# Computational Linguistics for Enhancing Scientific Reproducibility and Reducing Healthcare Inequities

**Julia Parish-Morris, PhD**[1,2,3]

[1] Center for Autism Research, Children's Hospital of Philadelphia (CHOP)
[2] Departments of Biomedical and Health Informatics and Child and Adolescent Psychiatry, CHOP
[3] Department of Psychiatry, Perelman School of Medicine of the University of Pennsylvania

## Abstract

Computational linguistics holds promise for improving scientific integrity in clinical psychology, and for reducing longstanding inequities in healthcare access and quality. This paper describes how computational linguistics approaches could address the "reproducibility crisis" facing social science, particularly with regards to reliable diagnosis of neurodevelopmental and psychiatric conditions including autism spectrum disorder (ASD). It is argued that these improvements in scientific integrity are poised to naturally reduce persistent healthcare inequities in neglected subpopulations, such as verbally fluent girls and women with ASD, but that concerted attention to this issue is necessary to avoid reproducing biases built into training data. Finally, it is suggested that computational linguistics is just *one* component of an emergent digital phenotyping toolkit that could ultimately be used for clinical decision support, to improve clinical care via precision medicine (i.e., personalized intervention planning), granular treatment response monitoring (including remotely), and for gene-brain-behavior studies aiming to pinpoint the underlying biological etiology of otherwise behaviorally-defined conditions like ASD.

## 1 Introduction

Humans are complex social beings, and the intricacies of language manifest this richness. Although language emanates from the brain, it has not yet been fully leveraged in the service of understanding brain-based psychiatric variation (e.g., disorders such as schizophrenia, bipolar disorder, and autism). Efforts to incorporate computational linguistics approaches into the mental health system have primarily focused on mining electronic medical records (Doshi-Velez, Ge, & Kohane, 2014; Lingren et al., 2016). While valuable, these efforts are often limited to analyzing text generated by doctors or other programs (Tran et al., 2014), rather than directly assessing specific psychiatric issues in patients themselves. This paper discusses ways in which analyzing spoken language in psychiatric contexts can move the needle on two persistent challenges: reproducibility in human social sciences (Section 2), and inequities in mental health care (Section 3).

## 2 Reproducibility

In 2015, an article appeared in the journal *Science,* which suggested that the majority of published experiments in psychology are not reproducible (Open Science Collaboration, 2015). Out of 100 experiments, only 39 replicated in a new sample, despite careful methods and communication with original authors (see (Gilbert, King, Pettigrew, & Wilson, 2016) for a comment, and (Anderson et al., 2016) for a response). In this and subsequent analyses, lack of scientific reproducibility has been argued to be due to a number of factors, including *p*-hacking, selective reporting of results, over-emphasis on innovation and novelty over stability, poor experimental training for scientists, lack of power (small sample sizes), and inadequate measurement (Button et al., 2013; National Science Foundation, 2015). The first part of this short paper focuses on reproducibility challenges that result from traditional methods of psychiatric diagnosis and symptom measurement, and proposes that computational linguistics is a promising tool for improving reliability and enhancing fine-grained characterization efforts.

### 2.1 Psychiatric Diagnosis

Reproducible methods in the field of clinical psychology and psychiatry require, first and foremost, accurate characterization of the



*Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 94–102
Minneapolis, Minnesota, June 6, 2019. 

condition under study. However, potential error is inherent in how psychiatric diagnoses are traditionally made. Although significant resources have been devoted to identifying biological causes of psychiatric conditions like schizophrenia, and some non-diagnostic brain-based (Ecker, Bookheimer, & Murphy, 2015; McDonald et al., 2005; Zalesky, Fornito, & Bullmore, 2010) and genetic (Geschwind et al., 2001) differences have been identified, the majority of mental health disorders are still diagnosed using behavior alone (American Psychiatric Association, 2013).

Whether or not a person has a psychiatric condition may seem obvious, but a number of factors complicate reliable diagnosis. First, in the absence of biological ground truth (e.g., a blood test or a brain scan), clinicians must grapple with wide behavioral heterogeneity that can cause two people with the same disorder to appear very different from one another. For example, ASD symptoms often manifest differently from one person to the next. Within a single subject, behavioral profiles may vary from week-to-week or even day-to-day. An individual may appear very typical in one context (e.g., familiar, low-stress environments), but their autistic behaviors could become very obvious in others (e.g., novel, high-stress environments). The consequences of this variability are measurable, such that a large, multi-site study of ASD found relatively low diagnostic agreement between expert clinicians at different sites (Catherine Lord, 2012).

