METHODS CORE BACKGROUND AND PRELIMINARY STUDIES

As we describe in associated Overall application, development of a learning mental health system will require a robust and efficient national infrastructure for practice-based mental health research, including:

- Rigorous research fully embedded in real-world practice
- Alignment of research goals with priorities of patient, family, and health system stakeholders
- Well-described, large-scale data infrastructure available for rapid analysis
- A culture of trust and transparency to facilitate collaborative learning and improvement

In pursuit of those goals, the Mental Health Research Network (MHRN) has developed a robust national research infrastructure and implemented a diverse program of mental health clinical and services research. We now propose to expand the existing MHRN to include 14 research centers embedded in large health systems serving a combined population of over 25 million in 16 states. Each health system provides mental health and general medical care to a defined patient population, including substantial numbers from all racial and ethnic groups and substantial numbers insured by Medicaid and other insurance programs for low-income members. Each system has organized longitudinal records data into compatible research data warehouses. The MHRN research portfolio now includes over 20 active projects, spanning a range of methods (qualitative research, observational epidemiology, machine learning, pragmatic clinical trials) and clinical or policy topics.

Following the principles of Engaged Scholarship [1, 2], MHRN research focuses on the priorities of patient, family, health system, and policy stakeholders. Rather than seeking use cases for our preferred methods, we seek to identify optimal methods or to develop innovative methods to address stakeholders' needs. Two MHRN projects illustrate this stakeholder-driven methodologic innovation:

- Computational modeling to predict suicidal behavior – MHRN's development of risk prediction tools was directly informed by health system leaders and front-line clinicians. Resulting design decisions included:
  - Selection of distinct real-world cohorts to match specific clinical decision scenarios
  - Inclusion of multiple visits per patient in order to accurately assess within-person changes in risk
  - Use of parametric models to increase transparency and speed electronic health record implementation

  These stakeholder-driven design decisions created computational and statistical challenges, prompting adaptation of machine learning methods to maximize relevance and facilitate rapid implementation.
- Adaptive treatment strategies for depression – This recently-funded project is developing analytic methods for observational studies comparing alternative depression treatments or treatment strategies – with a specific focus on dynamic treatment strategies in which treatment choices depend on favorable or unfavorable responses to prior treatment. This work aims to directly address a high-priority question identified by practicing clinicians and people living with chronic or treatment-resistant depression: How can second- and third-line treatment choices be informed by clinical history, including response to prior treatments? Methods development will use the real-world data typically available to patients and providers at the point of care so that results can directly inform real-world decision support.

Supported by funding from affiliated projects, MHRN investigators have made several additional contributions to research methods in the areas of informatics, design, and analysis. Examples include:

- Machine learning methods for classifying depression symptom trajectories[3, 4]
- Use of modified Zelen designs in pragmatic trials of prevention programs[5]
- Specifications for successful depression treatment outcomes using EHR data (Coley, under review)
- Sensitivity of machine learning suicide prediction models to data availability[6] (also Simon, under review)
- Simulation for sample size estimation in pragmatic trials using EHR data[7]
- Bayesian mixed-effects models to examine provider-level disparities in treatment (Merced, under review)

We now propose to formalize and expand this work, establishing a designated MHRN Methods Core.

METHODS CORE RESEARCH STRATEGY

1) Overview – The MHRN Methods Core will be based in the KPWHRI lead site, supporting scientific and technical personnel in each of the participating research centers. This core will support specific research projects proposed in this application, ongoing affiliated projects, and new projects led by MHRN-based or external investigators. Activities will include an Informatics Unit (continuing highly successful work of the past 8 years) and a newly-formed Scientific Analysis Unit. The Methods Core aims to:

- Support use of optimal existing tools and methods across the full range of MHRN infrastructure activities, core-funded research projects, and affiliated research projects.
- Develop and disseminate new tools and methods to address emerging questions and take advantage of emerging research opportunities

As described below, this work will integrate and expand MHRN expertise in medical informatics, "big-data" predictive analytics, and implementation science.

