## Title
Rationale and Design of the National Neuropsychology Network



## Journal
Journal of the International Neuropsychological Society, 28(1)

## ISSN
1355-6177

## Authors
Loring, David W
Bauer, Russell M
Cavanagh, Lucia
et al.

## Publication Date
2022

## DOI
10.1017/s1355617721000199

Peer reviewed

# Rationale and Design of the National Neuropsychology Network

David W. Loring,[1,2] Russell M. Bauer,[3] Lucia Cavanagh,[4] Daniel L. Drane,[1,2]

Kristen D. Enriquez,[4] Steven P. Reise,[5] KuoChung Shih,[4]

Laura Glass Umfleet,[6] Dustin Wahlstrom,[7] Fiona Whelan,[4] Keith F. Widaman,[8] Robert M. Bilder[4]

For the NNN Study Group *

[1] Department of Neurology, Emory University School of Medicine

[2] Department of Pediatrics, Emory University School of Medicine

[3] Department of Clinical & Health Psychology, University of Florida

[4] Department of Psychiatry & Biobehavioral Sciences, University of California, Los Angeles

[5] Department of Psychology, University of California, Los Angeles

[6] Department of Neurology, Medical College of Wisconsin

[7] Pearson Clinical Assessment

[8] Graduate School of Education, University of California, Riverside

Correspondence: David W. Loring
Emory Brain Health Center
12 Executive Park
Atlanta, GA 30329-2206
Email: dloring@emory.edu
https://orcid.org/0000-0003-2497-1074

**Abstract**

The National Neuropsychology Network (NNN) is a multi-center clinical research initiative funded by the National Institute of Mental Health (NIMH; R01 MH118514) to facilitate neuropsychology's transition to contemporary psychometric assessment methods with resultant improvement in clinical test validation and assessment efficiency. The NNN includes four clinical research sites (Emory University; Medical College of Wisconsin; University of California, Los Angeles (UCLA); University of Florida) and Pearson Clinical Assessment. Pearson Q-interactive (Q-i) is used for data capture for tests that it publishes, whereas UCLA programmed web-based data capture tools for other measures and serves as the Coordinating Center. The NNN is acquiring item-level data from 500-10,000 patients across 47 widely used NP tests and sharing these data via the NIMH Data Archive (NDA). Modern psychometric methods (including item response theory) will specify the constructs measured by different tests and determine their positive and negative predictive power regarding diagnostic outcomes and relationships to other clinical, historical, and demographic factors. The Structured History Protocol for Neuropsychology (SHiP-NP) helps standardize acquisition of relevant history and self-report data. The NNN is a proof-of-principle collaboration: by addressing logistical challenges, NNN aims to engage other clinics to create a national and ultimately an international network. The mature NNN will provide mechanisms for data aggregation enabling shared analysis and collaborative research. NNN promises ultimately to enable robust diagnostic inferences about neuropsychological test patterns, and to promote the validation of novel adaptive assessment strategies that will be more efficient, more precise, and more sensitive to clinical contexts and individual/cultural differences.

*Keywords:* Collaborative neuropsychology, psychological tests, clinical decision-making, modern psychometric methods

**Rationale and Design of the National Neuropsychology Network**

Neuropsychological practice typically involves manual administration of paper and pencil tests using methods and techniques developed during the mid-20th century, with some tests having historical roots in the 19th century (Bilder & Reise, 2019; Boake, 2000). Transition to newer methods that leverage the multiple advantages of computer-assisted testing has been limited despite a recognized need for method modernization (Marcopulos & Lojek, 2019) and despite the interest in computer-assisted testing associated with telehealth spawned by Covid-19 pandemic (Bilder et al., 2020a, 2020b; Hewitt, Rodgin, Loring, Pritchard, & Jacobson, 2020; Postal et al., 2020). Many computerized neuropsychological tests were designed to generate results comparable to their paper and pencil counterparts and are direct adaptions of those measures (e.g., Wisconsin Card Sorting Test, California Verbal Learning Test – Third edition). Although multiple cognitive assessment protocols have been developed specifically for computerized testing (e.g., CNS Vital Signs, Cogstate, CANTAB, ImPACT, NIH Toolbox Cognitive Battery), most of these procedures do not fully satisfy published standards for computerized testing (Bauer et al., 2012). More importantly, newer measures have not been optimized using advanced psychometric techniques to enhance test validity, increase efficiency, guide differential diagnoses, or suggest appropriate clinical recommendations.

Neuropsychology is criticized because of its lengthy testing sessions (Teng & Manly, 2005), which particularly disadvantage patients at-risk for fatigue (e.g., Parkinson Disease, Multiple Sclerosis). Lengthy assessment protocols further limit access because fewer patients can be tested each day, and long wait-times for appointments decrease timeliness and detract from patient care. Clinical test

validation has also been limited by variability in diagnostic criteria and selection biases associated with samples of convenience in which the base rates of neuropsychological performance patterns are unknown (Pawlowski, Segabinazi, Wagner, & Bandeira, 2013). Thus, many "rules of thumb" for diagnostic decision making have not been adequately validated across the spectrum of clinical conditions that may be referred for neuropsychological evaluation (Chaytor & Schmitter-Edgecombe, 2003; Duff, Suhrie, Dalley, Anderson, & Hoffman, 2019; Hoogland et al., 2018; Raspall et al., 2005).

The National Neuropsychology Network (NNN) is a multi-center clinical research initiative funded by the National Institute of Mental Health (NIMH; R01 MH118514). NNN was designed to facilitate neuropsychology's transition to contemporary psychometric assessment methods with resultant improvement in clinical test validation. Although previous collaborations have been successful in generating important clinical research findings (e.g., Bozeman Epilepsy Consortium, Loring, 2010), the absence of independent funding has not permitted the infrastructure development that is essential for uniform data acquisition and quality control nor the support for expertise in advanced psychometrics and test construction. The NIMH agreed to support a 5-year award (2019-2024) to achieve these objectives using the Multiple Program Director/Principal Investigator (PI) option (Overall PI: Robert M. Bilder; Clinical Site PIs: Russell M. Bauer, Daniel L. Drane, David W. Loring, Laura Glass Umfleet; Non-clinical Site PI: Dustin Wahlstrom).

The NNN collects item-level data on multiple neuropsychological measures to identify the most informative individual items that characterize relevant neurocognitive constructs. Item-level data analysis enables more efficient

assessment methods to be developed by applying modern psychometric analyses that either eliminate redundancies or specify adaptive strategies to efficiently answer diagnostic questions. These results are then used to establish robust estimates of positive and negative predictive power within relevant neurobehavioral domains. Derived NNN results will be rigorously characterized psychometrically and subjected to robust clinical validation to establish their incremental diagnostic validity over conventional approaches. NNN data will also be analyzed to identify differential item and test functioning to minimize inequities due to racial, ethnic, linguistic, and economic factors that may influence neuropsychological test performance (Zahodne, Manly, Smith, Seeman, & Lachman, 2017).

Although the need to improve existing measures or develop novel NP methods has long been recognized, clinicians are comfortable with existing measures and often reluctant to embrace new assessment methods. Consequently, NNN adopted an incremental approach to influence test usage starting with the "usual suspects" (Curtiz, 1942) to characterize how legacy measures perform in tightly characterized clinical environments, and then based on initial analyses, make recommendations for test modification and novel assessment method development. The initial database thus includes standard neuropsychological measures that are currently in widespread use (Rabin et al., 2016). Although most NP tests have either some psychometric or clinical validation, few data exist on how these measures perform when combined with other tests. The analytic approach used by NNN will address: (a) how do neuropsychological test findings, individually and in combination, provide unique information to establish a post-assessment diagnosis (i.e., by providing unique information relative to the pre-examination/referral diagnosis); and (b) what specific test items within

established measures are especially informative (and which are not) in this process. Following successful item-level analyses of existing neuropsychological tests, we will gradually implement cross-network data collection from a limited number of novel or experimental tests that have been developed using voice, video, and other unique formats. As additional clinical sites are added to NNN, rapid collection of validation data for new tests will be facilitated by relying on enrollment from multiple sites.

