AKRIVIA HEALTH | IQVIA | Cardiovascular Research Consortium

WHITE PAPER

The Transformative Potential of NLP Technology Within Psychiatric Clinical Practice:

The exploration of a US NLP Portability Pilot and its implications for Clinicians and wider Life Sciences

David Newton, Co-Founder & Chief Business Officer, Akrivia Health
Elias Kassis, Associate Director, Real World Networks and Partnerships, IQVIA
Gregory Folz, CCRP, Co-Founder, Cardiovascular Research Consortium, LLC

Transforming mental health & dementias research through data

Table of Contents

Executive Summary

Pilot Project Sample Size

Over
33,000
psychiatric encounters

From
500
patients

Sample Location

United States

This pilot project assessed the portability of Akrivia Health's (Akrivia) Natural Language Processing (NLP) technology — currently successfully implemented across several NHS organisations in the UK — for use on de-identified psychiatric Electronic Health Record (EHR) data from the United States. Specifically, the pilot aimed to evaluate whether Akrivia's NLP models, and in particular its pipeline for extracting clinical signs and symptoms could sustain high accuracy and clinical relevance when applied to US clinical data.

The pilot was conducted through a multi-institutional collaboration involving Akrivia (technology provider), the Cardiovascular Research Consortium, LLC (CRC, EHR data processor), Mary O'Daniel Stone and Bill Stone Center for Child and Adolescent Psychiatry at Indiana University School of Medicine - Evansville (EHR data provider) and IQVIA (scientific partner). The evaluation utilized a dataset comprising more than 33,000 psychiatric encounters from 500 patients. Akrivia's NLP models were tested on free-text clinical notes to extract signs and symptoms and contextual insights.

Initial validation indicated that the performance of the models on US data was comparable to the performance achieved on UK datasets. The Contextual Classification (CC) model achieved an F1 score of 88% without any alterations. After two refinement iterations, the Named Entity Recognition (NER) model demonstrated significant improvements—achieving a precision of 92% and a recall of 76%.

These findings demonstrate the potential of Akrivia's NLP technology to deliver reliable and generalizable performance across diverse international clinical settings.

Beyond technical validation, this pilot underscores the transformative potential of applying NLP technology to unstructured clinical text within US psychiatric care. The technology offers scalable, privacy-preserving solutions to persistent challenges in diagnostic accuracy, clinical documentation, and research, with particular promise for the early identification of complex clinical phenomena such as treatment-resistance.

This work establishes a foundation for broader adoption and integration of NLP technologies in US psychiatry practice and paves the way for a federated research network capable of linking clinical insights across the UK, US, and potentially other international settings. Future phases will aim to expand the dataset, validate predictive modeling approaches, and harness NLP-extracted data to accelerate research in psychiatric and neurodegenerative disorders.

Introduction

The global mental health crisis has intensified, placing unprecedented strain on clinicians and psychiatrists, particularly in the United States. Despite increased awareness and treatment efforts, mental health conditions are worsening.[1] Suicide rates have surged by 30% since 2000, reaching nearly 50,000 deaths in 2022—the highest since 1941.[2]

Clinicians face overwhelming demand: 73% of mental health practitioners report increased patient loads since the pandemic, with 45% having to turn away patients or place them on waitlists.[3] Concurrently, the US anticipated a shortage of up to 31,000 behavioral health professionals in 2024, affecting over 160 million Americans in areas with insufficient mental health services.[4]

The administrative burden exacerbates burnout; physicians spend two hours on clerical tasks for every hour of patient care, contributing to emotional exhaustion and reduced job satisfaction. This burnout not only affects clinicians' well-being but also compromises patient care quality.

But workload is not the only challenge being faced. Efficient access to key insights remains an issue. The increasing complexity of psychiatric disorders demands advanced tools to effectively utilize vast clinical data. Much of this data exists in unstructured Electronic Health Records (EHRs), making it difficult to leverage for clinical decisions and research.

