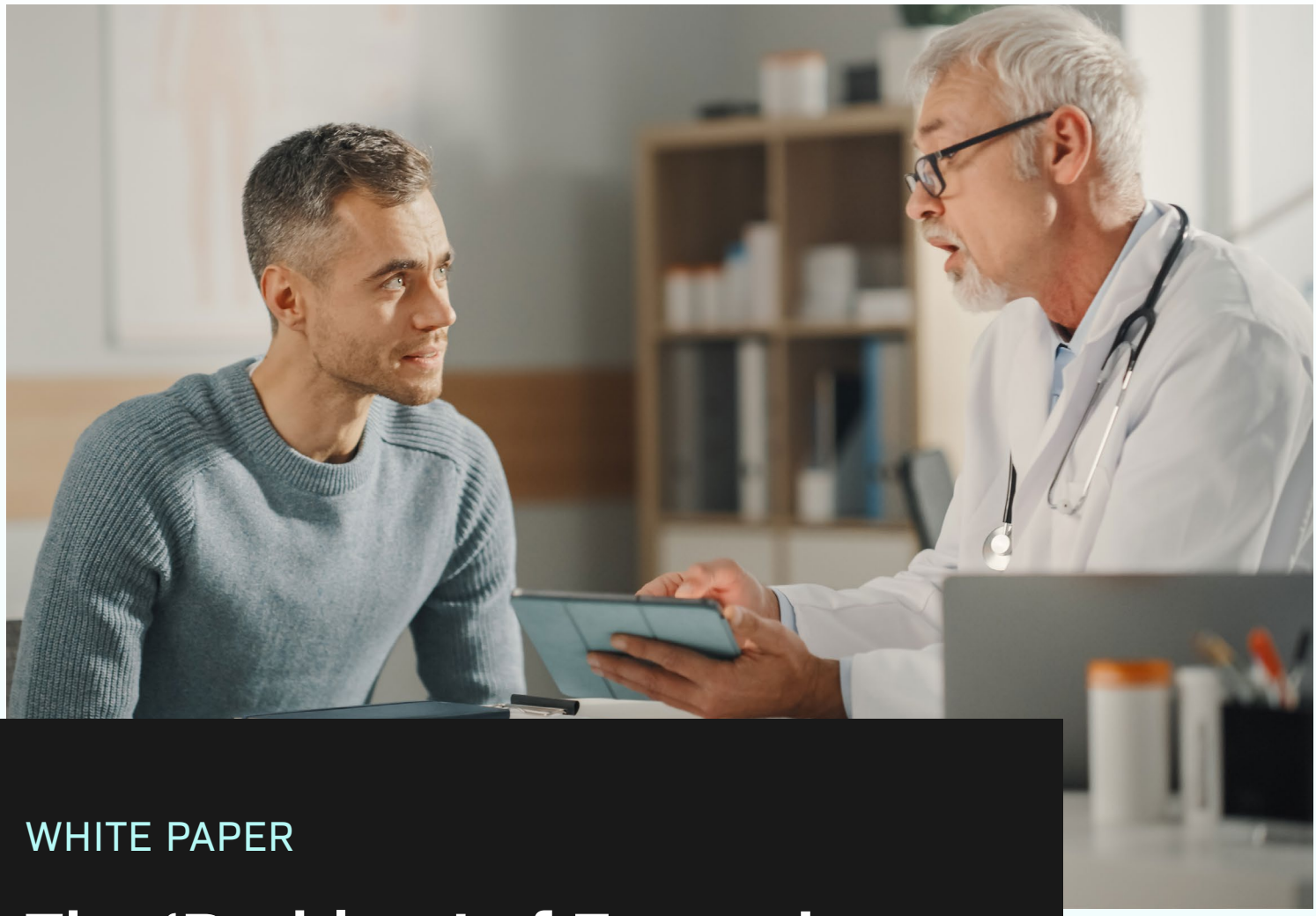
SIGNANT HEALTH
THE EVIDENCE GENERATION COMPANY®

PROOF AT THE SPEED OF LIFE™
SIGNANTHEALTH.COM



WHITE PAPER

The ‘Problem’ of Excessive Heterogeneity in Psychiatric Research and Solutions for Clinical Trials