Low diagnostic agreement has significant implications for the reliability of human scientific research. For example, in order to test whether ASD *causes* differences in executive function, a study should control every other variable except diagnosis. That is, two groups are assembled: individuals with ASD and neurotypical controls. Groups are matched on important variables like sex ratio, race/ethnicity, chronological age, full-scale IQ, verbal IQ, nonverbal IQ, maternal education (a strong predictor of offspring language ability, which has associations with executive function), etc. An executive function task is administered, and if the groups differ, it may be inferred that the difference is due to ASD. However, if the diagnostic category of ASD is in any way unreliable, another researcher following the exact same procedure with a new sample may not produce the same result due to differences in the ASD group.

Poor diagnostic reliability is a long-standing problem in psychiatric research. Some have suggested that larger sample sizes could reduce the impact of the problem, but the low incidence of ASD [current estimates suggest that approximately 1.5% of the population has ASD (Christensen, 2016)], in combination with long and expensive diagnostic processes, make it challenging to assemble high-powered samples. Recent research suggests that computational linguistics could provide objective diagnostic decision support (through direct measurement) in ways that might speed the process and make it more reliable.

### 2.2 Objective Measurement for Clinical Characterization

The process of making a mental health diagnosis is often mediated by language; primary diagnostic tools for many psychiatric conditions include structured or semi-structured interviews, wherein a clinical psychologist or psychiatrist asks patients about their thoughts, feelings, and experiences (Kaufman et al., 1997; Lord et al., 1989), comparing patterns of responding to diagnostic symptom checklists or scoring algorithms. After incorporating other relevant information (e.g., family/medical history, current stressors), clinicians use their best judgment to determine diagnostic category. When individuals are nonverbal or minimally verbal, these interviews may be conducted with family members who know the person well (Rutter, LeCouteur, & Lord, 2008). Characteristics of patient speech and language are often noted in the course of clinical evaluations, but they are often only minimally quantified; that is, presence or absence of atypical speech-language characteristics are noted, but highly detailed information is often not systematically gathered. Thus, one valuable application for computational linguistics within clinical psychology and psychiatry is to enhance existing phenotypic characterization methods by adding fine-grained measures of patient speech and language produced during diagnostic evaluations.

In recent years, linguists and computer scientists have begun to analyze clinical evaluations using computational approaches (Black et al., 2011; Kiss, Santen, Prud'Hommeaux, & Black, 2012; Kumar et al., 2016). For example, it has been shown that not only do children with ASD speak differently than neurotypical peers during diagnostic assessments (Parish-Morris et al., 2016), but characteristics of the interviewer's language predict children's symptom severity as well (Bone, Bishop, Gupta, Lee, & Narayanan, 2016).

Beyond applying computational linguistics approaches to audio recordings of clinical assessments (which remain expensive and complicated to collect, and are not very ecologically valid), researchers have begun to explore whether computational linguistics could be used to characterize psychiatric disorders using everyday language samples (Parish-Morris et al., 2018). Naturalistic samples are challenging to study for a variety of reasons, including the myriad uncontrolled (and perhaps uncontrollable) variables inherent in dynamic human interaction. Consider two people meeting each other for the first time. Each person's behavior is influenced not only by their genetically-linked dispositions, but also a lifetime of experiences, and immediate factors (e.g., did they eat breakfast that day?). When the two people begin to converse, their behavior becomes bi-directionally influential (e.g., each person dynamically reacts to the other in real time, which affects the next moment, and so on). When one or more participants brings extreme psychiatric variation (e.g., active psychosis) to the conversation – the interaction itself changes, and the course of the interaction will likely also fall outside the norm. Despite the challenges associated with measuring two people in an uncontrolled context instead of one person in a controlled context (as in a clinical evaluation), basing future research on naturalistic samples is key; the generalizability gap between research and the real world will shrink as we increase the ecological validity of our research samples.