Traditional methods for extracting clinically relevant data from EHRs are often manual, resource-intensive, and inconsistent, limiting their scalability and accuracy. These challenges highlight the need for robust NLP tools capable of accurately identifying and categorizing psychiatric signs and symptoms to enhance clinical practice and research.

NLP methodologies, specifically tailored NLP pipelines, offer potential solutions by transforming unstructured data into meaningful insights. Such technologies, like Akrivia Synapse™, offer promising avenues to alleviate clinician workload and uncover new insights by extracting structured insights from unstructured clinical notes.

This white paper explores the methodology and application of NLP solutions, highlighting their potential to significantly impact clinical workflows, patient care, and research efficiency. Specifically, it examines clinical perspectives, including real-world applications in UK and US psychiatric settings, the value of NLP in life sciences research, and broader future possibilities for NLP-driven advancements in care delivery.

NLP Definition

Natural Language Processing (NLP) is a branch of computer science and artificial intelligence (AI) focused on enabling machines to interpret, process, and generate human language. By combining computational linguistics, statistical modeling, and machine learning techniques, NLP empowers computers to understand both written and spoken language.[5]

In clinical settings—particularly psychiatry—NLP unlocks insights from unstructured text in electronic health records, enabling earlier detection of risk, more personalized care, and population-level analysis. It also supports proactive clinical decision-making, reduces administrative burden through automation, and accelerates research by identifying relevant patient cohorts.

Comparing EHRs in the US and the UK

EHRs are a core part of patient care in both the US and the UK, with shared challenges despite different healthcare models. In both jurisdictions — particularly in CNS and psychiatry — clinical documentation often includes rich narrative content, such as mental status exams, risk assessments, and general progress notes, which are not consistently templated or structured. While the US system places a greater emphasis on structuring data for billing and administrative purposes, the reality in psychiatric care is that many critical insights remain embedded in unstructured text. By contrast, the UK's NHS EHRs tend to be less driven by billing but similarly rely on narrative documentation in mental health, especially in clinical notes that fall outside the constraints of structured fields.

Despite differences in purpose and formatting, both systems face a common challenge: essential diagnostic and treatment-related information is often buried in unstructured text. This is particularly true in mental health, where nuances in clinical observations, patient history, and longitudinal care narratives are difficult to capture in standardised fields. An NLP model trained specifically on psychiatric data can unlock these hidden insights, transforming narrative content into analysable, actionable information. For US clinicians navigating complex psychiatric cases, this means improved care coordination, faster access to relevant clinical history, and enhanced decision-making — all without sacrificing the narrative richness essential to mental health care.

Data Source

The dataset for this pilot project was sourced from CRC/Sidus Insights, which has secure access to HIPAA-compliant EHR systems through a national network of community-based primary and specialty care providers across all 50 states.

De-identification of the CRC/Sidus Insights data was conducted by Datavant's Privacy Hub in accordance with the HIPAA Expert Determination method (45 CFR § 164.514[b][1]). The process included thorough redaction of both direct and indirect identifiers. An independent validation confirmed that over 99.9% of records had all direct identifiers successfully removed, with a recall rate of 99.99%, exceeding the compliance threshold defined by HIPAA. These results demonstrate the effectiveness and reliability of the de-identification process, ensuring the dataset meets stringent privacy standards for secure research use.

From a total population of 4.9 million patients, a random sample was selected, yielding 33,560 psychiatric encounters from 500 patients. To ensure alignment with the Akrivia's model's training, the dataset was filtered exclusively to psychiatric specialities with at least three visits to mental health specialists, as the models were initially trained on UK psychiatric notes. This filtering ensured that model performance was assessed on a comparable dataset. It is important to note that it is the NLP that was portable, not the data. Data remained securely within the US and adhered to required governance controls.