NNN has three specific aims: 1. Establish Network Infrastructure; 2. Data Collection and Deposit; and 3. Data Analyses (see supplemental material for complete specific aims as stated in the application). Creating appropriate network infrastructure is critical not only for project execution but also to provide the foundation for NNN expansion to new clinical sites that will permit data capture from larger and more diverse clinical settings. Included in infrastructure development is implementation of the Pearson Q-interactive (Q-i) system across existing NNN sites, and design and development of point-of-testing digital response capture of individual item responses for non-Q-i measures. Item-level data including response times are contributed to the NIMH Data Archive (NDA) and will comprise the largest single source of NP data at the item-level, broadly facilitating data analyses beyond the boundaries of the NNN project.

The NDA is a resource developed by NIMH to promote open access and use of shared information to accelerate scientific progress (see https://nda.nih.gov/about/about-us.html). Originally developed to support autism research, the NDA integrates several existing data repositories including the National Database for Autism Research (NDAR), the Research Domain Criteria Database (RdoCdb), the National Database for Clinical Trials related to Mental

Illness (NDCT), and the NIH Pediatric MRI Repository (PedsMRI). By placing data in the NDA, NNN data are available to researchers worldwide permitting additional independent analyses of archived NNN material. Many papers already have emanated from archival NDA data (e.g., Human Connectome Project, ABCD datasets) and we anticipate that NNN data will contribute to a similar trajectory of archival data analysis.

**NNN Neuropsychological Test Selection**

Data are obtained from the most frequently administered tests and are representative of national assessment trends (Rabin, Paolillo, & Barr, 2016). The clinical sites (Emory University, Medical College of Wisconsin, University of Florida, and UCLA) were selected since they represent geographic diversity, are nationally recognized programs, involve multiple board-certified clinical faculty, and have established track records of collaborative clinical research. Several NNN investigators have also been involved in formal test development initiatives learning the difficult lesson that funding, infrastructure, and dedicated/protected research time are essential elements for project success. These institutions provide neuropsychology training at the practicum, internship, doctoral, and postdoctoral levels and are well-positioned to influence practices and expectations of emerging neuropsychologists. Clinical sites are expected to enroll more than 10,000 participants spanning diverse neuropsychiatric and neurologic diseases/syndromes and to compile item level data on 500+ clinical cases for nearly 50 common neuropsychological measures during its 5-year NIMH funding period.

The NNN established a formal collaborative relationship with Pearson Clinical Assessment, the publisher of many of the most widely used NP tests (Rabin et al.,

2016). The NNN leadership plan includes Pearson in all project discussions although Pearson does not have voting rights in the governing board, which consists of leaders of the four clinical sites and an independent external advisor (Robert Heaton). The Pearson Q-i platform captures item-level responses on Pearson tests including measures of general cognition (Wechsler Adult Intelligence Scale-Fourth Edition; WAIS-IV), memory (Wechsler Memory Scale-Fourth Edition; WMS-4, California Verbal Learning Test-Third Edition; CVLT-3), executive function (Delis-Kaplan Executive Function System), and brief neuropsychological assessment screening (Repeatable Battery for the Assessment of Neuropsychological Status). The Q-i platform employs iPads for both stimulus presentation and for examiner recording of responses,

Because non-Pearson tests (e.g., Boston Naming Test) lacked a point-of-testing data entry capacity, the UCLA Semel Institute Biostatistics Core (SI-Stat) developed web-based data capture applications for these measures. Agreements to program data-entry applications have been executed with test publishers permitting appropriate per-use royalty payment for clinical test usage. Pearson Clinical Assessment is providing Q-i use for the NNN initiative as part of the research collaboration with support from NIMH. Neuropsychological tests included in NNN are listed in Table 1.

**Structured History Protocol for Neuropsychology (SHiP-NP) and Other Common Data Elements (CDEs)**

The Structured History Protocol for Neuropsychology (SHiP-NP) is a standardized history protocol developed to harmonize clinical data collection across NNN sites to facilitate data analysis including demographics, medication use, and

medical/psychiatric history. Demographic data elements are based on the PhenX Toolkit project (Hamilton et al., 2011), which developed consensus measures for "Phenotypes and eXposures", including age, race, ethnicity, sex, gender, marital status, educational attainment, annual family income, and child-reported parental educational attainment. Medical history, family history, and concomitant medications are recorded based on the NINDS Common Data Elements (CDE) conventions (Grinnon et al., 2012), and medications associated with increased risks for adverse cognitive effects (e.g., narcotics, benzodiazepines, anticholinergics, sedatives/hypnotics, and selected anti-seizure medications) are highlighted. The SHiP-NP includes a standardized assessment of features that may influence neuropsychological findings including developmental history, academic performance, legal and military history, mental health treatment, and social health determinants (e.g., financial strain). Appropriate follow-up questions are available for specific medical conditions (e.g., epilepsy, traumatic brain injury, stroke, cancer) to characterize symptom presentation, duration, and treatment history. In response to the COVID-19 pandemic, novel coronavirus exposure and associated distress queries were added that were obtained from the Montreal Behavioral Medicine Centre (https://mbmc-cmcm.ca/covid19/; Lavoie & Bacon, 2020). Additional clinical common data elements are linked to the SHiP-NP (see section below: **Clinical Interview**).

The SHiP-NP is completed on-line prior to the clinical appointment via a secure website hosted by SI-Stat, although a paper and pencil SHiP-NP version can be used by patients without internet access. The SHiP-NP relies on branching logic and forced response options to gather relevant information while minimizing overall assessment burden. SHiP-NP data are collected and stored securely on a secure

UCLA Si-Stat server without Protected Health Information (PHI), which is available only to patients and their clinicians. The SHiP-NP generates patient data sheets and narrative history text files that can be modified based upon specific needs of each clinical site and are easily integrated into clinical reports. For patients unwilling or unable to complete the SHiP-NP, NNN sites are collecting a minimal demographic dataset that includes age, gender, education, and handedness, with secondary questions documenting marital status, employment, and languages spoken other than English. The first 135 patients completing the SHiP-NP spent approximately 22 minutes completing the form (*Mdn* =22.4 min; *Range*=4.4–56.0 min).

**Milestones Achieved During the Initial Project Period**

The initial months of the project (March 2019 – August 2019) were dedicated to logistical problem-solving associated with collaboration across sites using different models of clinical service delivery, and addressed training and quality control, data capture, and options for obtaining participants' consent. Although NNN uses the SMART IRB National IRB reliance system, it was still necessary to obtain agreement from site IRBs prior to beginning subject enrollment. Although the Q-i platform allows individual item data capture for Pearson tests, the initial funding period was also dedicated to the development of individual item data capture for non-Q-i measures and to establish appropriate mechanisms for transfer of data to the NDA.

**Institutional Review Board**

UCLA serves as the IRB of Record, with all participating sites relying on the SMART IRB agreement (NIH's National Center for Advancing Translational Sciences (NCATS) Streamlined, Multi-site, Accelerated Resource for Trials (SMART) IRB

Reliance Platform). Future NNN sites will be able to sign on to SMART IRB and join this agreement, and if ineligible for SMART IRB, an alternative appropriate IRB agreement will be negotiated. This process coordinates, collects, and verifies information including: a) local context; b) site variations in areas such as recruiting, informed consent, HIPAA, populations; c) conflict of interest disclosure and management; d) completion of ancillary reviews; e) training and qualifications of study team; f) continuing review or closure information; and g) reportable events such as protocol deviations or adverse reactions. In response to the 2020 COVID-19 outbreak and the emergence of telehealth, NNN was first granted IRB approval to obtain verbal informed consent for study participation, and subsequently, waiver of informed consent was approved.

**Participant Recruitment**

All adult English-speaking patients referred for neuropsychological evaluation are considered potential NNN participants. After study participation is established, a Global Unique Identifier (GUID) is assigned (for details about GUIDs, please see https://nda.nih.gov/s/guid/nda-guid.html). The GUID enables research projects to share individual participant data without risk of exposing Protected Health Information (PHI). NP and SHiP-NP data are transferred to the NDA without PHI using only the GUID for subject identification. An NIH generated a GUID requires PHI (including date of birth, place of birth and middle name), and when that information is not available, a “pseudo-GUID” is generated. A pseudo-GUID has same structure as the GUID, but is a randomly generated ID that can only be used to identify an individual patient by the site that generated the pseudo-GUID (unlike the GUID, it cannot be used to coordinate participation across multiple independent research studies).

