SAS® Global Forum 2019: Identifying the Most Probable Principal Psychiatric Diagnosis using Administrative Data

Qingxian Chen, M.S., and Emily Leckman-Westin, Ph.D.,

New York State Office of Mental Health

ABSTRACT

Leveraging existing administrative data to better understand the course and consequences of psychiatric conditions is a potentially potent tool for medical researchers as well as those entities that manage and oversee health care for populations (states, plans, etc..). In psychiatry, some conditions are expected to remit, such that with appropriate treatment and care, the symptoms would be expected to be mitigated; while for other conditions, there is an expectation of a longer, more chronic course. In addition, people may express symptomatology from different conditions at different points in time, indicating that the primary presenting concern may fluctuate over time. Therefore, people with psychiatric conditions can present with a myriad of symptomatology and categorizing the underlying condition(s) can be difficult. Being able to identify the most likely probable principal diagnosis during a specific time-period using administrative records would help to identify cohorts of individuals with similar conditions, allow for the indication of most salient presenting symptomatology, and provide the best estimate of the most probable principal psychiatric diagnosis. This paper will review the macro developed and its steps to create a unique mental health diagnosis for a cohort of clients using mental health clinic services as well as compare the results to other, commonly used, diagnostic specification algorithms using administrative data.

INTRODUCTION

Psychiatric illnesses often share symptomatology, like depressed mood, which can make diagnosis of the presenting illness difficult (1). Furthermore, while early symptoms may suggest one diagnosis, as the disease progresses additional symptoms may suggest a different psychiatric condition. In fact, administrative claims' data often identify multiple diagnoses for a single individual over time. While this may represent true comorbidity, in some cases being able to identify the most probable principal diagnosis during a specific time-period would help researchers to identify cohorts of individuals with similar conditions. In this paper, we propose a preponderance method as a best estimate of the most probable principle psychiatric diagnosis in a given time period. Then, we compare the results of the preponderance with two commonly used diagnostic specification algorithms using claims data. Finally, we discuss the situations where the preponderance method may be most appropriate for answering questions regarding mental health diagnosis using administrative data.

IDENTIFYING THE BEST ESTIMATE MENTAL HEALTH DIAGNOSIS

The Preponderance method uses the primary diagnosis on the ten most recent mental health claims in the year, leveraging information on the frequency, recentness and intensity of mental health services received to assign a most probable principle psychiatric diagnosis. In this paper, we describe the data preparation (extraction/cleaning) and how to compile the data for the developed psychiatric diagnosis preponderance macro. Then we step through the macro, identifying all the data steps with explanatory details of the intermediate tables. Finally, we describe the final data set produced by the psychiatric diagnosis preponderance macro.

DATA EXTRACTION, PREPARATION AND CLEANING

Using the New York State Office of Mental Health Medicaid Claims data warehouse, we identified the population as Medicaid enrollees who received mental health clinic services during 4/1/2017-12/31/2017, were less than 65 years old, had continuous Medicaid eligibility, and were not eligible for Medicare in 2017 (N=292,945). To illustrate the preponderance method and compare to other diagnostic strategies, we took a random sample of 20,000 individuals from this population for diagnosis.

DATA EXTRACTION

Original claim number, primary diagnosis code, secondary diagnosis code, date of service (date of admission for inpatient records), invoice type, place of service and service setting fields were extracted from 2017 Medicaid claims for sample clients. Claims in dental, vision, radiology, medical equipment, laboratory and pathology, and ordered ambulatory services were excluded in the diagnosis assignment process. Secondary diagnosis codes were used only for the second two methods, invoice type and date of service fields were used to identify the population with at least 1 inpatient setting or 2 or more outpatient/emergency department/non-acute inpatient setting.

DATA PREPARATION

The Preponderance algorithm needs 7 variables to identify the most probable psychiatric diagnosis. The Client Id, Claim Number, and Date of Service (DOS) are taken directly from the source extraction file. Three variables are constructed from the primary diagnosis: the diagnosis category is derived from the primary chapter heading from the Diagnostic and Statistical Manual 5.0 based on the ICD-10 code[1]. There are a few exceptions that we made, for instance we choose to identify ADHD and Autism separately from Neurodevelopmental Disorders because of the prevalence and concerns for individuals with these diagnoses as compared to