NLP Methodology

This pilot project assessed the adaptability of Akrivia Synapse™, Akrivia's NLP pipeline, for analysing US psychiatric clinical notes. The primary objective was to extract and classify key psychiatric signs and symptoms using two of Akrivia's NLP models:

1 **Named Entity Recognition (NER)**

Model that extracts specific words or phrases in a sentence and identifies it as one of the signs and symptoms.

2 **Contextual Classification (CC)**

Model used to classify extracted signs and symptoms to specific classes based on its surrounding context and meaning.

The performance of the NER and CC models were evaluated using precision, recall, and F1 score, metrics that assess the models' predictions against human-annotated gold standards:

Key Metrics

Precision: Measures prediction accuracy. It is the proportion of correct predictions out of all predicted entities. Higher precision indicates fewer false positives.
Formula: Precision = TP / (TP + FP)

Recall: Measures coverage of relevant entities. It is the proportion of correct predictions out of all actual entities. Higher recall indicates fewer false negatives.
Formula: Recall = TP / (TP + FN)

F1 Score: A single metric that balances precision and recall, providing a comprehensive performance measure.
Formula: F1 Score = 2 × (Precision × Recall) / (Precision + Recall)

Confidence intervals at **95%** were calculated using the bootstrapping method, which involved resampling the dataset 10,000 times to estimate performance variability, providing a robust measure of results across different samples. Success criteria for model performance were jointly established by Akrivia, CRC, and IQVIA, with the following targets:

NER Model:
- Precision: **Above 80%**
- Recall: **Above 70%**

CC Model:
- Weighted Average F1 Score: **Above 80%**

Initial validation was conducted using random samples from the US psychiatric dataset with the NER model assessed on 300 randomly selected sentences and the CC model on 200 randomly selected sentences. Both NLP models performed similarly on UK and US datasets. The primary distinction was lower NER model recall on the US dataset. For complete initial validation results, see the appendix. Addressing this recall gap for NER has become a key focus of Akrivia's refinement phase.

During the refinement phase, enhancements were made to the NER model to expand clinical term coverage and to the evaluation methodology to improve performance measurement.

Table 1: NER Model Updates

Enhancement	Description	Example
New Labels	Additional labels were introduced to capture previously missed medical terms, expanding coverage and improving recall.	**“Wheezing”** was added to identify phrases like **“cardiac wheezes”**.
New Captures for Existing Labels	New phrases were added under existing labels to capture linguistic variations present in US notes.	**“AH”** was included for **“auditory hallucinations”**, and **“wt gain”** for **“weight gain”**

Table 2: Evaluation Methodology Improvements

To ensure performance metrics more accurately reflect model capabilities, two key enhancements were made:

Enhancement	Description
Stricter Overlap Verification	Only exact matches between model predictions and human annotations are automatically counted as true positives, while partial overlaps are flagged for human verification.
Comprehensive Recall Evaluation	Missed mentions within annotations are now counted as false negatives to improve recall measurement.

For the full details of the evaluation methodology improvements, please refer to the appendix.

These improvements enhance the NER model's coverage and ensure that precision and recall metrics more accurately represent its performance. The NER model underwent two iterations of refinement, with annotations conducted by three members of the R&D team and supported by consultancy advice from our in-house clinical psychiatrists. The annotations were performed independently, following established guidelines, with an Inter-Annotator Agreement (IAA) of approximately 90%. Any disagreements were resolved through group consensus.

A random sample of 300 psychiatric sentences was reviewed for every iteration. The table below shows the final refinement achieved significant performance improvements, particularly in NER model recall, addressing all previously identified gaps.

Table 3: US NER refinement iteration 1 and 2 results

NER Validation	Precision	Recall	F1
Iteration 1	**0.85** 95%CI [0.80,0.89]	**0.68** 95%CI [0.62,0.74]	**0.76** 95%CI [0.71,0.80]
Iteration 2	**0.92** 95%CI [0.89,0.95]	**0.76** 95%CI [0.70,0.81]	**0.83** 95%CI [0.79,0.87]

Overall, both the NER and CC models met or exceeded success criteria on US psychiatric data, demonstrating the robustness, reliability, and adaptability of Akrivia's NLP models for diverse clinical contexts.

