



Published in final edited form as:

*J Nurs Scholarsh.* 2015 September ; 47(5): 379–388. doi:10.1111/jnu.12155.

## Advancing Symptom Science Through Use of Common Data Elements

Nancy S. Redeker, PhD, RN, FAHA, FAAN<sup>1</sup>, Ruth Anderson, PhD, RN, FAAN<sup>2</sup>, Suzanne Bakken, PhD, RN, FAAN, FACMI<sup>3</sup>, Elizabeth Corwin, PhD, RN, FAAN<sup>4</sup>, Sharron Docherty, PhD, RN, FAAN<sup>5</sup>, Susan G. Dorsey, PhD, RN, FAAN<sup>6</sup>, Margaret Heitkemper, PhD, RN, FAAN<sup>7</sup>, Donna Jo McCloskey, PhD, RN, FAAN<sup>8</sup>, Shirley Moore, PhD, RN, FAAN<sup>9</sup>, Carol Pullen, EdD, RN<sup>10</sup>, Bruce Rapkin, PhD<sup>11</sup>, Rachel Schiffman, PhD, RN, FAAN<sup>12</sup>, Drenna Waldrop-Valverde, PhD<sup>13</sup>, and Patricia Grady, PhD, RN, FAAN<sup>14</sup>

<sup>1</sup>Professor, Yale School of Nursing, New Haven, CT, USA

<sup>2</sup>Duke University, Durham, NC, USA

<sup>3</sup>Columbia University, New York, NY, USA

<sup>4</sup>Emory University, Atlanta, GA, USA

<sup>5</sup>Duke University, Durham, NC, USA

<sup>6</sup>University of Maryland, Baltimore, MD, USA

<sup>7</sup>University of Washington, Seattle, WA USA

<sup>8</sup>National Institute of Nursing Research, NIH, Bethesda, MD, USA

<sup>9</sup>Case Western Reserve University, Cleveland, OH, USA

<sup>10</sup>University of Nebraska Medical Center, Omaha, NB, USA

<sup>11</sup>Albert Einstein College of Medicine, New York, NY, USA

<sup>12</sup>University of Wisconsin-Milwaukee, Milwaukee, WI, USA

<sup>13</sup>Emory University, Atlanta, GA, USA

<sup>14</sup>Director, National Institute of Nursing Research, NIH, Bethesda, MD, USA

### Abstract

**Background**—Use of common data elements (CDEs), conceptually defined as variables that are operationalized and measured in identical ways across studies, enables comparison of data across studies in ways that would otherwise be impossible. Although healthcare researchers are increasingly using CDEs, there has been little systematic use of CDEs for symptom science. CDEs are especially important in symptom science because people experience common symptoms across a broad range of health and developmental states, and symptom management interventions may have common outcomes across populations.

**Purposes**—The purposes of this article are to (a) recommend best practices for the use of CDEs for symptom science within and across centers; (b) evaluate the benefits and challenges associated with the use of CDEs for symptom science; (c) propose CDEs to be used in symptom science to serve as the basis for this emerging science; and (d) suggest implications and recommendations for future research and dissemination of CDEs for symptom science.

**Design**—The National Institute of Nursing Research (NINR)-supported P20 and P30 Center directors applied published best practices, expert advice, and the literature to identify CDEs to be used across the centers to measure pain, sleep, fatigue, and affective and cognitive symptoms.

**Findings**—We generated a minimum set of CDEs to measure symptoms.

**Conclusions**—The CDEs identified through this process will be used across the NINR Centers and will facilitate comparison of symptoms across studies. We expect that additional symptom CDEs will be added and the list will be refined in future work.

**Clinical Relevance**—Journal of Nursing Scholarship, 47:5, ©2015 Sigma Theta Tau International.

### Keywords

Cognition; fatigue; mood; pain; sleep disturbance; symptom management

---

Research on the management of symptoms is a primary focus of the National Institutes of Health (NIH) National Institute of Nursing Research (NINR, 2011). NINR-supported research has already contributed in major ways to managing symptoms and improving quality of life and overall function, especially among individuals and families living with acute and chronic conditions. Symptoms are a primary research focus of seven currently (2014) funded NINR-funded centers, including exploratory centers (P20) and centers of excellence (P30) designed to further build symptom science. Effective leveraging of resources and collaborations between researchers and across institutions are critical to a recently published Logic Model for Center Sustainability authored by the Center directors (Dorsey et al., 2014).

Sharing symptom data across studies is important to leveraging the results of research in a cost-effective manner. However, symptom data obtained in a single study may not be comparable across studies, thus hampering the ability to generalize research findings and compare the effects of treatments and characteristics of samples in different settings. Use of common data elements (CDEs), defined as fundamental, logical units of data pertaining to one kind of information that are clearly conceptualized (Cohen, Thompson, Yates, Zimmerman, & Pullen, 2015; Meredith, Zozus, Wilgus, & Hammond, under review), improves efficiency and data quality. CDEs are especially valuable in symptom science because people with a broad range of acute and chronic conditions, developmental stages, and ethnic backgrounds often experience common symptoms (e.g., anxiety, depression, fatigue, pain, sleep disturbance). Interventions (e.g., exercise) focused on symptoms may have common outcomes across populations (e.g., cancer, heart disease).