Juliet Brown, PhD
Clinical Principal, Mood & Anxiety, Science & Medicine

Introduction

The idea that every patient is different is one of the rare points on which clinicians and researchers hold consensus. Heterogeneity in psychiatric populations, or variation in the characteristics of individuals who ostensibly have the same psychiatric diagnosis, is vast and driven by many factors that may or may not be neatly measurable. The majority of psychiatric trials fail beyond phase 2. While an appropriate level of patient variability is essential for the generalizability of study findings, excessive differences among participants can compromise the success of psychiatric trials.

This paper examines the key factors contributing to patient variability, discusses the potential adverse effects of unaddressed variability in clinical research, organizes these factors into categories, and links targeted mitigation strategies to each. The overall objective is to provide practical guidance for improving the likelihood of trial success.

> **If it were not for the great variability among individuals, medicine might as well be a science, not an art."**
>
> - William Osler

Key heterogeneity factors

The spectrum of sources of heterogeneity is wide and reflects biological diversity as well as psychosocial and environmental influences. Below is a subset of factors commonly considered in the space of CNS clinical trials for psychiatric indications.

Genetics

Inherited genes and DNA sequences play a role in determining susceptibility to and manifestations of various psychiatric disorders, particularly schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorder, attention deficit disorder, and obsessive-compulsive disorder. Genetic variation and pleiotrophy, where a single genetic variant can be associated with several psychiatric traits, can complicate clinical trial diagnostics.[1] While genetic sub-studies are occasionally included in CNS trial protocols, most genotyping is still used for exploratory pharmacogenomics, dosing/safety optimization, and mechanistic insight, and few trials restrict randomization based on genotype.[2]

Brain morphology

Individuals vary in terms of innate and acquired brain structure. For example, individuals with repeated exposure to antipsychotics have demonstrated cortical thinning, and individuals with prior abuse of substances may have altered grey matter volume. Brain morphology factors may in turn drive variability in psychiatric illness manifestations.[3,4]

Epigenetics

Environmental factors and lifestyle play a role in the ultimate phenotypic expression (or lack thereof) of individual genetic variants, and context varies vastly between individuals. For example, trauma's effect on neurons, and the inflammatory effects of health choices such as excessive alcohol use can impact the ultimate manifestation of genetically-driven illness. In so far as such patient factors are often unknown, they generally cannot be controlled.

Psychiatric primary diagnosis and comorbidities

The rate of psychiatric comorbidities is high, particularly for anxiety disorders and mood disorders. Due to overlapping symptom criteria and standard of care treatments for various disorders, accurate differential diagnosis and determination of a patient's primary psychiatric disorder become much more challenging.[5]

Illness severity

Some psychiatric disorders represent "severe mental illness" regardless of the current illness status for a particular patient (e.g., schizophrenia), but understanding of precise current severity is key in ensuring that a trial sample has sufficient baseline range while avoiding enrollment of patients with an actual severity that is too high or too low for study entry. Severity can change significantly from prescreening to screening to baseline visit. Sensitive evaluations are key.

Medical comorbidities

Background medical illnesses vary among individuals with psychiatric disorders both in terms of presence and severity. Co-occurring medical illness not only increase patient risk but can drive variation in psychiatric presentation and treatment responsiveness. Fortunately, with thorough review of medical history and screening assessments, many medical factors can be controlled.

Age of illness onset

Earlier illness onset portends greater psychiatric comorbidity and a higher likelihood of prior inadequate treatment response or treatment resistance. Younger age of onset predicts higher severity for some disorders.[6] Failing to consider age of onset is a risk to study success.

