Exploring Use of NIH Common Data Elements (CDEs) in NIH-funded Research Projects

Project Lead

Erin Foster, NLM Associate Fellow

Project Sponsor

Jerry Sheehan, Assistant Director for Policy Development
Office of the Director (OD)

30 August 2015

Table of Contents

Abstract 3
Background 4
Project Objectives 5
Methods 5
Identifying initiatives 5
PubMed and Common Data Elements 5
Publication review 6
Website review 6
Results 7
Use of CDEs in NIH-funded research 7
Funding Opportunity Announcements (FOAs) sample 7
Publication & study lists sample 9
Existing guidance for citing use of NIH CDEs 10
NIH CDE Portal enhancements 11
Discussion 12
Recommendations 12
Limitations 13
Conclusion 14
References 15
Acknowledgements 15
Appendices 15
Appendix I – PubMed searches 15
Appendix II – Screen shot of FOA List 16
Appendix III – Screen shot of Summary Table 16

Abstract

Objective

The primary purpose of this project is to identify use of National Institutes of Health (NIH) Common Data Elements (CDEs) in published research to facilitate future tracking of that use. This exploratory research focuses on addressing a gap in knowledge about use of NIH CDEs in published research. This project also provides suggestions for future identification and tracking of research that uses CDEs.

Methods

The research used multiple methods: preliminary PubMed searches to identify use of CDEs in the literature, hand search of publications resulting from Funding Opportunity Announcements (FOAs) that encouraged use of NIH CDEs in research as well as publication or study lists included on NIH CDE initiative websites, and a review of NIH CDE initiative websites to determine existing guidance made available to researchers about citing use of CDEs in their research.

Results

The hand search of 573 publications resulted in a majority (46%) with no evidence of CDE use. Those publications that indicated use varied in their reference to CDEs in two ways – 28% of the sample mentioned use either at collection level or at instrument level. The remaining publications (26%) referred to CDEs in the context of their development and evaluation rather than their use in research. Currently no common language or guidance exists for citing usage of NIH CDEs in published research across the various NIH CDE initiatives. One action emerging from the results was to draft model language to guide future development of standardized language for citing use of NIH CDEs. Additional outcomes included enhancements to the existing NIH CDE Portal to reflect project research and further connect CDE efforts across NIH.

Discussion

The research conducted for this project clearly demonstrates the challenges of identifying use of NIH CDEs in published research. A primary recommendation is to standardize guidance provided by NIH CDE initiatives to researchers in how to cite use of NIH CDEs. This standardization should be considered not only for publications, but also for databases and/or data repositories that house datasets resulting from research.

Conclusion

Encouraging use of NIH CDEs is a trans-NIH effort and aligns strongly with other NIH efforts related to standardization and data-sharing. The National Library of Medicine (NLM) should continue its active involvement in existing NIH CDE collaborations and advocate for standardizing citation methods and tracking efforts of CDE use across NIH.

Background

Common Data Elements (CDEs) are variables that have been designated for use across sets of clinical research studies and registries in particular research domains.[1] Fundamentally, CDEs are data elements that are developed and used to allow data to be commonly collected across research studies. They are generally structured as a defined question with a specified set of permissible values. Ideally these elements have human and machine-readable definitions to allow for interoperability. However, CDEs do not just exist as individual data elements; frequently they are grouped into sets to form questionnaires, case report forms, or other validated instruments.

There are currently 11 existing initiatives at the National Institutes of Health (NIH) that develop and maintain CDEs. These initiatives are affiliated with institutes and are responsible for curating, vetting, and providing access to CDEs that are relevant to their domains of expertise.[2] In addition to these initiatives, there are trans-NIH collaborations that contribute to the development and promotion of CDEs at NIH. The National Library of Medicine (NLM) is an active member and facilitator in these collaborations, which include:

- NIH CDE Portal: a website that provides information about CDEs in clinical research, patient registries, and other human subject research at NIH. Includes summary information and links to existing CDE initiatives at NIH.[1] NLM hosts and maintains this website.
- NIH CDE Repository: provides access to CDEs that are developed and supported by NIH Institutes and Centers (ICs).[2] NLM hosts the repository and provides support for its continued development.
- NIH CDE Working Group: part of the Trans-NIH BioMedical Informatics Coordinating (BMIC) Committee. Exists to improve coordination & communication of efforts to identify, develop, and promote use of CDEs across NIH. NLM currently chairs and hosts Working Group meetings.

Recently, NIH Funding Opportunity Announcements (FOAs) have begun including language encouraging use of CDEs by researchers receiving grant funding.[3] For research purposes, use of CDEs is encouraged to promote higher data quality through standardization of data as well as enable data reuse and secondary analyses through data-sharing. Additionally, corresponding CDEs across studies allow for improved replication and reproducibility of studies. Encouragement for use of CDEs comes not only from the efforts at NIH, but more broadly in domains ranging from sharing of clinical trial data to data collected as part of systematic reviews.[4],[5]

In light of these existing methods to encourage use of CDEs, this project addresses a gap in knowledge about use of NIH CDEs in NIH-funded research. Motivations to demonstrate use of NIH CDEs in NIH-funded research are twofold: first, identifiable inclusion of CDEs in research demonstrates the functionality and use of CDEs to future researchers by showing higher data quality and the ability to perform strong secondary analyses on that data; secondly, identification of CDE use acts as a return on

[1] http://www.nlm.nih.gov/cde/
[2] https://cde.nlm.nih.gov/

investment for NIH CDE initiatives, and NIH at large, that have devoted considerable time and funds to developing, curating, and distributing CDEs for researcher use.

