Volume 14 Issue 1

JCS
Journal for Clinical Studies

PEER REVIEWED

Using Machine Learning to Identify At-Risk Sites in Acute Schizophrenia Clinical Trials

Optimising Early Clinical Development Strategies in Oncology

2022 Cold Chain Predictions: Creating a new Normal

Advancing Data from the Real World Novel Approaches to Clinical Studies

www.journalforclinicalstudies.com

PEER REVIEWED

Using Machine Learning to Identify At-Risk Sites in Acute Schizophrenia Clinical Trials

Data quality concerns are frequent in schizophrenia clinical trials, causing many to suffer from decreased drug placebo separation. Machine learning offers the opportunity to proactively identify raters and sites at risk of developing data quality concerns for early intervention.

CNS clinical trials tend to have a low success rate. In fact, only 1 in 14 molecules entering phase 1 clinical testing reaches the market.[1] Given the high risk associated with CNS drug development, it's no surprise that many pharmaceutical companies avoid researching this therapeutic area altogether. Schizophrenia clinical trials are no exception. These studies are particularly vulnerable to failure because they utilise relatively complex, subjective endpoints, the patient-reported outcomes are often inconsistent or unreliable, and drug placebo differences are modest at best. Successful development of new antipsychotics has been made even more difficult by increasing placebo response and diminishing effect sizes over the last two decades.[2] These trends are widely acknowledged to be multifactorial and have been attributed to causes including industry sponsorship, the growing number of sites involved, a higher probability of receiving medication over a placebo, and inadequate data quality. Of these issues, data quality requires continuous attention for the clinical trial's best hope of success.

To better understand how data quality frequently affects datasets, let's look at an example. Based on an internal analysis of 45,000 clinic visits collected from 6,500 patients in 17 acute schizophrenia clinical trials, the amount of clinically meaningful data quality concerns impact roughly 27% of all study visits. The presence of these data quality concerns has been shown to additionally decrease drug-placebo separation. For example, the presence of outlying variability led to a loss of signal in affected patients in an otherwise successful phase 2 clinical trial.[3] Similarly, the presence of erratic ratings in the negative factor of the Positive and Negative Syndrome Scale (PANSS) increased placebo response and decreased drug-placebo separation in a global, phase 3 negative symptom clinical program not only in affected patients but at sites, too.[4]

Often, data quality concerns accumulate at a small number of research sites.[5] Identification of these sites and timely intervention is paramount in maintaining the integrity of the trial, especially should any of these concerning sites recruit a disproportionally large number of patients. Data analytics implementing smart algorithms allow such an early identification of concerning sites and a timely targeted intervention in a form of retraining, remediation, rater replacement, or site closure.

Given the detrimental effect of data quality concerns on study outcomes, sponsors often implement a battery of solutions to identify and prevent these data concerns in addition to their data analytical programs. In acute schizophrenia clinical trials, these solutions typically consist of:

- site selection based on previous performance;
- pre-study calibration of interview and symptom severity measurement technique;
- placebo response mitigation training;
- operationalisation and monitoring of acuity criteria;
- enhanced instructions and data quality checks embedded in eCOA;
- recording and independent expert review of audio recorded PANSS interviews;
- rapid remediation of rating and interview errors; and
- site enrolment continually tied to data quality.

The application of all these measures has repeatedly shown to improve the overall quality of collected data,[6] yet a non-negligible proportion of data concerns remains in the datasets.

Ideally, an effective solution should offer not only a retrospective identification and resolution of data quality concerns but also a prospective identification of raters and sites at risk of developing these data quality concerns in the future. With the advances of machine learning this has now become possible. (Figure 1)

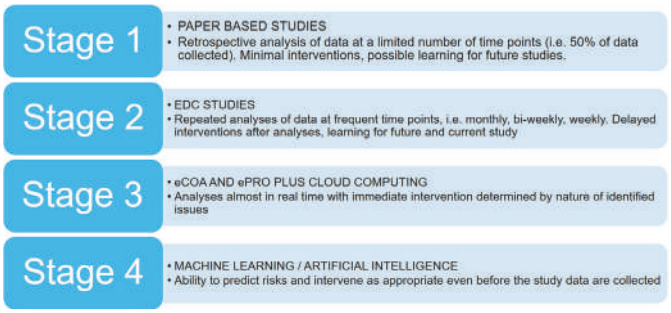


Figure 1: Paradigm Shift in Data Analytics

Leveraging Machine Learning

In clinical research especially, it is critical to apply machine learning models with extreme caution and realistic expectations. Machine learning models should strive to be generalisable, offer actionable results, and address only the clinically meaningful issues.