Although researchers across disciplines are increasingly using CDEs, as exemplified by initiatives of the National Institute of Neurological Disease and Stroke (NINDS, 2013),

National Cancer Institute (NCI, 2014), and the National Institute on Drug Abuse (NIDA, 2014), there has been little systematic use of CDEs to support symptom science.

The purposes of this article are to (a) recommend best practices for the use of CDEs for symptom science within and across centers; (b) evaluate the benefits and challenges associated with the use of CDEs for symptom science; (c) propose CDEs to be used in symptom science to serve as the basis for this emerging science; and (d) suggest implications and recommendations for future research and dissemination of CDEs for symptom science.

## **Best Practices for Developing, Managing, Selecting, and Using Common Data Elements**

Best practices for using CDEs include developing or identifying and selecting CDEs; creating or choosing and managing formats and electronic platforms for data collection and sharing; and assuring quality and administrative oversight to provide access to the data. These practices maximize the quality of the data and successful and efficient data sharing across studies.

The symptom CDE initiative described in this article addresses a unique unmet need. However, best practices used by other NIH units informed our work. Several organizations described collaborative CDE development processes (e.g., NIH-NINDS, NIH-NCI, NIH-NIDA, National Data Base for Autism Research) and identified CDEs appropriate to the research associated with their missions. Although some of the CDEs include symptoms, none of these initiatives focused specifically on symptom science.

The process of developing and making CDEs available is ideally transparent, inclusive, and involves identifying, developing, and vetting CDEs by national and international experts in the scientific community. For example, the iterative process that NINDS used to develop and refine CDEs (NINDS, 2013) occurred over 12 to 18 months. It included convening a working group, subdividing the working group based on areas of need, holding an introductory meeting, developing CDEs for assigned areas by subgroups, reviewing the work of all the subgroups, revising the CDEs based on feedback, obtaining public review of the identified CDEs, revising the CDEs based on feedback, and posting the first versions of the CDEs on the website. The iterative process and stakeholder engagement are critical elements to successful use of CDEs. The input of stakeholders in the development and use of CDEs promotes harmonization nationally and internationally (Choquet et al., 2014).

### **Identifying and Selecting CDEs**

Primary considerations when identifying and selecting CDEs include clear definitions of the concepts (i.e. symptoms); alignment between the selected measures and the concepts of interest; alignment between the selected CDEs and the study aims; and parsimony in the choice of measures to reduce costs and respondent burden (Cohen et al., 2015; Table 1). The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC; ISO, 2014) specified a framework for defining the types and quality of metadata (i.e., representations and meanings of the data) and methods of management and

administration of a metadata registry. Organizations who wish to develop and use CDEs use the ISO-IEC standards as the foundation for a conceptual understanding of metadata and metadata registries. Section 3.3.8 of the ISO-IEC 11179 defines a CDE as the fundamental unit of data, with a specified set of attributes that delineate the definition, identification, representation, and permissible values (ISO, 2014).

Data elements include concepts or abstract units of knowledge (Nadkarni & Brandt, 2006) and representations (value domains and units of measure). Value domains include permissible values, such as minimum and maximal values, decimals, type of data, and other characteristics. In some cases, only one value domain and set of permissible values are meaningful for a given concept (e.g., pain frequency: never, sometimes, often, always; Nadkarni & Brandt, 2006). In other cases, an abstract concept may be described by more than one domain (e.g., pain frequency: never, sometimes, often always; and severity: mild, moderate severe). These value domains may also be used across CDEs. For example, the value domains for frequency and severity might be used for both the symptoms fatigue and pain. Value domains, concepts, data elements, and choices (the possible list of enumerated values) are logically connected to one another (Nadkarni & Brandt, 2006).

### Electronic Platforms for Data Sharing

CDE websites should include lists of the CDEs; data dictionaries (readable and descriptive long names, abbreviated short names, specification of data types [single item, multiple item, bio specimen, biomarker], coding protocol [numeric, string, item, scale, other], and permissible values), case report form modules; references to instruments; and procedural or guideline documents. Best practices for electronic data sharing also include recording the conditions under which the CDEs should be collected and stored; recording the use of each CDE and any relevant conditions of use; and creation of a data file of CDEs used that is uploaded to the CDE web site. The data repository must be maintained, and data management and informatics expertise is required to collect and merge symptom data from multiple sources. The repository should be “user friendly” to maximize the benefits of unique queries of the data (Corwin et al., 2014).