Project Objectives

This project investigates the use of NIH CDEs in NIH-supported research projects. In particular, this project attempts to track use of CDEs in publications produced from NIH research projects. Some questions to be addressed in the analysis of these publications include – how are CDEs cited in research publications? If a citation is present, where is it located in the publication (e.g., methods, acknowledgements)? Is citing use of CDEs in publications standardized? Additionally, the project will provide suggestions for future tracking and identification of research that uses CDEs.

Methods

This project used multiple methods to explore the use of CDEs in NIH-funded research. The project focused on 3 CDE initiatives – the National Institute of Neurological Disorders and Stroke Common Data Elements (NINDS CDEs), Consensus Measures for Phenotypes and eXposures (PhenX), and Patient Reported Outcomes Measurement Information System (PROMIS). Searches of PubMed were conducted to explore the use of CDEs in the literature as well as the potential for indexing publications for use of CDEs. Following that, a hand search of 573 publications was performed to review for use of CDEs. In addition, the websites of NINDS CDEs, PhenX, and PROMIS were evaluated for guidance about citing use of CDEs.

Identifying initiatives

Three of the 11 current NIH CDE initiatives were selected to limit evaluation of research papers – NINDS CDEs, PhenX, and PROMIS. The NINDS CDEs initiative provides core or supplemental sets of CDEs to be used in NINDS funded disease-specific studies (e.g., Parkinson's Disease, Epilepsy). The PhenX initiative makes available standard measures to be used in genome-wide association studies (GWAS) as well as other large-scale genomic and epidemiologic research. The PROMIS initiative has item banks for use in measuring patient reported health status in clinical populations in the physical, mental, and social health domains. These 3 were chosen because these initiatives are well established and developed with collections of CDEs in a variety of subject areas.

PubMed and Common Data Elements

Initial exploration of PubMed for common data elements was conducted not only to find relevant background literature, but to explore how CDEs can currently be identified in the literature. In particular, attention was paid to how CDEs are indexed in PubMed.

The MeSH term “Common Data Elements” was released as part of 2015 MeSH. This MeSH term is currently treed in the MeSH vocabulary under “Vocabulary, Controlled” within the “Information Science Category”. At the outset of this project, the question of retreeing the MeSH term for “Common Data Elements” was raised because the term’s current position in the vocabulary does not accurately represent what CDEs are or what they are used for. Subsequently, a request was submitted, and

approved, to retree the MeSH term under “Data Collection” (L01.280) in the “Information Science Category” in the 2016 release of MeSH.

Despite the retreeing, the MeSH term “Common Data Elements” is less than a year old and currently only has 3 articles indexed under it. Additionally, the possibility of indexing CDE use raises some initial questions. The first issue relates to the function of indexing and applying MeSH – in general, “aboutness” rather than “use” determines the indexing of articles. A second issue is that more information is needed about *how* CDE use appears in the literature in order to make suggestions regarding its potential for indexing.

Given this information, searches were then conducted in PubMed to determine general literature available about CDEs. Searches were also run for NINDS CDEs, PhenX, and PROMIS as well to capture specific references to these initiatives. See Appendix I – PubMed searches for search strategies. While these search strategies identified articles, initial scans of the publications revealed that the majority of papers were in relation to the development of the 3 initiatives rather than use of these CDE collections in research studies.

Publication review

Following the exploration of PubMed, the Associate turned to FOA numbers, listed on the NIH CDE Portal, to identify a set of publications to analyze for use of CDEs. As discussed in the Background, FOAs are a primary way that NIH CDEs are promoted for use in research by different institutes. From the Portal list, 12 FOA numbers were identified that specified use of either the PhenX, PROMIS, or NINDS CDEs collections. Through use of the NIH RePORTER system, these 12 FOA numbers were connected to grants that then linked to associated publications included in PubMed and/or PubMed Central, which resulted in 1,126 publications. This sample was limited to exclude letters, comments, editorials, introductory journal articles, and news items, which reduced the sample to 1,073 articles. Through a combination of removal due to duplication, lack of access to full text, and convenience sampling; 819 publications were excluded resulting in a final sample of 254 publications.

In addition to locating a sample of publications through FOA numbers, NINDS CDEs, PhenX, and PROMIS each maintain a list of publications or studies on their websites related to their particular initiatives. Between these three lists, 797 publications were identified. Through a combination of removal due to duplication, lack of access to full text, and convenience sampling; 478 publications were excluded resulting in a final sample of 319 publications.

These two samples resulted in a total of 573 publications for review. Publications were initially evaluated for use of CDEs at the abstract level. If there was no mention in the abstract, the full text of the article was reviewed, in particular the Methods and Acknowledgements sections.

Website review

Following the examination of publications, NINDS CDEs, PhenX, and PROMIS websites were reviewed to determine whether guidance was provided to researchers as to how they should cite use of CDEs in

published research. Websites were evaluated based on the presence of language specified for inclusion in publications (or other mediums) about use of CDE collections in research.

Results

Use of CDEs in NIH-funded research

The majority of the project's results came from the review of publications. An examination of 3 NIH CDE initiative websites revealed if and how initiatives provide guidance for citing use of CDEs in published research. Finally, enhancements were made to the NIH CDE Portal as a result of the research conducted for this project.