Treatment history

Medical and non-pharmacological treatments can affect the severity of primary and secondary psychiatric disorders, and exposure to certain classes of pharmacological treatment can alter brain chemistry and structure (e.g., receptor upregulation). Patients may have real or pseudo treatment resistance that complicates clinical trial data interpretation, but definitions of treatment resistance and sensitivity of assessment tools to detect treatment resistance vary.[7] The effects of pharmacological washout that is sometimes too rapid further muddies data interpretation.

Illness course

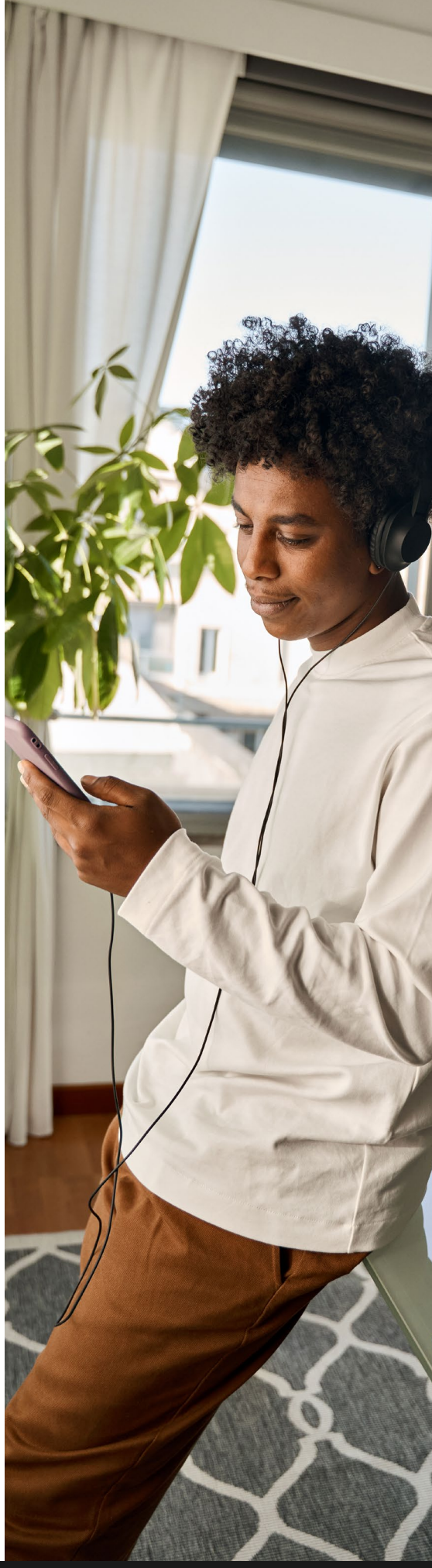
Patients with longer illness duration or a higher number of illness episodes (e.g., early onset schizophrenia or individuals with rapid cycling bipolar disorder) have been shown to have poorer treatment response, and the resulting increased data variability can adversely affect trial success.[8,9]

Psychosocial factors

Lifetime trauma experiences, substance use and abuse, extent of psychosocial support, and acute environmental stressors vary across patients and can affect illness presentation, leading to unpredictable in-study temporal data trends.

Expectations

Patients and their social networks will inevitably have differing ideas about how their primary illness will transpire over time (e.g., whether or not they believe they will recover from their current condition) and about potential benefits of trial treatments. While placebo response is to some degree a 'universal' problem, patients vary in terms of the degree to which they hold various cognitive biases about illness and interventions, and expectations vary over time for each individual patient.



What's the impact?

> **Not everything that can be counted counts, and not everything that counts can be counted."**
>
> - Albert Einstein

Patient characteristics and histories will inevitably vary, but a sufficient degree of variability is essential for signal detection and for ensuring that study samples reflect real-world populations, thereby supporting the generalizability of outcomes. Should patient diversity simply be accepted with the assumption that large sample sizes and regression to the mean will dilute any detrimental trial effects? No. The risks are too great as unmanaged heterogeneity can dilute true treatment effects. Potential consequences include:

Enrollment of inappropriate subjects

Randomization of subsyndromal patients (those lacking key pathology markers) has been consistently associated with exaggerated placebo response.[10] Those who have been mistakenly diagnosed with the primary illness under study, or individuals with psychiatric or medical comorbidities that impact baseline status or primary illness trajectory is associated with exaggerated placebo response, decreased data validity, inadequate investigational product benefit, and study failure.

