Brain-CODE Common Data Elements: Development of Core Demographic and Clinical Standards to Facilitate Data Aggregation, Sharing and Analyses

Report of the Brain-CODE Common Data Elements Committee

Summary

The Ontario Brain Institute's "Brain-CODE" is a large-scale informatics platform designed to support the collection, storage and integration of diverse types of data across several brain disorders as a means to understand common underlying causes of brain dysfunction and developing novel approaches to treatment. By providing access to aggregated datasets on patients with different brain disorders, Brain-CODE will allow analyses both within and across disease states. Brain-CODE will cover multiple brain disorders and encompass a wide array of data, including clinical, neuroimaging and molecular. To help achieve these goals, establishment of Common Data Elements (CDEs) within Brain-CODE will be critical to enable a high degree of consistency in data collection across studies. This will optimize the ability of investigators to analyze pooled data both within and across brain disorders, and facilitate federation with related external programs.

A set of General Core Demographic and Clinical CDEs have been developed. The present document provides a summary of approach, methodology, results and committee recommendations.

Introduction

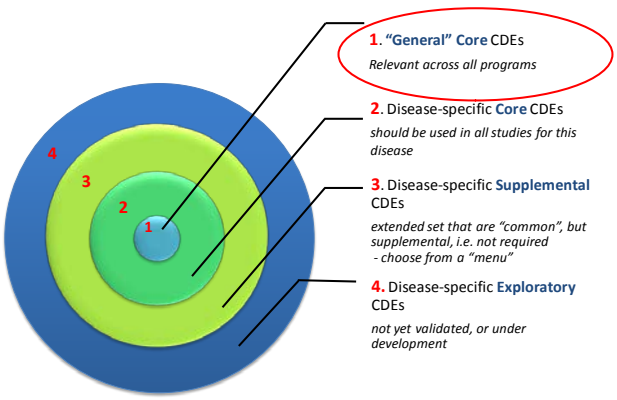
The principles of data sharing as a catalyst for scientific discovery are widely recognized by international organizations such as the National Institutes of Health [1], Canadian Institutes of Health Research [2] and Wellcome Trust [3]. Historically, however, research databases have existed in isolation with no practical avenue for sharing or pooling medical data into high dimensional "big" data sets that can be efficiently compared across databases. Databases have their own sets of data standards, database software and processes, thus limiting their ability to synthesize and share data with one another.

To address this challenge and allow researchers from across the province of Ontario to collaborate and work more efficiently, the Ontario Brain Institute's (OBI) “Brain-CODE” is designed to support the collection, storage and integration of diverse types of data across several brain disorders, including neurodevelopmental disorders, cerebral palsy, epilepsy, major depressive disorder, and neurodegenerative diseases [4]. By providing access to aggregated datasets on patients with different brain disorders, Brain-CODE will support scientific inquiry and analytics both within and across multiple brain diseases and modalities by integrating clinical, imaging, and molecular data. The Brain-CODE platform is highly secure and designed to provide linkages with provincial, national and international databases [4].

Given different research aims, study designs and technologies used across programs, establishment of a minimum set of clearly defined and standardized assessments to be used across studies is essential to facilitate data sharing and integration, and to conduct meaningful analyses. Indeed, in the absence of common measures and data standards it is difficult to compare results across studies.

In an effort to optimize the ability to aggregate and analyze data within Brain-CODE, Common Data Elements (CDEs) have developed to provide standard definitions and formats so that investigators collect data consistently across studies. This will reduce variability in data collection and ultimately facilitate comparisons across diseases, merging of data sets and meta-analyses. Using the framework of the National Institute of Neurological Disorders and Stroke (NINDS) CDE Project as guidance [5,6], General Core CDEs Demographic and Clinical CDEs have been developed. Critical to this process has been engagement of participating researchers through workshops and consensus methodologies. A summary of the approach, methodology, results and recommendations are presented here.

CDE classification in the (NINDS model)



Consensus development among stakeholders: Delphi surveys

To engage researchers and obtain feedback and opinion, a modified two-round Delphi survey [7,8] was used to identify core demographic and clinical variables to be collected across all participating programs. Participants from OBI-funded programs were invited to an online survey hosted through the Brain-CODE portal, detailing the variables and overview of Brain-CODE. Participants were asked to comment and respond to statements on a 5-point Likert scale regarding the collection of demographic and clinical variables, with possible responses ranging from Not Important to Very Important (example: How important is the collection of date of birth to achieving Brain-CODE goals?), Do Not Recommend to Highly Recommend (example: Please provide your recommendation for the GAD-7 to assess anxiety in adults across all programs) or Disagree to Strongly Agree (example: QIDS-SR is

appropriate to assess depression across all participating programs in adults and adolescents). The results were reviewed and anonymized results (aggregated ratings and comments) presented back to the participants in a follow-up survey to obtain additional opinion and clarification, as required. Participants were directed to consider the results of the first survey in their responses. Although the threshold for consensus is somewhat arbitrary, recommended criteria for Delphi consensus generally range from 70% to 80% of agreement within two categories.[7] In the present surveys, this would include ratings of Important/Very Important, Recommend/Highly Recommend or Agree/Strongly Agree. In the present exercise, consensus levels of >70% were considered, with other factors also weighted, including harmonization with existing relevant databases. When consensus was not achieved, the relevant programs were asked to discuss internally and provide their recommendations. Thirty-six researchers were invited to participate in the demographic surveys and 44 were invited to participate in the clinical surveys.