Practices for re-use of electronic clinical data for research have been identified (Safran et al., 2007). There is a need to identify the appropriate organization to manage and host the website and data activities and to develop rules and procedures for submitting and updating and refining the CDEs. Funding is required for sample banking, data banking, and data sharing systems. Current firewalls and other institutional limitations to data sharing need to be addressed. A system such as the University of Maryland Baltimore Novell Vibe software that allows groups to work in a secure, Health Insurance Portability and Accountability Act (HIPAA), compliant cloud may be useful in supporting collaboration on use of CDEs for symptom science.

Several U.S. initiatives stimulated the reuse of electronic clinical data in general, and CDEs in particular. These included requirements for meaningful use of electronic health records (EHRs), NIH’s Clinical and Translational Sciences program recommendation for the use of common evaluation metrics, Agency for Health Care Quality and Research (AHRQ) funded initiatives for comparative effectiveness research and patient registries, and PCORnet, a set

of clinical data research networks and patient-powered research networks funded by the Patient Centered Outcomes Research Institute. Hersh and colleagues (2013) emphasized the opportunities for using operational EHR data in comparative effectiveness research, while emphasizing the need for attention to the difference in goals of data collection for research in clinical practice, coding practices, and other issues of data quality that are directly relevant to CDEs for symptom science. Implementation of recommendations to assure data quality, relevance, and usefulness requires domain expertise (i.e., symptom expertise of a nursing scientist) and informatics expertise; collaboration among individuals with these areas of expertise is essential. The specific skills and activities needed are depicted in Table 2.

### **Data Quality and Administrative Oversight**

CDE databases and websites must have structures and administrative oversight that allow them to be easily available to end-users at low or no cost (Nadkarni & Brandt, 2006). There must be an approved vetting process for the addition of new or revised CDEs as they change over time and a quality assurance system to maintain standards and controls of the data. Common electronic data collection forms are needed to help end-users easily comply with the underlying standards for metadata (ISO, 2014).

Collaborative approaches are required to support large-scale data storage and management and multi-user access to CDEs. Some systems include features that support data collection. Two systems are particularly relevant to advancing symptom science: Research Electronic Data Capture (REDCap; 2014) and Patient Reported Outcomes Measurement Information System (PROMIS®; 2014).

Originally supported through the Clinical and Translational Award mechanism, REDCap (2014) is a secure web application designed for building and managing online surveys and databases. Surveys and databases can be designed by using the web-based online designer or by constructing a “data dictionary” template in Microsoft Excel that can be uploaded into REDCap. These two approaches can be combined as needed to support a specific project. Features that support collaborative research with CDEs include the ability to define access by multiple users for data capture, manipulation, and export. Automated export procedures for data downloads to Excel, PDF, and common statistical packages (SPSS, SAS, Stata, R) are available. Additional features include support for branching logic and calculated fields; scoring algorithms for common standardized instruments; and audit trails for tracking data manipulation and user activity. The REDCap Consortium includes more than 1,000 institutional partners in 85 countries and an active user group that continues to refine it in collaboration with REDCap developers.

Funded by the NIH, PROMIS (2014) is a collection of standardized measures of physical, mental, and social well-being. Many of these measures are designed to elicit symptoms. Although the measures can be administered in paper-based formats and stored and managed by individual researchers or institutions, the PROMIS Assessment Center is a web-based research management tool that provides functionality that can support cross-site collaboration. Through the Assessment Center’s software, the research team can design and implement study-specific web pages for data collection with PROMIS measures and other

CDEs. A significant advantage of the PROMIS Assessment Center is the availability of computerized adaptive testing. Both patient- and researcher-completed forms are supported. PROMIS also offers automated data export features, automated generation of NIH-compliant enrollment tables, and separate export of patient-reported outcome, registration, and consent data.

## Benefits of CDEs

CDEs have many benefits, including comparison of data across studies, access to larger samples than would otherwise be possible, efficiency, and cost-effectiveness. CDEs facilitate data sharing and aggregation within and across disciplines and within and across centers. Identifying and using CDEs will support secondary data use, that is “collect once and use many,” an approach to standardization that spans silos in primary and secondary data uses (Nahm et al., 2010). The ability to pool data across populations and settings will likely improve statistical power by increasing sample size and enable comparison across groups and subgroups, each of which might be too small for analysis if data were limited to studies conducted by individual research teams or at single sites. Use of larger samples will also permit use of more sophisticated multivariate statistical methods and data analytics than might otherwise be possible. These may facilitate subgroup analyses and evaluation of mechanisms and interactions, such as biobehavioral mediators and moderators that might not otherwise be possible with smaller samples, but are especially important to understanding the complex interactions among symptoms and health.