Obscuring of true treatment effects

Key factors that are unaccounted for or undetected may weaken effect sizes. For example, mood disorder studies may define treatment-resistance too liberally or may fail to effectively account for anxious distress or mixed features, thereby inadvertently enrolling more difficult-to-treat patients. Anxiety disorder trials may inadvertently enroll subjects with mild severity, resulting in a restricted severity range within which to detect change.

Amplification of noise and placebo response

Enrollment of misdiagnosed patients or those with syndromal and subsyndromal comorbidities is associated with reduced Clinician Reported Outcome (ClinRO) data reliability due to introduction of noise, and higher symptom severity temporal variability is associated with inflated placebo response and reduced drug-placebo separation.[11]

Misleading or ungeneralizable results

Unmanaged psychiatric heterogeneity can lead to misleading effect sizes and narrowed confidence intervals and may decrease the replicability of trial results.

Categorization of key heterogeneity factors

Heterogeneity factors can be conceptualized as either **inherent** patient characteristics (fixed biological and historical factors) that can guide patient eligibility decisions, identification of enrichment samples, or stratification but are immutable, and **manageable** factors that can be balanced through primary protocol design decisions, secondary data collection solutions, and tertiary data monitoring methodologies.

While heterogeneity factors cannot truly be "controlled," the categorical distinction - inherent vs. manageable - allows for clearer determination of where mitigation of study risks is feasible. *Table 1* provides a summary of inherent vs. manageable factors.

Table 1. Inherent vs. Manageable Heterogeneity Factors

INHERENT FACTORS	MANAGEABLE FACTORS
▪ Genetics ▪ Brain morphology ▪ Unmeasured epigenetics ▪ Undetected other patient characteristics	▪ Primary psychiatric diagnosis and comorbidities ▪ Severity of primary and secondary psychiatric illness ▪ Medical comorbidities ▪ Age of psychiatric onset ▪ Pharmacological treatment history ▪ Illness course ▪ Psychosocial factor ▪ Expectations

Mitigation approaches

Although psychiatric heterogeneity is a core feature of psychiatric disorders and is therefore unavoidable even in the context of the most conservative protocol inclusion/exclusion criteria and restrictions, along with placebo response inflation, heterogeneity plays a role in the high rate of CNS trial failure and cannot be ignored. Some mitigation strategies have tended to be avoided for fear of the impact on statistical power and trial duration, and concerns about decreased trial generalizability.

However, the ethical implications of failed psychiatric clinical trials suggest that management is imperative where possible. Below are mitigation strategies that can be implemented in association with manageable heterogeneity factors. Each mitigation strategy is categorized as *primary* - focused on randomization decisions, *secondary* - focused on protecting data integrity, and *tertiary* - focused on data monitoring.

Figure 1 maps mitigation strategies to manageable heterogeneity factors for practical application consideration. *Figure 2* provides a sample trial scenario where understanding of controllable patient variables drove customized study solutions.

Primary mitigation strategies

1. **Refined inclusion and exclusion criteria:** The addition of scientifically informed pre-randomization criteria focused on controllable but high-risk heterogeneity factors with the understanding that recruitment challenges and trial duration may increase (e.g., disallowing enrollment of patients with MDD onset after age 65 in an adult MDD study).

2. **Stratified randomization and planned subgroup or enrichment group analyses:** The grouping of participants by key background factors (e.g., age of onset) and current symptoms (e.g., psychosis severity) and pre-planned analysis of defined subgroups. Such strategies can apply to biomarker or clinical phenotypic characterization, but use of these factors will be limited until more progress is made in precision medicine.