Importantly, tools from computational linguistics might also be used to directly influence diagnostic decision making in ways that make it more reproducible. Rather than replacing clinicians, the current promise of computational linguistics is to develop objective and granular metrics for use as clinical decision *support* tools. For example, objective linguistic analysis could be used to flag subtle atypical patterns that are not perceptible to the naked ear [e.g., slightly elevated disfluency rates, or reduced lexical diversity; (Parish-Morris et al., 2017, 2018)]. Clinicians provided with this type of evidence could use it, in combination with other information like family history, as part of the diagnostic decision process.

In summary, using computational linguistics to more accurately specify behavioral phenotypes in psychiatry will not only improve our ability to quickly and objectively diagnose patients, but will also improve our efforts to understand the biological underpinnings of these disorders, by helping us identify diagnostic groups that can be carved along objective joints. Improved characterization of psychiatric conditions will allow researchers to assemble experimental groups that are more homogeneous than broad "ASD" vs. "neurotypical" designations. Reducing sample heterogeneity (noise) through improved characterization could increase the likelihood of identifying true signal in scientific studies, thus improving reproducibility. Finally, objective computational linguistics tools that do not require human intervention could be used by clinicians for clinical decision support, ultimately improving diagnostic reliability.

## 3 Healthcare Inequities

Computational linguistics has the potential improve *human behavioral science* by addressing problems with reproducibility, but it can also improve the state of *mental health care* by reducing inequities related to access and provider biases.

Persistent race-, sex-, and income-related inequities in health outcomes have been extensively documented across a wide variety of domains. These have been attributed, in part, to reduced access in some cases (Ahmed, Lemkau, Nealeigh, & Mann, 2001) and deep-seated provider biases in others (Burgess, van Ryn, Dovidio, & Saha, 2007; Chapman, Kaatz, & Carnes, 2013). This is especially problematic in psychiatry and clinical psychology, given recent estimates suggesting that nearly 1 in 5 people lives with a mental health condition (Hedden et al., 2015). Below, it is argued that some inequities could be addressed using tools developed jointly by computational linguists and clinicians.

### 3.1 Sources of Inequity: Access

Inter-related barriers to healthcare access include geographic distance, mental health provider shortages, and socio-economic disadvantages (expensive care). High-quality mental health care availability varies widely by region in the United States. Geographically remote individuals – those living far from a population center – currently have limited access to psychiatric screening and services (New American Economy, 2017). Even in population centers, a significant shortage of mental health providers leads to long wait lists for care (National Council for Behavioral Health, 2017). Given this shortage and lower reimbursement rates for mental vs. physical care (Melek, Perlman, & Davenport, 2017), many mental health providers choose not to accept insurance. Thus, if a patient does not have the economic resources to pay privately, they may not be able to receive care in their area, or may need to wait months to begin the intake and assessment process, much less engage in treatment.

### 3.2 Improving Access

Computational linguistics approaches, particularly when integrated into web- and phone-based telemedicine, could address some of these barriers to access. For example, long wait lists for screening or assessment of ASD could be shortened by the introduction of home- or school-based audio/video algorithms that measure how severely a person is impacted (and thus, help short-handed clinicians triage potential patients). Although this is not a complete fix (it addresses only one part of a larger problem), it could help overburdened clinicians organize their time and effort more efficiently to help those most immediately in need of assessment and services. Similarly, telemedicine approaches to depression monitoring could use vocal features (Yang, Fairbairn, & Cohn, 2013) alone or in combination with facial markers (Williamson, Quatieri, Helfer, Ciccarelli, & Mehta, 2014) to track change over time and signal the need for urgent intervention; moving people to the top of the waitlist. While expensive to initially build, these kinds of algorithms could reduce costs over time, as more people access health services through supportive automation.

### 3.3 Sources of Inequity: Biases

A growing body of research delineates deep and enduring biases within the medical and mental health treatment communities that negatively impact care for patients from racial/ethnic minority backgrounds, individuals born into poverty, immigrants/refugees/non-Western peoples, people with disabilities, gender minorities, and women (Conner et al., 2010; Fiscella, Franks, Doescher, & Saver, 2002; McCann & Sharek, 2016; Nadeem et al., 2007; Ojeda & Bergstresser, 2008; Puhl & Brownell, 2001; Sentell, Shumway, & Snowden, 2007; Winter et al., 2016). One potential source of bias is baked into mental health assessment tools: often, the tools used to assess, intervene, and monitor treatment response were not developed on the populations to whom they are currently being applied, and may therefore be inappropriate for entire segments of people. For example, when "depression inventories" were developed in the 1950s and 60s, who was included in the norming sample?