Given the negligible risks to patients of participation in the NNN, that clinical practice remains “as usual,” and that no PHI is included, we applied for and received approval for a waiver of informed consent from the UCLA IRB. This will allow new clinical sites to participate more easily since they will be able to use a GUID or pseudo-GUID to identify subjects without transmitting any PHI to the NNN database (and subsequently to the NDA). If the clinic does not enter a GUID or pseudo-GUID, the NNN registry will provide a pseudo-GUID and each clinic alone will have records of the relationship between the GUID and their patient’s data. This will enable clinics anywhere to: a) send a link to their patient so the patient can complete the SHiP-NP online; b) use the GUID or pseudo-GUID with Q-interactive, which will enable Pearson Clinical Assessment to send Q-i raw data (without identifiers) to NNN; and c) use the NNN point-of-testing data acquisition software (when available) to collect item-level data on other non-Q-i tests. In return, participating clinics will receive: a) SHiP-NP results, including history and other results of self-report scales from the common data elements (CDEs), in both tabular and narrative report forms; b and c) an integrated data summary sheet combining all Q-I and non-Q-I NP data elements, with normative references selected by clinic users, to create custom reports.

We initially aimed to enroll approximately 50 cases weekly across our network to reach 10,000 cases total in the database. Soon after starting enrollment on 7-31-2019, we reached this enrollment target. We experienced a sudden drop in recruitment in March 2020 at the onset of the COVID-19 pandemic, but since then as our clinics have returned to different models of practice, we have again reached our target recruitment levels and expect that to be surpassed in the coming months

(see Supplemental Figures 1a and 1b). Examinee characteristics are provided in Supplemental Table 1.

**Training and quality control**

Each site has a quality control officer, and each person performing NNN assessments is certified at both the test administration and Q-i assessment interface levels. The NNN has a library of training materials, with support from Pearson Clinical Assessment for Q-i training tools, and the NNN maintains a database documenting training outcomes for all personnel involved in data collection.

**Data Acquisition, Storage, and Transfer**

NNN protocols transmit only data without PHI to Pearson Clinical Assessment, and data received back from Pearson or SI-Stat is based upon assigned GUID or pseudo-GUID only. Thus, each site creates linking tables to connect GUIDs to other identifiers in their own clinical records.

Q-i data transmitted to Pearson Clinical Assessments are tagged based upon site-specific logins of de-identified data using GUID, which is also how data are stored in the Pearson database and subsequently transmitted to the NDA. Pearson executes their usual workflow to score the obtained information, creating data files for the sending site as is their usual reporting standard on Q-i. In parallel, Pearson sends complete data, including all trial timing and individual response selection variables, to the NDA. The test results and normative information generated by Pearson are provided back to the clinic sites in an Excel spreadsheet which facilitates preparation of the formal clinical report, a process that typically takes less than 30 minutes after data transmission to Pearson. UCLA Si-Stat is

responsible for data aggregation with non-Q-i measures and for transfer to the NDA. We continue to resolve issues related to the highly specific data elements that are enabled by the Q-i outputs, which not only include item-by-item scores and timings but also individual examiner annotations as image files (e.g., locations of each block in individual block designs). The complete data dictionary will be accessible on our website and from the NIMH Data Archive.

**Clinical Interview**

The NNN has adopted conventions for data collection for contemporary diagnostic and clinical status information while allowing individual clinicians to conduct interviews following their current practices. While diagnostic assessments will vary, specification of both pre-assessment and post-assessment diagnoses follows ICD-10-CM and is harmonized with the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) codes for psychiatric disorders. The same online platform developed for the SHiP-NP is used to acquire other Common Data Elements (CDEs) including multiple dimensional self-report rating scales. These CDEs were recommended by an NIMH workgroup (Barch et al., 2016) and include the DSM-5 Level 1 Cross-Cutting Symptom measure, a 23-item self-report that assesses 13 domains (depression, irritability, anxiety, mania, somatic symptoms, suicidality, psychotic symptoms, insomnia, memory problems, compulsions, derealization/depersonalization, personality functioning, and substance use disorders) (Available at: https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/assessment-measures). Participants who screen positive on the Level 1 assessment also receive the relevant DSM-5 Level 2 measures, which include the Patient-Reported Outcomes Measurement Information System (PROMIS) short-forms for

Depression, Anxiety, Anger, and Sleep Problems and self-report ratings of Mania, Substance Use, Repetitive Thoughts and Behaviors, and Somatic Symptoms. Finally, there is a Clinician-Rated Dimensions of Psychosis Symptom Severity, because this domain was determined to be unreliable by self-report. Many patients with neurologic referral diagnoses are not expected to screen positive on the Level 1 assessment, although this screening will permit detailed characterization of psychiatric comorbidities in chronic neurologic diseases such as epilepsy or Multiple Sclerosis or degenerative conditions such as Alzheimer's Disease or Parkinson's Disease.

**Dimensional ratings of everyday functioning and disability**

Disability ratings are provided by the 36-item self- and informant-reported World Health Organization Disability Assessment Schedule 2.0 (WHO-DAS 2.0; Available at: https://www.who.int/classifications/icf/whodasdownloads/en/ ), which are also incorporated as part of the SHiP-NP/CDE on-line platform. The WHODAS 2.0 follows the theoretical framework of the International Classification of Functioning, Disability and Health (ICF), permitting it to be used worldwide across all health conditions. Additional ratings of current functioning and quality of life for non-overlapping domains is gathered using validated short forms of the Quality of Life in Neurological Disorder (Neuro-QoL) battery (Cella et al., 2012), comprising a total of 65 additional items for Emotional and Behavioral Dyscontrol; Fatigue; Lower Extremity Function – Mobility; Positive Affect and Well-Being; Satisfaction with Social Roles and Activities; Sleep Disturbance; Stigma; and Upper Extremity Function - Fine Motor, Activities of Daily Living (ADL). These are all programmed in the SHiP-NP/CDE web application, with data collected either prior to each participant's visit

via their own web-enabled devices, or on site using one of the sites' iPads or other data entry devices.

**Data Analyses**

Initial NNN analyses will apply item response theory (IRT) to define more precisely the constructs measured by each test and how these constructs can be assessed more efficiently. Most IRT models are expected to be unidimensional, although multidimensional IRT (mIRT) models will be explored as appropriate. Following dimensionality determination, efficiency optimization will be explored by specifying fixed short forms, Computerized Adaptive Tests (CATs), or multidimensional adaptive tests (MATs). IRT application has led to efficiency gains of 50-95% relative to conventional method for many measures (Choi, Reise, Pilkonis, Hays, & Cella, 2010; Gibbons et al., 2008; Moore et al., 2015). IRT can also characterize the same construct using different item sets, which is particularly useful in longitudinal assessments to ensure that inclusion of a newly derived measure is back-compatible with earlier test versions that may have been used in previous testing, and to enable repeated measurement without using previously exposed items.

As a proof of principle demonstration that modern psychometric methods applied to existing measures can improve test efficiency, we applied IRT to WAIS-IV Matrix Reasoning (MR) data in a group of 549 NNN participants (Reise et al., 2021). The mean MR raw score was 15.4 (*SD* 5.6) and the mean MR scaled score was 9.9 (*SD* 3.2), similar to the standardization sample. The first 5 MR items were completed without error by 97% of the subjects, adding little to the measurement of the MR latent trait. The most difficult MR items also contributed little information at ability

levels greater than 1.5 SD above the test mean. Finally, there was considerable overlap in item information across the remaining MR items suggesting that a short form or adaptive version of MR could provide results with precision comparable to the full MR subtest. Using the standard administration start and stop rules, more than half of our sample (282/549 or 51%) were administered 23/26 MR items to obtain their total score. In contrast, a simulated computerized adaptive test (CAT) with only 10 MR items was almost perfectly correlated with the ability level estimated from the entire 26-item test ($r$ = .99). This represents more than a doubling of assessment efficiency.