During the review of publications, categories were developed to group publications based on similar themes. While the publication review was primarily intended to identify "use" of CDEs in the data collection process, there are additional contexts to which CDEs are referenced in the literature. The categories developed include:

- Collection level use – Indicates use of CDE collections in data collection through reference to the initiative name (i.e., NINDS CDEs, PhenX, or PROMIS).
- Instrument level use – Indicates use of CDEs in data collection through reference to instruments such as scored surveys, questionnaires, or validated measures included as part of the NINDS CDEs, PhenX, or PROMIS collections.
- Development – Describes the processes of selecting, identifying, and evaluating CDEs for NINDS CDEs, PROMIS, and PhenX. This category is also used to refer to efforts that used the processes developed by the 3 initiatives to develop their own collections of CDEs.
- Evaluation – Describes processes of validating and assessing CDEs in NINDS CDEs, PhenX, or PROMIS collections. This category is also used to refer to other efforts that duplicated these processes.
- Mention – Makes a brief reference to any of the 3 initiatives usually as examples of how to standardize data collection.

Funding Opportunity Announcements (FOAs) sample

As stated above, the final size for this publication sample was 254 publications. The majority of the 254 publications reviewed as part of this sample did not demonstrate evidence of CDE use in either the abstract or the full text of the publications (i.e., 187 publications). The remaining percentage of publications demonstrated use of two kinds – either at the collection level (i.e., 4 publications) or at the instrument level

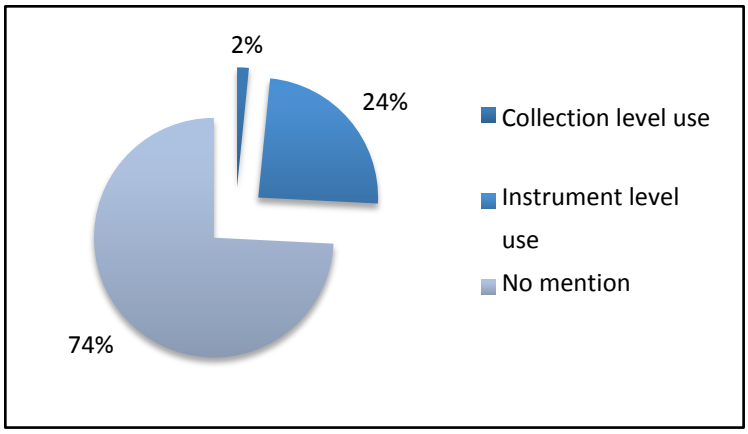

Figure 1: Publications referencing use of CDEs (%)

(i.e., 61 publications) (Figure 1).[3] As mentioned above, collection level use refers to publications that made reference to a CDE collection in gathering data during research. For example, a paper might state: "The NINDS CDE collection for Parkinson's Disease was used to collect information as to the cognitive function of study participants". Instrument level use refers to publications that cite use of a particular tool (included in CDE collections) in collecting information about study participants. Such as, "the Mini Mental Status Examination (MMSE) was used to collect information about the cognitive function of study participants".

Of the 65 publications that indicated use, further information was captured as to where in the publication (i.e., abstract or full text) a reference was made to use of CDEs (Figure 2). The majority of references to use were found in the full text of the article for these publications (i.e., 59 publications). Considering that the majority of mentions to use were at the instrument level for this sample, this finding is not too surprising. The abstracts often give broad overviews of data collection methods whereas within the article there is a higher likelihood of mention to specific tools used to collect data.

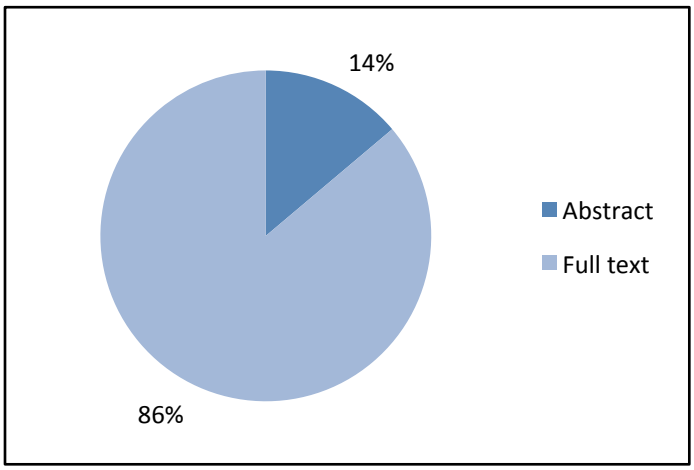

Figure 2: Reference to use in abstract or full text (%)

Figure 3 reflects where the reference to use (instrument or collection level) occurred, in either in the abstract or the full text. From this chart it is clear that references to instrument level use are more frequently found in the full text of the article. Since the number of publications referencing collection level use was so small in this sample (i.e., 4 publications), it is difficult to determine whether or not the abstract or article is a more likely place to find this type of reference.