Depression was once thought to be much more common in women than men, and thus "depression" was conceptualized using women as prototypical exemplars. However, research suggests that the stereotypical conceptualization of depression as feelings of extreme sadness, while true for many women, does not hold true for many men. For men, depression may be more likely to manifest as irritability and aggression (Martin, Neighbors, & Griffith, 2013), leading many men to live their lives undiagnosed and untreated.

On the flip side of the coin, autism was originally described in predominantly male samples (Asperger, 1944; Kanner, 1943). Subsequently, most established assessment tools are male-referenced. Unfortunately, failure to understand the female autistic phenotype has led to systematic *under-diagnosis* of girls and women with ASD, who are either missed entirely or misdiagnosed with other disorders instead (Loomes, Hull, & Mandy, 2017). Incorrect or missed diagnoses are a serious concern in ASD, as early intervention has been shown to improve later outcomes (Howlin, Magiati, & Charman, 2009). Although some researchers have developed sex-referenced norms for social characterization (Constantino, 2012), the primary diagnostic tools for ASD still do not acknowledge the ways in which the disorder may manifest differently in girls vs. boys (American Psychiatric Association, 2013; Lord, Risi, & Bishop, 2012; Rutter et al., 2008).

These two examples spark further questions: how might depression and autism look different in cultural subgroups, such as recent immigrants from various parts of the world? Questions about whether historical norming and development samples are truly representative of the diverse set of people now seeking help for mental health issues in the U.S. have significant implications for accurately identifying the needs of a diverse patient population, and for providing effective services.

### 3.4 Reducing Biases

Language is one of the primary mediums through which behavioral diagnoses like autism, ADHD, depression, and anxiety are made, so it is important to recognize that language is also one of the mediums through which biases operate most efficiently. Accents, grammar, prosody, and word choice are all features that may be associated with unconscious biases (e.g., negative stereotypes could be activated by accents typical of rural populations in the U.S., slang used in inner cities, upspeak/vocal fry, accents of individuals learning English as a second language, etc.).

The challenge that computational linguistics can address, at least in part, is to provide objective metrics for quantifying language in a way that could reduce the effects of these linguistic biases. Much like orchestral auditions that, when conducted behind a curtain, result in significantly more women being hired than when the judge sees the person performing (Goldin & Rouse, 2000), biases that affect clinician judgements could be significantly reduced – or perhaps even eliminated – through the application of more objective measurement approaches developed by computational linguists.

The goal of objective measurement is to circumvent identified problems with bias that affect the likelihood of understudied subgroups getting referred, evaluated, diagnosed, and treated appropriately (e.g., men with depression, girls and women with ASD). However, the promise of comprehensive digital phenotyping (to include audio, video, web- and phone-based methods, and wearables) is not that measurement in the social sciences will suddenly be perfect. Rather, it is hoped that the quest to develop objective metrics for use in mental health research and practice will shed light on biases that operate in assessment and treatment contexts, and will allow those biases to be purposefully counteracted. This effort has significant implications for how we detect and treat mental health conditions in diverse patient populations.

## 4 Limitations

Like humans, computerized algorithms and "objective" computational approaches for addressing mental health conditions are not without their weaknesses. For example, well-intentioned efforts to use machine learning in support of policing has led to unjust racial profiling; this profiling was largely due to racially-biased training data (Chander, 2017). If training data is biased, the algorithm will be biased too. In the case of ASD, labeled language training data is subject to the problems associated with systematic, long-term under-diagnosis of girls. This begs the question: How can we use computational linguistics or digital phenotyping to support clinician decision-making when available training data is biased against females, or racial/ethnic minorities, or economically disadvantaged individuals? It is critical to grapple with these questions while simultaneously forging ahead to collect new (less biased) data, and develop tools that purposefully counteract these biases while eliminating barriers to access for underserved populations.