2) Informatics Unit – Drs. Christine Stewart and Gregory Simon will continue to lead the Informatics Unit, responsible for the specific activities described below. Drs. Stewart and Simon have deep expertise regarding relevant data resources at each site, including standard HCSRN/MHRN Virtual Data Warehouse (VDW) tables, proprietary EHR databases, and other administrative data. Over the past 8 years, the multi-site community of MHRN programmers and analysts has developed strong collaborative relationships, facilitating efficient execution of standard programs and transparent discussions regarding data quality.
   a) Data Quality Assessment – We will continue frequent assessment of data quality in HCSRN/MHRN VDW tables, implementing error detection process developed by FDA Sentinel[8]. Programs developed by Dr. Stewart at the lead site will be executed at all sites, generating detailed reports of counts and rates for data quality indicators. Unexpected differences across sites or across time will prompt systematic exploration, including queries to primary data sources (e.g. insurance claims, EHR databases). Any errors discovered in each site's Extract-Translate-Load processes[9] will be corrected and documented.
   b) Routine Descriptive Analyses – In addition to identifying data quality concerns, the descriptive analyses described above also identify true variations in patterns of diagnosis or care across health systems, population subgroups, or time periods. Published examples from prior MHRN analyses include racial and ethnic variation in depression quality measures[10] and changes in coding of self-harm with the transition from ICD-9 to ICD-10 diagnoses[11]. These descriptive analyses also facilitate engagement with health system leaders and other stakeholders regarding variation in care, improvement opportunities, and priorities for future research. In addition, these analyses provide timely data regarding feasibility of new research projects. Plans in these areas are described in the associated Administrative Core application.
   c) Data Security and Privacy – We will continue existing MHRN standard practices to respect privacy and minimize risk of disclosing confidential information, including:
      - Use of standard methods to extract minimum necessary information from health system records
      - Maximizing use of distributed analytic methods[12, 13] to avoid sharing of confidential information
      - Sharing of data using a password-protected secure file transfer site

      In addition, we propose additional activities during a new funding cycle, including:
      i. Development of guidance regarding assessing and mitigating re-identification risk[14, 15] – Work in this area will include development and dissemination of toolkits for assessing re-identification risk (using quantitative assessment of k-anonymity thresholds and estimation of maximal risk[15-17]) and perturbing data to reduce risk while maximizing scientific value, using the Sdc Micro software package[18].
      ii. Engagement with patient stakeholders regarding benefits and risks of using records data in mental health research – In collaboration with the recently funded PCORI Engagement Award (led by Dr. Coleman), the Informatics Unit will engage with patient stakeholders regarding privacy concerns, including direct involvement of stakeholder representatives in data governance.
   d) Development of New Data Resources – The associated Overall application section describes improvements to data infrastructure accomplished during the current MHRN funding cycle. During a new funding cycle, we propose to focus on four data areas relevant to planned projects and NIMH research priorities. Work in each area will follow the process proven effective over the past 8 years, including:
      - Engagement with health system leaders to improve and standardize data capture in routine care
      - Regular and systematic assessment of data integrity and quality
      - Transparent reporting and documentation of data quality concerns
      - Engagement with relevant stakeholders (clinical leaders, EHR vendors, health system informatics leaders) to improve data quality and consistency

      Priority areas for data infrastructure development during the proposed new funding cycle include.
      - Suicide risk assessment and safety planning – As MHRN health systems implement Zero Suicide prevention programs, we will collaborate with clinical and informatics leaders to:

- o Promote use of standard measures (e.g. Columbia Suicide Rating Scale) for risk assessments
- o Promote use of standard processes (e.g. Stanley-Brown Safety Plan) for documenting safety plans or crisis response plans for patients at high risk of self-harm
- o Develop and disseminate standard tools (e.g. documentation flowsheets, EHR portal questionnaires) for recording those standard measures in health systems' EHRs
- o Develop and implement standard data extraction processes for identifying and organizing data regarding suicide risk assessments and safety plans.

In six MHRN health systems, work in these areas is already underway – supported by the MHRN Zero Suicide Implementation Evaluation project led by Drs. Ahmedani and Simon. We will extend this health system engagement and data infrastructure work to the remaining MHRN health systems.

- Perinatal mental health and prenatal exposures – Large-scale, population-based data regarding maternal mental health, prenatal exposures, birth outcomes, and early childhood development are essential to address a range of research topics, including: effects of maternal mental health on perinatal and developmental outcomes, safety of psychiatric medications in pregnancy, and influence of adverse prenatal exposures on early-onset mental disorders. To address these questions, we will collaborate with health system clinical and informatics leaders to improve quality of data regarding:
  - o Linkage of mother and infant electronic health records
  - o Standardized assessment of anxiety and depression at prenatal visits
  - o Standardized recording of adverse prenatal exposures (alcohol, tobacco, cannabis, etc.)
  - o Standardized recording of early developmental outcomes
- Expanding standardized data tables for patient-reported outcome (PRO) measures – In addition to more standard measures used in active MHRN research (PHQ9, GAD2/7, Columbia Suicide Rating Scale, AUDIT), we will expand PRO tables to other measures of interest to researchers and health system leaders – including measures relevant to pediatric mental health and severe mental illness.
- Adverse experiences and social determinants of health – MHRN health system leaders increasingly recognize the importance of assessing and addressing adverse experiences and social determinants of health. But health systems have not yet implemented standard processes for recording such data. We propose to collaborate with health system leaders and local informatics experts to:
  - o Identify priority constructs (e.g. Adverse Childhood Experiences, Trauma Exposure)
  - o Adopt standard measures
  - o Establish timing and procedures for routine assessment and recording
  - o Produce and deliver timely feedback regarding adherence to recommended procedures