These analyses can be further extended using the *nominal or graded response models*. In brief, the traditional analyses of the MR responses uses dichotomization, considering only whether each item was completed correctly or incorrectly. But because each response is chosen from a multiple-choice array that includes five (5) response options, each distractor can be considered independently as an indicator of ability on either the primary latent trait or on other factors that might be identified following data analysis. In the nominal response model, each response is considered independently, while in the graded response model, the response alternatives can be ranked on an ordinal scale from those that are "closest" to those that are most "distant" from the correct choice. For example, selection of a "close" response option that matches on some but not all item dimensions might be more frequent among those with higher ability, while selection of a "distant" option that does not follow the expected dimensions might be more frequent among those with lower ability or who may be following an unusual response strategy. This approach to response analysis may have special value in performance validity testing in which selection of particularly "distant" choices

might reveal a marked deviation from the best estimates of that individual's true abilities suggesting possible intentional response deficit exaggeration. Indeed, this method examining the relationship of individual responses to estimates of true ability (i.e., person-fit statistics) should allow the design of empirically derived embedded performance validity tests (PVTs) for many measures.

As part of quality control, data analyses will exam measurement invariance in an IRT framework to determine whether putatively identical tests behave comparably across sites, diagnostic boundaries, and groups defined by demographic characteristics. These analyses will establish to what extent data can be combined for subsequent analyses, and to what extent unique scoring and interpretation may be indicated for different groups defined by age, sex, education, racial/ethnic, linguistic, or cultural backgrounds. Novel methods for variable harmonization developed for the Whole Genome Sequencing in Psychiatric Disorders consortium (U01 MH105578) will be employed to combine data that were acquired using different instruments in different samples (with no overlap of items across samples, and no patient receiving both instruments) (Mansolf et al., 2020). This method analyzes the goodness-of-fit of correlations among variables to determine if two different variables are sufficiently similar to include in pooled analyses. Items are selected that best match items in another test based on the strengths of their correlations with all the other items in their respective test. When the loss function asymptotes, this process stops yielding the items that are considered equivalent across datasets.

Data analyses will include a combination of confirmatory and exploratory factor analyses. The degree to which alternate models may be more appropriate to our clinical samples, or specified using different tests, will be rigorously explored

and analyses will determine to what extent the additional tests measures add to or modify that structure. Alternate models using confirmatory factor analysis will establish how various models differ with respect to efficiency (i.e., assessment time) by determining the degree to which goodness of fit deteriorates with fewer variables or by substituting short-form test scores for long-form results. This permits the creation of short form tests, new short form batteries, and CATs to increase efficiency to measure the same constructs or arrive at the same diagnostic conclusions. We will also characterize short form use across different clinical diagnostic referrals since, for example, a short form assessment of confrontation naming may be appropriate in patients with referral diagnoses of Multiple Sclerosis but not for epilepsy patients referred for surgical evaluation. Thus, NNN will provide the evidence-base that enables multidimensional adaptive testing to maximize diagnostic information while minimizing burden to patients and facilitating new large-scale collaborative research projects (e.g., as in genomics and population behavioral health).

**Clinical Validation of Neuropsychological Procedures**

By employing adaptive assessment techniques, assessment protocols may eventually be personalized according to referral diagnoses and patient performance patterns in which both item and test selection maximizes the predictive power by sequentially selecting the most informative next item within a test, and the most informative next test within a battery. When specific test findings are obtained, prior probability estimates are updated, and the next most informative measure is selected. This process continues until desired levels of precision are obtained with respect to the outcomes of interest. With adaptive testing, assessment protocols

are not administered in the same pre-determined sequence, but rather change to the specific clinical context and adjust dynamically according to task performance.

The combination of tests, administration order, and all available sources of information are included to establish the best combination for accurate prediction, whether it be post-assessment diagnosis or risk of adverse treatment outcome. These analyses will generally use multinomial logistic regression (MLR) models if there are more than two categories, or logistic regression for cases with only two classes.

*Mild and Major Neurocognitive Disorder*. A common neuropsychological referral question is whether there is cognitive decline that exceeds that expected with normal aging, and if so, whether the pattern is consistent with a specific underlying etiology. Following DSM-5 nomenclature, three classes of neurocognitive function are characterized for all patients: No Neurocognitive Disorder (No NCD), Mild NCD, or Major NCD. This classification is applied for all NNN participants, not simply those referred due to age-related cognitive or memory concerns (e.g., includes Multiple Sclerosis, traumatic brain injury, epilepsy, stroke, etc.). his approach reflects the growing recognition that consistent application of cognitive taxonomies for all clinical conditions is necessary to better characterize and enable comparison across disease entities (e.g., presence of single vs. multi-domain impairments, natural history of disease, relationship to sociodemographic and psychological variables (Norman et al., Epilepsia 2020).

The inclusion of all diagnostic referrals will also help address the relationship of mild NCD and psychopathology. For example, the DSM-5 diagnosis of Mild NCD has been associated with more “anergia” and “observed slowness” while the

Petersen Mild Cognitive Impairment criteria have been more associated with neuro-vegetative symptoms and dysphoric mood (Lopez-Anton et al., 2015). Thus, there remains a major gap in understanding precisely how quantitative neurocognitive evidence, evidence of disruption in instrumental activities of daily living, and evidence of non-cognitive psychopathology (particularly mood and anxiety symptoms) all contribute to the ultimate diagnosis of Mild and Major NCD.

Multinomial logistic regression analyses will include estimated premorbid ability, neuropsychological performance in multiple discrete domains, psychopathology symptom ratings from the DSM-5 and PROMIS measures, and level of everyday functioning as measured by the WHODAS 2.0 and Neuro-QOL. Although this classification system does not assist in establishing diagnostic specificity regarding the presumed etiology of cognitive impairment, it provides a classification nosology to facilitate cross disease comparison that includes the important IADL component. The results will identify the relationship among premorbid, objective neurocognitive, and psychopathological features that contribute to impairment of everyday functioning and are associated with different dementia outcomes. The NNN final sample is anticipated to have sufficient sample size to characterize prediction of Alzheimer's Disease, Vascular Cognitive Impairment, Parkinson Disease Dementia, Dementia with Lewy Bodies, and mixed dementia syndromes.

*Epilepsy Lateralization***.** Epilepsy surgery candidates will be analyzed to characterize the diagnostic sensitivity of neuropsychological findings to confirm seizure onset lateralization and localization, particularly in patients with temporal lobe epilepsy. The primary approach will examine epilepsy patients who are candidates for surgical resection/ablation of the temporal lobe determined to be left hemisphere language dominant, although patients with mixed or right cerebral

language dominance provide a special subset to explore hypotheses related to cerebral “crowding” and other features of potential cerebral reorganization (Loring et al., 1999; Strauss, Satz, & Wada, 1990). Some data suggest greater risk of decline with atypical (particularly bilateral) language lateralization, as language functions tend to be complimentary across hemisphere rather than “redundant” processes (Drane & Pedersen, 2019)

The final consensus surgical decision is informed from results of ictal and interictal EEG recordings, PET, and MRI, and this serves as the gold standard for criterion classification. Typically, neuropsychological test results are considered confirmatory data, and when these results are inconsistent with other primary sources, patients are considered at greater risk of adverse cognitive outcome. This can lead to decisions to obtain additional data (e.g., Wada testing) to predict risk of cognitive decline (particularly to rule-out an elevated risk of amnesia). NNN analyses will examine neuropsychological predictors in logistic regression models and examine positive and negative predictive power of each test and of test combinations to classify patients into the four seizure groups - left TLE, right TLE, bilateral TLE, and extra-temporal seizure onset for patients experiencing seizure freedom following surgery. Predictors will include all leading verbal and nonverbal learning and memory tests, selected language measures, and site-specific measures that we can examine using our variable harmonization strategies. Secondary analysis will involve finer-grain localization of seizure onset based upon more precise regional differences in structure-function relationships (e.g., the temporal pole is more associated with proper nouns than common nouns as assessed with both category-related visual confrontation naming tasks and verbal

semantic fluency measures; Abel et al., 2015; Drane et al., 2013; Drane & Pedersen, 2019).

