



Rising Mental Health Incidence Among Adolescents in Westchester, NY

A. Ravishankar Rao¹ · Saroja Rao² · Rosy Chhabra³

Received: 29 June 2020 / Accepted: 29 January 2021 / Published online: 16 February 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

Context Many governments have publicly released healthcare data, which can be mined for insights about disease conditions, and their impact on society.

Methods We present a big-data analytics approach to investigate data in the New York Statewide Planning and Research Cooperative System (SPARCS) consisting of 20 million patient records.

Findings Whereas the age group 30–48 years exhibited an 18% decline in mental health (MH) disorders from 2009 to 2016, the age group 0–17 years showed a 5.4% increase. MH issues amongst the age group 0–17 years comprise a significant expenditure in New York State. Within this age group, we find a higher prevalence of MH disorders in females and minority populations. Westchester County has seen a 32% increase in incidences and a 41% increase in costs.

Conclusions Our approach is scalable to data from multiple government agencies and provides an independent perspective on health care issues, which can prove valuable to policy and decision-makers.

Keywords Big data analytics · Mental health

Introduction

Governments worldwide are releasing data about their functioning, consistent with the Sustainable Development Goals of the United Nations regarding accountability (<https://sustainabledevelopment.un.org/sdg16>). This created the Open Data Movement (Jetzek 2015), and expectations are high that this will lead to both accountability and effective management of government institutions to serve the people. Consequently, appropriate expertise and tools are required to explore government-released data, and disseminate the accompanying results to the public.

Open health data is the focus of this study, as healthcare is one of the largest expenditures for most governments,

especially in the United States. The big data revolution has made relevant data readily available, thereby allowing citizens to obtain an unbiased view of the health care status quo without interpretation by other reporting agencies (Kitchin 2014). In the U.S., the data sources include The Center for Medicare and Medicaid Services (CMS), which contains data about healthcare providers and hospital ratings (<https://data.medicare.gov/Physician-Compare/National-Downloadable-File/s63f-csi6>; <http://www.medicare.gov/hospitalcompare/data/total-performance-scores.html>); and the New York State (referred to as NY) Statewide Planning and Research Cooperative System (SPARCS) initiative (<https://www.health.ny.gov/statistics/sparcs/>), which collects data about de-identified patient discharges in NY. The CMS data covers nearly one million practitioners and three thousand hospitals in the U.S. The SPARCS data total approximately twenty million patient records in NY from 2009 to 2016. These data can provide important information about the patterns of disease outbreaks, trends in patient outcomes, and expected costs that can be used as a benchmark to determine a fair payment (Rao and Clarke 2017). The advantage of using the open health data sources is that they are released at regular intervals, and cover large patient populations.

Big data analytics is the use of advanced analytic techniques on very large data sets to discover hidden patterns and

✉ A. Ravishankar Rao
raviraodr@gmail.com

Saroja Rao
sarojara@buffalo.edu

Rosy Chhabra
rosy.chhabra@einsteinmed.org

¹ Fairleigh Dickinson University, Teaneck, NJ, USA

² State University of New York, Buffalo, NY, USA

³ Albert Einstein College of Medicine, New York, NY, USA

useful information. These techniques have been successfully applied in finance, marketing, and advertising (Minelli et al. 2012), and are now seeing widespread usage. These data sets provide a unique opportunity to determine trends in medical conditions within specific population demographics including race and ethnicity, age groups, geographical area, etc. Information at the local level can help policymakers take appropriate action to improve the health of a community. For public health researchers, it is often useful to connect the dots by comparing trends revealed by multiple data sources, as they may differ in terms of their sample sizes and areas of geographic coverage. In particular, big data techniques have substantial potential in mental health research (Russ et al. 2019). Figure 1 highlights the typical workflow, which includes: data cleansing/Extract Transform and Load (ETL), data-joining, feature engineering, clustering, classification and prediction, visualization of results, interpretation, and reporting. Data from multiple sources can be processed simultaneously this way. In the current study, we used big data analytics to understand open health data from the NY SPARCS repository.

Merikangas et al. (2009) observed that there is an “absence of empirical data on the magnitude, course and treatment patterns of mental disorders”, which has impeded efforts to establish mental health policy. Researchers have resorted to using meta-analyses of previously published work (Maura and Mamani 2017). Potential challenges for such meta-analyses include inconsistencies in the classifications of MH disorders, differences in time-periods for data collection, and differences amongst the regions studied. On the other hand, the NY SPARCS data provides a consistent

method to collect and disseminate the data through comprehensive clinical classifications software (CCS) diagnosis codes, annualized data collection, and geographical consistency by counties within the state. Hence, a systematic analysis of NY SPARCS data holds significant promise for MH research and for the guidance of MH policy. We applied our framework to examine the growth of MH issues amongst the adolescent population in NY and present a detailed view of important health trends in local populations.

According to the World Health Organization, “Mental health is a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her community”, and is considered an integral part of overall health (<https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>). MH issues have risen in the past decade in the U.S. (Patel et al. 2007) in the general population and impact the behavioral development of adolescents (Mills et al. 2012).

Children and adolescents go through different sensitive periods during their development (Andersen and Teicher 2008). For instance, the lifetime prevalence of major depression disorder increases from 1% of the population under age 12 to nearly 25% of the population by adolescence (Kessler et al. 2001). Kessler et al. (2005) provide a detailed report containing the ages at selected percentiles on the age-of-onset distributions of DSM-IV/WMH-CIDI Disorders. The age of 17 corresponds to a prevalence of 25% for drug abuse and for bipolar I-II disorders. The age of 11 corresponds to a prevalence of 50% of any impulse control disorder. By age 43, the prevalence of any mood disorder is 75% and by age 73 it reaches 99%. According to the CDC (<https://www.cdc.gov/childrensmentalhealth/data.html>), the rates of mental disorders change with age. The diagnoses of depression and anxiety increase steadily with age, whereas behavior disorders decline at ages 12–17 as compared with ages 6–11 years. Many of these figures and precise prevalence rates may change in the future based on a multitude of environmental factors including the use of digital technologies (Odgers and Jensen 2020).