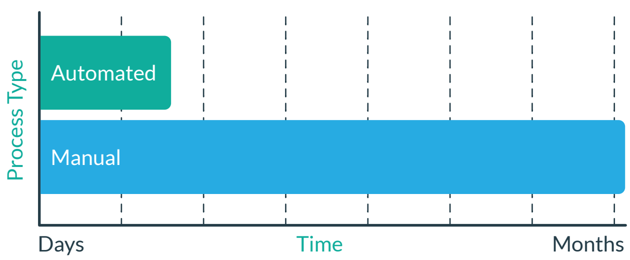
Clinical Perspectives on NLP Use

> From a clinical perspective, the capability to extract and synthesize information and insights from unstructured clinical notes provides substantial value to healthcare providers by improving clarity, efficiency, and decision-making in patient care. Early implementation in the UK, supported by feedback from frontline clinicians, has demonstrated both the immediate impact and future potential of this technology in transforming care delivery across diverse healthcare regions.

Current Insights in UK Psychiatry: Real World Applications

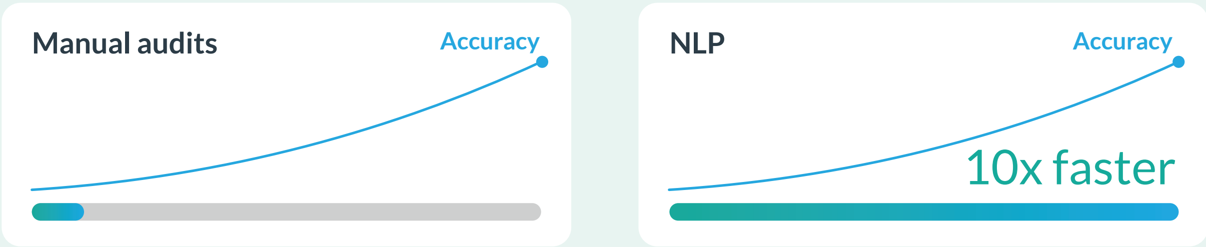
Clinicians routinely navigate vast amounts of unstructured data within electronic health records (EHRs), often manually reviewing patient histories and symptoms to infer treatment responses. This process is time-consuming, prone to inconsistencies, and limits large-scale insights.

Figure 1: **Akrivia's NLP technology reduces months of manual data reviewing to just a few days**



NLP is transforming clinical workflows by enabling faster, more comprehensive review of patient notes. In the UK, NLP-powered tools like Akrivia Synapse are streamlining audits and research, **reducing months of manual data reviewing to just a few days**. The application of NLP has created new opportunities for conducting high-impact audits that support evidence-based policy and research. A national audit examining clozapine prescription patterns utilised NLP-structured data from Akrivia, enabling faster identification of clinical trends, improved data accuracy, and more efficient benchmarking across healthcare providers.

Figure 2: Comparable Accuracy: Manual vs NLP-enhanced audit



In the UK, the national audit examining clozapine prescription patterns revealed a significant gap in care, where 65% of patients with treatment-resistant schizophrenia had no recorded mention of clozapine prescription in their clinical records, despite its established effectiveness. Traditionally, manual audits of this scale required many months to complete. **By contrast, NLP achieved comparable accuracy at over ten times the speed**, allowing clinicians to analyse prescribing trends, demographic disparities, and treatment pathways with unprecedented efficiency.

Beyond audits, NLP enhances clinical research by **supporting predictive models that identify high-risk patients.** One ongoing study leverages NLP to predict whether patients with first-episode psychosis will require clozapine, optimising early intervention strategies. Similarly, NLP is being applied to investigate adult risk factors for dementia, **extracting meaningful patterns from years of free-text notes.**

By integrating NLP into clinical practice, clinicians can **move beyond manual data review and focus on delivering proactive, data-driven care.** Whether identifying treatment gaps, improving documentation, or conducting large-scale audits, NLP is directly shaping the future of psychiatric care.