Epilepsy patients will also be examined as an initial group to evaluate possible reduction in test administration length. The BNT is sensitive to seizure onset lateralization in TLE patients and may be superior in diagnostic sensitivity than material-specific performance discrepancies of verbal learning and memory (Busch, Frazier, Iampietro, Chapin, & Kubu, 2008). In the standard administration, presentation begins with item 30 although anecdotal clinical evidence suggests that earlier items are sensitive to both seizure onset laterality effects and may demonstrate post-operative decline following open resection of the anterior temporal lobe (Strauss et al., 2000). In addition, there are also likely items that are unfamiliar to younger patients compared to when the test was initially published in 1978 (e.g., yoke, compass) making IRT analysis a particularly valuable approach to improve BNT validity and improve testing efficiency.

*Psychiatric Contributions to Cognitive Function and Disability***.** Comorbid psychiatric disorders frequently complicate neurological disease and diagnosis. Depression is both a risk factor for, and a prodromal feature of, dementia (Brzezińska et al., 2020). Similarly, depression is common in temporal lobe epilepsy and is related to outcomes following temporal lobe resection with evidence suggesting a bidirectional relationship between factors (Hermann, Loring, & Wilson, 2017). Whether these relationships simply reflect brain changes affecting both domains, there remains a critical need to identify and then explore whether modulation of comorbidity affects disease progression. Complex effects of anxiety on cognitive test performance are also present (Mella et al., 2020). There is limited information, however, on the disorders as either comorbid conditions or

complications of neurologic disease. Increased accuracy in identifying treatable mental disorders may have a marked impact on future incidence of neurocognitive dysfunction and disability. NNN will examine these relationships using appropriate diagnostic categorization and offer a unique and valuable resource for testing diverse hypotheses about those NP features that distinguish comorbid psychiatric diagnoses in neurological contexts.

In addition to categorical diagnoses, the NNN will provide one of the largest consistent collections of dimensional psychiatric symptoms collected contemporaneously with neuropsychological assessments and clinical interviews following structured history-taking. This enables analyses of covariation among cognitive and psychiatric symptom indicators not previously possible. NNN will characterize robust assessments of symptoms assessed by DSM-5 Level 1 measures, with many subjects also assessed with DSM-5 Level 2 characteristics. These data will reveal the degree to which symptoms of depression, anxiety, or psychosis either influence neuropsychological function or are at least co-morbid factors, informing treatment options designed to maximize cognitive outcome. These models will determine how neuropsychological measures covary with psychiatric symptoms, and the degree to which those relations are observed consistently across different syndromes and levels of ability. While we are currently recording ICD-10-CM/DSM-5 diagnostic information, our long-term strategy is agnostic about the validity of these diagnostic taxonomies, and we are eager to determine if alternative systems for dimensional or categorical representation of neuropsychiatric syndromes (e.g., the Research Domains Criteria (RDoC; Cuthbert et al., 2013; Cuthbert, 2020) or Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017)) may possess greater validity and provide deeper insights into

our understanding of neurocognitive disorders. Finally, the NNN dataset will provide a unique opportunity to examine models in which NP measures are mediators or moderators of relations between psychiatric symptoms and everyday functioning. These models will explore how neuropsychiatric syndromes exert their effects through neurocognitive features and, more importantly, if psychiatric symptoms have direct effects on outcomes. These analyses will generate evidence-based hypotheses about the relationship of psychiatric symptoms to real world functioning, the role of neurocognitive assessment in the understanding of these effects, and the degree to which effective treatments for mental illness can yield major benefits in disability reduction.

**Summary**

The NNN will provide a major resource to advance understanding of neurocognitive function through refined neurocognitive assessment. NNN is designed to establish a strong foundation capable of engaging neuropsychologists across a variety of practices. By contributing neuropsychological findings to the NDA, we aim to stimulate development of the next generation of evidence-based assessments. The NNN will provide foundational data to establish incremental validity of current, reconfigured, and ultimately novel neuropsychological measures for characterizing our emerging understanding of functional systems in the brain and real-world outcomes. We hope the ultimate result will be marked improvements in our ability to assess brain functions, leading to both improved understanding and treatment of neuropsychological impairment and mental illness in the United States. By enhancing the efficiency of assessment and directly addressing the problems of measurement invariance that have long plagued neuropsychological assessment, we anticipate that NNN will promote the development of methods that improve

access globally and reduce current inequities in neuropsychological service delivery.

Our vision is to create a platform that has a foundation forged from “classic” NP tests with widespread use in the NP community, and to build upon this foundation a new generation of tasks that may incorporate a range of current technologies to explore modern theories of brain structure-function relationships. By supplementing current test batteries with additional procedures to examine concurrent and predictive validity of newer measures, we believe the field can benefit both from back-compatibility with established standards, and future-directed extensions that will be more efficient and possess greater utility than the current methods. The new methods may include the application of virtual and augmented reality, “the internet of things,” videography, wearables, and eye tracking, among other current and anticipated innovations – all of which may enhance ecological validity relative to current methods. New assessment methods combined with multimodal neuroimaging or electrophysiological techniques (e.g., cortico-cortical evoked potentials in the context of stereoelectroencephalography), and advanced computational processing (e.g., artificial intelligence) may further hold the promise of moving assessments beyond neuropsychological “domains” (e.g., motor, visuo-perceptual, and language processes such as word-retrieval) to a deeper understanding of complex neural circuits (e.g., distributed neural processing, connectivity metrics)(Fox & Friston, 2012; Gonzalez-Castillo & Bandettini, 2015). These developments will ultimately be needed if we hope to understand major unanswered questions such as those that center on the complexities of memory (e.g., how is novel information integrated into a sense of autobiographical self and integrated semantic knowledge over time and space), consciousness and its

distortions (e.g., déjà vu, jamais vu, reduplicative paramnesia), socio-emotional functions, and default mode processing.

To facilitate engagement of the larger neuropsychology community, neuropsychologists are encouraged to register at the NNN website https://www.sistat.ucla.edu/NNNWeb/index.html, by providing their name, email address, phone number, and organization with which interested person is affiliated. The NNN will provide registered members of the neuropsychology community with progress updates, invitations to provide feedback on new initiatives, and access to data, assessment tools and algorithms developed by the NNN. We also hope soon to be able to extend registration of clinical patients with the NNN so that the network can grow beyond the current proof-of-concept to include multiple sites that can contribute data and benefit from sharing novel normative and clinical validity data that are rapidly being aggregated by the network nationwide. The mature NNN will help assure that clinical neuropsychology evolves to include the most accessible, efficient, and valid methods for the assessment of brain-behavior relations.

* **NNN Study Group Members**

Russell M. Bauer, Robert M. Bilder, Lucia Cavanagh, Daniel L. Drane, Kristen Enriquez, Felicia C. Goldstein, Jude Henry. Kelsey C. Hewitt, David W. Loring, Stephen P. Reise, Kuo Chung Shih, Catherine Sugar, Sean Turner, Laura Glass Umfleet, Dustin Wahlstrom, Patricia Walshaw, Fiona Whelan

## Author Notes

We have no known conflicts of interest to disclose.

This project is funded by NIMH R01 MH118514.

Partial support for development of the SHiP-NP was provided by a grant from the National Academy of Neuropsychology (NAN) to Lucia Cavanagh.