Secondary mitigation strategies

1. **Centralized comprehensive eligibility review:** Third-party review of relevant site-generated information including key aspects of developmental history, psychiatric course, pharmacological and non-pharmacological treatment history, subsyndromal and syndromal medical and psychiatric comorbidities, potentially confounding life circumstances, and psychosocial variables. Reviews are generally done at the screening visit but can be applied at the baseline visit to enhance sound randomization decisions.

2. **Centralized diagnostic or severity validation:** Third-party clinicians review rater diagnostic conclusions or severity scale data and generate an unbiased second opinion. Reviews are generally done at screening but can be done at the baseline visit as well. Concordance between site rater and reviewer scores confirms confidence in data and randomization decisions.

3. **True tandem diagnostics:** Independent clinician- or computer-administered interviews or AI-driven assessments are used to generate scores that serve as unbiased indexes for gauging confidence in the diagnostic conclusions of site investigators prior to randomization. Site raters can be queried when diagnostic discrepancies are found.

4. **True tandem severity ratings:** Independent illness severity assessments are administered directly to participants by third-party clinicians. The resulting second set of scores aids in evaluation of the accuracy of site rater-generated data. Scores from AI- or computer-administered severity assessments that are independent of site severity scales can be used similarly for quality oversight. Consensus discussions between site raters and 3rd party clinicians can be incorporated to ensure an accurate analysis dataset.[12]

5. **Central rating:** The use of off-site, blinded clinicians to administer scales via remote videoconference or teleconference to ensure assessment standardization across patients and visits and control the problem of cognitive biases that are more likely in the site rater-patient relationship. Although well-known for use in studies of treatments with high risk of functional unblinding (e.g., psychedelic studies), expert central rating can also mitigate placebo response and ensure data integrity when site raters have insufficient scale experience. Central ratings represent source for the study.

6. **Central scoring:** Centralized clinicians who are disaffiliated from sites provide a second ClinRO score set based on unbiased review of audio/video recordings of site-based interviews throughout the trial. The secondary score set can be used as the main analysis dataset, can drive consensus scoring via discussions between central scorers and site raters, or can be used purely for site rater performance quality assessment purposes. While central scoring datasets do not always represent source, they provide a means of adjusting the degree of data noise.

7. **Placebo response mitigation training:** Pre-study training for raters and staff that can be standardized in content and timing of provision (generally at site start-up), or provision of tools to enable site staff to apply mitigation strategies through the entire life cycle of a study in an individualized manner that adjusts to each patient over time. Trial subjects can be socialized to the concept of placebo response and the importance of providing frank evaluation responses that will be valued whether representative of improvement, worsening, or no change from baseline. Training can be enhanced by the addition of vendor-hosted site staff practice sessions at investigator meetings or at individual trial sites.

8. **Electronic clinical outcome assessment (eCOA):** Reported and observed manifestations of patient variability can lead to alterations in rater in-study assessment behaviors. For example, patients in a post-traumatic stress disorder trial may vary in terms of the profundity of their index traumas, and the tenacity of recovery of patients in a cocaine use disorder trial may vary. Raters may be more or less inclined to probe sufficiently, or may alter the way they probe, depending on these background characteristics, sometimes in ways that affect data validity and reliability. Electronic scales provide a structure that ensures standardization of evaluation and reduction of background noise. eCOA can be built with scale training conventions provided as 'tips,' and with edit checks that fire when there are logical inconsistencies in data entered by a rater.

Tertiary mitigation strategies

1. **Data Quality Monitoring:** Ongoing, visit-level scale data can be reviewed for administration and scoring quality by trained clinicians or automated AI-driven methodologies. Site raters can accordingly be queried and remediated in terms of training conventions to mitigate the problems of mismeasurement and data noise.

2. **Blinded Data Analytics:** Periodic blinded data analysis can be employed throughout a trial using scientifically-informed algorithms tailored to protocol details and clinical outcome assessment expectations, allowing for characterization of patients based on pre-randomization factors known to vary and tracking of unusual post-baseline data variability and anomalous data results throughout trials.