---

[1] Depending on the use case and type of diagnoses available in the administrative data, the diagnosis categories can be manipulated at this step. Specifically, one could choose to use the ICD-10 major diagnostic categories, see the ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research

individuals with only intellectual disability and/or other rare conditions, like Tourette's Disorder. These decisions should be based on the use case and needs of the underlying analysis and question of interest, see footnote. The Mental Health Score identifies the diagnostic category of the primary diagnosis: 1 - Schizophrenia, 2 - Bipolar, 3 - Depression, 4 - PTSD, 5 - Trauma and stress related disorders, 6 - Anxiety disorder, 7 - Autism, 8 - Conduct disorder, 9 - ADHD, 10 - Neurodevelopmental Disorders, 11 - Any MH diagnosis not listed before, 12 - Substance -Related and Addictive Disorders. The sequence number identifies whether the diagnosis is primarily mental health (1) or substance related (2). Depending on the analytic needs, the sequence number is used so the substance-related disorder can be assigned after MH disorder, if the mental health conditions are decided to be the primary analytic focus.

The macro also takes into consideration the service intensity level an individual received, thus each claim was assigned a service intensity score based on service type. Intensity service scores were given to each claim: 1 - Inpatient, 2 – Mental Health Specialty Service (e.g. Assertive Community Treatment), 3 – Mental Health Specialty Clinic Service, 4 - Emergency Room Visits, 5 - Outpatient Mental Health Service in a setting other than Mental Health Specialty Clinic, 6 - Other Outpatient Services except Nursing Home, and Home Care, 7 - Nursing Home and Home Care.

The final input dataset for preponderance diagnosis algorithm macro includes following fields: Recip ID, original claim number, date of service, primary diagnosis category, mental health score, MH/SU sequence score, and service intensity score.

DATA CLEANING

Administrative billing data can be messy. To avoid over counting diagnoses, we identified two cases that require data cleaning. First, we kept the unique diagnosis from different services captured on the same day, but when an individual received a diagnosis on the same day from different services, only the claim associated with the most intensive service was kept (case 1) to avoid over counting of the occurrence of the diagnosis. Second, some episodes of service identified multiple visits on different days and billed for each date of service, only the claim associated with the most recent date of service is kept in the algorithm to avoid over representativeness of a diagnosis (case 2).

SAS® code for data cleaning:

```
* step 1: data cleaning;
* a. For multiple claims on same date and same diagnosis, keep the record with most intense service;
proc sort data=diag_claims;
by recipID DOS diag_dsm service_score;
data CleanData1;
  set diag_claims;
  by recipID DOS diag_dsm service_score;
  if first.diag;
run;
```

```
* b. For claims on different date with same claim number, keep most recent and intense service claim;
proc sort data=CleanData1;
by recipID claimNO seqn descending DOS diag_dsm service_score diag_seqn;
data CleanData;
  set CleanData1;
  by recipID claimNO seqn descending DOS diag_dsm service_score diag_seqn;
  if first.claimNO;
run;
```

Table 1: Sample case from data cleaning (case 1-2): Highlighted rows are kept in the data set.

Client ID	Date of Service	Original Claim Number	Primary Diagnosis	Service Intensity Score
1	1/18/2017	XXXXXXXXXXXXXX1	Depressive Disorders	1
1	1/18/2017	XXXXXXXXXXXXXX2	Depressive Disorders	4
2	5/10/2017	XXXXXXXXXXXXXX3	Bipolar and Related Disorders	2
2	5/11/2017	XXXXXXXXXXXXXX3	Bipolar and Related Disorders	2
2	5/15/2017	XXXXXXXXXXXXXX3	Bipolar and Related Disorders	2

MOST PROBABLE PRIMARY PSYCHIATRIC DIAGNOSIS: PREPONDERANCE MACRO

A macro was created to assign a most probable primary psychiatric diagnosis based on the ten most recent service claims, such that the most frequently listed diagnosis would be assigned to the individual (case 3). When there is a tie in frequency, the most recent diagnosis is assigned (case 4). If the diagnoses both occur on the same day, then a diagnosis associated with the most intense service would be assigned (case 5). If both have same intensity of service, the diagnosis is based on hierarchical order would be assigned (case 6). If the individual had less than ten claims during the year, then starting with the most recent claim, the algorithm reviews backwards until a match is found. If no match is found, then the diagnosis associated with most intensive service claim is assigned. If there is a tie on service intensity, then the most recent diagnosis would be assigned. If both occur on same day, diagnosis will be based on mental health score. In the case of a single claim with a diagnosis, the individual would be assigned to that diagnosis. The macro to assign the most probable primary psychiatric diagnosis is provided below.