One in four to five youth in the U.S. is affected by a MH disorder (Merikangas et al. 2010). Between 1995–1998 and 2007–2010, MH visits leading to diagnoses increased faster for youths (from 7.78 to 15.30 visits) than for adults (from 23.23 to 28.48 visits) per 100 people (interaction: $P < 0.001$) (Olfson et al. 2014).

The most common MH disorders among youth are anxiety disorders with the highest prevalence (31.9%), behavior disorders (19.1%), mood disorders (14.3%), and substance abuse disorders (11.4%) (Merikangas et al. 2010; <https://www.cdc.gov/childrensmentalhealth/features/school-aged-mental-health-in-communities.html>). About 40% of youth diagnosed with one of these disorders are also prone to

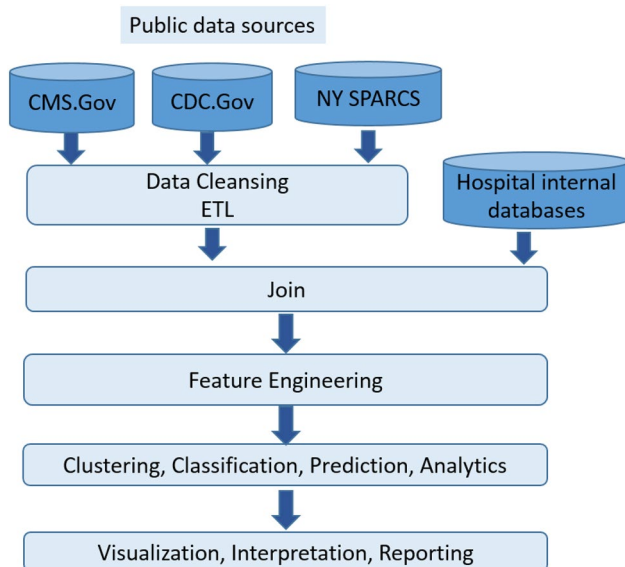


Fig. 1 Shows the data processing framework we have developed for analyzing open health care data

another MH disorder in the future (Merikangas et al. 2010). Nearly half the children with a MH disorder did not receive treatment or counseling from a MH professional (Whitney and Peterson 2019).

There are reported gender differences in MH related disorders with males exhibiting more ‘externalization’ versus females who exhibit more ‘internalization’ although females tend to report a higher prevalence of treated MH needs as compared to males (Mackenzie et al. 2006). Among racial and ethnic minorities, Latinos and African Americans report lower access to MH treatment than whites (Alegria et al. 2002).

Stewart and Davis (2016) surveyed the use of big data in MH research and showed that the most common MH issues addressed were unipolar depression and dementia. Furthermore, the bulk of this literature is devoted to medication-oriented research questions. Hence, there is a gap in the literature in using big data and open data to understand MH issues in the youth, and to explore questions related to prevalence and trends in this population. We address this gap in the current paper.

Research Questions

Though there are studies of national trends in MH (Olson et al. 2014), there may be different trends operating at state and county levels. Furthermore, the trends may change over time as new data is collected. Hence, it is important to explore the trends at different scales, continually update them, and communicate these findings to policy makers. The two main research questions addressed in this paper are as follows. What are the significant MH trends in NY? Which regions in NY show the largest increases in MH issues?

An examination of these trends should provide guidance to policy makers about the types of MH interventions required at appropriate locations.

Methods

Our specific software framework builds upon existing open-source components based on the Python programming language, which is becoming increasingly popular in data science and machine learning applications (Dhar 2013). We utilized Pandas (McKinney 2012), SciPy, Scikit-Learn, and Matplotlib (Hunter 2007). Pandas conveniently integrates tabular and statistical modeling with visualization packages such as Matplotlib. Our previous (Rao and Clarke 2017) and current research has successfully applied Pandas to health-care datasets. Our software framework is termed BOAT (Rao and Clarke 2017), for Big Data Open Source Analytics Tool,

and is available freely at <https://github.com/fdudataming/framework>.

We used de-identified data describing patient discharges and the cost of medical procedures in NY as reported by SPARCS (<https://www.health.ny.gov/statistics/sparcs/>) on an annual basis from 2009–2016. Each patient discharge comprises 40 fields containing values such as hospital, county, zip code, age group, ethnicity, CCS diagnosis code and description, duration of hospital stay, type of insurance coverage, and total cost. The age groups in the data consist of 0–17, 18–29, 30–49, 50–69, and 70 or older. We used discharge record information from the SPARCS dataset to create our cohort for analysis, consisting of children/youth (0–17 years) treated for MH disorders in NY, which includes basic demographics (age group, gender, race, and ethnicity) and county/hospital names. Our outcome data was based on the diagnosis description (focus was on adjustment disorders, anxiety disorders, behavior disorders, childhood disorders, developmental disorders, impulse control disorders, miscellaneous disorders, mood disorders, and schizophrenia) and the cost of each incident. Figure 2a shows sample data for a single patient record.

We processed this data with operations such as grouping by labels, binning, sorting, and statistical analysis as shown in Fig. 2b. For example, upon filtering the NY SPARCS data in 2014 containing approximately 2 million de-identified patient records, we obtained 115,680 records for Clinical Classification Software (CCS) codes representing mental disorders such as ‘Mood disorders’, ‘Anxiety disorders’, ‘Adjustment disorders’, and so on. Subsequently, we examined the prevalence of different types of MH disorders in the age group 0–17 in NY. There were 10,017 records representing MH disorders in the age group 0–17 in the NY SPARCS dataset in 2014. We carried out relevant statistical analyses to determine the relationships between race, ethnicity, and the prevalence of MH disorders. We further analyzed patients treated in Westchester County, which has the largest percentage increase in MH disorders in NY. There were approximately 1,984 cases for the age group 0–17 years in Westchester.

In order to compute trends, we used the values in the year 2009 as the baseline. Values for subsequent years, from 2010 to 2016 were compared with the value in year 2009 to compute percentage changes appropriately. For the computation of counts, we used the raw values and did not perform any normalization. Other views of the data, such as pie charts can be computed easily. The grouping of data was done through the use of text labels and the Pandas “groupby” command (McKinney 2012).