## 5 Conclusion

Objective phenotyping approaches based in computational linguistics will likely prove useful for scientific reasons like reproducibility and measurement granularity. Importantly, these methods also hold promise as tools to improve healthcare access and equity. Groups that have been historically understudied, subject to bias, and otherwise disenfranchised from getting early accurate mental health screening and personalized treatment, with negative impacts on long-term outcomes, stand to benefit from carefully implemented digital phenotyping efforts that identify/correct deeply problematic biases and barriers to equitable research and care.

## Acknowledgments

This work was supported by an Autism Science Foundation postdoctoral fellowship to J.P.M., and generous gifts from the Eagles Charitable Foundation and the Allerton Foundation to R.T. Schultz at the Center for Autism Research, CHOP.

## References

Ahmed, S. M., Lemkau, J. P., Nealeigh, N., & Mann, B. (2001). Barriers to healthcare access in a non-elderly urban poor American population. *Health & Social Care in the Community*, *9*(6), 445–453.

https://doi.org/10.1046/j.1365-2524.2001.00318.x

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5* (5 edition). Washington, D.C: American Psychiatric Publishing.

Anderson, C. J., Bahník, Š., Barnett-Cowan, M., Bosco, F. A., Chandler, J., Chartier, C. R., … Zuni, K. (2016). Response to Comment on “Estimating the reproducibility of psychological science.” *Science (New York, N.Y.)*, *351*(6277), 1037. https://doi.org/10.1126/science.aad9163

Asperger, H. (1944). Die "Autistischen Psychopathen” im Kindesalter. *Archiv für Psychiatrie und Nervenkrankheiten*, *117*(1), 76–136. https://doi.org/10.1007/BF01837709

Black, M., Bone, D., Williams, M. E., Gorrindo, P., Levitt, P., & Narayanan, S. S. (2011). The USC CARE Corpus: Child-Psychologist Interactions of Children with Autism Spectrum Disorders. *INTERSPEECH*, 1497–1500. Retrieved from http://www.researchgate.net/profile/Daniel_Bone/publication/221485841_The_USC_CARE_Corpus_Child-Psychologist_Interactions_of_Children_with_Autism_Spectrum_Disorders/links/09e4150c125896017d000000.pdf

Bone, D., Bishop, S., Gupta, R., Lee, S., & Narayanan, S. (2016). Acoustic-prosodic and turn-taking features in interactions with children with neurodevelopmental disorders. *Interspeech 2016*, 1185–1189.

Burgess, D., van Ryn, M., Dovidio, J., & Saha, S. (2007). Reducing Racial Bias Among Health Care Providers: Lessons from Social-Cognitive Psychology. *Journal of General Internal Medicine*, *22*(6), 882–887. https://doi.org/10.1007/s11606-007-0160-1

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365. https://doi.org/10.1038/nrn3475

Chander, A. (2017). The Racist Algorithm? *Michigan Law Review*, *115*, 24.

Chapman, E. N., Kaatz, A., & Carnes, M. (2013). Physicians and Implicit Bias: How Doctors May Unwittingly Perpetuate Health Care Disparities. *Journal of General Internal Medicine*, *28*(11), 1504–1510. https://doi.org/10.1007/s11606-013-2441-1

Christensen, D. L. (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR. Surveillance Summaries*, *65*. Retrieved from http://www.cdc.gov/mmwr/volumes/65/ss/ss6503a1.htm

Conner, K. O., Copeland, V. C., Grote, N. K., Koeske, G., Rosen, D., Reynolds, C. F., & Brown, C. (2010). Mental Health Treatment Seeking Among Older Adults with Depression: The Impact of Stigma and Race. *The American Journal of Geriatric Psychiatry : Official Journal of the American Association for Geriatric Psychiatry*, *18*(6), 531–543. https://doi.org/10.1097/JGP.0b013e3181cc0366

Constantino, J. N. (2012). *SRS-2 (Social Responsiveness Scale, Second Edition)*. Retrieved from http://www4.parinc.com/Products/Product.aspx?ProductID=SRS-2

Doshi-Velez, F., Ge, Y., & Kohane, I. (2014). Comorbidity Clusters in Autism Spectrum Disorders: An Electronic Health Record Time-Series Analysis. *Pediatrics*, *133*(1), e54–e63. https://doi.org/10.1542/peds.2013-0819

Ecker, C., Bookheimer, S. Y., & Murphy, D. G. M. (2015). Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. *The Lancet Neurology*, *14*(11), 1121–1134. https://doi.org/10.1016/S1474-4422(15)00050-2

Fiscella, K., Franks, P., Doescher, M. P., & Saver, B. G. (2002). Disparities in Health Care by Race, Ethnicity, and Language among the Insured: Findings from a National Sample. *Medical Care*, *40*(1), 52–59. Retrieved from JSTOR.