SAS MACRO

```
%macro MostProbDiag
(sample,   /* The data set contains a list of individuals need to be assigned a diagnosis */
infile,   /* Service claims data for sample clients in the period of observation*/
recipID, /* Client ID   */
claimNo  /* claim number associated with the claim, if using netted data may not be
              necessary    */
DOS,     /* Date of Service associated with the service claim  */
```

```
Diag,    /* DSM 5.0 Chapter heading Diagnosis associated with the claim(some
                exceptions noted above – could use alternatives, like ICD Major
                Diagnostic Categories */
seqn,    /* set MH diagnosis=1, SU=2, can be set as null, if not focused mainly on
                mental health diagnosis relative to substance or reversed, if more
                interested in substance use diagnoses */
svc_score,/* service intensity score derived from service type
                , lower number equates more intense service      */
diag_seqn,/* diagnosis hierarchy sequence on severity of the diagnosis, lower number
                indicates more severe diagnosis   */
outfile   /* final data set with a most probable diagnosis assigned */
);


* Step 2:
a. Keep recipients with at least 10 claims,
b. Count number of claims per diagnosis, date of service and intensity of service,
c. Order claims by number of claims (highest first), Date of service (most recent  first),and Intensity
of service (Most intense first),
d. Assign the most frequent diagnosis, if there is a tie on the number of occurrence, keep the most
recent diagnosis, if there is a tie on the most recent diagnosis, keep the one with most intense
service;

* pick those had at least 10 claims and order them in the order of number of claims per diagnosis,
date of service, intensity of service;

proc sql;
   create table recipient_over_10claims as
   select distinct &recipID., &seqn., &DOS., &diag., &svc_score., &diag_seqn.
        from &infile.
    group by &recipID.
   having count(distinct &ClaimNO.)>=10
   order by &recipID., &seqn., &DOS. desc, &svc_score., &diag_seqn.;
quit;

* keep the most recent 10 claims for recipients with ge 10 claims;
data recipient_10claims;
  set recipient_over_10claims;
  by &recipID. &seqn. descending &DOS. &svc_score. &diag_seqn.;
     cnt+1;
  if first.&recipID. then cnt=1;
  if cnt<=10;
run;

* count number of claims per diagnosis for each recipient;
proc sql;
   create table claims_per_diagnosis as
   select &recipID., count(*) as claim_cnt, &diag., &DOS., &svc_score., &seqn.,  &diag_seqn.
     from recipient_10claims
    group by &recipID., &diag.
    order by &recipID., &seqn., claim_cnt desc, &DOS. desc, &svc_score., &diag_seqn.;
quit;

* assign the most frequent diagnosis, if there is a tie on the number of diagnoses, keep the most
recent diagnosis, if there are multiple diagnoses on the same day, keep the one with most intense
service;
data assigned_diag_over10claims;
  set claims_per_diagnosis;
    by &recipID. &seqn. descending claim_cnt descending &DOS. &svc_score. &diag_seqn.;
    if first.&recipID.;
```

```
   keep &recipID. &diag.;
run;

* Step 3:
a. For recipients with fewer than 10 claims, starting with most recent claim, work backwards until a
match is found.
b. If no match is found, then the most intensive service and most recent trumps,
c. If the recipient had only one claim, then the patient gets assign the diagnosis associated with the
one claim;
proc sql;
   create table recipient_less_10claims as
   select &recipID., count(*) as claim_cnt, &diag., &seqn., &DOS., &svc_score., &diag_seqn.
         from &infile.
   where &recipID. not in (select distinct &recipID. from recipient_over_10claims)
   group by &recipID., &diag.
   order by &recipID., &seqn., &DOS. desc, &svc_score., &diag_seqn.;
quit;

* a. find the list of claims with matches on diagnosis, eg. claim_cnt>=2;
proc sql;
   create table claims_w_match as
   select distinct * from recipient_less_10claims
   where claim_cnt>=2
   order by &recipID., &seqn., &DOS. desc, &svc_score., &diag_seqn.;
quit;

* for diagnoses having two or more claims, the most recent diagnosis is the first diagnosis finding a
match when working backwards and will be assigned as the most probable diagnosis;
data claims_w_match2;
  set claims_w_match;
  by &recipID. &seqn. descending &DOS. &svc_score. &diag_seqn.;
  if first.&recipID.;
run;

* c. find the list of recipients only had one claim;
data single_claim;
   set recipient_less_10claims;
   by &recipID. &seqn. descending &DOS. &svc_score. &diag_seqn.;
   if first.&recipID. and last.&recipID.;
run;

* combine the list with match and single claims and order by date of service;
proc sql;
  create table claim_w_match_or_single_claim as
  (select distinct * from claims_w_match2)
union
  (select distinct * from single_claim)
  order by &recipID., &seqn., &DOS. desc, &svc_score., &diag_seqn.;
quit;

* pick the diagnosis on most recent claim and assign to the recipient;
data assigned_diag_less10a;
   set claim_w_match_or_single_claim;
   by &recipID. &seqn. descending &DOS. &svc_score. &diag_seqn.;

   if first.&recipID.;
   keep &recipID. &diag.;
run;
```

```
* b. find the claims for the recipients with no matching on diagnosis;
proc sql;
   create table claims_wo_match as
   select distinct * from recipient_less_10claims
    where &recipID. not in (select distinct &recipID. from assigned_diag_less10a)
    order by &recipID., &seqn., &svc_score., &DOS. desc,  &diag_seqn.;
quit;

* assign the diagnosis according to the intensity of service;
data assigned_diag_less10b;
   set claims_wo_match;
   by &recipID. &seqn. &svc_score. descending &DOS.  &diag_seqn.;
   if first.&recipID.;
run;

* combine all the assigning results get the distribution of BH diagnosis for all recipients;
data assigned_diag;
  set assigned_diag_over10claims
     assigned_diag_less10a
     assigned_diag_less10b;
  keep &recipID. &diag.;
run;

* find the list of recipients with any alcohol/substance abuse disorder;
proc sql;
        create table dx_alsa AS
        select distinct &recipID., 1 as dx_alsa from &infile.
    where &diag.='Substance-Related and Addictive Disorders'
    order by &recipID.;
quit;

proc sort data=&sample.; by &recipID.;
proc sort data=assigned_diag; by &recipID.; run;

data &outfile.;
  merge &sample.
     assigned_diag
     dx_alsa;
        by &recipID.;
run;
%mend MostProbDiag;

%MostProbDiag (SamplePOP, CleanData, recipID, claimNO, DOS, Diag_dsm, seqn, service_score,
diag_seqn, MostProbDiag_assigned);
```