Our initial approach (Rao et al. 2015) represents one of the early efforts to use big data techniques on publicly available open health datasets. This approach has been gaining acceptance in the health data analytics literature (Rao

(A)

Hospital County	Westchester
Facility Name	New York Presbyterian
Age Group	0 to 17
Gender	M
Race	White
Ethnicity	Spanish/Hispanic
CCS Diagnosis Description	MOOD DISORDERS
Total Costs	\$30,123.33

(B)

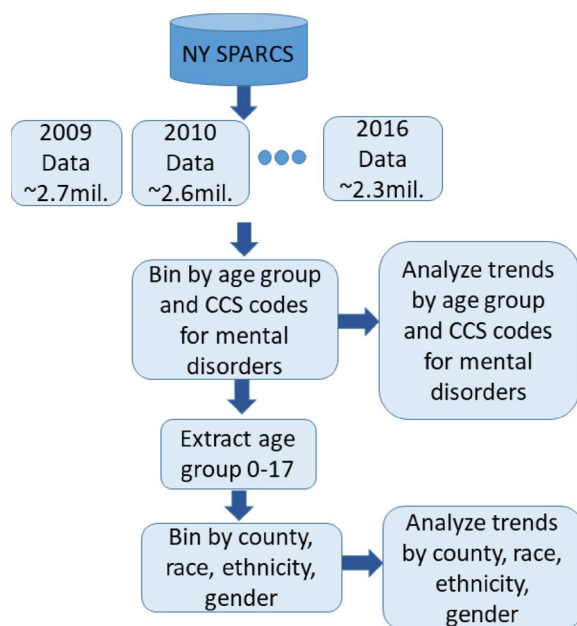


Fig. 2 a Sample data for a single patient visit from the year 2009 SPARCS repository. b The workflow used to process the data. SPARCS contains multiple datasets from 2009–2016. The size of the dataset varies from 2.7 million patient records in 2009 to 2.3 million patient records in 2016

et al. 2018, 2020, 2018; Rao and Clarke 2019). The techniques we used, including combining large datasets from multiple sources, the split-apply-combine paradigm, and open-source software tools have not been reported in the journal *Community Mental Health*. Of the current 4062 articles in the searchable database for the journal *Community Mental Health*, the term “big data” does not occur, and the term “data analytics” occurs only twice (Shukor et al. 2019; Mahmoud et al. 2020). Hence, the discussion of more recent techniques in the area of big data analytics could be

beneficial to the readers of this journal. Mikk et al. (2017) observe that considerable value can be obtained by merging disparate health data from multiple sources. Martin and Begany (Martin and Begany 2017) enlist the benefits of opening government health data to the public, including improved health literacy, consumer engagement, and community empowerment. The interpretation of open health data can lead to an understanding of community mental health issues (Rich et al. 2015).

Hafferty et al. (2017) made the following observation about big data in mental health research: “Familiarity with this toolbox of data science techniques will become a critical attribute for the psychiatric researcher of the twenty-first century. It should also be central to the strategies of psychiatric research funders.”

Results

Figure 3a shows the trends in mental illnesses/disorders by age group from 2009 to 2016.

From Fig. 3a we note that age groups 30–49 and 70 or older have been trending downwards, while the other age groups are trending upwards. The two age groups with rising trends from 2013 to 2016 are shown in bold lines and constitute the age groups 0–17 and 18–29. We will focus our attention on the age group 0–17 as this is a vulnerable age group consisting of school-going children.

The trends in Fig. 3b and c show that the percentage increase in the number of incidences and costs of MH disorders in Westchester county is significantly higher than other counties. The number of incidences increased by 30% and the costs increased by 40% between 2009 and 2016.

In NY, the total amount spent on MH in 2016 for children ages 0–17 was \$119,216,840 and \$29,983,002 of this amount (25%) was spent in Westchester county, although the population of Westchester County composes only 4.9% of the total population of NY (US Census Bureau, 2016).

Prevalence of Mental Health Disorders in NY for Ages 0–17 by CCS Diagnosis code and Demographics

We compare trends in the prevalence of mental disorders from years 2009–2016 as follows. Figure 4a shows the distribution and trends among the top six CCS diagnosis codes. The relative ordering of the top six codes is maintained between 2009 and 2016, where ‘mood disorders’ are the most prevalent, and increase the most rapidly, by 19%. Figure 4b shows a similar distribution and trend in Westchester County, with a rise in ‘mood disorders’ of 67%. Figure 4c shows that minorities (‘Blacks/African Americans’ and ‘other’) have a consistently higher prevalence of MH

disorders (54% in 2016) compared to the majority white population. Specifically, ‘Blacks/African Americans’ comprised 27.5% of the MH disorders in 2016, whereas their proportion of the New York population was 19.3%. Figure 4d shows a steadily increasing prevalence of MH disorders amongst females, rising from 48.5% in 2009 to 58% in 2016. Figure 4e shows that Spanish/Hispanics accounted for 15% of the MH disorders in 2016, whereas they comprised 18.8% of New York’s population.

Trends in cases for Mood Disorders in NY and Westchester County

Figure 4b shows that the rate of increase in mood disorder cases in Westchester County is significantly higher (67%) than the rest of NY (19%) during the period 2009–2016.

Discussion

Our finding in Fig. 3a is consistent with the observation in (Olfson et al. 2014) that between 2007 and 2010, mental disorder diagnoses increased significantly for youths compared to adults. Although the study in (Olfson et al. 2014) covers the period 2007–2010, our results show that this trend continues until 2016.

From Fig. 4a the relative occurrences of the MH disorders in 2016 are as follows: mood disorders (62%), behavior disorders (including attention deficit) (11.8%), schizophrenia or other psychological disorders (5.4%), adjustment disorders (5%), anxiety disorders (4.8%), and impulse control disorders (4.05%). Previous research (Olfson et al. 2014) using outpatient visits to physicians in office-based practice for the 2007–2010 period found that disruptive behavior disorders were the most common amongst youth (age < 21 years) at a relative percentage of 48%, followed by mood disorders (20%), anxiety disorders (13%), other (11%), and psychoses (6%). The data we analyzed are for inpatient visits. It may be necessary to combine an analysis of inpatient and outpatient visits to determine a holistic view of MH needs.