Geschwind, D. H., Sowinski, J., Lord, C., Iversen, P., Shestack, J., Jones, P., … Spence, S.

J. (2001). The Autism Genetic Resource Exchange: A Resource for the Study of Autism and Related Neuropsychiatric Conditions. *The American Journal of Human Genetics*, *69*(2), 463–466. https://doi.org/10.1086/321292

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science." *Science (New York, N.Y.)*, *351*(6277), 1037. https://doi.org/10.1126/science.aad7243

Goldin, C., & Rouse, C. (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review*, *90*(4), 715–741. https://doi.org/10.1257/aer.90.4.715

Hedden, S. L., Kennet, J., Lipari, R., Medley, G., Tice, P., Copello, E. A. P., & Kroutil, L. A. (2015). *Key Substance Use and Mental Health Indicators in the United States: Results from the 2015 National Survey on Drug Use and Health*. 74.

Howlin, P., Magiati, I., & Charman, T. (2009). Systematic Review of Early Intensive Behavioral Interventions for Children With Autism. *American Journal on Intellectual and Developmental Disabilities*, *114*(1), 23–41. https://doi.org/10.1352/2009.114:23-41

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*(3), 217–250.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., … Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980–988. https://doi.org/10.1097/00004583-199707000-00021

Kiss, G., Santen, J. P. van, Prud'Hommeaux, E., & Black, L. M. (2012). Quantitative analysis of pitch in speech of children with neurodevelopmental disorders. *Thirteenth Annual Conference of the International Speech Communication Association*. Retrieved from http://people.rit.edu/emilypx/papers/Interspeech12-GK.pdf

Kumar, M., Gupta, R., Bone, D., Malandrakis, N., Bishop, S., & Narayanan, S. S. (2016, September 8). *Objective Language Feature Analysis in Children with Neurodevelopmental Disorders During Autism Assessment*. 2721–2725. https://doi.org/10.21437/Interspeech.2016-563

Lingren, T., Chen, P., Bochenek, J., Doshi-Velez, F., Manning-Courtney, P., Bickel, J., … Savova, G. (2016). Electronic Health Record Based Algorithm to Identify Patients with Autism Spectrum Disorder. *PLOS ONE*, *11*(7), e0159621. https://doi.org/10.1371/journal.pone.0159621

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(6), 466–474. https://doi.org/10.1016/j.jaac.2017.03.013

Lord, C., Risi, S., & Bishop, S. L. (2012). *Autism diagnostic observation schedule, second edition (ADOS-2)*. Torrance, CA: Western Psychological Services.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, *19*(2), 185–212.

Lord, Catherine. (2012). A Multisite Study of the Clinical Diagnosis of Different Autism Spectrum Disorders. *Archives of General Psychiatry*, *69*(3), 306. https://doi.org/10.1001/archgenpsychiatry.2011.148

Martin, L. A., Neighbors, H. W., & Griffith, D. M. (2013). The Experience of Symptoms of Depression in Men vs Women: Analysis of the National Comorbidity Survey Replication. *JAMA Psychiatry*, *70*(10), 1100–1106. https://doi.org/10.1001/jamapsychiatry.2013.1985

McCann, E., & Sharek, D. (2016). Mental Health Needs of People Who Identify as Transgender: A Review of the Literature. *Archives of Psychiatric Nursing*, *30*(2), 280–285. https://doi.org/10.1016/j.apnu.2015.07.003

McDonald, C., Bullmore, E., Sham, P., Chitnis, X., Suckling, J., Maccabe, J., … Murray, R. M. (2005). Regional volume deviations of brain structure in schizophrenia and psychotic bipolar disorder. *British Journal of Psychiatry*, *186*(05), 369–377. https://doi.org/10.1192/bjp.186.5.369

Melek, S. P., Perlman, D., & Davenport, S. (2017). *Addiction and mental health vs. physical health: Analyzing disparities in network use and provider reimbursement rates* (pp. 1–56) [Milliman Research Report].