Unique to MHRN, our informatics work includes both improving methods for extracting and translating data from existing records AND direct engagement with health systems to improve and standardize data capture at the point of care. In the previous funding cycle, This engaged approach has created a database of over 14 million PHQ-9 depression assessments across MHRN health systems.

e) Consultation to Core and Affiliated Projects – The Informatics Unit will provide consultation and technical support to the 4 research projects proposed in this application as well as to current and future affiliated research projects. Key areas for consultation and technical support will include:

- Applicability of existing MHRN specifications, code lists, and computable EHR phenotypes
- Procedures for validation of proposed new computable EHR phenotypes
- Tools and procedures for identifying and extracting PRO data from EHRs
- Mapping of text labels to ICD-9 and ICD-10 diagnosis codes in health system EHRs
- Defining denominator or at-risk populations in complex or mixed-model health systems

f) Resource Sharing – The Informatics Unit will share all resources with researchers and operations analysts within and outside MHRN health systems and will support adaptation of MHRN resources to other settings. The MHRN public repository (https://github.com/MHResearchNetwork/MHRN-Central) will continue to serve as the primary channel for resource sharing.

3) Scientific Analysis Unit – Drs. Susan Shortreed and Patrick Heagerty will lead the newly-formed Scientific Analysis Unit. Dr. Shortreed's relevant expertise includes assessment of error and bias in data derived from health records, methods for estimating and comparing adaptive treatment strategies, application of machine learning methods to clinical data and clinical questions, and the design, implementation, and analysis of

pragmatic clinical trials. Dr. Heagerty is Chair of Biostatistics at the University of Washington School of Public Health. His areas of expertise include marginal models and random effects models for longitudinal data, estimation of prediction models, and design of pragmatic clinical trials. This unit will provide consultation to MHRN core and affiliated projects, disseminate best practices regarding analytic methods, with the goal of maximizing the rigor and reproducibility of MHRN research. Targeted methods development will address unmet needs in two high-priority research areas (listed below).

a) Consultation to Core and Affiliated Projects – In preparation for this proposal, we surveyed biostatisticians and data scientists in MHRN research centers to identify areas of expertise as well as priorities for training and resource development. Regarding existing expertise: respondents identified existing expertise in design and analysis of traditional randomized clinical trials, design and analysis of pragmatic trials, design and analysis of stepped wedge trials, design-based and analytic-based approaches to accounting for confounding, analysis of complex longitudinal data - including strategies for addressing missing data, methods for estimating effects of adaptive interventions, and machine learning and prediction modelling approaches. Regarding areas for training or resource development: several survey respondents expressed interest in additional training or mentoring regarding all the specific areas listed above. Most endorsed using the Scientific Analysis Unit for feedback on study design and analytic plans, as an informal mentoring group, and for developing projects advancing statistical methodology. These survey results inform our plan to both take advantage of existing expertise and disseminate more advanced methods to research centers now lacking those resources. To promote optimal use of existing methods, we propose to develop a learning community within MHRN to include:
   - Analytic methods interest group – Following the model used by other MHRN interest groups, an analytic methods interest group will include analysts and biostatisticians from MHRN research centers as well as external collaborators. This group will serve as a forum to identify priorities for methods development, share best practices across MHRN projects, and identify opportunities for new external collaborations. This group will communicate through a monthly teleconference and email listserve.
   - Project-specific consultation – The scientific analysis unit will provide both scheduled and as-needed consultation to MHRN core and affiliated projects to promote use of optimal research designs and analytic methods. While this consultation will be led by Drs. Shortreed and Heagerty, other MHRN biostatisticians and analysts will also provide consultation in areas of expertise. Each project will receive scheduled consultation during proposal development and again prior to initiation of primary analyses. Following Dr. Shortreed's experience supporting MHRN projects and in the NIH Collaboratory, these consultations will be structured as supportive rather than directive or regulatory. This work will both draw from and add to the MHRN resource sharing repository described below.
   - Training opportunities – While this funding opportunity does not support formal training activities, the Scientific Analysis Unit will expand our existing practice of welcoming trainees to participate in MHRN projects and supporting trainees via other affiliated funding. As described in the Overall application, 20 MHRN investigators hold graduate faculty appointments at local academic institutions.