Use of CDEs enables faster start-up of a study because there is no need to create new measures, unique data collection forms, determine acceptable response ranges, reliability, or validity, or to develop new logical data checks. Availability of standardized data frees up oversight committees (e.g., data and safety monitoring boards) to focus on safety and other important study design issues, rather than measurement concerns. Cost effectiveness is likely to improve given the increased feasibility of access to data from large samples without the need to recruit additional subjects at a specific site.

Ultimately, the use of CDEs will help determine how symptoms or symptom clusters vary depending on the disorder or other specific characteristics of the respondents and whether common biological or behavioral factors across conditions contribute to those symptoms. CDEs will allow researchers to identify genotypes or phenotypes of risk and better estimate the influence of baseline participant characteristics on the risk of developing a particular symptom. For instance, if three separate research teams focused on cognitive decline in different populations (e.g., adults with heart failure, adults with HIV/AIDS, and cancer survivors), together they might be able to identify a phenotype of risk for cognitive decline by assessing CDEs of altered cognition across these groups. Lack of family support, stress, gender, unemployment, comorbidity, or other factors may explain group-related differences. If common genetic polymorphisms related to neuroplasticity (Hariri et al., 2003) were available with CDEs for cognitive function, genetic risks for cognitive decline related to chronic conditions might also be identified. Targeted and early interventions might be focused on risk phenotypes or genotypes identified a priori.

## Challenges to The Use of CDEs for Symptom Science

Challenges to the development and use of CDEs for symptom science are conceptual and pragmatic. Although it might be tempting to approach the measurement of symptoms by selecting items that demonstrate high correlation with objective, biological, or observable indicators, this approach conflicts with the subjective and perceptual nature of symptoms and with science that has often found that symptoms are not closely correlated with “objective” indicators. For example, it might seem desirable to select symptom measures that correlate well with measures of actual performance, such as time taken or level of effort required. However, if patient ratings could be expressed in terms of objective performance and health indicators, self-reported information on symptoms would not be necessary. Differences in patient ratings of symptoms, given similar severity of illness, are not errors of measurement. Rather, understanding how and why patients experience similar health states in different ways is a central challenge to symptom science.

Over the past decade or so, researchers have developed CDEs based on the analysis of item banks, expert judgment, and empirical examination of the psychometric properties of symptom measures. For example, in the NIH’s PROMIS initiative, investigators used item response theory (IRT) to establish symptom measures that map onto specific symptoms or quality-of-life dimensions, while eliminating items that were not monotonically associated with the underlying dimensions. Items selected using IRT are associated with different levels of item difficulty that elicit an individuals’ level of depression or other symptoms with a high degree of accuracy and efficiency. While measurement bias can be removed by eliminating items whose performance differs between populations, this may result in eliminating items that reflect dimensions of symptoms not used to characterize severity. For example, depressive symptoms include cognitive (e.g., negative affect) and somatic (e.g., fatigue, loss of appetite) symptoms. These variations may be missed when using a monotonic rather than a multidimensional approach, and may mask the complex nature of symptoms.

Item development with IRT may also obscure the multidimensional nature of cultural factors that may contribute to symptoms. Elimination of items that function differently across cultures may yield a scale free of “measurement bias,” but may also distort meaning and impede communication about multidimensional symptoms. Rather than selecting CDEs focused solely on narrowly defined symptoms, it will likely be necessary to include broader sets of indicators that represent the multidimensional nature of experience. This trade-off has been described as the “bandwidth-fidelity” problem in psychological assessment (Schwartz & Rapkin, 2004).

Understanding of individuals’ histories and perspectives is needed to interpret their current concerns. For example, an individual may come to tolerate a level of pain that was previously highly distressing and debilitating as time progresses. Although the objective level of pain did not change, expectations, habituation, or ability to cope may change and contribute to perception. Individuals may also adjust self-reported symptoms according to the demand characteristics of the clinical situation. Self-report may depend on understanding the context and implications of individuals’ symptoms: What were the behaviors occurring

at the time of the symptoms (e.g., physical activity, eating)? Will greater symptoms lead to increased care or discontinuation of treatments due to toxicity? Will caregivers be activated or disappointed? Will explanations to new, unfamiliar care providers differ from the shorthand used by patients in longer-term relationships? Nurses recognize the necessity of accounting for these nuances when communicating with individuals about their health. These complexities cannot and should not be ignored when measuring symptoms. Thus, the nature of people's responses to symptom scales must be an intrinsic part of the assessment of the usefulness of CDEs.