Table 2: Sample cases from preponderance diagnosis algorithm (case 3-6)

Client ID	Date of Service	Primary Diagnosis	Number of Claims per Diagnosis	Service Intensity Score	Diagnosis Hierarchy Number
3	12/1/2017	Bipolar and Related Disorders	5	6	5
3	11/1/2017	Bipolar and Related Disorders	5	6	5
3	10/1/2017	Bipolar and Related Disorders	5	6	5
3	9/1/2017	Bipolar and Related Disorders	5	6	5
3	8/31/2017	Bipolar and Related Disorders	5	6	5
3	11/28/2017	PTSD	4	6	2
3	10/31/2017	PTSD	4	6	2
3	9/21/2017	PTSD	4	6	2
3	8/22/2017	PTSD	4	6	2
3	8/24/2017	Schizophrenia Spectrum and Other Psychotic Disorders	1	5	1
4	12/15/2017	PTSD	5	3	5
4	12/8/2017	Schizophrenia Spectrum and Other Psychotic Disorders	5	3	1
4	12/1/2017	PTSD	5	3	5
4	11/17/2017	Schizophrenia Spectrum and Other Psychotic Disorders	5	3	1
4	11/10/2017	Schizophrenia Spectrum and Other Psychotic Disorders	5	3	1
4	11/3/2017	Schizophrenia Spectrum and Other Psychotic Disorders	5	3	1
4	10/23/2017	Schizophrenia Spectrum and Other Psychotic Disorders	5	3	1
4	10/2/2017	PTSD	5	3	5
4	9/22/2017	PTSD	5	3	5
4	9/15/2017	PTSD	5	3	5
5	12/1/2017	Depressive Disorders	5	3	3
5	12/1/2017	Anxiety Disorders	5	6	6
6	11/30/2017	Schizophrenia Spectrum and Other Psychotic Disorders	5	3	1
6	11/30/2017	Depressive Disorders	5	3	3

Table 3 provides the prevalence of each diagnosis in the random sample of 20,000 mental health clients. The most common most probable psychiatric diagnosis was depression, with 28% (n=5,593). This was followed by schizophrenia and other psychotic conditions, with 13.1% of the population identified. This was followed by anxiety (11.5%), bipolar (10.6%), and ADHD (9%). Overall, these prevalence proportions are consistent with the mental health clinic clients in New York State. Specifically, among the most recent Patient Characteristics Survey for all clients (all individuals in the public mental health system), mood disorders were the most common, followed by schizophrenia and anxiety (2). They did not break out by clinic services or more refined than these 3 conditions.

ADDITIONAL DIAGNOSTIC METHODS

We identified two additional diagnostic strategies commonly used in administrative data to compare the preponderance method.