A study of emergency department data in NY showed that mood disorders were more prevalent than anxiety disorders (Lin et al. 2016). Another study using national data found that anxiety disorders were the most common MH disorders, followed by behavior disorders, mood disorders, and substance use disorders (Merikangas et al. 2009). Estimates of the prevalence rate of anxiety disorders are known to vary with an extremely wide range, from 2 to 24% (Merikangas et al. 2009). The NY SPARCS data provides a consistent source of data collected annually, hence serving as a useful benchmark for analysis.

The observation in Fig. 4d that minorities (‘Blacks/African Americans’ and ‘other’) have a higher prevalence of MH

disorders (54% in 2016) is consistent with prior research which shows that African Americans and Hispanics have lower access to MH care compared to whites (Maura and Mamani 2017; Alegria et al. 2002). It appears that instead of utilizing office visits (Olfson et al. 2014), this segment is using in-patient hospital services to treat MH disorders. A consistent and regular measurement of such differences between population segments is crucial in shaping public health policy (McGuire and Miranda 2008).

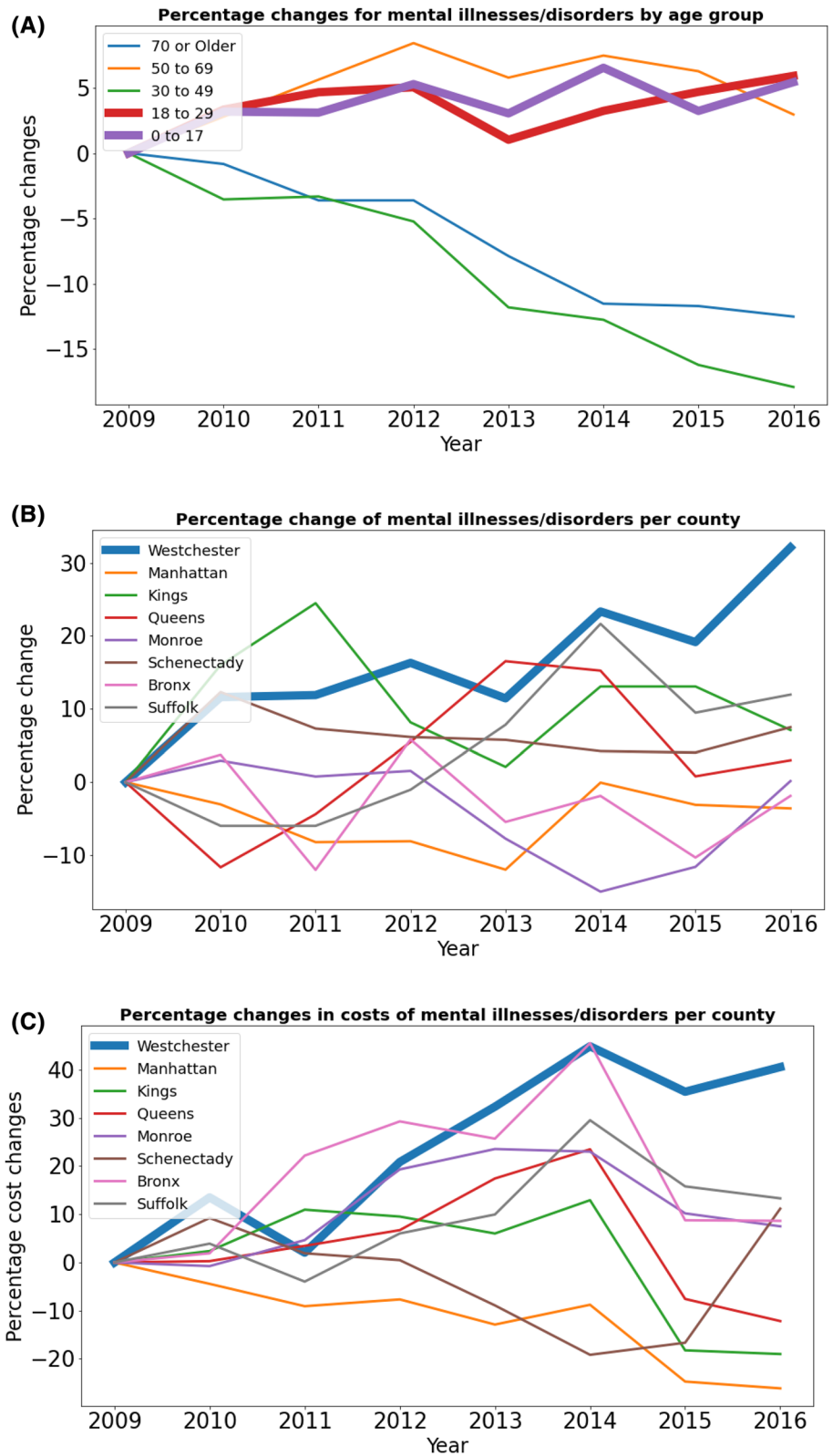
Figure 4e shows that females have greater in-patient visits for MH than males. This is consistent with research that shows that the incidence of depressive disorders rises sharply in pubescent females (LeGates et al. 2019). Other research with younger children (birth—age 7) also shows greater incidence of MH disorders amongst females (Pennap et al. 2018). Another factor is that men are less likely to communicate about their MH issues compared to women (Mackenzie et al. 2006).

Figures Fig. 4a and b show that the most prevalent MH issue is ‘mood disorders’, which also increased the fastest from 2009 to 2016. For NY as a whole, it increased by 19%. This increase is not uniform across all counties, and in Westchester County it increased by 67%. This finding is alarming in view of current research, and might demonstrate that the need is much higher than reflected by the numbers reported, since this data is based on in-patient care (discharge). Earlier research based on data from the National Comorbidity Study- Adolescent Supplement (NCS-A), a nationally representative survey of more than 10,000 teens, has already shown that only about 36% of youth with any lifetime mental disorder receive professional MH treatment, while teens with ADHD (60%) and behavior disorder (45%) utilize most services. Only about 38% with mood disorders receive services (Merikangas et al. 2014). Despite the availability of programs designed to improve MH services for youth, such as the State Children’s Health Insurance Program and the federal Children’s Mental Health Initiative, the majority of children and adolescents that need MH services are not receiving treatment (Kazdin 2019).