Nadeem, E., Lange, J. M., Edge, D., Fongwa, M., Belin, T., & Miranda, J. (2007). Does Stigma Keep Poor Young Immigrant and U.S.-Born Black and Latina Women From Seeking Mental Health Care? *Psychiatric Services*, *58*(12), 1547–1554. https://doi.org/10.1176/ps.2007.58.12.1547

National Council for Behavioral Health. (2017). *The Psychiatric Shortage: Causes and Solutions*. Retrieved from https://www.thenationalcouncil.org/wp-content/uploads/2017/03/Psychiatric-Shortage_National-Council-.pdf

National Science Foundation. (2015). *Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science*. Retrieved from https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

New American Economy. (2017). *The Silent Shortage: How Immigration Can Help Address the Large and Growing Psychiatrist Shortage in the United States* (pp. 1–31) [Health]. Retrieved from http://www.newamericaneconomy.org/wp-content/uploads/2017/10/NAE_PsychiatristShortage_V6-1.pdf

Ojeda, V. D., & Bergstresser, S. M. (2008). Gender, Race-Ethnicity, and Psychosocial Barriers to Mental Health Care: An Examination of Perceptions and Attitudes among Adults Reporting Unmet Need. *Journal of Health and Social Behavior*, *49*(3), 317–334. Retrieved from JSTOR.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Parish-Morris, J., Liberman, M., Ryant, N., Cieri, C., Bateman, L., Ferguson, E., & Schultz, R. T. (2016). Exploring Autism Spectrum Disorders Using HLT. *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, *3*, 74–84. Retrieved from http://languagelog.ldc.upenn.edu/myl/CLPsych2016_FINAL1.pdf

Parish-Morris, J., Liberman, M. Y., Cieri, C., Herrington, J. D., Yerys, B. E., Bateman, L., … Schultz, R. T. (2017). Linguistic camouflage in girls with autism spectrum disorder. *Molecular Autism*, *8*(1). https://doi.org/10.1186/s13229-017-0164-6

Parish-Morris, J., Sariyanidi, E., Zampella, C., Bartley, G. K., Ferguson, E., Pallathra, A. A., … Tunc, B. (2018). Oral-Motor and Lexical Diversity During Naturalistic Conversations in Adults with Autism Spectrum Disorder. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 147–157. https://doi.org/10.18653/v1/W18-0616

Puhl, R., & Brownell, K. D. (2001). Bias, Discrimination, and Obesity. *Obesity Research*, *9*(12), 788–805. https://doi.org/10.1038/oby.2001.108

Rutter, M., LeCouteur, A., & Lord, C. (2008). *Autism Diagnostic Interview - Revised (ADI-R)*. Los Angeles: Western Psychological Services.

Sentell, T., Shumway, M., & Snowden, L. (2007). Access to Mental Health Treatment by English Language Proficiency and Race/Ethnicity. *Journal of General Internal Medicine*, *22*(S2), 289–293. https://doi.org/10.1007/s11606-007-0345-7

Tran, T., Luo, W., Phung, D., Harvey, R., Berk, M., Kennedy, R. L., & Venkatesh, S. (2014). Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry*, *14*(1), 76. https://doi.org/10.1186/1471-244X-14-76

Williamson, J. R., Quatieri, T. F., Helfer, B. S., Ciccarelli, G., & Mehta, D. D. (2014). Vocal and Facial Biomarkers of

Depression based on Motor Incoordination and Timing. *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge - AVEC '14*, 65–72. https://doi.org/10.1145/2661806.2661809

Winter, S., Diamond, M., Green, J., Karasic, D., Reed, T., Whittle, S., & Wylie, K. (2016). Transgender people: health at the margins of society. *The Lancet*, *388*(10042), 390–400. https://doi.org/10.1016/S0140-6736(16)00683-8

Yang, Y., Fairbairn, C., & Cohn, J. F. (2013). Detecting Depression Severity from Vocal Prosody. *IEEE Transactions on Affective Computing*, *4*(2), 142–150. https://doi.org/10.1109/T-AFFC.2012.38

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. *NeuroImage*, *53*(4), 1197–1207. https://doi.org/10.1016/j.neuroimage.2010.06.041