**Table 1.** Measures being obtained as part of the National Neuropsychology Network

| Battery or Domain | Test |
|---|---|
| Achievement | Wechsler Individual Achievement Test Second Edition |
| D-KEFS | Color-Word Interference Test |
| D-KEFS | Design Fluency Test |
| D-KEFS | Trail Making Test |
| D-KEFS | Verbal Fluency Test |
| Executive | Symbol Digit Modalities Test |
| Executive | Wisconsin Card Sorting Test |
| General | ACS-Test of Premorbid Function |
| General | Montreal Cognitive Assessment |
| General | Repeatable Battery for the Assessment of Neuropsychological Status** |
| General | SHiP-NP (PhenX Toolkit) |
| Language | Boston Naming Test |
| Language | Columbia Auditory Naming Test |
| Language | Emory Semantic Fluency Paradigm |
| Language | Test of Memory Malingering |
| Memory | California Verbal Learning Test -3 Standard Form |
| Memory | California Verbal Learning Test Alternate Form |
| Memory | California Verbal Learning Test Brief Form |
| Memory | Brief Vis Memory Test-Revised |
| Memory | Hopkins Verbal Learning Test |
| Memory | Rey Auditory Verbal Learning Test |
| Memory | Rey Complex Figure Test |
| Motor | Finger Tapping Test |
| Motor | Grooved Pegboard Test |
| PVT | Medical Symptom Validity Test |
| PVT | Word Memory Test |
| Symptom | Beck Anxiety Inventory-II |
| Symptom | Beck Depression Inventory |
| Visuospatial | Facial Recognition Test |
| Visuospatial | Judgment of Line Orientation |
| WAIS-IV | Arithmetic |
| WAIS-IV | Block Design |
| WAIS-IV | Coding |
| WAIS-IV | Comprehension |
| WAIS-IV | Digit Span |
| WAIS-IV | Information |
| WAIS-IV | Letter-Number Sequencing |
| WAIS-IV | Matrix Reasoning |
| WAIS-IV | Picture Completion |
| WAIS-IV | Similarities |

| | |
|---|---|
| WAIS-IV | Symbol Search |
| WAIS-IV | Visual Puzzles |
| WAIS-IV | Vocabulary |
| WMS-IV | Design Memory I |
| WMS-IV | Design Memory I - Older Adults |
| WMS-IV | Design Memory II |
| WMS-IV | Design Memory II - Older Adults |
| WMS-IV | Logical Memory I |
| WMS-IV | Logical Memory I - Older Adults |
| WMS-IV | Logical Memory II |
| WMS-IV | Logical Memory II - Older Adults |
| WMS-IV | Verbal Paired Associates I |
| WMS-IV | Verbal Paired Associates I - Older Adults |
| WMS-IV | Verbal Paired Associates II |
| WMS-IV | Verbal Paired Associates II - Older Adults |
| WMS-IV | Visual Reproduction |
| WMS-IV | Visual Reproduction - Older Adults |
| Achievement | Wechsler Individual Achievement Test Second Edition |

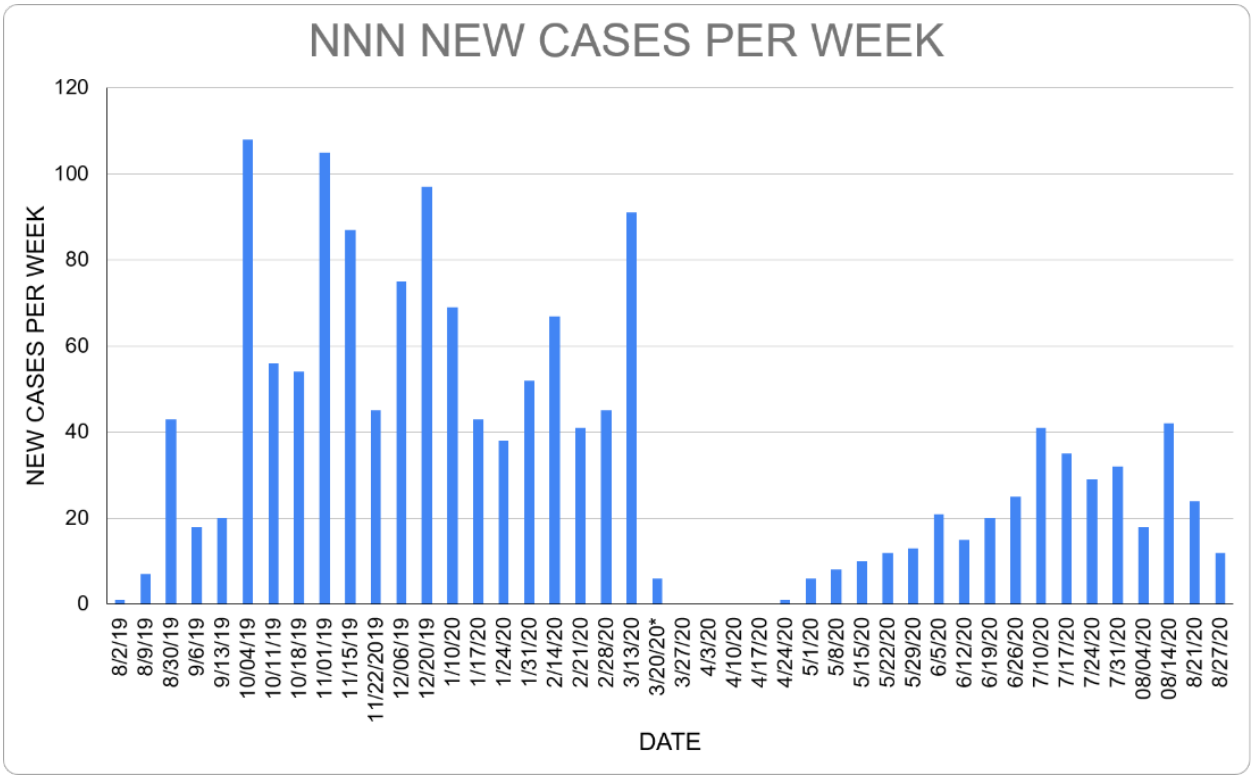
** *Selected RBANS subtests administered after TeleNP implementation due to COVID-19 pandemic*

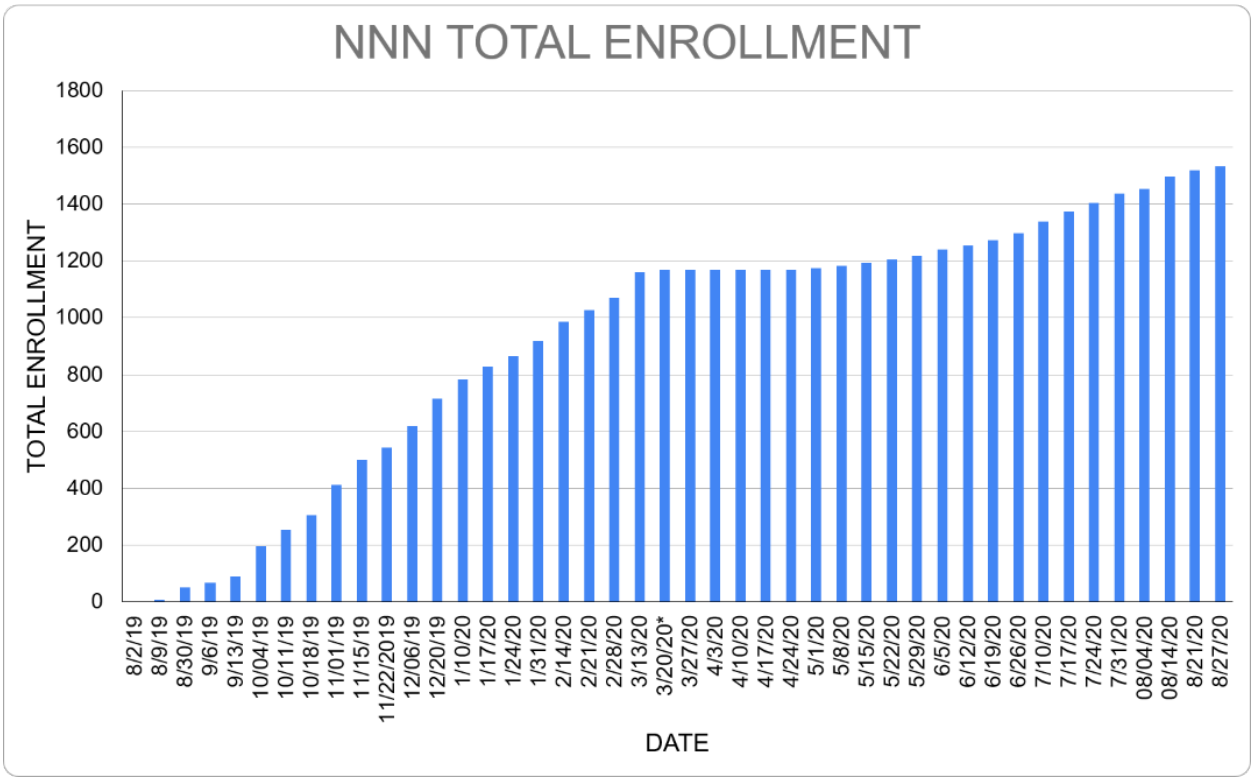
Supplemental Table 1. Demographic Characteristics of Participants in the NNN