HEDIS METHOD: 1 INPATIENT OR 2 OUTPATIENT DIAGNOSES

Many investigations leverage at least 1 diagnosis in an inpatient setting or 2 or more diagnoses in outpatient/emergency department/non-acute inpatient settings to identify cohorts of individuals with a given diagnosis. We have termed this the "HEDIS method". In the *Adherence to Antipsychotic Medications for Individuals with Schizophrenia"* measure, HEDIS defined the population with schizophrenia as having at least two encounters with a diagnosis of schizophrenia/schizoaffective disorder with different dates of service in an outpatient setting, emergency department setting, or non-acute inpatient setting, or at least one encounter with a diagnosis of schizophrenia/schizoaffective disorder in an acute inpatient setting in the measurement year (3). Leveraging this base definition, we identify the primary and secondary diagnoses from at least 1 inpatient setting or 2 or more outpatient/emergency department/non-acute inpatient setting in 2017 were used to identify the presence of schizophrenia, bipolar, depression, trauma and stress related disorders, PTSD, anxiety disorder, autism, conduct disorder, and ADHD. Each member in the sample was flagged for the above diagnoses with either a 1(met the diagnostic specifications) or 0 (did not met the diagnostic specfications).

SAS code for assigning HEDIS diagnosis code:

```
* &schiz -- &ALSA macro variables contain the list of DX codes for each diagnostic
category based on DSM 5 Chapter Heading, the list can contain whatever diagnosis
needed for analysis;

* flag clients with specific diagnosis using inpatient claims;
data inpt_diag;
   set inpt_claims;
   if PRIMARY_DIAG in (&schiz)   or SEC_DIAG in (&schiz) then schiz=1;
   if PRIMARY_DIAG in (&bipolar) or SEC_DIAG in (&bipolart) then bipolar=1;
   if PRIMARY_DIAG in (&depress) or SEC_DIAG in (&depress) then depress=1;
   if PRIMARY_DIAG in (&trauma)  or SEC_DIAG in (&trauma) then trauma=1;
   if PRIMARY_DIAG in (&PTSD)    or SEC_DIAG in (&PTSD) then PTSD=1;
   if PRIMARY_DIAG in (&anxiety) or SEC_DIAG in (&anxiety) then anxiety=1;
   if PRIMARY_DIAG in (&autism)  or SEC_DIAG in (&autism) then autism=1;
   if PRIMARY_DIAG in (&conduct) or SEC_DIAG in (&conduct) then conduct=1;
   if PRIMARY_DIAG in (&ADHD)    or SEC_DIAG in (&ADHD) then ADHD=1;
   if PRIMARY_DIAG in (&ALSA)    or SEC_DIAG in (&ALSA) then ALSA=1;
run;

proc sql;
   create table inpt_diag_recip as
   select distinct RecipID, max(schiz) as schiz,max(bipolar) as bipolar,
   max(depress) as depress, max(trauma) as trauma, max(PTSD) as PTSD,
   max(anxiety) as anxiety, max(autism) as autism, max(conduct) as conduct,
   max(ADHD) as ADHD, max(ALSA) as ALSA
   from inpt_diag group by recipID;
quit;
```

```
* flag clients with specific diagnosis using outpatient claims;
proc sql;
create table outpt_diag as
select  recipID, sum(schiz) as schiz,sum(bipolar) as bipolar,
        sum(depress) as depress, sum(trauma) as trauma, sum(PTSD) as PTSD,
        sum(anxiety) as anxiety, sum(autism) as autism, sum(conduct) as conduct,
        sum(ADHD) as ADHD, sum(ALSA) as ALSA
from
(select recipID, DATE_OF_SERVICE,
max(case when PRIMARY_DIAG in (&schiz)  or SEC_DIAG in (&schiz)
        then 1 else 0 end) as schiz,
max(case when PRIMARY_DIAG in (&bipolar) or SEC_DIAG in (&bipolar)
        then 1 else 0 end) as bipolar,
max(case when PRIMARY_DIAG in (&depress) or SEC_DIAG in (&depress)
        then 1 else 0 end) as depress,
max(case when PRIMARY_DIAG in (&trauma)  or SEC_DIAG in (&trauma)
        then 1 else 0 end) as trauma,
max(case when PRIMARY_DIAG in (&PTSD)    or SEC_DIAG in (&PTSD)
        then 1 else 0 end) as PTSD,
max(case when PRIMARY_DIAG in (&anxiety) or SEC_DIAG in (&anxiety)
        then 1 else 0 end) as anxiety,
max(case when PRIMARY_DIAG in (&autism)  or SEC_DIAG  in (&autism)
        then 1 else 0 end) as autism,
max(case when PRIMARY_DIAG in (&conduct) or SEC_DIAG  in (&conduct)
        then 1 else 0 end) as conduct,
max(case when PRIMARY_DIAG in (&ADHD)    or SEC_DIAG  in (&ADHD)
        then 1 else 0 end) as ADHD,
max(case when PRIMARY_DIAG in (&ALSA)    or SEC_DIAG  in (&ALSA)
         then 1 else 0 end) as ALSA
from Outpt_claims group by recipID, DATE_OF_SERVICE)
group by recipID;
quit;


* assign clients with the specific diagnosis if the client had at least one inpatient
or 2 outpatient claims with the diagnosis;
proc sql;
  create table diag_hedis as
  select distinct a.recipient_id_1010,
        case when b.schiz=1 or c.schiz >=2 then 1 else 0 end as schiz,
         case when b.bipolar=1 or c.bipolar>=2 then 1 else 0 end as bipolar,
        case when b.depress=1 or c.depress >=2 then 1 else 0 end as depress,
            case when b.trauma=1 or c.trauma>=2 then 1 else 0 end as trauma,
            case when b.PTSD=1 or c.PTSD >=2 then 1 else 0 end as PTSD,
            case when b.anxiety=1 or c.anxiety>=2 then 1 else 0 end as anxiety,
        case when b.autism=1 or c.autism >=2 then 1 else 0 end as autism,
            case when b.conduct=1 or c.conduct>=2 then 1 else 0 end as conduct,
            case when b.ADHD=1 or c.ADHD>=2 then 1 else 0 end as ADHD,
            case when b.ALSA=1 or c.ALSA>=2 then 1 else 0 end as ALSA
   from SamplePOP a left join inpt_diag_recip b on a.recipID=b.recipID
                 left join outpt_diag c on a.recipID=c.recipID;
quit;
```