Rapkin and Schwartz (2004) developed a model to account for interindividual differences and intraindividual changes in appraisal of quality of life (Wyrwich & Tardino, 2006) that is relevant to the use of symptom measures. They posited that constructs that involve subjective evaluation, such as symptoms, have four aspects: goals, priorities, and concerns that comprise the individuals' frames of reference (Jobe, 2003); the ways that individuals sample experiences within pertinent time frames of reference (Alfano et al., 2009; Cohen et al., 1998; Mezuk et al., 2010; Rabiau, Knauper, & Miquelon, 2006; Suls & Fletcher, 1985); the standards of comparison that they consider in evaluating these experiences, including past history (Allison, Locker, & Feine, 1997; Bernhard, Hurny, Maibach, Herrmann, & Laffer, 1999), perceptions of salient others, social norms (Hoeymans, Feskens, Kromhout, & van den Bos, 1997; Stanton, Danoff-Burg, Cameron, Snider, & Kirk, 1999), and personal ideals; and the ways that they formulate summary judgments and reconcile disparate experiences (Rabiau et al., 2006). These parameters are likely to be useful in understanding symptoms when unidimensional measures such as PROMIS are used. They will allow investigators to zero in on the parts of the broader multidimensional symptom scope for each individual. Current assessment of appraisal parameters is lengthy, entails specialized interviewer training, and uses open-ended qualitative responses that require coding. There is a continuing need for the development of appraisal measures that are more suitable for broad dissemination as CDEs and can be used to explain the multidimensional symptom experience.

Investigators may need to employ measures that they have used in the past (i.e., "legacy measures") to permit comparison with past or ongoing research in which the specific CDEs were not employed, as well as CDEs to enable comparison across new studies. The addition of CDEs may result in additional respondent burden or excessive costs, especially when participants are very ill or recovering from surgery or an illness. This is an ethical concern, and potential subjects may decide not to participate in a study if they believe that the assessments are too time-consuming or burdensome. Institutional review boards may also decline to approve studies that have high levels of participant burden relative to benefit. Taken together, none of these challenges are insurmountable, but must be carefully considered by individual investigators and centers and addressed as the emphasis on CDEs for symptom science progress. We expect that there will be efforts to harmonize the CDEs with legacy measures of symptoms in the future.



## Common Data Elements for Symptom Science

In 2014, the NINR Center directors, including the authors of this article, proposed an initial set of CDEs for symptom science. Through a series of telephone conferences and a workshop, the Center directors, including experts on specific symptoms, arrived at consensus regarding recommendations for CDEs for symptom science. We used best practices (see Table 1) and criteria for CDE identification and selection, based on expert advice, previous NIH initiatives, and the literature. We initiated our work by selecting a list of symptoms of primary relevance to our centers (pain, sleep, fatigue, and affective and cognitive symptoms), developed extensive lists of reliable and valid measures for each symptom from those used in our centers, and organized them into a matrix. The team reviewed and revised the list of measures in an interactive and iterative process. The initial list, developed over several phone conferences, included scales measuring pain ( $n = 13$ ) and self-reported or objective measures of sleep ( $n = 11$ ), fatigue ( $n = 4$ ), and affective ( $n = 15$ ) and cognitive symptoms ( $n = 15$ ). The committee organized these lists into submatrices and worksheets to support the next stage of the process.

The directors worked in subgroups to refine a list of common measures for each specific symptom. The goal of each group was to choose one or two “best” measures that could be used across studies for symptoms of interest. In addition to documented best practices, work groups considered the need for alignment with study aims; participant burden (e.g., length, number of items); cost; usability; feasibility; and relevance across diagnostic, developmental, and cultural subgroups. We did not select instruments that can only be administered by registered and trained providers or those that require payment for a license because these requirements could increase costs and may be prohibitive in some studies or settings. For example, the gold standard for cognitive assessment is a battery of tests that require training for those who administer them and often large amounts of time to administer. Depending on available resources, it may not be feasible or cost effective to use this battery. In this case, simpler and less expensive common measures were identified for studies that do not require this detailed cognitive assessment.

We considered a wide range of possible study aims for measuring symptoms in nursing science and evaluated the potential consistency of chosen measures for their fit across study aims. For example, if the purpose of a particular study was to obtain detailed knowledge about a particular symptom, the investigators would use the recommended common measure and additional measures to explore multiple attributes of the symptom (e.g., intensity, frequency, burden). On the other hand, if the study aim was to understand symptom responses to a chronic condition or a cluster of symptoms, single common measures of several symptoms (e.g., pain, fatigue, and sleep disturbance) might be more useful.

The Center directors also recognized that a host of demographic, clinical environmental, and social factors influence symptoms and are important to understanding the data obtained with CDEs. They identified age, gender, race, ethnicity, education, socio-economic status, and occupation as a minimum set of sociodemographic variables that should be collected and reported in a consistent format along with CDEs.

The recommended symptom CDEs are presented in Table 3. This is a working document that the Center directors expect to be refined over time, with additions, deletions, and other changes. A first attempt at this refinement was the presentation of a podium session and request for feedback at the 2014 Council for Advancement of Nursing Science Scientific Sessions. Audience members supported the CDE efforts, and many indicated that it would benefit their work and the advancement of symptom science. For example, greater scrutiny regarding the value of inclusion of both the PROMIS Well Being and Positive Affect measures was requested, as was a concern that neither of these measures was developed using IRT. Additional feedback was solicited from scientists at the Eastern, Western, Southern, and Midwest Nursing Research Societies' Annual Scientific Sessions in 2015, and the directors plan to continue to develop and evaluate the use of CDEs in the future.