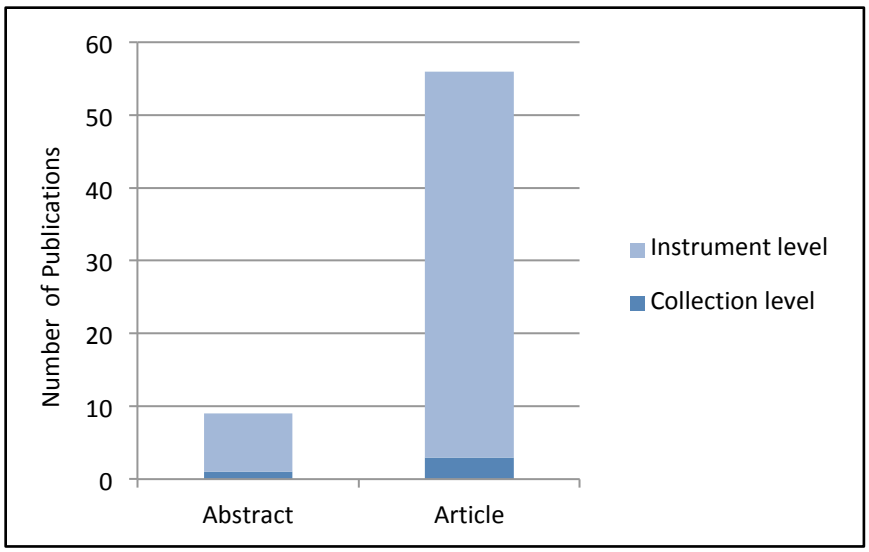

Figure 3: Type of use in abstract or article

Finally, numbers were collected as to where the references appeared in the publications that

[3] Two additional publications fell into an "other mention", non-use category.

required full text appraisal. An overwhelming majority of references to use in full text (95%) were found in the Methods section.

Publication & study lists sample

For this sample, the final sample size was 319 publications. The review of these publications resulted in 25% demonstrating no use of CDEs, 11% demonstrating collection level use, 13% demonstrating instrument level use, and 46% having “other” mentions of CDEs in articles (Figure 4). No mention, again, refers to those publications that did not show evidence of CDE use in either the abstract or the full text. Once again, collection level use garnered the smallest number of references in publications slightly bested by publications referencing instrument level use. The majority of publications in this sample fell into the "Other mention" category. This category refers to publications that mention PROMIS, PhenX, or NINDS CDEs as part of research, but not as part of the data collection process.

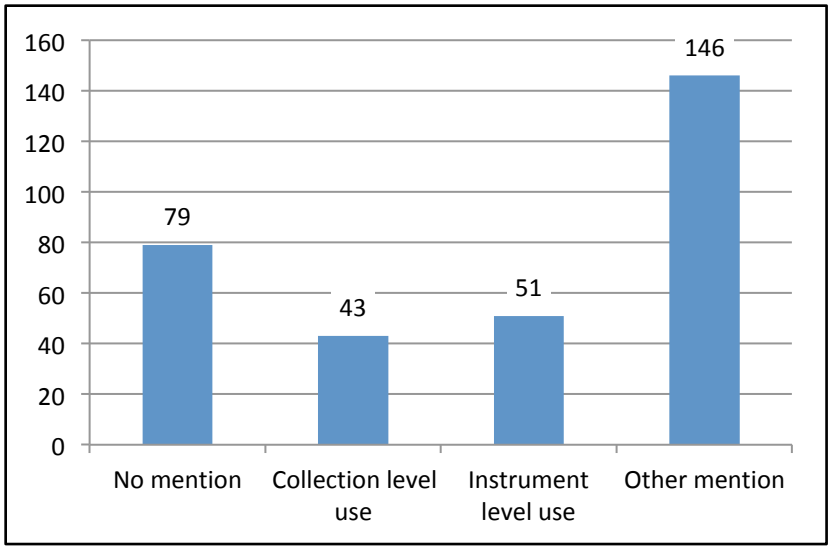

Figure 4: Publications referencing use of CDEs

Furthermore, "Other mention" is made up of three categories: Mention, Evaluation, and Development (Figure 5). Sixty-nine publication (47%) mentioned PhenX, PROMIS, or NINDS CDEs initiatives briefly in the publication generally as a suggested method for standardizing data collection. For example, a publication in this category might state: "In collecting information about study participants in GWAS studies, we recommend use of the PhenX measures". These brief mentions call attention to the initiatives and promote them for use in data collection, but are not papers that demonstrate use of CDEs in research. The second category, Evaluation (29%), contains 42 publications that describe processes of validating and assessing

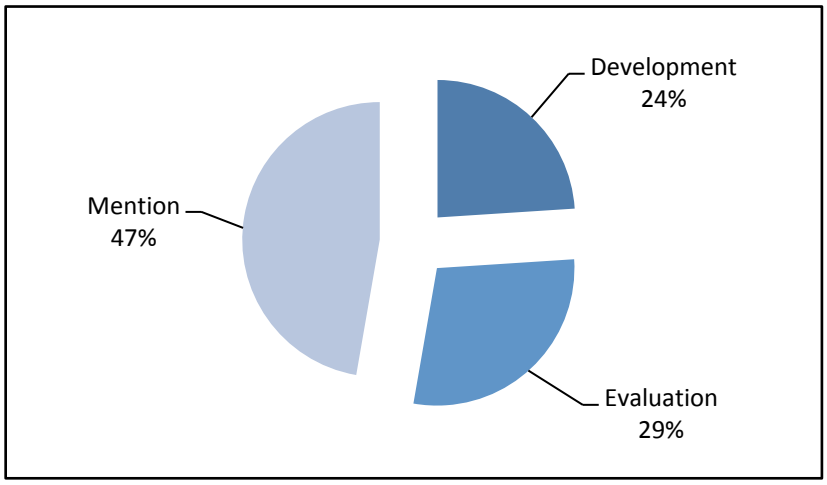

Figure 5: Breakdown of "Other mention" category (%)

CDEs for use in research in the PhenX, PROMIS or NINDS CDEs collections. Finally, the Development category (24%) is made up of 35 publications that describe processes of selecting, identifying, and evaluating CDEs for inclusion in the PhenX, PROMIS, or NINDS CDEs collections. These other types of mentions to CDEs are an important finding because they demonstrate the diversity of reference to CDEs in the literature.