HIERARCHICAL METHOD

Another method to assign one diagnosis for each member is to apply a hierarchy order based on the severity of the diagnosis. Diagnoses are ordered from most severe to less severe in the order of: Schizophrenia, bipolar, depression, trauma and stress related disorders, PTSD, anxiety disorder, autism, conduct

disorder, ADHD (4). This hierarchical method leveraged the results from HEDIS method, reported above. If a person presented multiple diagnoses, we assign the person the most severe diagnosis. For example, when both schizophrenia and bipolar are presented, we assigned a diagnosis of schizophrenia. The full hierarchy is 1 - Schizophrenia, 2 - Bipolar, 3 - Depression, 4 - PTSD, 5 - Trauma and stress related disorders, 6 - Anxiety disorder, 7 - Autism, 8 - Conduct disorder, 9 - ADHD, 10 - Neurodevelopmental Disorders, 11 - Any MH diagnosis not listed before, 12 - Substance -Related and Addictive Disorders.

RESULTS AND DISCUSSION

Number and the prevalence proportion of each diagnosis using each method are presented in table 3. In general, all three methods identify depressive disorders as the most common. Schizophrenia followed for both the most probable and the hierarchy methods, while anxiety was the second most common in HEDIS (see table 3). Some of the differences identified in Table 3 are method dependent – for instance, the hierarchy definition forced the identification of schizophrenia and bipolar over depression and anxiety. In addition, the HEDIS method allowed clients to be assigned to more than one diagnosis, while the preponderance and hierarchy method only assign one diagnosis to each client. Given this, the HEDIS method identified more individuals in each category that the other methods that allow a single diagnosis per person.

Table 3: Diagnostic Prevalence Comparison in 20,000 Mental Health Clinic Clients

Diagnoses (N=20,000)	Most Probable Diagnosis		Hierarchy Diagnosis		HEDIS Diagnosis	
	n	%	n	%	n	%
Depressive Disorders	5595	28.0%	6452	32.3%	7772	38.9%
Schizophrenia Spectrum and Other Psychotic Disorders	2604	13.0%	3143	15.7%	3143	15.7%
Anxiety Disorders	2295	11.5%	1587	7.9%	4785	23.9%
Bipolar and Related Disorders	2112	10.6%	2570	12.9%	2997	15.0%
Trauma- and Stressor-Related Disorders	1824	9.1%	1775	8.9%	2693	13.5%
Attention Deficit Hyperactivity Disorder	1813	9.1%	1262	6.3%	2814	14.1%
Post Traumatic Stress Disorder (PTSD)	1039	5.2%	779	3.9%	2011	10.1%
Disruptive, Impulse-Control, and Conduct Disorders	669	3.3%	704	3.5%	1340	6.7%
Autism Spectrum Disorder	153	0.8%	177	0.9%	383	1.9%
No Diagnosis Above	1111	5.6%				
Substance-Related and Addictive Disorders	370	1.9%				
No Primary BH Diagnosis	415	2.1%				
Any Substance-Related and Addictive Disorders	3325	16.6%			3705	18.5%

Just about every individual given a most probable primary diagnosis was also identified by the HEDIS method (see table 4). The few cases that were differently assigned is mainly due to the use of any 1 outpatient record, in the absence of other information by the preponderance method.