Our results show that the incidences of MH disorders have been steadily increasing. We offer a hypothesis for why Westchester County in particular has seen a dramatic increase. Previous research has shown that students are increasingly facing pressure to get into college and stay there (Mowbray et al. 2006). The competition begins as early as middle school (Hoff 2002), and is likely to be high in affluent counties such as Westchester. There may be other factors at play, and this needs further investigation. Active research is being conducted to find new methods for treatment, including social media (Kazdin 2019).

It is likely that children who develop MH disorders will continue to face challenges when they transition to college. The 2014–2015 report from the Center for Collegiate Mental

Fig. 3 **a** Trends in mental illnesses/disorders in NY for different age groups from 2009–2016. The year 2009 is treated as a baseline and percentage increases for the number of patient records from this year are calculated over subsequent years. **b** The trends in number of cases for age group 0–17 are examined by county. The top 8 counties with the most cases of MH disorders are depicted. The curve in bold, Westchester County, shows the largest percentage increase from 2009–2016. **c** The trends in cost increases are shown, and the curve for Westchester County shows the largest percentage increase



Health (<http://ccmh.psu.edu/>) states that college enrollment grew by 5.6% between 2009 and 2015. During this period, the number of students seeking MH services rose by 29.6%,

while the number of attended appointments rose by 38.4%. It is interesting and concerning that both the NY SPARCS data and the Center for Collegiate Mental Health data show

similarly large double-digit increases in MH issues for the younger generation. Though the data are measuring two different quantities, one dealing with hospital admissions and the other dealing with MH outpatient services at colleges, the trends are consistent with each other. Recent studies by Johnson (Johnson 2018) indicate that colleges continue to be overwhelmed by students seeking MH services. For instance, UCLA offered free online screenings for depression to incoming freshman in September 2017, where one quarter of the freshmen took the screening. It is important to introduce interventions early at the school level, as the growing potential for suicides is of great concern (Mackenzie et al. 2011).

The NY SPARCS data is available freely to the public and contains fine-grained de-identified information about individual patient characteristics including race, gender, and diagnosis. The Center for Collegiate Mental Health contains aggregate data that presents a bird's eye view of trends. It seems that the alarming rise in MH issues was hiding in plain sight in the NY SPARCS data. This visible trend has important policy implications, as schools/colleges are struggling to meet the increased demand for MH services for youth (Simon 2017).

Data released from public health records such as NY SPARCS could potentially provide leading indicators of health conditions that may face vulnerable sections of the populations such as the youth. Appropriate intervention programs in schools and colleges, such as MH services, should take these indicators into account for planning and budgeting purposes. Since the number of mood disorder cases in Westchester County surged by 67% between 2009 and 2016, and this was the highest increase among all counties in the NY SPARCS data, this may be taken as a “worst case” scenario for capacity planning purposes. Schools and colleges may consider planning for a similar capacity surge for their MH services by adding interventions at early stages to treat mood disorders. For instance, Mahopac High School in Westchester county added a new course to educate students about substance abuse in 2020 (Grosserode 2020). The school system (Glied et al. 1997) continues to be the most important single provider of services for children. School based MH interventions (Sanchez et al. 2018; Atkins et al. 2017) should be encouraged in the future. There is also room to integrate care across multiple systems, including family, school, and health systems (Power et al. 2019). Though our research and analysis was conducted on data gathered before the COVID-19 pandemic, there is concern about a cascading crisis in MH care in the future (Ramachandran 2020).

Big data techniques have not been used effectively in the service of minority populations. Hence, a comprehensive report from the National Institute on Minority Health and Health Disparities examined issues surrounding the use of

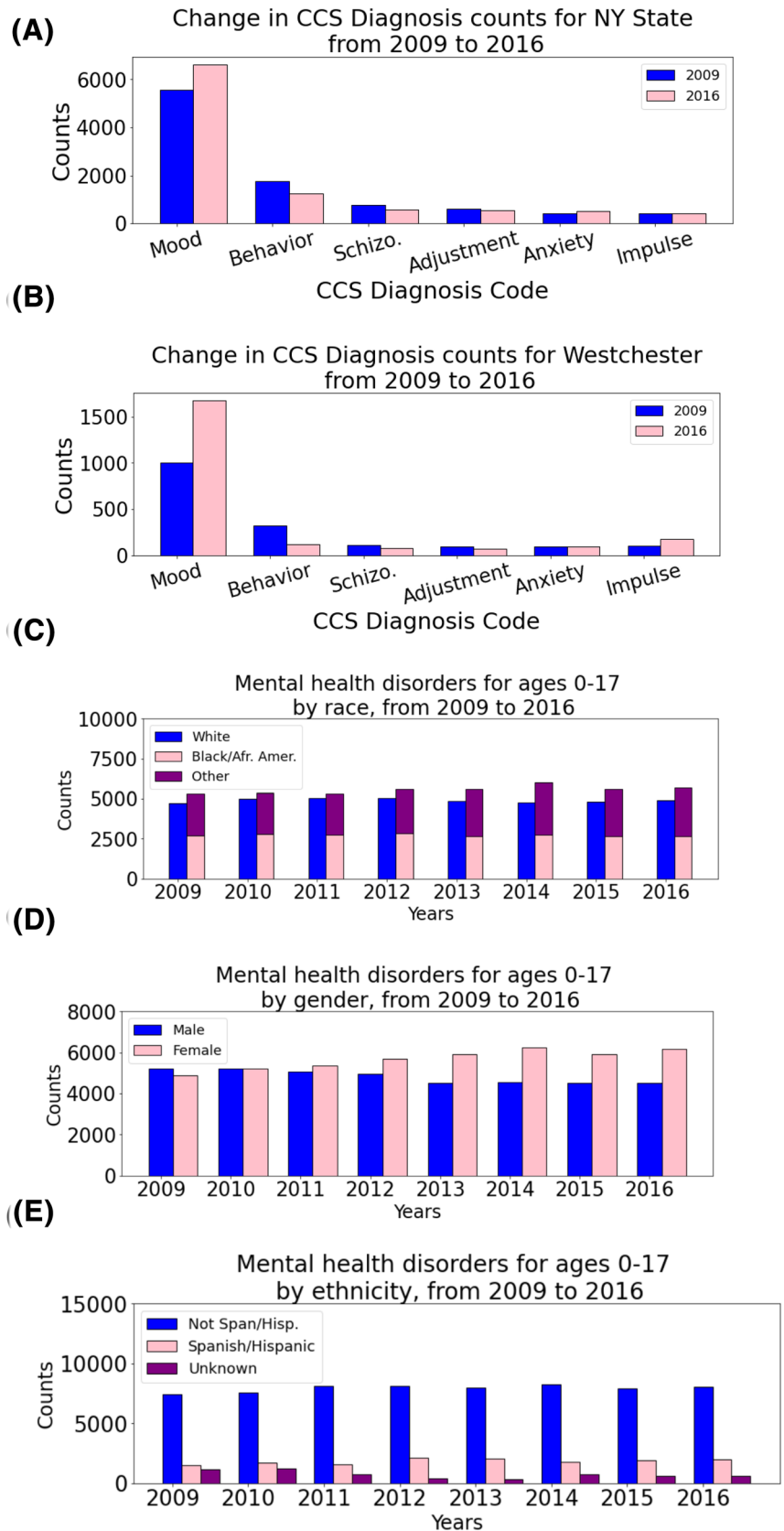
big data science to overcome minority health and health disparities (Zhang et al. 2017). The report recommended that big data be used to combine clinical data with environmental and policy data to conduct spatio-temporal surveillance and monitoring. This allows agencies to track trends in disparities, and allocate resources to ensure that all patients are getting the medical services they need. The report specifically recommended improving data access and sharing through open data policies.