Figure 1. Mapping Primary, Secondary, and Tertiary Mitigation Strategies to Manageable Heterogeneity Factors

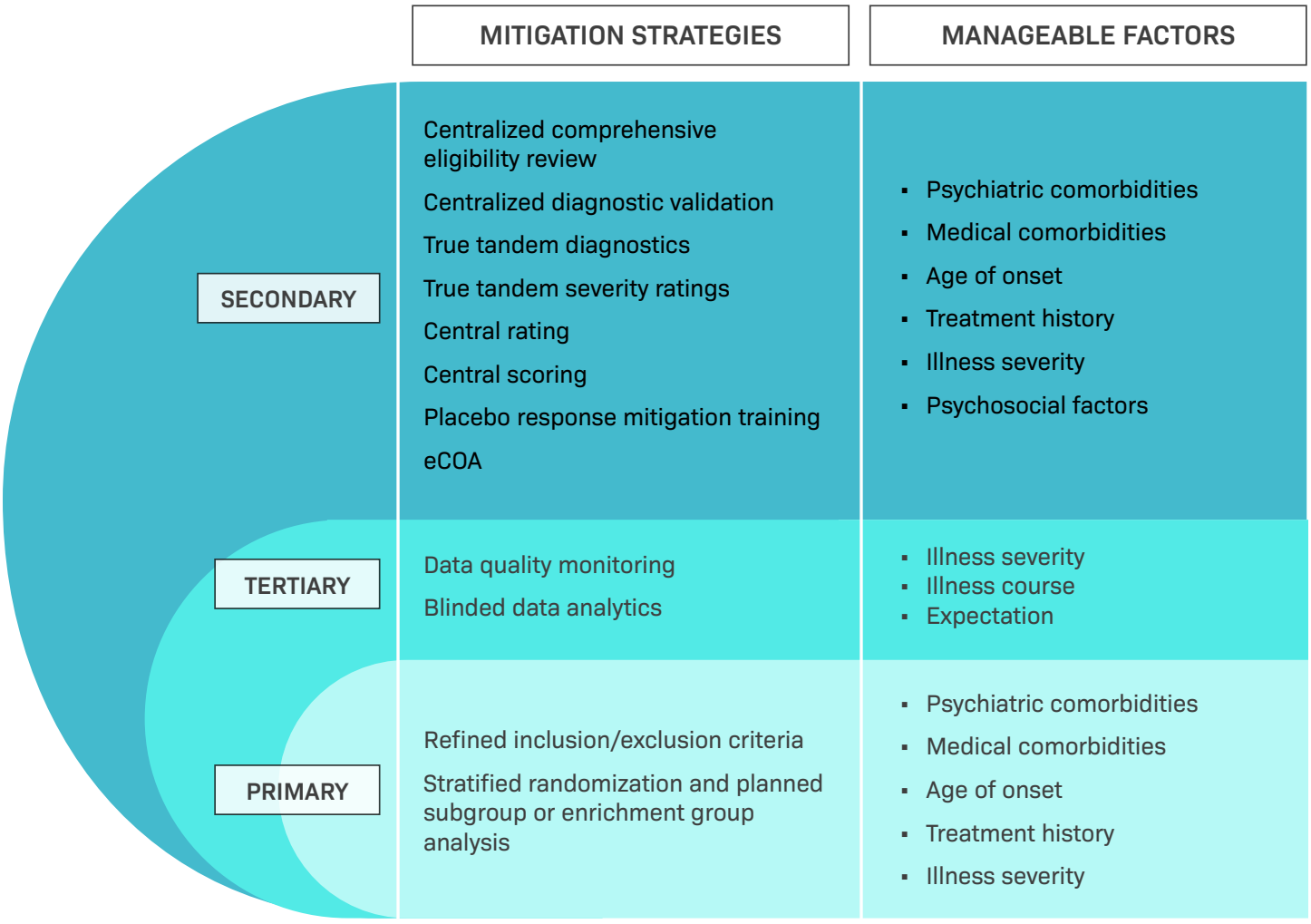


Figure 2. Treatment Resistant Depression Trial Example: Targeted Application of Solutions