Benefits of Applying NLP Solutions in US Primary Care and Psychiatric Care Settings

Many of the challenges faced by US clinicians parallel those observed in the UK. There is a growing gap between the numbers of specialized mental health care workers and mental health needs. The Association of American Medical Colleges (AAMC) has projected a shortage of up to 31,109 psychiatrists in the US in future years.[6] In the US, the primary care setting has become critical to delivery of mental healthcare. Nearly 60% of patients receive mental health diagnosis and treatment from a primary care provider (PCP). While PCPs often provide the first point of contact for individuals exhibiting psychiatric symptoms, they frequently lack the resources or specialized training necessary for early and accurate diagnosis. This gap contributes to delays in appropriate intervention and poorer patient outcomes.

Further, mental health care in the US is hindered by fragmented care delivery and poor communication between providers. Patients commonly interact with multiple providers for various symptoms before receiving a formal psychiatric diagnosis. Variability in clinical documentation, development of condition over time, inconsistent use of validated screeners (e.g. PHQ-9), and comorbidities may contribute to missed diagnoses and inaccurate diagnoses. Estimates suggest that general practitioners correctly identified depression in less than half of cases — which leads to suboptimal prescribing practices, ultimately exacerbating downstream care complexities.[7]

Integrating NLP into US psychiatric care may present an effective and sustainable solution. PCPs and psychiatrists must interpret extensive patient histories, including symptom onset, medication adherence, comorbid physical health conditions, social determinants of health, and critical safety concerns (e.g., suicidality, psychosis). NLP can synthesize this information into structured, longitudinal views of patient journeys which can highlight red flags, provide diagnostic criteria already met and prompt for further diagnostic questioning, identify unmet needs, and enable more informed clinical decision-making.[8]

By equipping clinicians with tools that simplify data extraction and interpretation, NLP facilitates more accurate diagnosis and intervention in primary care settings and faster transitions of more complex into specialized psychiatric care.

This not only enhances the clinician's experience but also fundamentally reshapes patient outcomes, enabling more effective, personalized interventions rooted in best-practice guidelines, such as the Collaborative Care Model (CoCM). While consultation services of this nature are critical, it often takes a lot of time to review and interpret medical records. This further highlights the value that NLP brings to practice.

Life Sciences Perspectives

EHRs contain rich, longitudinal data on patient symptoms, treatments, outcomes, and psychosocial factors—critical for advancing life sciences, especially in fields like psychiatry where therapeutic progress has stagnated. Precision research increasingly depends on the ability to subtype patients based on features not captured by standard diagnoses, such as cognitive impairment in schizophrenia. With ~85% of relevant data locked in unstructured clinical notes, AI techniques like NLP are essential to unlock this information at scale, turning EHRs into powerful, research-ready datasets.

Scaling NLP: Maximizing Efficiency in Time and Cost

Precision in identifying patient subgroups from EHR data paradoxically requires significant scale, demanding the curation of large datasets—often thousands to millions of records—to reliably differentiate statistically meaningful signals from noise. Traditional manual chart reviews quickly become unsustainable due to high costs where studies suggest that human-led chart review can process approximately 5–10 records per hour, depending on complexity.

In contrast, computationally driven NLP methods offer a scalable, cost-effective, and highly efficient alternative, processing thousands of records per second with consistent accuracy. This automation drastically reduces the time from weeks or months to mere hours and **significantly lowers operational overhead by minimizing the need for human annotators**, thus **improving cost-efficiency** as dataset size increases.