Returning to the use of CDEs, as with the FOA papers, information was collected about the location of references to CDE use in the publications (i.e., collection and instrument level). Figure 6 indicates the breakdown of references in abstract and full text. In contrast to the FOA publication sample, the majority of references to use were found in the abstracts of this sample. The higher instance of "collection level use" might provide insight into this difference. This type of mention to CDE use is referenced at a higher level as opposed to instrument level use, which is more commonly found in the full text of an article.

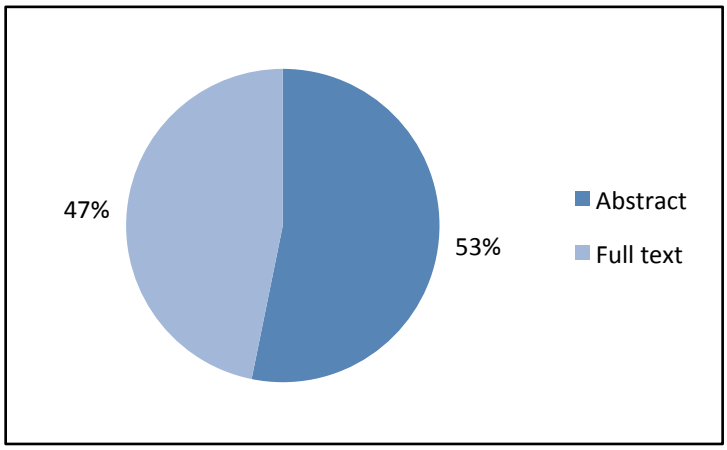

Figure 6: Reference to use: abstract or full text (%)

This is illustrated in Figure 7 – references in abstract broke down as 39 publications that referenced collection level use and 4 publications that referenced instrument level use. In the case of the article, 4 publications were identified as collection level use while 40 publications were identified as instrument level use.

As with the FOA publications, numbers were collected as to where references to use appeared in the publications that required full text appraisal. An overwhelming majority of references to use in full text (91%) were found in the Methods section.

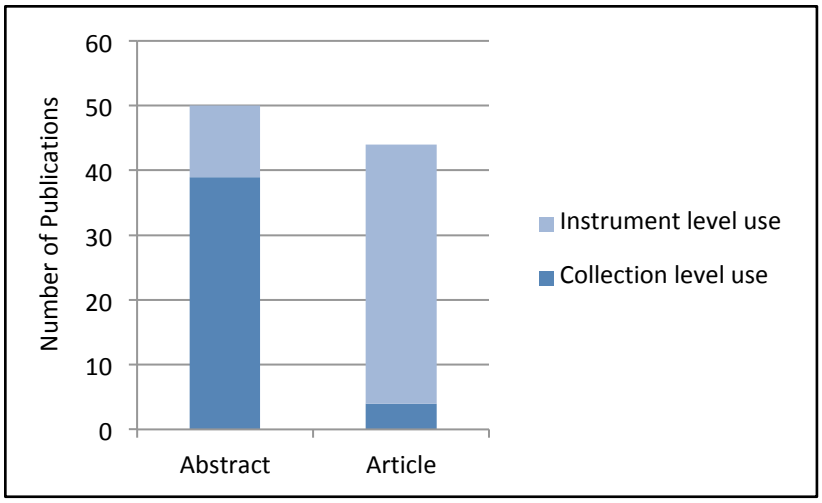

Figure 7: Type of use in abstract or article

Existing guidance for citing use of NIH CDEs

Following the review of publications, the websites for the NINDS CDEs, PhenX, and PROMIS initiatives were evaluated to determine whether or not they included guidance for citing use of their CDE collections in published research. All 3 initiatives provide some type of guidance in citing use of their efforts; however, none of the 3 are standardized.

PROMIS provides language to be included in the Acknowledgments if researchers use their data - PROMIS provides access to public use, research data that can be requested by researchers. The PROMIS websites provides instruction on citing use of that data, but does not provide guidance on citing use of PROMIS CDEs (e.g., surveys, instruments).

PhenX lists on the front page of the PhenX Toolkit website – which provides access to PhenX CDEs – brief instructions for citing in publications which version of the Toolkit was used to find and select PhenX CDEs.

NINDS CDEs has the most comprehensive guidance with instructions for citing in presentations, publications, and a database. In publications, researchers should include a sentence acknowledging use of NINDS CDEs in either their Methods or Acknowledgement sections. Additionally, they should try to incorporate "NINDS CDEs" as a keyword in the publication. In presentations, researchers should acknowledge use of NINDS CDEs at some point. Finally, if depositing their research into CLinicalTrials.gov, researchers should include "NINDS CDEs" as a keyword if they were used in collecting data for the trial.

After reviewing these websites, the Associate also explored whether guidance was provided by the other 8 NIH CDE initiatives. Currently, excepting the 3 already mentioned, only 2 other initiatives seem to have guidance for researchers in citing use of CDEs. This information was obtained through direct communication with program managers as no other NIH CDE initiatives had guidance information available on their websites.