In addition to the CDEs themselves, understanding the data obtained would likely benefit from data about the context of symptom assessment that may affect interpretation. For example, it is important to consider the demographic composition of the community in which the research is conducted; how and where measures were administered; the nature of the treatment events or the patient behaviors at the time of data collection; privacy; involvement of family members or proxies in interview; placement of CDEs in a battery including other measures; overall survey length; and whether and when respondents previously completed the CDEs. Measures of the assessment situation may vary from study to study and even among participants in the same study, but should be relatively straightforward to report. These data may help to explain unexpected findings or differences across samples and can potentially be used as covariates in the analyses of factors that contribute to symptoms.

Given the importance of biomarkers to emerging symptom science, the Center directors agreed that CDEs should be developed to address these critical data elements. However, the current state of the science does not permit identification of a parsimonious set of standard biological data elements. Therefore, the directors recommend obtaining biological specimens (i.e., blood, urine, saliva, stool, hair) that can be used for future analysis to better understand the role of specific biomarkers to symptom phenomena. For example, there are specific gene-gene interactions that place individuals at risk for increased pain sensitivity or greater fatigue after chemotherapy at particular times of day. The development of CDEs for these biomarkers will be addressed in more detail as the science advances.

The work reported in this article represents a first step in the use of CDEs for symptom science. Use of CDEs for symptom science by NINR centers is at a formative stage, and more work will be required to finalize, add additional CDEs for other symptoms, and refine the CDEs through further dialogue and stakeholder feedback; develop websites with specified CDE definitions and protocols for data collection and variable coding; and develop and implement quality assurance strategies and administrative processes to manage and coordinate the CDE data and related activities.

## Implications for Future Research and Use of CDEs for Symptom Science

Understanding the nature of symptoms, their contributions to health and quality of life, and the development of symptom interventions is a critical element of the NINR strategic plan, and NINR-funded scientists have made enormous advances in this area. Refinement of the CDEs described here and development of additional symptom CDEs, including CDEs for symptoms among children, are goals for future work. Given differences in contextual variables and appraisal of symptoms between individuals, there is a need to develop ways of assuring that these differences are accommodated. This may include measurement of CDEs at the point of care or “bedside” with electronic technology or other methods with documentation about the conditions under which the CDEs were obtained. Given the widespread use of “legacy” measures of specific symptoms, there is also a need to harmonize the use of these measures with emerging CDEs.

Development of an infrastructure to support the CDEs is a critical step in advancing this line of research. A data repository will be needed for the collection of CDEs across centers that will provide effective ways to ask and answer questions that benefit public health. Given the high costs of developing and maintaining the infrastructure to support a CDE repository, it is possible that one or more institutions could leverage existing and new technology, in partnership with NINR, to support national efforts. This effort might be similar to the Human Microbiome Project, in which data are housed at one institution, but freely available to all collaborative partners. In the context of NIH research and training initiatives related to Big Data to Knowledge, it is vital that data central to the symptom experience, its management (including self-management), and symptom outcomes are captured and analyzed to complement other data sources.

While the CDE initiative described here was driven by the goals of the centers funded by the U.S. NINR, the need for CDEs for symptom science is international in scope. Use of electronic platforms is likely to facilitate international use of CDEs for symptom science, such as the efforts of the international CDE effort on traumatic brain injury (NINDS, 2015). Translation of the symptom measures into multiple languages will be needed, and understanding the healthcare contexts in which the symptoms are elicited will be particularly important, given the differences in healthcare delivery models and access around the world.

## Conclusions

The use of CDEs for symptom science presents important opportunities to leverage resources across NINR centers and beyond. Advancing this work will involve continued refinement of the measures, harmonization with legacy measures, and providing electronic platforms for data acquisition and analysis. CDEs will advance the goals of improving symptoms, health promotion, and improving quality of life for people with acute and chronic conditions.

## Acknowledgments

This work was supported through the following grants: P30NR014134 (Waldrop-Valverde, principal investigator [PI]); P30NR010677 (Bakken, PI); P20NR01426 (Redeker, PI); P30NR011396 (Dorsey, PI); P30NR014129 (Dorsey, PI); P30NR011400 (Heitkemper, PI); and P30NR014139-01 (Anderson/Docherty, PI).