Standardized, Measurable, and Self-Improving NLP Approach

NLP offers a standardized and highly reproducible methodology for extracting structured data from EHR, significantly reducing the variability, fatigue, and cognitive biases inherent in manual abstraction. NLP models include highly efficient regular-expression-(regex)-based models and advanced deep learning models, such as transformer-based architectures (e.g., BERT). These models achieve high performance, measurable through objective metrics including precision, recall, and F1 score.

> *Additionally, privacy-preserving NLP techniques, such as automated de-identification, **maintain patient confidentiality and foster public trust, enabling responsible, scalable, and accelerated health data research.***

Impact on Patient Cohorts

Many clinically relevant patient cohorts lack explicit labels in structured EHR data. Within psychiatry, conditions such as cognitive impairment associated with schizophrenia (CIAS) have no specific DSM or ICD codes, leaving many diagnoses undocumented in structured fields. This gap is often due to structural healthcare system factors rather than data quality issues alone.

Clinicians frequently cite:

- Lack of approved treatments, reducing the incentive to formally document diagnoses.
- Stigma associated with certain conditions.
- Diagnostic uncertainty.
- Additional workload required for structured data entry.

However, critical details such as symptoms, clinical reasoning, and suspected diagnoses are commonly documented in unstructured clinical notes.

NLP enables extraction of these insights, allowing researchers to define patient cohorts systematically using operational criteria beyond structured data alone.

NLP-based approaches can rapidly analyze extensive patient records to:

- Identify accurate patient cohorts based on symptoms, risk factors, and clinical patterns.
- Estimate prevalence and incidence, enhancing epidemiological research and disease burden analysis.
- Characterize sub-groups by disease trajectories, severity markers, and novel phenotypic clusters.

These insights support health economic modeling, risk stratification, and decision-support systems.

For instance, NLP can profile schizophrenia patients at initial assessments to identify those at risk of treatment resistance, facilitating earlier intervention and optimized treatment plans.

Broader Use Cases and Future Possibilities

Psychiatry and neurodegeneration research faces challenges due to condition heterogeneity, overlapping symptoms, and complex biopsychosocial causes. There is an increasing need for longitudinal, detailed EHR data. NLP methods tailored to specific diseases help extract valuable clinical phenotype data hidden within EHRs, enabling access to extensive, longitudinal, and diverse datasets crucial for advancing research.

Through NLP enrichment and targeted analytics, these datasets become valuable resources for life sciences, facilitating deeper understanding and progress in the field.

Collaborations among Akrivia, CRC, and IQVIA, facilitate a range of advanced research opportunities aimed at enhancing the understanding and treatment of psychiatric conditions. These include:

- Identifying sub-groups and transdiagnostic clusters, supporting symptom-driven research.
- Modeling disease progression, delineating stages and markers of severity.
- Optimizing clinical trial recruitment through pre-identification of eligible participants thereby reducing screen failure rates .
- Developing personalized treatment models to predict patient responses to specific interventions.
- Conducting epidemiological and burden-of-disease studies to assess prevalence, incidence, and resource use.
- Integrating clinical phenotypes with multi-omics data (genomic, transcriptomic, proteomic) to discover new therapeutic targets.

Conclusion

This pilot demonstrated the strong adaptability and performance of Akrivia's NLP technology on US psychiatric EHRs, confirming its potential to transform care delivery across geographies. By enabling structured insights from unstructured clinical notes, NLP empowers clinicians with faster, more accurate decision-making and supports researchers in unlocking critical data for precision psychiatry and drug development. As adoption grows, NLP stands to bridge clinical and life sciences innovation where it has high potential in reshaping psychiatric care, accelerating research, and ultimately improving outcomes for complex patient populations worldwide.