| **Age (years)** | ***Parameter*** | **Value** |
|---|---|---|
| | *Mean (SD)* | 57.2 (17.6) |
| | *Range* | 18-90+ ** |
| **Sex** | | |
| | *Male N (%)* | 742 (49.2%) |
| | *Female N (%)* | 765 (50.7%) |
| | *Prefer not to answer N (%)* | 2 (0.1%) |
| **Race** | | |
| | *Group Label* | *N (%)* |
| | White | 1199 (79.5%) |
| | Black | 175 (11.6%) |
| | Asian | 25 (1.7%) |
| | Native Hawaiian/Other Pacific Islander | 1 (0.1%) |
| | Native American/Alaskan Native | 5 (0.3%) |
| | Other | 32 (2.1%) |
| | Unknown | 57 (3.8%) |
| | Prefer not to answer or Declined to Specify | 15 (1.0%) |
| **Ethnicity** | | |
| | Hispanic | 40 (2.7%) |
| | Not Hispanic<br>Unknown | 1391 (92.2%)<br>78 (5.2%) |

** *Participants over 90 years of age do not have specific age breakdown to due to Protected Health Information standards.*

**Supplemental Figures 1a (top) and 1b (bottom)**



**Supplemental Text**

**National Neuropsychology Network (NNN) Specific Aims**

**(NIMH; R01MH118514)**

The proposed National Neuropsychology Network (NNN) will contribute clinical diagnostic information and item-level data on the most widely used neuropsychological (NP) tests to the National Institute of Mental Health (NIMH) Data Archive (NDA). Data analyses will identify the latent constructs underlying these tests, increase efficiency of NP measurement, determine which NP measures are most informative with respect to key diagnostic questions, and examine the relations of psychiatric diagnoses and symptoms to cognitive impairment and disability.

**Aim 1**. **Establish Network Infrastructure:** We will launch the NNN with four sites, comprising major teaching clinics nationwide, and implement a shared clinical protocol and technological infrastructure for assessment. Sites include: Emory University, Medical College of Wisconsin, UCLA, and University of Florida. Goals for this aim include:

- Establishing the technological infrastructure for the network, including implementation of the Q-interactive platform for Pearson measures, and a new point-of-testing digital platform for additional measures. Pipelines will be developed to transmit data from both platforms to the NDA/Research Domains Criteria Data Base (RDoCdb).

- Collecting data on NP tests that are the most widely used in the United States in real-world clinic samples, comprising a diversity of neuropsychiatric syndromes that raise complex differential diagnostic questions.
- Implementing a structured clinical protocol to include demographics, diagnostics, and dimensional ratings of symptoms and disability using instruments proposed as NIH Common Data Elements, emphasizing those endorsed by the NIMH Research Panel, and developing a short, Structured History Protocol for Neuropsychology (SHiP-NP) to promote standardization and serve as a core transdiagnostic instrument specifically for the NP exam.

**Aim 2: Data Collection and Deposit:** The NNN will enroll 10,000 cases and deposit item-level data in RDoCdb. The cases and tests used will represent clinical NP services nationally. The sites span general outpatient and multiple specialty clinics, including those focused on dementia and degenerative conditions, epilepsies (including psychogenic non-epileptic seizures [PNES]), movement disorders, and other complex neuropsychiatric disorders. In these syndromes, mental illnesses (prominently depression, anxiety, or psychotic symptoms) are either directly part of the differential diagnosis (e.g., “dementia vs depression”) or the psychiatric symptoms may be critical moderators of cognitive impairment. The NNN aims to deposit in the NDA item-level data records for each of 47 widely administered instruments in more than 500, up to 10,000 participants each.

**Aim 3**. **Data Analyses:** The NNN aims to execute analyses of high value to the field, to:

- Identify the latent constructs measured by the NP tests and determine the most efficient measurement methods to identify these constructs, leading to proposal

for new tests and batteries. We hypothesize that this project will yield proposals to cut administration time by 50% for widely used “core” batteries.

- Determine how original and proposed novel measures relate to diagnostic outcomes. We will examine hypotheses that examine the utility of NP measures in: (a) the differential diagnoses of Mild and Major Neurocognitive Disorders; (b) the lateralization of seizures in focal epilepsies; and (c) determining the impact of comorbid mood, anxiety and psychotic disorders, and the relations of dimensional mood, anxiety, and psychotic symptoms to neurocognitive dysfunction, problems with everyday functioning, and disability.

## References

Abel, T. J., Rhone, A. E., Nourski, K. V., Kawasaki, H., Oya, H., Griffiths, T. D., . . . Tranel, D. (2015). Direct physiologic evidence of a heteromodal convergence region for proper naming in human left anterior temporal lobe. *J Neurosci, 35*(4), 1513-1520. doi:10.1523/jneurosci.3387-14.2015

Barch, D. M., Gotlib, I. H., Bilder, R. M., Pine, D. S., Smoller, J. W., Brown, C. H., . . . Farber, G. K. (2016). Common Measures for National Institute of Mental Health Funded Research. *Biological Psychiatry, 79*(12), e91-96. doi:10.1016/j.biopsych.2015.07.006

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist, 26*(2), 177-196. doi:10.1080/13854046.2012.663001

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R., . . . Woodhouse, J. (2020a). Inter Organizational Practice Committee recommendations/guidance for Teleneuropsychology in response to the COVID-19 pandemic. *Archives of Clinical Neuropsychology, 35*(6), 647-659. doi:10.1093/arclin/acaa046

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R., . . . Woodhouse, J. (2020b). InterOrganizational practice committee recommendations/guidance for Teleneuropsychology (TeleNP) in response to the COVID-19 pandemic. *The Clinical Neuropsychologist, 34*(7-8), 1314-1334. doi:10.1080/13854046.2020.1767214

Bilder, R. M., & Reise, S. P. (2019). Neuropsychological tests of the future: How do we get there from here? *The Clinical Neuropsychologist, 33*(2), 220-245. doi:10.1080/13854046.2018.1521993

Boake C. (2000). Edouard Claparède and the Auditory Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*, *22*(2), 286–292. doi.org/10.1076/1380-3395(200004)22:2;1-1;FT286

Brzezińska, A., Bourke, J., Rivera-Hernández, R., Tsolaki, M., Woźniak, J., & Kaźmierski, J. (2020). Depression in dementia or dementia in depression? Systematic review of studies and hypotheses. *Current Alzheimer Research, 17*(1), 16-28. doi:10.2174/1567205017666200217104114

Busch, R. M., Frazier, T. W., Iampietro, M. C., Chapin, J. S., & Kubu, C. S. (2008). Clinical utility of the Boston Naming Test in predicting ultimate side of surgery in patients with medically intractable temporal lobe epilepsy: A double cross-validation study. *Epilepsia, 17*, 17.

Cella, D., Lai, J. S., Nowinski, C. J., Victorson, D., Peterman, A., Miller, D., . . . Moy, C. (2012). Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. *Neurology, 78*(23), 1860-1867. doi:10.1212/WNL.0b013e318258f744

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181-197.

Choi, S. W., Reise, S. P., Pilkonis, P. A., Hays, R. D., & Cella, D. (2010). Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Quality of Life Research, 19*(1), 125-136. doi:10.1007/s11136-009-9560-5

Curtiz, Z. (Director). (1942). *Casablanca* [film]. Warner Bros. Pictures.