Table 4: Almost All Individuals identified with a diagnosis using the Most Probable Psychiatric Diagnosis were also identified with the diagnosis using the HEDIS method

Diagnoses	Identified by Most Probable Dx and HEDIS DX	
	n	%
Schizophrenia Spectrum and Other Psychotic Disorders	2580	99.0%
Bipolar and Related Disorders	2074	98.3%
Depressive Disorders	5411	96.7%
PTSD	1000	96.2%
Trauma- and Stressor-Related Disorders	1642	90.1%
Anxiety Disorders	2196	95.7%
Autism Spectrum Disorder	150	98.7%
Disruptive, Impulse-Control, and Conduct Disorders	631	95.0%
Attention Deficit Hyperactivity Disorder	1776	98.0%

Table 5 identifies the overlap between the Most Probable Primary Psychiatric Diagnosis and HEDIS. This table highlights the amount of comorbidity that exists in our data. In general, the literature suggests that a diagnosis of schizophrenia is reasonably stable over time, providing some support for using a hierarchy that puts some preference on the identification of schizophrenia over other diagnoses (5,6). That said, even with a diagnosis of schizophrenia, a number of individuals did change to other mental health conditions over a ten-year period, suggesting that even for those identified with schizophrenia, it is important to allow for changing diagnoses. The level of overlap between the most probable primary psychiatric diagnosis and the HEDIS method was quite high, with over 82%. The evidence is more mixed for a diagnosis of bipolar. While an acceptable amount of stability was identified, there were multiple factors identified that affect the stability of bipolar over time, most notably, substance use (7).

Anxiety, autism and substance use disorders had the least overlap or identified the most comorbidity. The substance use disorders overlap proportion was expected, as the macro assigned a mental health condition prior to allowing the substance use disorder. However, the macro also allowed for the identification of any substance use disorder, as the comorbidity of both mental health and substance use conditions is high and related to the expected course of illness (8,9). When we look at those with any substance diagnosis (how the preponderance method also looks at substance), we found a 96.6% overlap. Autism highlights an interesting aspect of the diagnosis spectrum itself- many conditions, like Rett syndrome, other pervasive developmental disorders would be captured in the other

category, but could also commonly be used to identify individuals on the spectrum. In addition, our data consists of clients who are engaged with a specialty mental health clinic, so it may be expected that Autism wouldn't be as prevalent in this population. This suggests that if autism is the primary diagnosis of interest, the most probable psychiatric diagnosis would need to be tweaked a bit to ensure that all the diagnoses of interest were given preference over some of the other diagnoses. In some literature, primarily in children, autism is higher in the hierarchy and this would be one approach to better identify this population. Overall, the prevalence of autism in the mental health clinic population was quite low, with 1.9% identified with the HEDIS definition, to less than 1% in the other two methods. Finally, the differences in anxiety are important. Anxiety is commonly comorbid with other psychiatric conditions, most notably with depressive disorders (1,10). In our data, the preponderance method identified over a quarter of the HEDIS identified individuals with depression over anxiety, as the primary presenting condition. Further, in literature reviewing the validity of administrative data to identify psychiatric illness, identifying true cases of anxiety presented as a particular concern (11). Given this, it would be helpful to conduct a validation analysis, with a standardized diagnostic assessment to identify whether which is the best estimate of the most probable primary diagnosis.

Table 5: Identification of Comorbidity - Most Probable Primary Psychiatric Diagnosis and HEDIS Method