References

1. Cantor, J. H., McBain, R. K., Ho, P.-C., Bravata, D. M., & Whaley, C. (2023). Telehealth and In-Person Mental Health Service Utilization and Spending, 2019 to 2022. JAMA HealthForum, 4(8), e232645.

https://doi.org/10.1001/jamahealthforum.2023.2645

2. Curtin, S., Garnett, M., & Ahmad, F. (2023). Vital Statistics Rapid Release Provisional Estimates of Suicide by Demographic Characteristics: United States, 2022.

https://www.cdc.gov/nchs/data/vsrr/vsrr034.pdf

3. Mental Health Practitioners Seeing Increasing Number of Patients, Experiencing Burnout. (2020). Wiley.com.

https://newsroom.wiley.com/press-releases/press-release-details/2023/Mental-Health-Practitioners-Seeing-Increasing-Number-of-Patients-Experiencing-Burnout/default.aspx?

4. Satiani, A., Niedermier, J., Satiani, B., & Svendsen, D. P. (2018). Projected Workforce of Psychiatrists in the United States: A Population Analysis. Psychiatric Services, 69(6), 710–713.

https://doi.org/10.1176/appi.ps.201700344

5. Stryker, C., & Holdsworth, J. (2024, August 11). Natural language processing. IBM.

https://www.ibm.com/think/topics/natural-language-processing

6. Weiner, S. (2022, August 9). A growing psychiatrist shortage and an enormous demand for mental health services. Association of American Medical Colleges.

https://www.aamc.org/news/growing-psychiatrist-shortage-enormous-demand-mental-health-services

7. Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: a meta-analysis. The Lancet, 374(9690), 609–619.

https://doi.org/10.1016/s0140-6736(09)60879-5

8. Cerimele, J. M., Fortney, J. C., Pyne, J. M., & Curran, G. M. (2018). Bipolar disorder in primary care: a qualitative study of clinician and patient experiences with diagnosis and treatment. Family Practice, 36(1), 32–37.

https://doi.org/10.1093/fampra/cmy019

Appendix

Table 4: US psychiatric sentences NER results on initial validation

Score	Point Estimate	95% Confidence Interval
Precision	0.92	[0.87, 0.96]
Recall	0.60	[0.54, 0.65]
F1	0.60	[0.68, 0.76]

Table 5: US CC Results

Score (Weighted Average)	Point Estimate	95% Confidence Interval
Precision	0.89	[0.85, 0.94]
Recall	0.87	[0.82, 0.91]
F1	0.88	[0.83, 0.92]

Stricter Overlap Verification

The previous methodology identified any overlap between a model-predicted extraction and the human gold standard annotation as a true positive. While this approach was effective in many cases, it occasionally overestimated performance metrics by not adequately distinguishing between contextually correct and incorrect extractions or by failing to account for partial misses. Consider the following sentence example below:

Sentence: *"His sleep is unstable."*

Gold Standard Annotation: The phrase "*sleep is unstable*" was identified as a capture under "Poor sleep" label.

Model Output:

1. *"unstable"* (identified as a capture under the label "Labile Affect"): The model extracted the word "*unstable*" independently and incorrectly associated it with the label "Labile Affect," which does not match the intended context. In this case, "*unstable*" on its own carries a different meaning.
2. "*sleep is unstable*" (identified as a capture under the label "Poor sleep"): The model also extracted the full phrase "sleep is unstable" and assigned it the correct label, "Poor sleep," aligning with the gold standard.

Previous Methodology Outcome: Both extractions were treated as true positives because they overlapped with the gold standard annotation, even though only the second extraction was contextually accurate.

Improved Methodology Outcome: Model-predicted extractions that partially overlap with the gold standard are no longer automatically classified as true positives. Instead, overlaps are classified as a disagreement and are manually reviewed to determine whether they fully align with the intended scope of the gold standard. In the example above, the extraction "*sleep is unstable*" remains as a true positive. However, the extraction "*unstable*" will now be treated as a false positive on the improved methodology.