Cuthbert, B.N. (2020). The role of RDoC in future classification of mental disorders. Dialogues Clin Neurosci. 22(1):81-85. doi: 10.31887/DCNS.2020.22.1/bcuthbert. PMID: 32699508; PMCID: PMC7365298.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC medicine, 11(1), 126.

Drane, D. L., Ojemann, J. G., Phatak, V., Loring, D. W., Gross, R. E., Hebb, A. O., . . . Tranel, D. (2013). Famous face identification in temporal lobe epilepsy: support for a multimodal integration model of semantic memory. *Cortex, 49*(6), 1648-1667. doi:10.1016/j.cortex.2012.08.009

Drane, D. L., & Pedersen, N. P. (2019). Knowledge of language function and underlying neural networks gained from focal seizures and epilepsy surgery. *Brain & Language, 189*, 20-33. doi:10.1016/j.bandl.2018.12.007

Duff, K., Suhrie, K. R., Dalley, B. C. A., Anderson, J. S., & Hoffman, J. M. (2019). External validation of change formulae in neuropsychology with neuroimaging biomarkers: A methodological recommendation and preliminary clinical data. *The Clinical Neuropsychologist, 33*(3), 478-489. doi:10.1080/13854046.2018.1484518

Fox, P. T., & Friston, K. J. (2012). Distributed processing; distributed functions? *Neuroimage, 61*(2), 407-426. doi:10.1016/j.neuroimage.2011.12.051

Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., . . . Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59*(4), 361-368. doi:10.1176/ps.2008.59.4.361

Gonzalez-Castillo, J., & Bandettini, P. A. (2015). What Cascade Spreading Models Can Teach Us about the Brain. *Neuron, 86*(6), 1327-1329. doi:10.1016/j.neuron.2015.06.006

Grinnon, S. T., Miller, K., Marler, J. R., Lu, Y., Stout, A., Odenkirchen, J., & Kunitz, S. (2012). National Institute of Neurological Disorders and Stroke Common Data Element Project: Approach and methods. *Clincal Trials, 9*(3), 322-329. doi:10.1177/1740774512438980

Hamilton, C. M., Strader, L. C., Pratt, J. G., Maiese, D., Hendershot, T., Kwok, R. K., . . . Haines, J. (2011). The PhenX Toolkit: get the most from your measures. *American Journal of Epidemiology, 174*(3), 253-260. doi:10.1093/aje/kwr193

Hermann, B., Loring, D. W., & Wilson, S. (2017). Paradigm shifts in the neuropsychology of epilepsy. *Journal of the International Neuropsychological Society, 23*(9-10), 791-805. doi:10.1017/S1355617717000650

Hewitt, K. C., Rodgin, S., Loring, D. W., Pritchard, A. E., & Jacobson, L. A. (2020). Transitioning to telehealth neuropsychology service: Considerations across adult and pediatric care settings. *The Clinical Neuropsychologist*, 1-17. doi:10.1080/13854046.2020.1811891

Hoogland, J., van Wanrooij, L. L., Boel, J. A., Goldman, J. G., Stebbins, G. T., Dalrymple-Alford, J. C., . . . Weintraub, D. (2018). Detecting mild cognitive deficits in Parkinson's Disease: Comparison of neuropsychological tests. *Movement Disorders, 33*(11), 1750-1759. doi:10.1002/mds.110

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., Miller, J. D., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. Journal of abnormal psychology, 126(4), 454–477. doi.org/10.1037/abn0000258

Lavoie, K., & Bacon, S. for the iCARE Study Team (2020). iCARE Study Questionnaire.

Lopez-Anton, R., Santabárbara, J., De-la-Cámara, C., Gracia-García, P., Lobo, E., Marcos, G., . . . Lobo, A. (2015). Mild cognitive impairment diagnosed with the new DSM-5 criteria: Prevalence and associations with non-cognitive psychopathology. *Acta Psychiatrica Scandinavica, 131*(1), 29-39. doi:10.1111/acps.12297

Loring, D. W., Strauss, E., Hermann, B. P., Perrine, K., Trenerry, M. R., Barr, W. B., . . . Meador, K. J. (1999). Effects of anomalous language representation on

neuropsychological performance in temporal lobe epilepsy. *Neurology, 53*(2), 260-264.

Mansolf, M., Vreeker, A., Reise, S. P., Freimer, N. B., Glahn, D. C., Gur, R. E., . . . GROUP2", W. C. a. (2020). Extensions of multiple-group item response theory alignment: Application to psychiatric phenotypes in an international genomics consortium. *Educational and Psychological Measurement*.

Marcopulos, B., & Lojek, E. (2019). Introduction to the special issue: Are modern neuropsychological assessment methods really "modern?" Reflections on the current neuropsychological test armamentarium. *The Clinical Neuropsychologist, 33*(2), 187-199. doi:10.1080/13854046.2018.1560502

Mella, N., Vallet, F., Beaudoin, M., Fagot, D., Baeriswyl, M., Ballhausen, N., . . . Desrichard, O. (2020). Distinct effects of cognitive versus somatic anxiety on cognitive performance in old age: The role of working memory capacity. *Aging and Mental Health, 24*(4), 604-610. doi:10.1080/13607863.2018.1548566

Moore, T. M., Scott, J. C., Reise, S. P., Port, A. M., Jackson, C. T., Ruparel, K., . . . Gur, R. C. (2015). Development of an abbreviated form of the Penn Line Orientation Test using large samples and computerized adaptive test simulation. *Psychological Assessment, 27*(3), 955-964. doi:10.1037/pas0000102

Norman, M., Wilson, S. J., Baxendale, S., Barr, W. B., Block, C., Busch, R. M., . . . Hermann, B. P. (in press). Addressing neuropsychological disorders in adults with epilepsy: The International League Against Epilepsy (ILAE)/International Neuropsychological Socieity (INS) initiative. . *Epilepsia Open*.

Pawlowski, J., Segabinazi, J. D., Wagner, F., & Bandeira, D. R. (2013). A systematic review of validity procedures used in neuropsychological batteries. *Psychology & Neuroscience, 6*, 311-329.

Postal, K. S., Bilder, R. M., Lanca, M., Aase, D. M., Barisa, M., Holland, A. A., . . . Salinas, C. (2020). Inter Organizational Practice Committee guidance/recommendation for models of care during the novel coronavirus pandemic. *Archives of Clinical Neuropsychology*. doi:10.1093/arclin/acaa073

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology, 31*(3), 206-230. doi:10.1093/arclin/acw007

Raspall, T., Donate, M., Boget, T., Carreno, M., Donaire, A., Agudo, R., . . . Salamero, M. (2005). Neuropsychological tests with lateralizing value in patients with temporal lobe epilepsy: reconsidering material-specific theory. *Seizure, 14*(8), 569-576. doi:10.1016/j.seizure.2005.09.007

Reise, S. P., Widaman, K., Bauer, R. M., Drane, D. L., Loring, D. W., Umfleet, L. G., . . . BIlder, R. M. (2021). *Item response theory analyses of Matrix Reasoning: Towards a new short form or adaptive Test?* Paper presented at the Annual Meeting of the International Neuropsychological Society, San Diego, CA.

Strauss, E., Satz, P., & Wada, J. (1990). An examination of the crowding hypothesis in epileptic patients who have undergone the carotid amytal test. *Neuropsychologia, 28*(11), 1221-1227.

Strauss, E., Semenza, C., Hunter, M., Hermann, B., Barr, W., Chelune, G., . . . Westerveld, M. (2000). Left anterior lobectomy and category-specific naming. *Brain & Cognition, 43*(1-3), 403-406.

Teng, E. L., & Manly, J. J. (2005). Neuropsychological testing: helpful or harmful? *Alzheimer Disease and Associated Disorders, 19*(4), 267-271.

Zahodne, L. B., Manly, J. J., Smith, J., Seeman, T., & Lachman, M. E. (2017). Socioeconomic, health, and psychosocial mediators of racial disparities in cognition in early, middle, and late adulthood. *Psychology & Aging, 32*(2), 118-130. doi:10.1037/pag0000154