b) Knowledge Management and Resource Sharing– MHRN's resource sharing portal on Github will be expanded to include a public repository for analytic code across MHRN projects. Transparent sharing of detailed analytic methods will both support dissemination of best practices with the MHRN statistical community and promote the rigor and reproducibility of MHRN research. All MHRN-supported projects will be expected to deposit final analytic code at time of manuscript publication.

c) Other expertise – Others contributing to, but not directly supported by, the methods core will include:
   - Jyotishman Pathak PhD, Weill Cornell Medicine, will contribute expertise regarding health informatics via his participation in the Administrative Core and his to-be-funded R01 project with MHRN.
   - Rinad Beidas PhD, Perelman School of Medicine, will contribute expertise in implementation science via involvement in Pilot Project #2 and MHRN's Zero Suicide Evaluation project.
   - Noah Simon PhD, Univ. of Washington Dept. of Biostatistics, will contribute expertise in machine learning and penalized regression via participation in FDA contract 223201810201C regarding effects of medical products on suicidal ideation and behavior.
   - Darren Toh PhD, Harvard School of Public Health Dept. of Epidemiology will contribute expertise in privacy-preserving analysis of high-dimensional data via involvement in FDA Contract 223201810201C.

- Zac Imel PhD, Univ. or Utah will contribute expertise regarding multi-level modeling to identify provider-level effects on racial/ethnic disparities via involvement in the MHRN minority health supplement

d) *Methods Focus Area: Evaluating adaptive treatment strategies*: Treatments for mental health conditions include a wide range of pharmacotherapies, other somatic therapies, and psychotherapies – often provided in sequence or combination. People with more severe or treatment-resistant conditions often follow complex treatment pathways in which poor response or adverse effects with one treatment prompts a treatment change or augmentation. Even when available treatments for a specific condition differ little in average effectiveness, beneficial or adverse effects of treatments may vary widely among individuals. Ideally, selection of any next-step treatment would consider any individual's response to prior treatment exposures, including beneficial and adverse effects (i.e. adaptive treatment strategies). Given the large number of possible treatment sequences, traditional clinical trials are simply unable to identify optimal treatment pathways. The EHR and insurance claims data available in MHRN health systems, including detailed records of treatment exposures and sequential assessments of patient-reported outcomes, provide a unique resource to identify optimal treatment pathways for treatment-resistant depression, bipolar disorder, and psychotic disorders. To take advantage of this remarkable data resource, the MHRN Scientific Analysis Unit brings expertise in patient engagement, medical informatics, psychiatric pharmacoepidemiology, and biostatistics. Development and implementation of optimal analytic methods must be informed by each of these related disciplines. In addition to Drs. Shortreed and Heagerty, other MHRN investigators participating in this work will include Gregory Simon, Christine Stewart, Karen Coleman, and Brian Ahmedani. Specific work in this area will include:
   - Patient stakeholder engagement – Activities described in the Administrative Core will identify treatment choices of highest interest to patients and identify factors influencing "real world" treatment selection.
   - Health system stakeholder engagement – Activities described in the Administrative Core will identify clinical questions of highest interest to front-line clinicians and clinical leaders, especially questions regarding emerging new treatments, such as the glutamate receptor modulator (ketamine-like) drugs.
   - Patient-reported outcome measures – The engagement and informatics work described above will promote more systematic use of current standard outcome measures (PHQ9, GAD7) and promote use of standard outcomes for psychotic disorders as well as common childhood mental health conditions.
   - Analytic methods development: Dr Shortreed recently received funding from the NIMH for an R01 entitled *Improved Tailoring of Depression Care using Customized Clinical Decision Support*. This project will develop, evaluate and disseminate methods for estimating adaptive treatment strategies from EHRs with a focus on improving care for individuals with treatment-resistant depression. This work is complementary to ongoing work within the MRHN, and investigators will be in a unique position to use MHRN data to estimate optimal adaptive treatment strategies for a variety of mental health conditions using the methods developed in Dr. Shortreed's project. The MHRN Statistical Analysis unit will extend work in Dr. Shortreed's R01 regarding accounting for complex time-dependent confounding, handling missing data, and accommodating data measured with error.
   - Integration with other MHRN components – Methods development in this area will both support and be supported by other MHRN components, including: Signature Project #2 (Rapid evaluation of emerging treatments for suicide risk), the Administrative Core Outreach and External Collaborations Unit, and the Methods Core Informatics Unit. This work will also support and be supported by affiliated projects including Dr. Shortreed's R01 and the All of Us consortium led by Dr. Ahmedani.
   - Alignment with NIMH Priorities – This work directly addresses NIMH Strategic Priority 3.2b regarding personalized strategies for sequencing or combining interventions and Strategic Priority 3.2c regarding alternative designs and analytic approaches to test precise interventions