	HEDIS Identified Total	Most Probable Primary Psychiatric Diagnosis									
		Schizophrenia and Other Psychotic Disorders	Bipolar and Related Disorders	Depressive Disorders	PTSD	Trauma- and Stressor-Related Disorders	Anxiety Disorders	Disruptive, Impulse-Control, and Conduct Disorders	ADHD	Other DX	Substance-Related and Addictive Disorders
Schizophrenia and Other Psychotic Disorders	3143	82.09%	5.15%	5.15%	1.65%	0.45%	1.72%	0.22%	0.32%	2.83%	0.10%
Bipolar and Related Disorders	2997	6.54%	69.20%	9.71%	3.70%	1.17%	3.67%	0.73%	1.84%	2.97%	0.10%
Depressive Disorders	7772	3.91%	5.15%	69.62%	4.05%	2.78%	8.03%	0.94%	1.97%	2.97%	0.21%
Post Traumatic Stress Disorder (PTSD)	2011	4.08%	9.50%	21.73%	49.73%	2.93%	5.27%	0.60%	2.93%	2.83%	0.10%
Trauma- and Stressor-Related Disorders	2693	1.45%	2.90%	11.55%	2.82%	60.97%	5.35%	2.30%	5.90%	5.94%	0.04%
Anxiety Disorders	4785	3.74%	7.88%	26.65%	4.43%	3.20%	45.89%	0.82%	2.57%	3.74%	0.29%
Disruptive, Impulse-Control, and Conduct Disorders	1340	2.54%	3.51%	7.31%	1.42%	3.73%	3.36%	47.09%	18.88%	10.67%	0.00%
Attention Deficit Hyperactivity Disorder (ADHD)	2814	0.78%	2.99%	5.97%	1.60%	4.37%	4.37%	4.37%	63.11%	10.16%	0.11%
Substance-Related and Addictive Disorders	3705	14.28%	17.06%	31.20%	6.83%	5.29%	11.26%	1.21%	1.27%	2.13%	9.42%

These results highlight the need to consider the analytic questions at hand. Overall, if you need to compare a cohort of individuals with a diagnosis of schizophrenia to those with bipolar or depression, the most probable primary psychiatric diagnosis may be your best bet, as the HEDIS definition may be over-inclusive. If you want to better understand comorbidity of psychiatric illnesses, you may want to use a combined approach, with the most probable primary diagnosis used to identify the primary diagnosis, and the HEDIS method would identify any

comorbidity. Finally, the HEDIS method may be computationally easier and depending on the level of resources, may provide an effective way to identify the cohort of individuals based on their diagnostic profile.

REFERENCES

1. American Psychiatric Association: Diagnostic and statistical manual of mental disorders, 5th edition: DSM-5 [Internet]. Washington, DC, 2013.Available from: http://www.academia.edu/download/38718268/CSL6820_21.pdf
2. New York State Office of Mental Health: Tableau: New York State Office of Mental Health patient characteristics survey [Internet], 2017.Available from: https://omh.ny.gov/omhweb/tableau/pcs.html
3. National Committee for Quality Assurance: HEDIS 2018 technical specifications for physician measurement (epub) [Internet], 2018.Available from: http://store.ncqa.org/index.php/catalog/product/view/id/2869/s/hedis-2018-technical-specifications-for-physician-measurement-epub/
4. Cardno AG, Owen MJ: Genetic relationships between schizophrenia, bipolar disorder, and schizoaffective disorder. Schizophrenia Bulletin 40: 504–515, 2014.
5. Heslin M, Lomas B, Lappin JM, et al.: Diagnostic change 10 years after a first episode of psychosis. Psychological Medicine 45: 2757–2769, 2015.
6. Tsuang MT, Woolson RF, Winokur G, et al.: Stability of psychiatric diagnosis: Schizophrenia and affective disorders followed up over a 30- to 40-year period. Arch Gen Psychiatry 38: 535–539, 1981.
7. Cegla-Schvartzman FB, Ovejero S, López-Castroman J, et al.: Diagnostic stability in bipolar disorder: A narrative review. Harvard Review of Psychiatry 27: 3–14, 2019.Available from: http://insights.ovid.com/crossref?an=00023727-900000000-99957
8. Jorgensen KB, Nordentoft M, Hjorthoj C: Association between alcohol and substance use disorders and psychiatric service use in patients with severe mental illness: A nationwide Danish register-based cohort study. Psychological Medicine 48: 2592–2600, 2018.
9. Hjorthøj C, Østergaard MLD, Benros ME, et al.: Association between alcohol and substance use disorders and all-cause and cause-specific mortality in schizophrenia, bipolar disorder, and unipolar depression: A nationwide, prospective, register-based study. The Lancet Psychiatry 2: 801–808, 2015.
10. Kessler RC, Chiu TW, Demler O, et al.: Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the national comorbidity survey replication (NCS- R). Archives of General Psychiatry 62: 617–627, 2005.
11. Davis KAS, Sudlow CLM, Hotopf M: Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses. BMC Psychiatry 16: 263, 2016.Available from: http://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-016-0963-x

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Qingxian Chen
New York State Office of Mental Health
Qingxian.Chen@omh.ny.gov