Demographic CDEs

As a first step, a Brain-CODE CDE Committee identified demographic variables collected, or planned to be collected, by participating programs. Following review of IDP data dictionaries, study protocols, and through interactions with program researchers, a Demographic Variables Catalog identified demographic variables of relevance across diseases. This catalog recorded for each program whether or not a specified demographic variable was collected (or was planned to be collected); and if available the format in which it was recorded. In addition, demographic variables collected by related programs were also reviewed and considered, including the NINDS CDE Project which supports similar research areas [5,6]. The following demographic variables were considered: *Sex, Date of Birth, Handedness, Ethnicity, Race, Education Level, Marital Status, Primary Language, Place of Birth, Geographic Region, Weight/Height.*

Core Demographic CDEs recommended to be collected across participating programs:

Sex. There was consensus that the subject's sex should be recorded across programs (>90% agreement). Sex-related differences in brain and behavior are often reported and required to perform analyses based on population stratification and for reporting in peer-reviewed publications.

Date of Birth. There was consensus that full DOB should be recorded across programs (>80% agreement), with participants indicating that collection of full DOB was less critical in studies involving adults. In the follow-up survey, participants were queried regarding collection of partial DOB. Although participants re-iterated distinctions with respect to study populations, there was support for capturing full DOB as it provides the most precise and source-verifiable information. As full DOB would provide the most detailed information required for calculation of age and could potentially be used for Brain-CODE linkage purposes, it is recommend that full DOB be recorded across all participating programs. It is important to note that because full DOB is considered a personal identifier it is subject to research ethics oversight and adherence to Personal Health Information Protection Act (PHIPA) [9]. As such, full DOB will be de-identified for data sharing in Brain-CODE, or only partial DOB (month and year) or calculated age will be stored.

Handedness. In Survey 1, 64% agreed that handedness should be recorded across programs. The primary reason for support was that handedness is routinely collected in brain imaging studies, as it reflects cerebral lateralization. In survey 2, when considering the collection of imaging data as part of the Brain-CODE platform, it was recommended that handedness be recorded across all programs (73% agreement).

Race and Ethnicity. In Survey 1, 74% agreed that ethnicity should be recorded across all programs, whereas only 55% supported the collection of race. Although race and ethnicity are core NINDS CDEs, harmonizing with US standards is not recommended given differences in Canadian and US census-based categorizations. It was

therefore recommended that self-identified ethnicity should be recorded using Canadian census-based categories. For pediatric studies, parental (biological) ethnicity should also be recorded.

Social Economic Status (SES). Although not originally considered, respondents commented that SES should also be collected, as it is an important predictor of developmental disorders and is used in many of their analyses. SES includes assessment of education, marital status, occupation and income. There was clear support for including education level (>80% agreement) but not marital status (<50% agreement). However, given the support for collection of SES and considering that SES is a core NINDS Demographic CDE, it was recommend that SES be recorded across all programs. For pediatric studies, primary caregiver's SES should be recorded. Primary language, place of birth, geographic region and height/weight were not supported as core CDEs.

Harmonizing Brain-CODE Demographic CDEs

Harmonization of the Brain-CODE CDEs with other large centralized data repositories and CDE initiatives is also a key consideration. This would enable data sharing across research addressing similar topics, within the larger research community, and would allow for adequate comparison of results across studies. Where appropriate, therefore, demographic CDEs developed within Brain-CODE were harmonized with the NINDS core demographics domains. In addition, data standards provided by the Clinical Data Interchange Standards Consortium (CDISC) were also considered, including alignment with relevant Clinical Data Acquisition Standards Harmonization (CDASH) -recommended standards and guidelines [10].

Clinical CDEs

Prior to sending out Delphi surveys, a preliminary online survey was sent to program representatives to identify clinical domains that may be of relevance to the goals of Brain-CODE. The results were presented at a follow-up workshop, and following group discussion there was agreement that *Comorbidity, Depression, Anxiety, Sleep, Quality of Life and Activities of Daily Living* should be assessed across programs. There was also agreement that when possible, the measure should be patient-reported, brief and easy to administer, widely used and validated, and available in the public domain. In addition, although there was consensus that *Cognitive Function* should be assessed across all programs, it was agreed that implementing a common battery across all programs was not practical. Following the workshop, Delphi surveys were sent out to identify the common outcome measures that should be used to assess these domains (Table 1). Participants were presented a copy and detailed summary of the scales to consider in their responses.