To maximise the benefits machine learning offers over traditional statistical methods, the technology should be generalisable and allow seamless implementation across clinical trials from the first patient visit. This proves especially challenging in CNS clinical trials due to the variability in study design, treatment allocation, efficacy outcomes, and visit schedules between clinical trials. Researchers are continually refining study designs for the best chance at providing valid readouts. Thus, two clinical trials are rarely alike and the number of useful commonalities between trials is limited, even within the same indication. This makes the development and implementation of generalisable complex models challenging, especially since they rely on a plethora of data types and features (variables) unique to the trial. While various data manipulation and feature engineering techniques could possibly overcome some of these issues, these approaches would unnecessarily increase the complexity of the models and the risk of inaccurate results.

In addition, machine learning models should offer actionable outcomes. If a rater or a site is identified as a data quality risk, there should be a clear set of actions defined upfront with customised

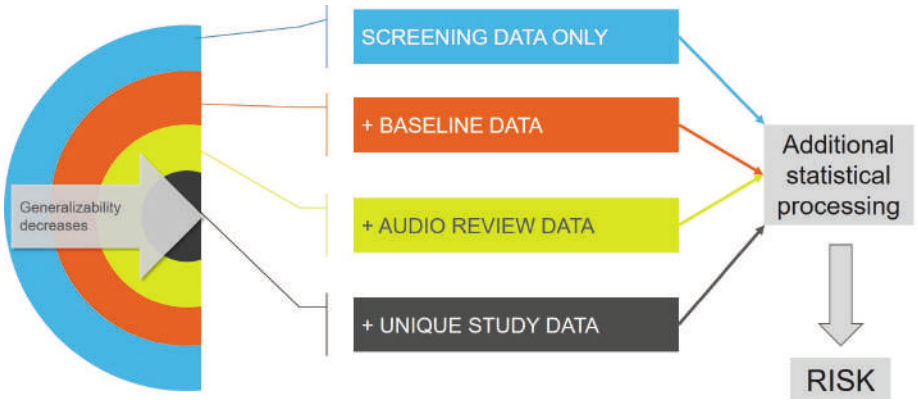


Figure 2: Building the Pipeline

interventions automatically carried out. Such an approach leaves no room for interpretation and allows a uniform response to identified risks.

Finally, and arguably most importantly, the considered models should address only those data quality concerns that are clinically relevant and have the largest potential to significantly affect study outcomes. While correct identification of minor clinically insignificant issues could theoretically improve the performance characteristics of the overall machine learning pipeline, the consequences of such an approach could be risky. Sites and raters would likely become overburdened and alienated by frequent, unnecessary contacts, while sponsors would experience frustration over the program's inability to discern important from unimportant issues, ultimately resulting in the analytical solution being removed from the study and negatively impacting the data. It is thus critical to carefully select the issues that have repeatedly shown to strongly impact placebo response and drug placebo separation, or the issues that are clinically improbable and are associated with data tampering or fraud.

Successful Implementation for PANSS Inconsistencies

The Positive and Negative Syndrome Scale (PANSS) is one of the most frequently used primary efficacy outcomes in schizophrenia research. The scale consists of 30 items, each of which is rated from 1 (absent) to 7 (extreme). Many items measure similar symptoms and are therefore expected to show a strong degree of association. For example, a patient suffering from severe persecutory delusions should have a high score on both, generic Delusions (P1) and specific Suspiciousness/Persecution (P6) items. A failure to do so is an indication of an error (logical within PANSS inconsistency) that could originate in incorrect or idiosyncratic application of PANSS scoring guidelines, inadequate understanding of underlying psychopathology, rating sloppiness, or, in the worst-case scenario, data fabrication.

The presence of these logical inconsistencies within PANSS has been previously shown not only to significantly increase placebo response in both affected patients and sites but also eliminate the drug-placebo separation.[7] These findings thus make logical inconsistencies a good target for the application of machine learning pipelines.

Given the obstacles and goals mentioned previously, Signant Health's team built a layered machine learning pipeline. (Figure 2)

The first defensive layer is a parsimonious model that uses only the most common screening data available (in this case just the PANSS scale), so that as the patient is screened the data is initially processed by this screening model only. As the patient progresses to baseline, a second more complex model will be utilised that includes both screening and baseline data. If audio review of the study visits and PANSS administration were available, once the reviewer data become available, a third model would be implemented. Lastly, with critical mass of data within the study available, a fourth model could be implemented that would rely not only on the data commonly collected across trials but also the study-specific measures.