The investigation of MH issues in children continues to be an active area for both researchers and policy makers. Childhood adversity is pervasive in the U.S., and is a contributor to mental health problems (Bartlett 2020). Policy-makers are taking note, and are recommending that children be screened routinely for adversity during pediatric care, and in community mental health settings. An example is the \$160 million initiative in California started in January 2020 to screen 7 million children on the Medi-Cal insurance plan for adverse childhood experiences (ACEs) (Underwood 2020). This plan is California's Medicaid program covering low-income residents, which amounts to one-third of the state's population. According to the Kaiser Family Foundation (Foundation 2013), in 2013, 56% of the Medicaid recipients in California were Hispanic, 9% were Blacks, and 19% were White. The plan to screen for ACEs is controversial. Though ACEs affect health in adulthood, it is not clear how to develop adequate programs that show clear benefit (Underwood 2020). Most prior studies have involved hundreds to tens of thousands of children (Underwood 2020). However, the current plan is aimed at millions of children. Tracking the progress of such a large population necessitates the use of big data technologies. The reporting of the results through open data portals similar to New York SPARCS will facilitate broad exploration by the scientific community to assess the efficacy and impact of such initiatives in the future.

There is a growing interest in applying big data techniques for public health. Dolley (Dolley 2018) provides a comprehensive review of this area and discusses public health crises that have been analyzed through big data, including the opioid epidemic, diabetes, and HIV. Dolley (2018) specifically highlights the importance of creating and mobilizing open data, open science, and open collaboration platforms. One such example is the Observational Health Data Sciences and Informatics collaborative (ohdsi.org), which focuses on large-scale population health outcomes. This collaborative enables practitioners to share best practices, data and software.

Martin et al. (2015) observed that the majority of health researchers did not use open data portals. Our paper demonstrates that the full potential of open data could be realized through continuing research and dissemination. There are several significant advantages (Martin et al. 2015) to

Fig. 4 Comparison of trends for the age group 0–17. **a** Changes in incidences of the top six CCS diagnosis codes for MH disorders. Years 2009 and 2016 are compared. **b** Changes in incidences for the same CCS diagnosis codes for Westchester county. **c** Changes in MH disorders by race. **d** Changes in MH disorders by gender. **e** Changes in MH disorders by ethnicity



creating and promoting the use of open health data, including conducting exploratory analyses to formulate research questions, conducting pilot research, using data for teaching undergraduate and doctoral students, and merging data sources. Many studies have demonstrated the value of merging health data sources using big data techniques, and deriving new insights (Rao et al. 2018; Rao and Clarke 2019).

Limitations

All public data sets might have potential limitations that determine the kind of analysis that can be accomplished. For example, the NY SPARCS data consists of de-identified patient data, where every incident is a separate record. This makes it hard to distinguish individual cases versus the numbers of incidents and that limits the ability to analyze the data set. In addition, the categorization of age groups within the dataset is broad and decided by NY. The age group 0–17 is quite broad, and different MH conditions may be prevalent at specific ages, but it is not possible to break the data set further down to age groups to look at those differences. Hence, it is difficult to develop a fine-grained analysis for this entire age group. Nevertheless, this implies that in future, agencies such as NY SPARCS could consider making more detailed information about age groups available, say in 5-year bins so that more nuanced research can be facilitated.

SPARCS data is for patient discharge only, meaning that it does not include outpatient data and thus limits our ability to compare our results with other findings. Westchester is one of the richest counties in the state and its residents largely do have access to healthcare. The general population in other counties in the state may not have the access and/or the financial ability to get inpatient services for MH conditions such as mood disorders.

Conclusion

Our study was able to demonstrate the usefulness of open big data sets and big-data analytics to identify significant trends in mental health to amend service needs and fiscal planning. Our results shed light on the spectrum of issues and the possibility of generating evidence-based recommendations/plans. We highlight the fact that there is much insight that remains to be gleaned from existing public data sources in the health care domain. Indeed, this is being recognized by funding agencies such as NIH (Department of Health and Human Services, Secondary Analyses of Existing Data Sets and Stored Biospecimens to Address Clinical Aging Research Questions (R01); Department of Health and Human Services, Secondary Dataset Analyses in Heart, Lung, and Blood Diseases and Sleep Disorders (R21)), the Robert Wood Johnson Foundation, and the Russell Sage

Foundation (RUSSELL SAGE FOUNDATION, Funding Opportunity: The Social, Economic, and Political Effects of the Affordable Care Act) that are offering grants to support the analysis of existing healthcare datasets and novel methods for their analysis and interpretation. Our approach using open health data is scalable to multiple states that release such information, and provides an independent perspective on health care issues, including costs.

We have specifically identified the vulnerability of age group 0–17 to mental health disorders. To address the growing incidence of these disorders, we recommend targeted interventions at the school level in order. The dissemination of our results will prove valuable to policy and decision makers in capacity planning, and the creation of relevant interventions.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by ARR and SR. RC contributed to the formal analysis and investigation. All authors were involved in writing, review and editing.