NIH CDE Portal enhancements

As a result of the research conducted, two enhancements were made to the NIH CDE Portal. The first is to the list of current, active FOAs maintained on the NIH CDE Portal. Prior to this project, the list was comprised of IC name, FOA titles, and FOA numbers. During the process of identifying research publications through FOA numbers, the Associate determined which NIH CDE collections were specified for use in the FOAs listed. All FOAs encourage/require researchers to use NIH CDEs in their research, but not all provide direct indication of what CDE collections to use. Thirteen FOAs include general calls for CDE use and indicate the Portal as a place to decide which CDEs to use. The remaining 28 include language specifying use of one or more NIH CDE initiatives (e.g., PhenX, NIH Toolbox, Neuro-QoL). In particular to this project, as stated above, 12 FOAs specify use of NINDS CDEs, PhenX, or PROMIS collections. See Appendix II – Screen shot of FOA List.

The other enhancement was to incorporate a column to the existing Summary Table that links to Studies and Publications made available on NIH CDE initiative websites. Of the 11 initiatives at NIH, 8 have lists of publications or studies pertinent to their CDE initiative. As the research showed, these publications are a mixed bag – most are about development or evaluation of CDEs with a minority demonstrating clear use of CDEs in research. However, linking to these lists from the Portal allows for greater connection to the numerous CDE efforts taking place at NIH and also, moving forward, provides a space for linking to publications or studies that use NIH CDEs in research. See Appendix III – Screen shot of Summary Table.

These enhancements are important because they fit into the goal of the Portal, which is to provide information about ongoing NIH CDE efforts and, in doing so, encourage widespread adoption of CDEs amongst researchers.

Discussion

The results from the publication review clearly demonstrate the difficulties of tracking use of NIH CDEs in the published literature. If the results from the two samples (FOAs and publication/study lists) are aggregated, 46% of the publications reviewed show no evidence of CDE use, 8% show collection level use, 20% show instrument level use, and 26% are categorized as "other mention". In regards to no evidence of use, either researchers are not using CDEs or they are not talking about them and even if they are referencing use it is not in an immediately apparent way. The instrument level use references are at a level of granularity that is not conducive to tracking – particularly, since there are thousands of individual CDEs and hundreds of instruments in NIH CDE initiative collections. Collection level use is the easiest to track due to the presence of the initiative name and the tendency of the reference to be in the abstract. Yet, as the results indicate, this type of use only makes up 8% of the publication samples analyzed. In addition, 26% of the publications reviewed fell into a category that was not about use, but rather CDEs in other aspects of research.

These varied references to CDEs in the literature are one aspect that complicates the process of tracking. Another difficulty is due to the complexities of CDEs themselves. As part of collections, CDEs are curated and vetted to ensure they are relevant to the area of focus. This process is repeated at points to guarantee that the CDEs in initiative collections remain pertinent. As a result, CDEs are often assigned a "version" - versions identify potential changes in the CDEs and are helpful in tracking what iteration of CDEs were used in research. This adds complexity in referencing CDE use because while it may be possible to identify the CDE used, often the version of the CDE is not captured.

Despite these difficulties, this research provides insight into the variety of ways in which CDEs (use or otherwise) are currently referenced in the literature. Additionally, in considering tracking CDE use moving forward, the following recommendations build upon the knowledge gained from this project and provide suggestions to address the variety of ways CDEs appear in research. These recommendations also raise additional questions and challenges with the intent of encouraging further research into this topic.

Recommendations

As this report indicates, there are current efforts by the 3 initiatives explored for this project as well as 2 others to encourage inclusion of language acknowledging use of initiative CDEs. However, none of these efforts are standardized, which presents difficulty if attempts are made to track publications this way. Additionally, based on the publications reviewed, there is no evidence that the guidance provided is followed by researchers. Based on the findings of this project, an immediate recommendation is to develop and implement standardized language for citing use of NIH CDEs in publications. This can take the form of a brief sentence following the format of the NINDS CDEs guidance for publications: “this study used common data elements from [insert initiative]”. This language should also be accompanied

by instruction about where to place the citation in the publication (e.g., Methods, Acknowledgements) – based on the results, the Methods section seems to be the place where reference to CDEs fits most logically. Having this type of standardization would enable use of PubMed and PMC – more likely PMC – to target use of CDEs in the published literature.

Though the project did not focus on CDE use as it appears in datasets, CDEs by definition are pieces of data and are found in databases and repositories that contain datasets. Particularly in light of the increasing NIH requirements for data deposits to further data-sharing, it is important to consider ways in which to standardize tracking of CDEs in databases and repositories that receive data deposits. This could be as simple as having the NIH CDE initiative listed as a key word in the database (e.g., NINDS guidance on ClinicalTrials.gov). However, it may be worthwhile to think about integrating tracking more deeply into these systems. Perhaps having a metadata element that indicates which data standard was used to collect the data deposited into these databases and repositories. There are some questions raised with this suggestion such as when would these metadata elements be integrated? Would they be part of the data submission process as the researcher enters the data? Or applied post-submission as part of the data clean-up process? Further consideration of these questions, amongst others, is necessary in order to seriously consider and employ this recommendation.

In considering the potential for tracking CDEs in databases and repositories, this project also raised some questions around the possibility of assigning unique IDs to CDEs. Having unique identifiers could contribute towards standardized tracking of CDEs in databases and perhaps assist with indexing CDEs in PubMed should that be of interest moving forward. However, this raises additional question as well: how would these unique IDs be assigned? At the collection level? At the instrument level? Who (or what institute) would be in charge of assigning, maintaining, and keeping track of these unique IDs? How would these unique IDs addresses versioning of CDEs? Again, these types of questions would need to be considered before moving forward.