You can appreciate that with each layered model the complexity increases, and the generalisability of the models decreases. Once the unique study data is used, these models become study specific, and thus, not reusable.

If we assess the performance of the pipeline (Figure 3), it can be noted that the performance improves as expected with the more complex models, however the gains in the performance decrease with the additional complexity of the models.

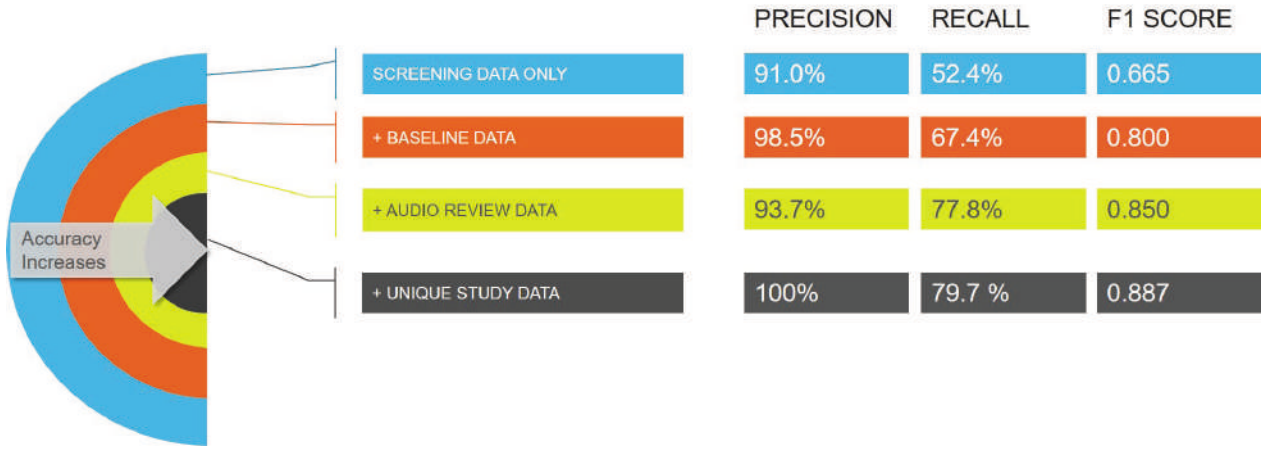
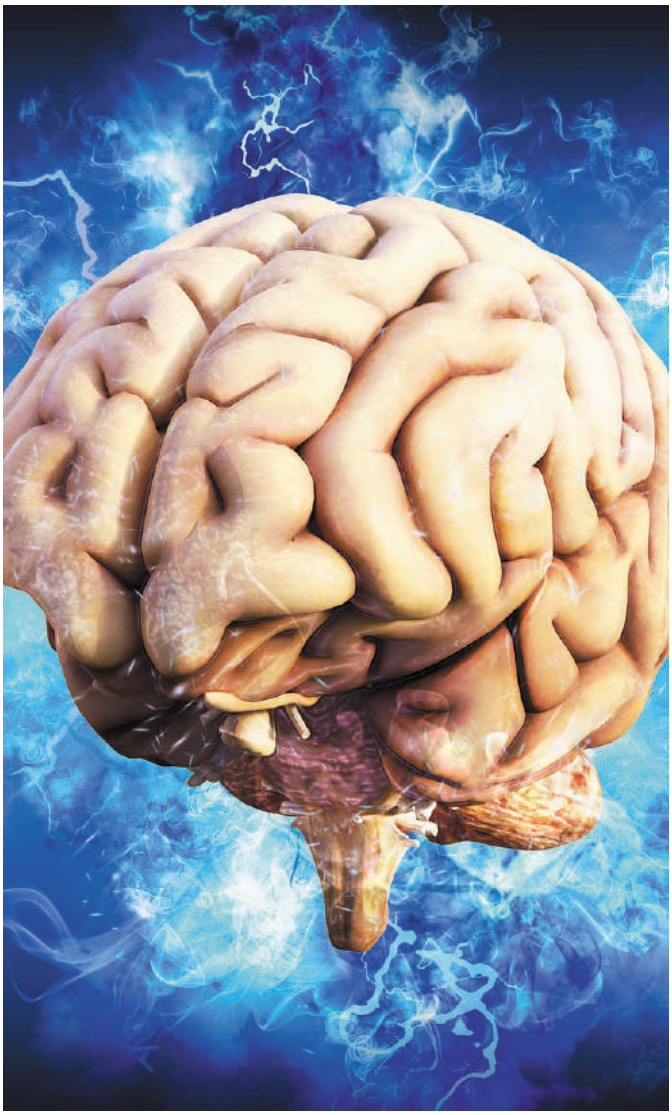


Figure 3: How Does the Pipeline Perform?