MANAGEABLE HETEROGENEITY FACTORS	CUSTOMIZED MITIGATION STRATEGIES
Pre-Baseline Variability Considerations: ▪ Confidence in the primacy of major depressive disorder (MDD) and presence of a current major depressive episode ▪ Treatment Resistant Depression (TRD) diagnostic certainty ▪ Age of MDD onset ▪ Psychiatric comorbidities ▪ Medical comorbidities ▪ Pharmacological treatment history ▪ Baseline illness severity	Targeted protocol criteria: ▪ Lower inclusion threshold for age of onset and severity ▪ Clearly defined inclusion criteria for current major depressive episode and TRD ▪ Clearly defined exclusion criteria related to psychiatric and medical comorbidities ▪ Objective definition of treatment resistance ▪ Clear specification of allowed current and prior antidepressants and washout periods Psychiatric eligibility review of pre-randomization data: ▪ Centralized reviewers evaluate MDD and TRD diagnostic confidence and general psychiatric appropriateness of patients based on review of audio recordings and data from SCID and ClinROs, plus focused review of investigator-provided history and psychosocial information ▪ Central reviewers determine confidence in illness severity at baseline visits based on review of outputs from an independent computer-administered depression severity scale
Post-Baseline Variability Considerations: ▪ Expectation bias related to investigational product media advertisements ▪ Variation in depression severity course from baseline to end or treatment visit resulting from historical and ongoing medical and psychosocial factors	Placebo response mitigation training ▪ Didactic training for patients, raters, and other site staff ▪ Investigator meeting workshops for site staff to allow practice in maintaining neutrality and mitigating potential bias in a patient-centric manner Customized visit-level monitoring of endpoint scale administration and rating quality ▪ Discordances between site ratings and outputs from an independent, computer-administered depression severity scale detect potential rating errors ▪ Centralized reviewers evaluate site MADRSs for administration and rating quality based on review of audio recordings and associated data, and raters are contacted for remediation provision to prevent rater drift ▪ A study-customized blinded data analytics methodology is implemented for detection of key within- and between-visit data anomalies at the rater, site, protocol, and regional levels; expert clinicians interpret analytical results; periodic meetings are held for discussion of findings and determination of needed queries and remedial actions for re-aligning raters with scale conventions

Conclusion

Failing to address psychiatric heterogeneity in clinical trials undermines both the scientific validity and clinical utility of research findings and may contribute to the grave problem of failed CNS clinical trials. Unmanaged heterogeneity can obscure true treatment effects, mask clinically important subgroup responses that would otherwise fuel gains in precision medicine and ultimately result in misleading conclusions.

In contrast, proactive management of psychiatric heterogeneity factors where possible enhances trial rigor and ensures that samples are composed of appropriateness patients and are balanced, and ultimately improves data accuracy, study generalizability, and the likelihood of patient access to precision treatments. Mitigation strategies should be targeted based on an understanding of the drivers of variability in the population under study and protocol-specific considerations. Customized psychiatric heterogeneity countermeasures are crucial for translating clinical trial results into meaningful improvements in real life patient care.

Integrated services for every stage of your trial

Scientific Consulting

Eligibility Review

Central Quality Review

PureSignal Analytics

Placebo Response Mitigation

Juliet Brown, PhD

Juliet Brown, Clinical Director, Science & Medicine, has been with Signant Health for over 10 years and has over 25 years of clinical and research experience. Her specialties are evaluation and treatment in Psychotic Spectrum Disorders, Mood and Anxiety Disorders, and Substance Use Disorders, as well as Cognitive Behavioral Psychotherapy.

