental health level into the file ccc_physician_measure4b.txt. Third, we extract all professional claims for members and collapse these records at the physician-year-age bracket level into the file ccc_physician_measure7c.txt. Fourth, we identify the members between the ages of 10 and 18 with a diagnosis of depression, anxiety, and/or an adjustment disorder. Because the claims data use the plain ID, we extract the subset of these members with a universal ID (where applicable) into the file jpe_bp_kids_wmmi2.txt. For these members, we extract both their pharmacy claims into the file jpe_bp_pharm.txt and their coverage information into the file jpe_bp_kidscov_postdiag2.txt.

- ccc_physician_measure1e.txt includes variables NPI (npi_prscrb_prov_id_cd), AGE (age), NDC (prod_serv_id_cd), YEAR (script_year), and NEW ID (new_id).
- ccc_physician_measure4b.txt includes variables NPI (npi_prscrb_prov_id_cd), YEAR (script_year), PLAIN ID (patientid0), NON-MH PRESCRIPTIONS (non_mh_claims), MH PRESCRIPTIONS (mh_claims), and ID (patientid).
- ccc_physician_measure7c.txt includes variables NPI (npi_rndrng_prov_id_cd), YEAR (service_year), NUMBER OF PATIENTS (patients), and AGE BRACKET (age_bracket). The variable age_bracket can take on values of ‘child’ (age<18) or ‘adult’ (age≥18).
- jpe_bp_kids_wmmi2.txt includes variables ID (patientid) and PLAIN ID (patientid0).

- jpe_bp_pharm.txt includes variable PLAIN ID (patientid0), DATE EARLIEST MH DIAGNOSIS (earliest), ID (patientid), NDC (prod_serv_id_cd), NPI (npi_prscrb_prov_id_cd), and DATE (serv_dt).
- jpe_bp_kidscov_postdiag2.txt includes variables PLAIN ID (patientid0), START COVERAGE (begin_dt), CURRENT ZIP CODE (mbrs_crnt_prim_zip_cd), PHARMACY CARVE OUT (phrmcy_carve_out_subm_ind), DATE EARLIEST MH DIAGNOSIS (earliest), and ID (mmi_id). The variable PHRMCY_CARVE_OUT_SUBM_IND takes on values of C=CARVED OUT, I-INVALID, N-NOT CARVED OUT, X-NO DRUG BENEFIT, and Y=CARVED OUT.

Finally, in order to construct our mental health care markets, we compile several additional data extracts. In each case, we identify all BCBS enrollees who are between the ages of 10 and 18 with a claim for (i) therapy, (ii) a mental health prescription, and/or (iii) an antidepressant prescription into the files network_therapy.txt, network_mhs.txt, and network_ads.txt, respectively. Because the claims data use the plain ID, we extract the universal ID (where applicable) into the file network_ids_mmi.txt.

- network_therapy.txt includes variables PLAIN ID (patientid0), ZIP CODE (memb_clm_zip_cd), PROCEDURE CODE (cpt_and_hcpcs_cd), DATE (serv_end_dt), and NPI (npi_rndrng_prov_id_cd).
- network_ads.txt includes variables PLAIN ID (patientid0), DATE (serv_dt), NDC CODE (prod_serv_id_cd), NPI (npi_prscrb_prov_id_cd), and ZIP CODE (memb_clm_zip_cd).
- network_mhs.txt includes variables PLAIN ID (patientid0), DATE (serv_dt), NDC CODE (prod_serv_id_cd), NPI (npi_prscrb_prov_id_cd), and ZIP CODE (memb_clm_zip_cd).
- network_ids_mmi.txt includes variables PLAIN ID (patientid0), ID (patientid), and DATE (dates).

Notes:

1. In what follows, the variable name is listed in parentheses. Any variable with the prefix "i_" indicates a dummy variable.
2. There are two ID variables in the data: ID (patientid) and PLAIN ID (patientid0). The variable ID is consistent *across* plans *within* member whereas the variable PLAIN ID is consistent *within* plan-member. We define the variable NEW ID, which is equal to the ID where applicable and to the PLAIN ID otherwise.
3. Throughout, the date variable corresponds to the date of service rendered.

Restricted Dataset #2: IQVIA's Longitudinal Prescription data (LRx)

IQVIA Longitudinal Prescription Data (LRx) is a longitudinal patient prescription dataset based on retail pharmacy data. It enables the longitudinal tracking of patient prescription activity. Data is captured on co-prescribing as well as new, switch and repeat prescriptions that allows the monitoring of brand performance and analysis of market dynamics. IQVIA (formerly known as IMSQuintiles) is a public company specializing in pharmaceutical market intelligence. The IQVIA data is available for purchase to qualified researchers. For further information, contact Allen.Campbell@iqvia.com.