This does not mean that the more complex models should be discounted; it merely informs us that the simpler models are perfectly capable of identifying risks without a huge loss in the model performance.

In the final step, the outcomes of the machine learning pipeline are statistically analysed, and at-risk raters and sites are retrained. This step further improves the performance of the system and decreases the risk of possible over contacting.

In conclusion, machine learning offers a viable analytical solution allowing researchers to predict and address future data quality concerns throughout a study. However, one needs to be realistic and cautious about the implementation of machine learning. Only highly accurate and clinically relevant models should be considered, as the consequences of inaccurate or irrelevant models may ultimately result in data quality deterioration. A layered, onion-like structure allows studies to implement parsimonious models collecting from the clinical trial's first patient visit and add more complex, study-specific models as subsequent data becomes available to further improve the pipeline performance.

REFERENCES

1. Thomas, David; Chancellor, Daniel; Micklus, Amanda; LaFever, Sara; Hay, Michael; Chaudhuri, Shomesh et al. (2021): Clinical Development Success Rates and Contributing Factors 2011–2020. Biotechnology Innovation Organization (BIO); Informa Pharma Intelligence; QLS. Available online at https://pharmaintelligence.informa.com/~/media/informa-shop-window/pharma/2021/files/reports/2021-clinical-development-success-rates-2011-2020-v17.pdf, checked on 1/3/2022.
2. Leucht, Stefan; Leucht, Claudia; Huhn, Maximilian; Chaimani, Anna; Mavridis, Dimitris; Helfer, Bartosz et al. (2017): Sixty Years of Placebo-Controlled Antipsychotic Drug Trials in Acute Schizophrenia. Systematic Review, Bayesian Meta-Analysis, and Meta-Regression of Efficacy Predictors. In The American journal of psychiatry 174 (10), pp. 927–942. DOI: 10.1176/appi.ajp.2017.16121358.
3. Kott, Alan; Brannan, Stephen; Wang, Xingmei; Daniel, David (2021): The Impact of Aberrant Data Variability on Drug–Placebo Separation and Drug/Placebo Response in an Acute Schizophrenia Clinical Trial. In Schizophrenia Bulletin Open 2 (1), p. 295. DOI: 10.1093/schizbullopen/sgab037.
4. Umbricht, Daniel; Kott, Alan; Daniel, David G. (2020): The Effects of Erratic Ratings on Placebo Response and Signal Detection in the Roche Bitopertin Phase 3 Negative Symptom Studies—A Post Hoc Analysis. In Schizophrenia Bulletin Open 1 (1), p. 203. DOI: 10.1093/schizbullopen/sgaa040.
5. Daniel, David G.; Kott, Alan (2014): Risk Based Data Quality Monitoring Utilizing Data Analytics and Recorded PANSS Interviews in Global Schizophrenia Trials. Poster presentation at the 10th Anniversary International Society of Clinical Trials Methodology (ISCTM) Meeting, Philadelphia, PA, 18–20 February 2014.
6. Kott, Alan; Brannan, Steven K.; Wang, Xingmei; Murphy, Christopher; Targum, Steven D.; Daniel, David G. (2020): Procedures to Optimize Endpoint Data Quality in an Acute Schizophrenia Study. Presented at the ISCTM 2020 Autumn Virtual Conference, September 21 – 25, 2020., 9/21/2020.
7. Kott, Alan; Lee, Jeniffer; Forbes, Andy; Pfister, Stephanie; Ouyang, John; Wang, Xingmei; Daniel, David G. (2016): Logical inconsistencies among PANSS items are associated with greater placebo response in acute schizophrenia trials – A post-hoc analysis. Poster presentation. Philadelphia, PA, USA, 9/26/2016.

Dr. Alan Kott

Dr. Kott oversees the design and reporting of Signant Health's data analytics in large schizophrenia studies. For the past seven years, he has also lent his valuable expertise to training investigators based on best practices.

Andrei Iacob

As Associate Clinical Data Scientist, Andrei supports Signant Health's Blinded Data Analytics and related initiatives. He continues to lend his years of experience to support various machine learning projects.

Emanuel Pintilii

Emanuel is serves as a Clinical Data Scientist for Signant Health and has been instrumental in the company's development of different machine learning systems.

Xingmei Wang

Xingmei Wang is a Senior Statistician on Signant Health's Science & Medicine team, providing company-wide, statistical support.