Core Clinical CDEs recommended to be collected across all participating programs:

Comorbidity. At the workshop, there was consensus to use the NINDS Medical History form to assess medical comorbidity and thus not followed-up by survey. When surveyed regarding instruments to assess comorbid psychiatric symptoms and asked to choose among the available instruments, both the Symptom Checklist-90-R (SCL-90-R) and Brief Symptom Inventory (BSI) were equally endorsed. As the BSI covers the same domains as the SCL-90 and requires less time to administer, the BSI was recommended.

Depression and Anxiety. For adults and adolescents, the QIDS-SR was recommended for assessment of depression with 90% agreement and GAD-7 to assess anxiety with 83% agreement. Although consensus was not achieved for assessment depression and anxiety in children, following further internal evaluation by the relevant programs, the Revised Children's Anxiety and Depression Scale (RCADS) was recommended for children and adolescents.

Sleep Disturbances. The PSQI was recommended to assess sleep adults and adolescents with 88% agreement. Although consensus was not achieved by survey for assessment of sleep in children, following further internal evaluation by the relevant programs, the CSHQ was recommended.

Quality of Life and Activities of Daily Living. At a Brain-CODE Clinical CDE Workshop, there was general agreement to use World Health Organization Quality of Life Short Version (WHO-QoL-BREF) to assess QoL in adults, as it is a valid, reliable, brief and commonly used QoL measure. Although consensus was not achieved by survey for assessment of QoL in children and adolescents, following further internal evaluation by the relevant programs, the KINDL-R was recommended. With respect to ADLs, it was recognized that the specific ADL scale used may vary given the population/disability studied. Therefore, although the Sheehan Disability Scale was recommended, there was agreement that additional scales should be used, as appropriate.

Conclusions

Developing a set of standardized assessments to be adopted across different studies is a challenging endeavor that must consider the goals of individual research programs. Indeed, as a matter of good scientific research practice, measurements selected should be scientifically valid and justified to support specified aims. Critical to the success of this initiative, therefore, has been buy-in and cooperation from participating programs, including engagement of representative researchers through participation in workshops and agreement on a recommended set of common assessment and standards. In developing the present core set of demographic variables and clinical outcome measures, consensus-based methodology was used to inform participants, gain their input and opinion, and arrive at levels of consensus. In addition, by including open-ended questions and comments, suggestions and opinions were not restricted to the predefined variables, thus allowing for broadening of opinion and consideration of program-specific needs and challenges.

Brain-CODE includes multidisciplinary collaborative research networks across multiple brain diseases. Establishing of CDEs within Brain-CODE has been a critical step towards enabling a high degree of consistency in data collection, optimizing the ability of investigators to analyze data both within and across brain disorders, and facilitating federation with related external programs. These core Demographic and Clinical CDEs (Table 2) have been successfully implemented within all participating programs and will be available to the research community for analyses. Brain-CODE CDEs[1] are available at www.braincode.ca.

[1] ***Case Report Forms and Data Dictionaries*** are available and provide the framework for the development of eCRFs (*electronic* case report forms) allowing the input of data into Electronic Data Capture systems housed on Brain-CODE, including OpenClinica and REDCap. The development of eCRFs will be facilitated by the OBI and the InDOC Consortium, which is responsible for the technical development of Brain-CODE. The CRFs are available in the public domain without cost (permission may be required for some). Users fees apply to the BSI (not included)

References

1. National Institutes of Health: NIH Data Sharing Policy and Implementation Guidance. 2003. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

2. Canadian Institutes of Health Research: CIHR open access policy. 2013. http://www.cihr-irsc.gc.ca/e/46068.html

3. Wellcome Trust: Data Sharing. http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing

4. Ontario Brain Institute: Brain-CODE. http://www.braininstitute.ca/brain-code

5. Grinnon ST, Miller K, Marler JR, Lu Y, Stout A, Odenkirchen J, Kunitz S. National Institute of Neurological Disorders and Stroke Common Data Element Project - approach and methods. Clin Trials. 2012;9:322-329.

6. National Institutes of Health: NINDS CDE Catalog. http://www.commondataelements.ninds.nih.gov/CDE.aspx

7. Dalkey NC, Helmer O. An experimental application of the Delphi method to the use of experts. Management Science. 1963;9:458-467.

8. Hsu CC, Standford, BA. The Delphi Technique: Making Sense Of Consensus. Practical Assessment, Research & Evaluation, 2007;12:1-8.