A final recommendation is to the importance of continuing to develop the NIH CDE Portal as a collecting ground for information about NIH CDE efforts. Several outcomes of this project contributed directly to the Portal by increasing and enriching the information available related to different CDE initiatives and funding opportunities. Moving forward, the Portal could be enhanced further with more information about literature published related to CDEs. This could take the form of a bibliography of CDE research - including articles about the development of initiatives as well as links to use of NIH CDE collections in research. The work done are part of this project, including the search strategies listed in Appendix I – PubMed searches, can help gather publications to make up this bibliography. Additionally, it may be of use to develop custom PubMed and PMC search strategies for each NIH CDE initiative – this is currently challenging with some of the initiatives, but may be of use as these initiatives continue to develop and use increases.

Limitations

This research provides strong insight into how CDEs are currently cited and the potential for tracking moving forward. It also demonstrates the difficulties and complexities of tracking CDE use in research, which shows the limitations of the project, primarily in regards to the hand review of publications. The

review of publications was made up of two samples, the FOAs and the publication or study lists on the 3 initiative websites, and each sample has its own set of limitations. With FOAs it is challenging to identify which associated publications document processes of data collection rather than other aspects of the research. Additionally, some of the FOAs that encourage use of PhenX, PROMIS, or NINDS CDEs collections are for sub-projects on grants and it is not possible to limit within NIH RePORTER (or PubMed for that matter) to publications coming from a specific sub-project – instead all publications associated with the grant number (linked to an FOA number) are retrieved when searching using NIH RePORTER.

The publication or study lists made available on the PhenX, PROMIS, and NINDS CDEs websites brought to light other types of mentions to CDEs in the literature aside from use; however, these lists are biased in the manner in which they are compiled. After speaking with people at the 3 initiatives that are responsible for gathering the publications or studies for these lists, it became clear that these lists are developed one of two ways: through keyword searches in major databases or through progress reports. This means that the lists are targeted to the initiatives, which was why this sample was chosen, but it also means that these publications have a higher likelihood of listing the initiative name and not necessarily reflecting "use" of CDEs in data collection (as the results demonstrated). In short, this sample helps reflect the diversity of reference to CDEs in literature, but may be limited in its reflection of CDEs in the published literature because of the manner in which these lists were compiled.

Finally, the category of "instrument level use" was developed to group publications as part of the review of publications. This category, while helpful in identifying a type of reference to CDEs, is difficult to apply because it depends on knowledge of the instruments included in NIH CDE collections. The Associate, while familiar with many instruments in the PhenX, PROMIS, and NINDS CDEs collections due to this research, was not aware of all instruments in these collections. Therefore, the "instrument level use" category may be inconsistently applied to publications in the sample reviewed.

Conclusion

Common data elements strongly correlate to current NIH initiatives related to standardization and data-sharing and, as a result, will continue to develop and grow at NIH. NLM has the opportunity to shape how reference to and tracking of CDEs proceeds through encouraging standardization of citation. This standardization would enable NLM to leverage existing tools and resources to make tracking of CDE use possible. This project also raises questions around the role of NLM in the tracking of CDEs – not only on questions of how to track, but at what level NLM should be involved in this process. NLM continues to be involved with the NIH CDE Portal, the NIH CDE Repository, and the NIH CDE Working Group, but further discussions will be necessary to determine how best to proceed with the tracking of CDE use.

At the NIH level, there are current opportunities to track CDE use through the development of resources for the research community. Several of these come as part of recent Big Data to Knowledge (BD2K) initiatives. The development of the Data Discovery Index (through bioCADDIE) presents an opportunity to integrate acknowledgment of data standards (e.g., CDEs) into databases and repositories. Additionally, with the growth of the Standards Coordinating Centre (SCC), there is the potential to align

CDEs with existing NIH standards related to data and data collection. This association would increase visibility of CDEs at NIH and contribute towards encouraging use of CDEs in research.

Though all of these opportunities exist, they still remain challenging undertakings and will require further research before moving forward. Yet, these are challenges and difficulties that must be met and overcome in order to increase harmonization of data and promote a strong data-sharing culture at NIH.

References

[1] NIH CDE Resource Portal. (2012). *Glossary*. Retrieved from http://www.nlm.nih.gov/cde/glossary.html#cdedefinition.

[2] NIH CDE Resource Portal. (2012). *Summary Table for NIH CDE Initiatives*. Retrieved from http://www.nlm.nih.gov/cde/summary_table_1.html.

[3] NIH CDE Resource Portal. (2015). *Guidance to encourage use of CDEs*. Retrieved from http://www.nlm.nih.gov/cde/policyinformation.html.

[4] Institute of Medicine. (2015). Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25590113.

[5] Agency for Healthcare Research and Quality (AHRQ). (2015). *Systematic Review Data Repository (SRDR)*. Retrieved from http://srdr.ahrq.gov/.

Acknowledgements

Thank you to Jerry Sheehan for his support, guidance, and knowledge offered as the project sponsor. Thank you to Lou Knecht and Dr. Michael Huerta for their guidance and support as project consultants. Thanks also to Liz Amos, Joanna Widzer, Olga Printseva, Mary Ann Hantakas, and the NIH CDE Working Group for contributing their expertise and support as needed during the course of this project. Finally, thank you to Dr. Kathel Dunn, Maureen Madden, Ariel Deardorff, Kristina Elliott, and Lori Harris for their encouragement and guidance.