Comprehensive Recall Evaluation

When performing annotations, it is sometimes necessary for multiple mentions to be included within the same annotation due to the placement of a shared modifier. We define *mention* as an individual reference of a sign and symptom within a sentence. An *annotation* is a human-annotated phrase or text that may encompass one or more mentions. Consider the following sentence example below:

Sentence: *"He has been eating and drinking well."*

Gold Standard Annotation: The full phrase "*eating and drinking well*" is an annotation containing two mentions: *eating* and *drinking*. The word '*well'* modifies and applies to both *eating* and *drinking* which overall reflects the patient's nutritional and hydration status.

Model Output: The model only extracted "*drinking well*"

Previous Methodology Outcome: Under the previous methodology, the partial match "*drinking well*" was counted as a true positive, missing the additional component "*eating well*".

Improved Methodology Outcome: In the improved methodology, any missed mentions within a gold standard annotation are recorded as false negatives. In the example above, the missing mention "*eating well*" is appropriately classified as a false negative. This ensures that the recall metric reflects the model's ability to capture all elements of the gold standard annotations.

Sample Inclusion Criteria:

Patients have at least three encounters with a specialty provider of the following:

- Addiction Medicine (Psychiatry & Neurology) Physician
- Adult Psychiatric/Mental Health Registered Nurse
- Child & Adolescent Psychiatry Physician
- Geriatric Psychiatry Physician
- Psychiatric/Mental Health Nurse Practitioner
- Psychiatric/Mental Health Registered Nurse
- Psychiatry
- Psychiatry Physician

The sample has a net cumulative of 10,000 sentences within the unstructured clinical notes over a total of 500 patients resulting in 37,662 encounters.

The extracted data included the following information for each patient record:

- Patient Demographics
- Diagnosis History (all)
- Clinical Notes for each Encounter
- Signs and Symptom
- Present Illnesses
- Current Medications
- Past Medical History
- Review of Systems
- Social History
- Family History
- Physical Exam
- Assessment
- Plan
- Other Notes:
- Age or Gender Specific Comments
- Call Back Comments
- Pregnancy Comments

Authors and contributors:

David Newton, Co-founder & Chief Business Officer, Akrivia Health

Greg Folz, CCRP, Co-founder, Cardiovascular Research Consortium, LLC

Manash Sahoo, Chief Data Architect, Cardiovascular Research Consortium, LLC

Elias Kassis, Associate Director, Real World Networks & Partnerships, IQVIA

Julianne Giust, MD, Associate Director of the Mary O'Daniel Stone and Bill Stone Center for Child and Adolescent Psychiatry at Indiana University School of Medicine; Assistant Professor of Clinical Psychiatry, Indiana University

Dr Maxime Taquet, NIHR Academic Clinical Lecturer, University of Oxford

Gloria Roque, AI Product Owner, Akrivia Health

Tarso Franarin, Principal AI Scientist, Akrivia Health

Jack Richmond, AI Scientist, Akrivia Health

Laura Akonor, Marketing Manager, Akrivia Health

Contact us

To learn more about how Akrivia Health, CRC and IQVIA can advance your research, support evidence generation and accelerate clinical trials in psychiatry and CNS please reach out to:

David Newton
Co-Founder & Chief Business Officer, Akrivia Health
david.newton@akriviahealth.com

Dr Benjamin Fell
Director of R&D, Akrivia Health
benjamin.fell@akriviahealth.com

akriviahealth.com

Elias Kassis
Associate Director, Real World Networks & Partnerships, IQVIA
elias.kassis@iqvia.com

iqvia.com

Gregory Folz, CCRP
Co-Founder, Cardiovascular Research Consortium
greg.folz@cardiorc.com

cardiorc.com

For inquiries relating to the EHR data used in this study, please contact:

Steven G. Becker, MD
Associate Dean and Director, Koch Professor of Medicine
Executive Director, Mary O'Daniel Stone and Bill Center for Child and Adolescent Psychiatry
Assistant DIO, Professor of Clinical Radiology and Imaging Sciences
IU School of Medicine, Evansville
sgbecker@iu.edu

iu.edu