e) Methods Focus Area: Stakeholder-Driven Predictive Analytics: As described above, recently completed MHRN work has used EHR data to develop and validate models predicting suicide attempt and suicide death after mental health outpatient visits. Work in progress will use a larger sample (approximately 30 million visits) to develop improved models, using a wider range of potential predictors and more detailed encoding of predictor timing. Unique among modeling efforts in this area, MHRN work has used a cohort (rather than case-control) design and has developed models accommodating multiple observations (visits) per person. These design choices introduced significant methodologic and computational

challenges, requiring adaptation of standard machine learning methods. But these choices were necessary to address stakeholder priorities regarding predictive analytics. Choosing a cohort design focused on specific visits as prediction occasions addresses stakeholder requests for accurate prediction of absolute risk (rather than relative risk) at time of a specific visit. Including multiple observations per person addresses stakeholder requests to accurately estimate increases and decreases in risk within individual between visits. Primary modeling focused on parametric methods (penalized lasso logistic regression) rather than tree-based methods (such as random forests) or multi-level models (such as Bayesian neural networks). While more complex model-fitting methods could improve accuracy of prediction, the resulting models cannot be readily implemented within widely used commercial EHR platforms such as Epic or Cerner. This interplay between technical and practical aspects of predictive analytics illustrates a core principle guiding current and future MHRN work in this area: Decisions regarding design and modeling methods must be driven by stakeholder requirements and potential for rapid implementation. To accomplish this balancing of analytic and implementation priorities, the MHRN Scientific Analysis Unit brings a unique range of expertise regarding health system engagement, quality improvement, implementation science, medical informatics, and machine learning. Development and implementation of stakeholder-driven predictive analytics must be informed by each of these disciplines. In addition to Drs. Shortreed and Heagerty, other MHRN investigators working in this area will include Gregory Simon, Rebecca Yates Coley, Jean Lawrence, and Stacy Sterling, and Bobbi Jo Yarborough. Specific new work will include:

- Health system engagement – The engagement activities described above will support ongoing consultation with health system leaders regarding implementation of existing suicide risk prediction models, priorities for extension or expansion of suicide risk prediction models, and priority areas for development of new predictive analytic tools.
- Patient/consumer engagement – The engagement activities described above will support ongoing consultation with patient and family stakeholders regarding protection of privacy in development of predictive analytic tools and appropriateness of algorithm-driven outreach efforts.
- Engagement with EHR vendors – The engagement activities described above will support ongoing consultation with EHR vendors regarding the capability of existing and anticipated EHR products to accommodate complex temporal relationships and implement non-parametric prediction algorithms.
- Development and evaluation of modeling methods - Ongoing and future work will compare non-parametric machine learning approaches (e.g. random forests, neural networks) with more familiar statistical learning approaches (e.g. logistic regression with lasso for model selection). We will continue to improve upon existing predictive models and analytic modelling strategies, aiming to balance the desire for more accurate predictions within homogenous identified risk sets with easily computable models that can be implemented and updated in real-time in the EHRs of real-world health systems. Current collaboration with University of Washington Assistant Professor of Biostatistics, Dr Noah Simon, will implement convolutional neural networks to develop deep-learning risk prediction models, which will leverage information about the severity of mental health condition as well as the temporal patterns of prior care received to better identify when an individual is at increased suicide risk. We will compare the performance of convolutional neural networks, random forests, and lasso-based logistic regression to assess which approach provides a pragmatic approach to implementing suicide risk prediction models in health systems to improve mental health care.
- Integration with other MHRN components – Methods development in this area will both support and be supported by several other components, including Pilot Project 1 regarding implementation of risk prediction models, the Administrative Core Outreach and External Collaborations Unit, and the Methods Core Informatics Unit. This work will also support and be supported by FDA Contract 223201810201C regarding effects of medical products on suicidal ideation and behavior.
- Alignment with NIMH Strategic Priorities – The proposed work directly addresses NIMH Strategic Objective 3.3c to evaluate how patient, provider, and organizational factors affect intervention effectiveness and Strategic Objective 4.2 to establish research-practice partnerships to disseminate and implement evidence-based services.