Compliance with Ethical Standards

Conflict of interest The authors have no known conflicts of interest.

Ethics Approval No specific ethics approval was required for our study. We used de-identified patient data made available by the New York State SPARCS program. This dataset is made available freely to the public without any restrictions. We use the data on an “as-is” basis and have not made any attempts to identify the patients. We have followed the New York State SPARCS policies for use of their data. In this age of greater scrutiny of data sources, we hope the editors and reviewers appreciate the value brought by such freely available public data sources, and our efforts to analyze such data.

References

- Alegria, M., Canino, G., Ríos, R., Vera, M., Calderon, J., Rusch, D., et al. (2002). Mental health care for Latinos: Inequalities in use of specialty mental health services among Latinos, African Americans, and non-Latino Whites. *Psychiatric Services*, 53(12), 1547–1555.
- Andersen, S. L., & Teicher, M. H. (2008). Stress, sensitive periods and maturational events in adolescent depression. *Trends in Neurosciences*, 31(4), 183–191.
- Atkins, M. S., Cappella, E., Shernoff, E. S., Mehta, T. G., & Gustafson, E. L. (2017). Schooling and children’s mental health: Realigning resources to reduce disparities and advance public health. *Annual Review of Clinical Psychology*, 13, 123–147.
- Bartlett, J. D. (2020). Screening for childhood adversity: contemporary challenges and recommendations. *Adversity and Resilience Science*, 2020, 1–15.
- Data and Statistics on Children’s Mental Health. Retrieved February 15, 2021 from <https://www.cdc.gov/childrensmentalhealth/data.html>.

- Department of Health and Human Services, Secondary Analyses of Existing Data Sets and Stored Biospecimens to Address Clinical Aging Research Questions (R01).
- Department of Health and Human Services, Secondary Dataset Analyses in Heart, Lung, and Blood Diseases and Sleep Disorders (R21).
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64–73.
- Dolley, S. (2018). Big data's role in precision public health. *Frontiers in Public Health*, 6, 68.
- Foundation, K. F. (2013). Medicaid enrollment by race/ethnicity. Retrieved from: <https://www.kff.org/medicaid/state-indicator/medicaid-enrollment-by-raceethnicity/>.
- Glied, S., Hoven, C. W., Moore, R. E., Garrett, A. B., & Regier, D. A. (1997). Children's access to mental health care: Does insurance matter? *Health Affairs*, 16(1), 167–174.
- Grosserode, S. (2020). Mahopac High School offers state's first substance abuse course. *lohud.com*, 2020 March 5.
- Hafferty, J. D., Smith, D. J., & McIntosh, A. M. (2017). Invited commentary on Stewart and Davis "'Big data' in mental health research—current status and emerging possibilities." *Social Psychiatry and Psychiatric Epidemiology*, 52(2), 127–129.
- Hoff, D. (2002). College Board to extend reach to middle school. *Education Week*, 21(27), 10. <http://www.medicare.gov/hospitalcompare/data/total-performance-scores.html>.
- <https://data.medicare.gov/Physician-Compare/National-Downloadable-File/s63f-csi6>.
- <https://sustainabledevelopment.un.org/sdg16>.
- <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>. (2019).
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3), 90–95.
- Jetzek, T. (2015). The sustainable value of open government data: Uncovering the generative mechanisms of open data through a mixed methods approach: Copenhagen Business School Copenhagen Business School, Institut for IT-Ledelse/Department of IT Management.
- Johnson, S. R. (2018 March). Growing demand for behavioral health services overwhelms colleges. *Modern Healthcare*, 24, 2018.
- Kazdin, A. E. (2019). Annual Research Review: Expanding mental health services through novel models of intervention delivery. *Journal of Child Psychology and Psychiatry*, 60(4), 455–472.
- Kessler, R. C., Avenevoli, S., & Merikangas, K. R. (2001). Mood disorders in children and adolescents: An epidemiologic perspective. *Biological Psychiatry*, 49(12), 1002–1014.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593–602.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequence*. New York: Sage.
- LeGates, T. A., Kvarita, M. D., & Thompson, S. M. (2019). Sex differences in antidepressant efficacy. *Neuropsychopharmacology*, 44(1), 140–154.
- Lin, S., Lu, Y., Justino, J., Dong, G., & Lauper, U. (2016). What happened to our environment and mental health as a result of Hurricane Sandy? *Disaster Medicine and Public Health Preparedness*, 10(3), 314–319.
- Mackenzie, C., Gekoski, W., & Knox, V. (2006). Age, gender, and the underutilization of mental health services: The influence of help-seeking attitudes. *Aging and Mental Health*, 10(6), 574–582.
- Mackenzie, S., Wiegel, J. R., Mundt, M., Brown, D., Saewyc, E., Heiligenstein, E., et al. (2011). Depression and suicide ideation among students accessing campus health care. *American Journal of Orthopsychiatry*, 81(1), 101.
- Mahmoud, H., Naal, H., & Cerda, S. (2020). Planning and implementing telepsychiatry in a community mental health setting: a case study report. *Community Mental Health Journal*, 2020, 1–7.
- Martin, E. G., & Begany, G. M. (2017). Opening government health data to the public: Benefits, challenges, and lessons learned from early innovators. *Journal of the American Medical Informatics Association*, 24(2), 345–351.
- Martin, E. G., Helbig, N., & Birkhead, G. S. (2015). Opening health data: what do researchers want? Early experiences with New York's open health data platform. *Journal of Public Health Management and Practice*, 21(5), E1–E7.
- Maura, J., & de Mamani, A. W. (2017). Mental health disparities, treatment engagement, and attrition among racial/ethnic minorities with severe mental illness: a review. *Journal of Clinical Psychology in Medical Settings*, 24(3–4), 187–210.
- McGuire, T. G., & Miranda, J. (2008). New evidence regarding racial and ethnic disparities in mental health: Policy implications. *Health Affairs*, 27(2), 393–403.
- McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. New York: O'Reilly Media Inc.
- Mental health symptoms in school-aged children in four communities 2020. Retrieved from: <https://www.cdc.gov/childrensmentalhealth/features/school-aged-mental-health-in-communities.html>.
- Merikangas, K. R., He, J.-P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., et al. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Survey Replication-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(10), 980–989.
- Merikangas, K. R., He, J.-P., Burstein, M., Swendsen, J., Avenevoli, S., Case, B., et al. (2011). Service utilization for lifetime mental disorders in US adolescents: results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, 50(1), 32–45.
- Merikangas, K. R., Nakamura, E. F., & Kessler, R. C. (2009). Epidemiology of mental disorders in children and adolescents. *Dialogues in Clinical Neuroscience*, 11(1), 7.
- Mikk, K. A., Sleeper, H. A., & Topol, E. J. (2017). The pathway to patient data ownership and better health. *JAMA*, 318(15), 1433–1434.
- Mills, K. L., Lalonde, F., Clasen, L. S., Giedd, J. N., & Blakemore, S.-J. (2012). Developmental changes in the structure of the social brain in late childhood and adolescence. *Social Cognitive and Affective Neuroscience*, 9(1), 123–131.
- Minelli, M., Chambers, M., & Dhiraj, A. (2012). *Big data, big analytics: Emerging business intelligence and analytic trends for today's businesses*. New York: Wiley.
- Mowbray, C. T., Mandiberg, J. M., Stein, C. H., Kopels, S., Curlin, C., Megivern, D., et al. (2006). Campus mental health services: Recommendations for change. *American Journal of Orthopsychiatry*, 76(2), 226–237.
- New York State Department of Health, Statewide Planning and Research Cooperative System (SPARCS). <https://www.health.ny.gov/statistics/sparcs/>.
- Ogders, C. L., & Jensen, M. R. (2020). Annual Research Review: Adolescent mental health in the digital age: Facts, fears, and future directions. *Journal of Child Psychology and Psychiatry*, 61(3), 336–348.
- Olfson, M., Blanco, C., Wang, S., Laje, G., & Correll, C. U. (2014). National trends in the mental health care of children, adolescents, and adults by office-based physicians. *JAMA Psychiatry*, 71(1), 81–90.
- Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: A global public-health challenge. *The Lancet*, 369(9569), 1302–1313.