## References

- Alfano CM, Day JM, Katz ML, Herndon JE 2nd, Bittoni MA, Oliveri JM, Paskett ED. Exercise and dietary change after diagnosis and cancer-related symptoms in long-term survivors of breast cancer. *Psychooncology*. 2009; 18:128–133. [PubMed: 18536022]
- Allison PJ, Locker D, Feine JS. Quality of life: A dynamic construct. *Social Science and Medicine*. 1997; 45:221–230. [PubMed: 9225410]
- Barton C, Kalleem C, Van Dyke P, Mon D, Richesson R. Demonstrating "collect once, use many"—Assimilating public health secondary data use requirements into an existing domain analysis model. *AMIA Annual Symposium Proceedings*. 2011:98–107. 2011. [PubMed: 22195060]
- Bernhard J, Hurny C, Maibach R, Herrmann R, Laffer U. Quality of life as subjective experience: Reframing of perception in patients with colon cancer undergoing radical resection with or without adjuvant chemotherapy. *Annals of Oncology*. 1999; 10:775–782. [PubMed: 10470423]
- Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. *Journal of the American Medical Informatics Association*. 2014; 22(1):76–85. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/25038198>. [PubMed: 25038198]
- Cohen MZ, Thompson CB, Yates B, Zimmerman L, Pullen CH. Implementing common data elements across studies to advance research. *Nursing Outlook*. 2015; 63(2):181–188. [PubMed: 25771192]
- Cohen S, Frank E, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM Jr. Types of stressors that increase susceptibility to the common cold in healthy adults. *Health Psychology*. 1998; 17:214–223. [PubMed: 9619470]
- Corwin EJ, Berg JA, Armstrong TS, DeVito Dabbs A, Lee KA, Meek P, Redeker N. Envisioning the future in symptom science. *Nursing Outlook*. 2014
- Dorsey SG, Schiffman R, Redeker NS, Heitkemper M, McCloskey DJ, Weglicki LS. National Institute of Nursing Research Center Directors. National Institute of Nursing Research Centers of Excellence: A logic model for sustainability, leveraging resources, and collaboration to accelerate cross-disciplinary science. *Nursing Outlook*. 2014
- Hariri AR, Goldberg TE, Mattay VS, Kolachana BS, Callicott JH, Egan MF, Weinberger DR. Brain-derived neurotrophic factor val66met polymorphism affects human memory-related hippocampal activity and predicts memory performance. *Journal of Neuroscience*. 2003; 23:6690–6694. [PubMed: 12890761]
- Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, Saltz JH. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical Care*. 2013; 51(8, Suppl. 3):S30–S37. [PubMed: 23774517]
- Hoeymans N, Feskens EJ, Kromhout D, van den Bos GA. Ageing and the relationship between functional status and self-rated health in elderly men. *Social Science and Medicine*. 1997; 45(10): 1527–1536. [PubMed: 9351142]
- International Organization for Standardization. ISO/IEC 11179-1:2004 Information technology—Metadata registries (MDR). 2014. Retrieved from [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=35343](http://www.iso.org/iso/catalogue_detail.htm?csnumber=35343)
- Jobe JB. Cognitive psychology and self-reports: Models and methods. *Quality of Life Research*. 2003; 12:219–227. [PubMed: 12769134]
- Meredith N, Zozus MN, Wilgus RI, Hammond WE. Considerations for data element identification and definition. *Methods of Informatics in Medicine*. (Under review).
- Mezuk B, Rafferty JA, Kershaw KN, Hudson D, Abdou CM, Lee H, Jackson JS. Reconsidering the role of social disadvantage in physical and mental health: Stressful life events, health behaviors,

race, and depression. *American Journal of Epidemiology*. 2010; 172:1238–1249. [PubMed: 20884682]

Nadkarni PM, Brandt CA. The common data elements for cancer research: Remarks on functions and structure. *Methods of Informatics in Medicine*. 2006; 45(6):594–601.

Nahm M, Walden A, McCourt B, Pieper K, Honeycutt E, Hamilton CD, et al. Standardising clinical data elements. *International Journal of Functional Informatics and Personalised Medicine*. 2010; 3:314–341.

National Cancer Institute. ca-BIG NCI data standards. 2014 Retrieved November 2, 2014, from ca-BIG NCI Data Standards.

National Institute on Drug Abuse. Common data elements. 2014. Retrieved November 2, 2014, from <http://www.drugabuse.gov/about-nida/organization/cctn/ctn/resources/common-data-elements-cde>

National Institute of Neurological Diseases and Stroke. NINDS common data elements. 2014. Retrieved from <http://www.commondataelements.ninds.nih.gov/#page=Default>

National Institute of Neurological Diseases and Stroke. Common data elements for traumatic brain injury. 2015. Retrieved from [http://www.commondataelements.ninds.nih.gov/tbi.aspx#tab=Data\\_Standards](http://www.commondataelements.ninds.nih.gov/tbi.aspx#tab=Data_Standards)

National Institute of Nursing Research. Bringing science to life: National Institute of Nursing Research strategic plan (11-7783). Bethesda, MD: 2011. Retrieved from <https://www.ninr.nih.gov/sites/www.ninr.nih.gov/files/ninr-strategic-plan-2011.pdf>