Appendices

Appendix I – PubMed searches

(common data element[mh] OR common data element[ot] OR common data elements[ot] OR common data element[tiab] OR common data elements[tiab])

(common data element[mh] OR common data element[ot] OR common data elements[ot] OR common data element[tiab] OR common data elements[tiab]) AND ninds

(promis[tw] OR promis[tiab] OR patient reported outcome measurement information system[tw] OR patient reported outcome measurement information system[tiab])

(phenx[tw] OR phenx[tiab] OR consensus measures for phenotypes and exposures[tw] OR consensus measures for phenotypes and exposures[tiab] OR phenx toolkit[tw] OR phenx toolkit[tiab])

Appendix II – Screen shot of FOA List

Home

Guidance to Encourage the Use of CDEs

NIH Institutes and Centers (ICs) use a range of approaches to encourage use of CDEs by funded investigators. Such encouragement may be included in policy documents, NIH Guide Notices, and/or Funding Opportunity Announcements (Requests for Application and Program Announcements). This list below provides a summary of such directives, including a list of active Funding Opportunity Announcements. It will be updated on a periodic basis to reflect new announcements.

Policy Statements

- NCATS Global Rare Diseases Patient Registry Data Repository (GRDRSM) Program Policy Guidance
- Federal Interagency Traumatic Brain Injury Research Informatics System (FITBIR) Data Sharing Policy (27 March 2014)

NIH Guide Notices

- NIDA - Substance Abuse and Addiction (NOT-DA-12-008)
- NIMH – Data Sharing Expectations for NIMH-funded Clinical Trials (NOT-MH-14-015)
- NIMH - Notice Announcing Data Harmonization for NIMH Human Subjects Research via the PhenX Toolkit (NOT-MH-15-009)
- NIMH - National Database for Autism Research (NOT-MH-09-005)
- NINDS - Notice of Intent to Publish a Funding Opportunity Announcement for Detect, Define and Measure the Progression of Chronic Traumatic Encephalopathy (NOT-NS-14-031)

Requests for Applications and Program Announcements

Show 100 entries Search:

IC/Project	Title	FOA Number	NIH CDEs Specified
BD2K	Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science	PA-14-155	Select from Portal
BD2K	Extended Development, Hardening and Dissemination of Technologies in Biomedical Computing, Informatics and Big Data Science	PA-14-156	Select from Portal
Multiple ICs	Connectomes Related to Human Disease	PAR-14-281	Select from Portal
NCATS	Rare Diseases Clinical Research Consortia (RDCRC) for Rare Diseases Clinical Research Network	RFA-TR-13-002	Neuro-QoL, PROMIS
NCI	Molecular and Cellular Characterization of Screen-Detected Lesions	RFA-CA-14-010	Select from Portal
NCI	Small Grants for Behavioral Research in Cancer Control	PAR-12-035	PhenX, PROMIS
NHGRI	Human Heredity and Health in Africa (H3Africa): Collaborative	RFA-RM-12-006	PhenX
NHGRI	Human Heredity and Health in Africa (H3Africa): H3Africa Research Grants	RFA-RM-12-007	PhenX
NHGRI	Genome-wide Association Studies of Treatment Response in Randomized Clinical Trials - Coordinating Center	RFA-HG-08-005	PhenX
NHGRI	Genome-wide Association Studies of Treatment Response in Randomized Clinical Trials-Study Investigators	RFA-HG-08-004	PhenX
NHGRI	The Electronic Medical Records and Genomics (eMERGE) Network, Phase II Coordinating Center	RFA-HG-10-010	PhenX

Appendix III – Screen shot of Summary Table

NIH Common Data Element (CDE) Resource Portal

Home | Resource Summaries | Glossary

Home

Summary Table for NIH CDE Initiatives

This table lists summary information for NIH CDE Initiatives. More information on NIH CDE Initiatives: Subject Areas, Detailed Summaries.

Show 50 entries Search:

Link to Homepage	Link to CDEs	Brief Summary	Number of Elements	Studies and Publications	CDE Resource Contact
Standardized Asthma Outcomes for Clinical Research	Asthma CDEs	The standardized asthma outcomes for clinical research represent recommendations for core (required in future studies), supplemental (to be used according to study aims), and emerging (requiring validation and standardization) outcomes for 7 domains of asthma clinical research outcome measures. Subject Areas | More...	10 (adults), 25 (children)	--	NHLBI, NIAID
Chronic Low Back Pain CDEs	cLBP	Recommended minimum dataset for research on chronic low back pain. Subject Areas | More...	40	--	NCCAM
Early Detection Research Program	EDRN	CDEs for use in describing samples and data collected as part of cancer biomarker research. Subject Areas | More...	1,600	Publications	NCI
eyeGENE	eyeGENE	As part of eyeGENE, common data elements have been developed for collecting phenotypic data associated with more than 30 inherited ophthalmic diseases. Subject Areas | More...	300+	Studies	NEI
Global Rare Diseases Patient Registry and Data Repository	GRDR	CDEs to facilitate standardized data collection into the GRDR and to assist organizations in establishing rare disease registries that contribute information to GRDR. Subject Areas | More...	70	Publications	ORDR
Quality of Life Outcomes in Neurological Disorders	Neuro-QOL	A core set of quality-of-life questions that address chronic neurologic disorders, plus sets of supplemental questions specific to targeted diseases or subgroups of patients. Subject Areas | More...	500	Publications	NINDS
NIDA Substance Abuse Electronic Health Record Data Elements	NIDA EHR	A set of brief screening and initial assessment tools for substance use disorders (SUDs) for use in general medical settings. Subject Areas | More...	80+	--	NIDA