9. http://www.e-laws.gov.on.ca/html/statutes/english/elaws_statutes_04p03_e.htm.

10. CDISC CDASH core and domain teams: Clinical Data Acquisition Standards Harmonization (CDASH). 2008. http://www.cdisc.org/stuff/contentmgr/files/0/9b32bc345908ac4c31ce72b529a3d995/misc/cdash_std_1_0_2008_10_01.pdf

Table 1. Clinical Scales Considered

Domain	Age	Scales
Quality of Life	Child	Infant Toddler Quality of Life Questionnaire KINDL-R Health Related Quality of Life Child Health Questionnaire KIDSCREEN Pediatric Quality of Life Inventory NIH Neuro-QOL
	Adolescent	KINDL-R Health Related Quality of Life Child Health Questionnaire KIDSCREEN Pediatric Quality of Life Inventory NIH Neuro-QOL
	Adult	Quality of Life and Satisfaction World Health Organization Quality of Life Short Version 36-Item Short Form Health Survey EuroQoL-5D NIH Neuro-QOL
Activities of Daily Living	Child	Klein-Bell Activities of Daily Living Scale Patient Reported Outcomes Measurement Information System
	Adolescent	Klein-Bell Activities of Daily Living Scale Patient Reported Outcomes Measurement Information System
	Adult	Lawton Instrumental Activities of Daily Living Katz Index of Independence in Activities of Daily Living modified Rankin Scale Activities of Daily Living Questionnaire Alzheimer's Disease Co-operative Study - ADL Inventory Disability Assessment for Dementia Klein-Bell Activities of Daily Living Scale Sheehan Disability Scale Patient Reported Outcomes Measurement Information System
Depression	Child	Reynolds Adolescent Depression Scales-2 Children's Depression Inventory Center for Epidemiologic Studies Depression Scale for Children Reynolds Child Depression Scale Revised Children's Anxiety and Depression Scale Patient Reported Outcomes Measurement Information System
	Adolescent	Center for Epidemiologic Studies Depression Scale Beck Depression Inventory Reynolds Adolescent Depression Scales-2 Children's Depression Inventory Center for Epidemiologic Studies Depression Scale for Children Revised Children's Anxiety and Depression Scale *Patient Reported Outcomes Measurement Information System

	Adult	Quick Inventory of Depressive Symptomatology Inventory of Depressive Symptomatology Zung Self-Rated Depression Scale Hospital Anxiety Depression Scale Patient Health Questionnaire Center for Epidemiologic Studies Depression Scale Beck Depression Inventory Patient Reported Outcomes Measurement Information System
Anxiety	Child	Revised Children's Manifest Anxiety Scale 2 Spence Children's Anxiety Scale Beck Anxiety Inventory for Youth Patient Reported Outcomes Measurement Information System
	Adolescent	Revised Children's Manifest Anxiety Scale 2 Patient Reported Outcomes Measurement Information System
	Adult	Generalized Anxiety Disorder-7 Hospital Anxiety and Depression Scale Beck Anxiety Inventory State-Trait Anxiety Inventory Patient Reported Outcomes Measurement Information System
Sleep	Child	BEARS Sleep Screening Assessment Sleep Disturbance Scale for Children Tayside Children's Sleep Questionnaire Children's Sleep Habits Questionnaire Brief Infant Sleep Questionnaire Patient Reported Outcomes Measurement Information System
	Adolescent	BEARS Sleep Screening Assessment Sleep Disturbance Scale for Children Patient Reported Outcomes Measurement Information System
	Adult	Pittsburgh Sleep Quality Index Insomnia Severity Index Sleep Quality Scale Stanford Sleepiness Scale Patient Reported Outcomes Measurement Information System
Psychiatric Comorbidity	Child	none
	Adolescent	Symptom Checklist-90-R Brief Symptom Inventory Behavior and Symptom Identification Scale
	Adult	Symptom Checklist-90-R Brief Symptom Inventory Behavior and Symptom Identification Scale

Table 2 Summary of Brain-CODE Core Demographic and Clinical CDEs

DOMAIN	SUB-DOMAIN	Brain-CODE CDE
Patient Characteristics	Demographic	Brain-CODE Demographic Form
	SES	Brain-CODE Demographic Form
Physical and Mental Health	Quality of Life	WHO-Qol-BREF (adult) KINDL-R (child & adolescent)
	Activities of Daily Living	Sheehan Disability Scale (adult)
	Medical Comorbidity	NINDS Medical History
	Psychiatric Comorbidity	BSI (adolescent & adult)
Clinical Endpoints	Depression	QIDS-SR (adolescent & adult) RCADS (child & adolescent)
	Anxiety	GAD-7 (adolescent & adult) RCADS (child & adolescent)
	Sleep	PSQI (adult) CSHQ (adolescent & adult)