- Pennap, D., Zito, J. M., Santosh, P. J., Tom, S. E., Onukwugha, E., & Magder, L. S. (2018). Patterns of early mental health diagnosis and medication treatment in a Medicaid-insured birth cohort. *JAMA Pediatrics*, *172*(6), 576–584.
- Power, T. J., Mautone, J. A., Blum, N. J., Fiks, A. G., & Guevara, J. P. (2019). Integrated behavioral health: Coordinating psychosocial and pharmacological interventions across family, school, and health systems. *School Psychopharmacology*: Springer, pp. 195–212.
- RUSSELL SAGE FOUNDATION, Funding Opportunity: The Social, Economic, and Political Effects of the Affordable Care Act.
- Ramachandran, S. (2020). A hidden cost of Covid: Shrinking Mental-Health Services. *Wall Street Journal*. 2020 October 9.
- Rao, A. R., & Clarke, D. (2019). Exploring relationships between medical college rankings and performance with big data. *Big Data Analytics*, *4*(1), 3.
- Rao, A. R., Clarke, D., & Vargas, M. (2018). Building an open health data analytics platform: A case study examining relationships and trends in seniority and performance in healthcare providers. *Journal of Healthcare Informatics Research*, *2018*, 1–27.
- Rao, A. R., Chhabra, A., Das, R., & Ruhil, V. (Eds.) (2015). A framework for analyzing publicly available healthcare data. In 17th International Conference on E-health Networking, Application & Services (IEEE HealthCom). IEEE.
- Rao, A. R., & Clarke, D., (Eds.). (2018). A comparison of models to predict medical procedure costs from open public healthcare data. In *IEEE IJCNN International Joint Conference on Neural Networks*. Rio de Janeiro, Brazil: IEEE.
- Rao, A. R., Garai, S., Dey, S., & Peng, H. (2020). Building predictive models of healthcare costs with open healthcare data. In *IEEE conference on healthcare informatics* (submitted Feb 2020).
- Rao, A. R., & Clarke, D. (2017). Facilitating the exploration of open health-care data through BOAT: A big data open source analytics tool. *Emerging Challenges in Business, Optimization, Technology and Industry: Third International Conference on Business Management and Technology*. Vancouver: Springer Verlag.
- Rich, E., Andersen, D. F., Augustine, W., Cronemberger, F., Hull, K., & Luna-Reyes, L., et al., (Eds.). (2015). An experimental platform for interpreting open-source health data though integration with dynamic disease models and geoplots. In *2015 17th International conference on E-health networking, application & services (HealthCom)*. IEEE.
- Russ, T. C., Woelbert, E., Davis, K. A., Hafferty, J. D., Ibrahim, Z., Inkster, B., et al. (2019). How data science can advance mental health research. *Nature Human Behaviour*, *3*(1), 24–32.
- Sanchez, A. L., Cornacchio, D., Poznanski, B., Golik, A. M., Chou, T., & Comer, J. S. (2018). The effectiveness of school-based mental health services for elementary-aged children: A meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(3), 153–165.
- Shukor, A. R., Joe, R., Sincaian, G., Klazinga, N., & Kringos, D. S. (2019). A multi-sourced data analytics approach to measuring and assessing biopsychosocial complexity: The Vancouver Community analytics tool complexity module (VCAT-CM). *Community Mental Health Journal*, *55*(8), 1326–1343.
- Simon, C. (2017). More and more students need mental health services. But colleges struggle to keep up. *USA Today* 2017 May 4.
- Stewart, R., & Davis, K. (2016). ‘Big data’ in mental health research: Current status and emerging possibilities. *Social Psychiatry and Psychiatric Epidemiology*, *51*(8), 1055–1072.
- Underwood, E. (2020). California has begun screening for early childhood trauma, but critics urge caution. *Science*.
- Whitney, D. G., & Peterson, M. D. (2019). US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. *JAMA Pediatrics*, *173*(4), 389–391.
- Zhang, X., Pérez-Stable, E. J., Bourne, P. E., Peprah, E., Duru, O. K., Breen, N., et al. (2017). Big data science: Opportunities and challenges to address minority health and health disparities in the 21st century. *Ethnicity & Disease*, *27*(2), 95.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.