Dr. Brown received her Ph.D. and Master's Degrees in Clinical Psychology at Drexel University, United States. Prior to joining Signant Health, she provided psychotherapy to individuals with Severe Mental Illness (SM), in community and inpatient settings, performed federally-funded evaluations for individuals with SMI in supportive housing programs, treated individuals with Substance Use Disorders in a stipulated residential setting, served as a trial Central Rater for various indications, and was most recently a site investigator.

At Signant, she provides scientific oversight for phase 1-3 global CNS psychiatry trials, is a Subject Matter Expert for the Commercial and Business Development Teams, and provides thought leadership for company clinical initiatives.

References

1. Lee, P. H., Feng, Y. A., & Smoller, J. W. (2021). Pleiotropy and cross-disorder genetics among psychiatric disorders. Biol Psychiatry. 89(1): 20-31.

2. Alchakee, A., Ahmed, M., Eldohaji, L., et al. Pharmacogenomics in psychiatry practice: The value and the challenges (2022). Int J Mol Sci.;23(21):13485.

3. Reng, R., Womer, F. Y., Edmiston, E., K., et al. (2020). Antipsychotic effects on cortical morphology in schizophrenia and bipolar disorders. Front Neurosci. 14: 579139.

4. Jedema, H. P., Song, X., Aizenstein, H. J., et al. (2021). Long-term cocaine self-administration produces structural brain changes that correlate with altered cognition. Biol Psychiatry. 15;89(4): 376-385.

5. McGrath, J. J., Lim C. C. W., Plana-Ripoll, O., et al. (2020). Comorbidity within mental disorders: a comprehensive analysis based on 145 990 survey respondents from 27 countries. Epidemiol Psychiatr Sci. 29: e153.

6. Musket, C. W., Kuo, S. S., Rupert, P. E., et al. (2020). Why does age of onset predict clinical severity in schizophrenia? A multiplex extended pedigree study. Am J Med Genet B Neuropsychiatr Genet. 183(7): 403-411.

7. Howes. O. D., Thase, M. E., & Pillinger, T. (2022). Treatment resistance in psychiatry: State of the art and new directions. Mol Psychiatry. 27(1):58-72.

8. Peters, A., Sylvia, L. G., Magalhães, P. V., et al. (2014). Age at onset, course of illness and response to psychotherapy in bipolar disorder: results from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). Psychol Med. 44(16): 3455-67.

9. Correll, C. U., & Rubio, J. M. (2017). Duration and relevance of untreated psychiatric disorders, 1: Psychotic Disorders. J Clin Psych. 78(3), 358-359

10. Rutherford, B. R. & Roose, S. P. (2015). Placebo effects in psychiatry: mediators and moderators. The Lancet Psychiatry. 2(7):654–662.

11. Kott, A., Brannan, S., Wang, X., & Daniel, D. (2021). The impact of aberrant data variability on drug-placebo separation and drug/placebo response in an acute schizophrenia clinical trial. Schizophr Bull Open. 7;2(1).

12. Sachs, G. S., & Peters, A. T. (2016). What blinded raters don't see: Relapse events in a bipolar study detected by site-based raters versus a computer simulated rater. European Neuropsychopharmacology, 26(Supplement 2), S476–S477.

WHO IS SIGNANT HEALTH?

SIGNANT HEALTH
THE EVIDENCE GENERATION COMPANY

Signant Health is the evidence generation company, leveraging software, deep therapeutic and scientific knowledge, and operational expertise to deliver quality evidence across traditional, virtual, and hybrid trial models. For over 25 years, 600+ sponsors and CROs of all sizes – including all Top 20 pharma – have trusted Signant's comprehensive eClinical solutions. Our clinical outcome assessments (eCOA, clinician ratings, cognitive testing) and wearable digital health technologies powered by Ametris provide the industry's most comprehensive evidence generation capabilities, alongside EDC, eConsent, RTSM, supply chain management, and data quality analytics. Learn more at www.signanthealth.com.

06-2026