We reference the following restricted data files, where AD refers to antidepressants and AA refers to anti-anxiety medications:

- AD_Patient.dta and AA_Patient.dta which include variables PATIENT ID (PATIENT_ID), PATIENT BIRTH YEAR (PAT_BRTH_YR_NBR), GENDER (PAT_GENDER_CD), PATIENT FIRST FILL DATE (PAT_FIRST_FILL_DT), and PATIENT LAST FILL DATE (PAT_LAST_FILL_DT).
- AA_PatientActivityYYYY.dta and AD_PatientActivityYYYY.dta which include variables PATIENT ID (PATIENT_ID), MONTH IDENTIFIER (MONTH_ID), and RX COUNT (RX_CNT). There are data extracts for YYYY in {2006, 2007, … 2018}.

- AA_FactRXYYYY.dta and AD_FactRXYYYY.dta which include variables MONTH IDENTIFIER (MONTH_ID), PATIENT IDENTIFIER (PATIENT_ID), PATIENT ZIP-3 CODE (PAT_ZIP3), RX TYPE (RX_TYP_CD), and PROVIDER ID (PROVIDER_ID). The variable RX_TYPE_CD takes on values of 0 and 1, which are defined as new and refilled prescriptions, respectively.
- AA_Provider.dta and AD_Provider.dta which include variables PROVIDER ID (PROVIDER_ID), PHYSICIAN SPECIALTY DESCRIPTION (PRI_SPCL_DESC), PROVIDER TYPE (PROVIDER_TYP_ID), and NPI (NPI). The variable PROVIDER_TYP_ID takes on values of 1-PROFESSIONAL, 2-ORGANIZATION, -1-UNKNOWN, -2-ENCUMBERED, -3-PATIENT ZIP DEFAULT, -4-STORE ZIP DEFAULT, and -5-PRESCRIBER ZIP DEFAULT.

Unrestricted Datasets: External data files referenced in do-files

- ndc_data_info.xlsx: Wolters Kluwer Medi-Span Data file which includes the NDC, drug name, and generic product identifiers (08, 14). More information is available here: https://www.wolterskluwer.com/en/solutions/medi-span/about.
- product.txt: FDA National Drug Code Directory file from April 2018. Available for download here: https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory.
- PRINCETON_06192019.zip_to_zcta_2018.xlsx: HUD USPS ZIP code crosswalk file from 2018. Available for download here: https://www.huduser.gov/portal/datasets/usps_crosswalk.html.
- ZIP_COUNTY_12year.xlsx: HUD USPS ZIP code crosswalk files from 2012, 2013, 2014, 2015, 2016, and 2017. Available for download here: https://www.huduser.gov/portal/datasets/usps_crosswalk.html.
- gaz2016zcta5distance50miles.dta: NBER ZIP code distance file from 2016. Available for download here: https://www.nber.org/research/data/zip-code-distance-database.
- nppes_102018.dta: NPPES data file from October 2018. Using the standardized taxonomy. Available for download here: https://download.cms.gov/nppes/NPI_Files.html.

Code Files------

Data management:

- Iqvia_instrument.do uses the data from **IQVIA's LRx Data** to construct the IQVIA-based instrument. This file is called All_instrZ_noPAT.dta, and it includes variables PATIENT ZIP-3 CODE (PAT_ZIP3), YEAR (year), and NPI (npi).
- Intermediate_files.do uses the data from **IQVIA's LRx Data**, the **BCBS Axis Data**, and the **External Data** to produced many intermediate files which are referenced in the analysis code files. The code is separated into 13 sections (A-M), where each section defines the files which are created in that section, as well as a brief description of the code itself.
- Create_single_crosswalk_file.do uses the **External Data** to produce a multi-year crosswalk between zip and county codes.
- 1.child.do uses **IQVIA's LRx Data**, the **BCBS Axis Data**, and the **External Data** in order to produce a monthly patient panel of claims.
- 2.regressions.do builds on 1.child.do in order to generate other variables and create a 24-month post initial-MH episode panel.
- 3.regression_setup.do builds on 2.regressions.do in order to generate the data used in the regression analysis.

Analysis:

1. Table1.do performs the calculations in Table 1 in the paper.
2. Table2.do performs the calculations in Table 2 in the paper.

3. Table3.do performs the calculations in Table 3 in the paper.
4. Table4_5.do performs the calculations in Tables 4 and 5 in the paper.
5. Table6.do performs the calculations in Table 6 in the paper.
6. Tablea3.do performs the calculations in Appendix Table A3 in the paper.
7. Tablea4.do performs the calculations in Appendix Table A4 in the paper.
8. Tablea5.do performs the calculations in Appendix Table A5 in the paper.
9. Bootstrap_outcomes.do performs the bootstrap analysis referenced in Footnote 18. The results are plotted in Appendix Figure A1.
10. Calculate_outcome_means_table_by_group.do calculates the outcome means listed in Tables 4 and 5.