National Library of Medicine. Common data elements resource portal. 2014. Retrieved November 2, 2014, from <http://www.nlm.nih.gov/cde/>

Patient Reported Outcomes Measurement Information System. PROMIS. 2014. Retrieved November 2, 2014, from [www.nihPROMIS.org](http://www.nihPROMIS.org)

Rabiau M, Knauper B, Miquelon P. The eternal quest for optimal balance between maximizing pleasure and minimizing harm: The compensatory health beliefs model. *British Journal of Health Psychology*. 2006; 11:139–153. [PubMed: 16480560]

Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health and Quality of Life Outcomes*. 2004; 2:14. [PubMed: 15023229]

Research Electronic Data Capture. REDCap. 2014. Retrieved November 2, 2014, from <http://www.project-redcap.org/>

Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Expert P. Toward a national framework for the secondary use of health data: An American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*. 2007; 14:1–9. [PubMed: 17077452]

Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes*. 2004; 2:16. [PubMed: 15038830]

Stanton AL, Danoff-Burg S, Cameron CL, Snider PR, Kirk SB. Social comparison and adjustment to breast cancer: An experimental examination of upward affiliation and downward evaluation. *Health Psychology*. 1999; 18:151–158. [PubMed: 10194050]

Suls J, Fletcher B. Self-attention, life stress, and illness: A prospective study. *Psychosomatic Medicine*. 1985; 47:469–481. [PubMed: 4059480]

Wyrwich KW, Tardino VM. Understanding global transition assessments. *Quality of Life Research*. 2006; 15:995–1004. [PubMed: 16900280]

**Table 1****Best Practices for Identifying and Managing Common Data Elements (CDEs)**

<p>Criteria for selection of CDEs (Barton, Kallem, Van Dyke, Mon, &amp; Richesson, 2011; International Organization for Standardization, 2014)</p> <p>Conceptual consistency of measure with conceptualization of symptom of interest. Parsimony in measures. Fit with study purpose.</p>
<p>Practices for group processes in identifying CDEs (Choquet et al., 2014; Nahm et al., 2010)</p> <p>Convene a working group and subdivide based on areas of need. Hold an introductory meeting followed by regular subgroups meetings to develop CDEs. Full group review work of all subgroups, specifying selected CDEs. Public review of the CDEs. Revision of CDEs based on feedback from public review. Post CDEs on a website with protocols for use.</p>
<p>Practices for developing protocols for CDE use and management (Choquet et al., 2014)</p> <p>Clearly define the CDE concept. Identify a long name for the CDE that is readable and descriptive. Identify a short name for the CDE that is an abbreviated form of the long name. Specify the measure of the CDE with the data type (single item, multiple item scale, bio specimen, biomarker), coding protocol (for item, scale, other), permissible values. Develop a protocol for data collection and storage, if appropriate. Collect the CDE as specified by the protocol. Record the relevant conditions. Create a data file as specified and upload to a CDE shared site.</p>

**Table 2****Practical Recommendations for Use of Operational Electronic Health Record Data in Research**

<b>Recommendations</b>	<b>Domain expertise</b>	<b>Informatics expertise</b>
Apply an evidence-based approach—ask an answerable question, find the best EHR data (evidence), appraise the data, apply data to the question	X	
Use tools for searching, browsing, and extracting data	X	
Evaluate and manage data—assess availability, completeness, quality (validity), and transformability of data	X	
Create tools for data management—create software (especially pipelines) for data aggregation, validation, and transformation		X
Develop methods for comparative validation—develop tools that support analysis of multisite data collections	X	X
Develop a methodology knowledge base—develop a data catalogue that relates data elements to recommended transformations	X	
Standardize reporting methods—provide details of data sources, provenance, and manipulation to support data comparisons	X	X
Engage informatics expertise—ensure validity of findings derived from data collected from disparate sources	X	X
Include an informatics research agenda—generate systematic studies of inherent biases in EHR and data collection methods, such as data entry user interfaces	X	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

## Recommended Common Data Elements for Symptom Studies

<b>Symptom</b>	<b>First choice CDE recommendation</b>	<b>Reasoning for selection</b>
Pain	PROMIS Pain	Full PROMIS includes domains that cover other symptoms with interference and intensity.
Fatigue	PROMIS Fatigue	Adult and pediatric, available, flexible in number of items
Sleep disturbance	PROMIS+ additional duration question	Multidimensional, well validated, flexible, very broad
Affective disturbance <i>Mood</i>	PROMIS Positive Affect & PROMIS Depression	Multiple forms, options for adult and pediatric populations, and covers wide population
Affective <i>Anxiety</i>	PROMIS Anxiety	
Affective <i>Well-being</i>	Psychological Well-being Scale SF-36	
Cognitive disturbance	PROMIS applied cognition & general concerns	Free, multiple languages available

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript