

Distributed Cognition Artifacts on Clinical Research Data Collection Forms

Meredith Nahm, MS^{1,2}, Vickie D. Nguyen, MA¹, Elie Razzouk, MD¹, Min Zhu, MD, MS¹,
Jiajie Zhang, PhD¹

¹University of Texas, School of Health Information Sciences, Houston TX, ²Duke
Translational Medicine Institute, Durham, NC

Abstract

Medical record abstraction, a primary mode of data collection in secondary data use, is associated with high error rates. Cognitive factors have not been studied as a possible explanation for medical record abstraction errors. We employed the theory of distributed representation and representational analysis to systematically evaluate cognitive demands in medical record abstraction and the extent of external cognitive support employed in a sample of clinical research data collection forms.

We show that the cognitive load required for abstraction in 61% of the sampled data elements was high, exceedingly so in 9%. Further, the data collection forms did not support external cognition for the most complex data elements. High working memory demands are a possible explanation for the association of data errors with data elements requiring abstractor interpretation, comparison, mapping or calculation. The representational analysis used here can be used to identify data elements with high cognitive demands.

Introduction

Data collection in clinical research, both retrospective and prospective, relies on the abstraction of data from medical records^{1,2}. Abstraction is a time and resource intensive task^{3,4} and is associated with high error rates⁵. However, little is known about the causes and mitigators of these errors⁶. Over time, authors have suggested that the design of the data collection form is a significant factor in the accuracy of abstracted data^{7,8,9,10}. Although data collection forms are widely touted as a key factor in data quality, little evaluative work has been done to understand the mechanism and impact of data collection form design on data accuracy. Today, the design of data collection forms is guided by primarily a-theoretical lists of things that form designers should and should not do^{8,15,16}.

While the role of paper-based patient records in clinician cognition has been studied¹⁴, the extent to which data collection forms impact cognition in clinical research data collection has not yet been investigated. Furthermore, cognitive science models

and methodology have yet to be applied to medical record abstraction in clinical research or other secondary data use settings.

From cognitive science we know that how information is distributed across internal and external representations, *i.e.*, in the user's mind and in the world, affects human task performance¹³. Additionally, representation can extend human performance through external cognition^{12,13}. Thus, one of the ways in which data collection forms may impact data accuracy is through form representation that supports distributed, *i.e.*, external cognition.

We applied the distributed cognition framework¹² and adapted Gong's information search model¹¹ to medical record abstraction, and applied them through a representational analysis to perform a systematic evaluation of data collection forms to 1) identify the type and extent of internal cognition required in medical record abstraction, and 2) to characterize the extent of support for external cognition in data collection forms.

Background

Medical record abstraction entails the identification of required data in the medical record, transformations of that data, and recording the data onto data collection forms. While two representations, 1) the source medical record, and 2) the destination data collection form, may impact data accuracy, secondary data users usually cannot impact the manner in which data are represented in the medical record. However, secondary users can control the representation of their data collection forms. Often, data collection forms employ form instructions, prompts, and structural graphical elements to guide form completion^{8,15,16}. This information is represented on the data collection forms to different extents¹⁵. Since data collection forms are present during the abstraction, and within control of the secondary data users, there is reason to believe that they may provide a mechanism to decrease cognitive load by increasing the extent to which they support external cognition during the abstraction process.

In his 2006 work, Gong applied the theory of distributed representation to explore how information distribution between internal and external representations affects information search performance. He showed that search task performance increased with increasing amounts of information represented externally¹¹. Further, the work of Gong and others has shown that search task performance improves when the scales between the task and the data representation match^{11, 12}.

Because medical record abstraction is both a search and a cognitively intense process, the Gong model has particular utility for exploring and characterizing the extent to which data collection forms support distributed cognition in medical record abstraction. As such, we adapted Gong’s model to the task of medical abstraction (Figure 1).

We extended the data mapping portion of Gong’s model for medical record abstraction, as shown in Figure 1. Representation boxes were added for medical record and data collection form representation. Task boxes were added for both documentation and abstraction tasks. *Remember*, *transform*, and *transcribe* are shown at the sub-task level, clearly delineating them from the *search* task. In addition, *localize* from Gong’s model was considered a direct search task, while *compare* and *calculate* were relocated to the transform task where along with additional transformations “*interpret*” and “*map*”. Importantly, all tasks presented opportunities for distributed cognition. Light grey boxes were added for completeness but are not evaluated here.

In medical record abstraction, information is represented both in the medical record and on the data collection form. Therefore, there are opportunities for mismatch between 1) the representing medical record and the represented information, 2) the representing data collection form and the represented information, and 3) the representing medical record and the representing data collection form. Scale mismatch may increase working memory load. Moreover, the search, remember, transform, and transcribe tasks are performed internally unless external cognition artifacts exist.

While the medical record may have artifacts that enable external cognition for search tasks, the remember, transform, and transcribe tasks are unique to each secondary data use. Therefore, we expect the medical record representation will not provide significant opportunities for external cognition for these tasks. As a result, we concentrated on the evaluation of the data collection form for external cognition artifacts.

In medical record abstraction, virtually every data element by definition has a search task. Each data element may or may not have form artifacts supporting external cognition. Further, for each data element, zero to multiple transform tasks may apply. Each transform task required for a data element may or may not have an external cognition artifact

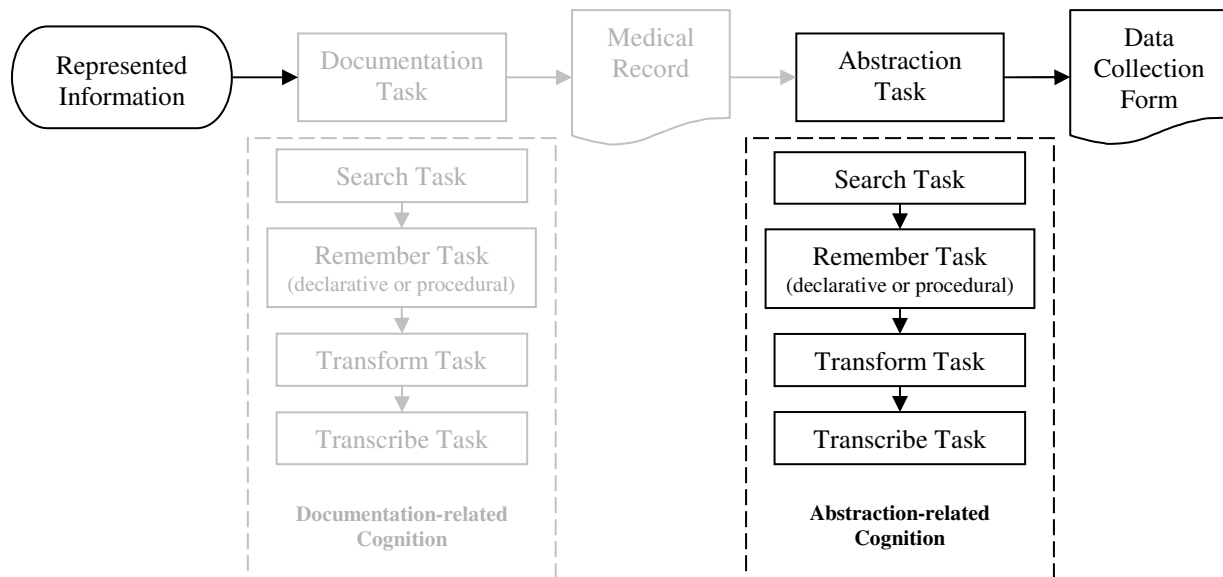


Figure 1. Model of Cognition in Medical Record Abstraction

Methods

We employed a representational analysis to evaluate the medical data collection forms. Our unit of analysis was the data element, *i.e.*, a form question and the associated response field¹. We captured data on eight aspects of data elements with respect to their representation and cognitive demands. Analysis of these items measured the following, 1) the extent of data reduction, *i.e.*, scale downshift, between the represented value and the data collection form representation, 2) the scale mismatch between the abstraction task and the scale of the represented value, 3) the scale mismatch between the abstraction task and the data collection form representation, 4) presence of a search task and whether an external artifact was present for the search task, 5) the type and number of transform tasks required for abstraction of the data element, 6) the dimensions required for abstraction of the data element, and 7) whether the rule representation for the transform task was internal or external.

Fifteen structured data collection form modules² were randomly selected from the data collection form library at the Duke Clinical Research Institute. The library houses data collection forms, many of which have been broken out by modules. We sampled the 256 available modules, randomly selecting 15 modules. Once nine unique trials were obtained, the remaining five modules were accepted sequentially only if they were from a trial already selected for the sample. This allowed comparison between forms within a trial.

The fifteen modules were from nine different clinical trials completed from 1992-2004. The module types and number of data elements per module are listed in Table 1. A total of 250 data elements were assessed in this study.

Module Type	Data Elements*
Patient status (Trial 1)	25
Drug administration (Trial 1)	9
Canadian Cardiovascular Society Class (Trial 2)	4
Pacemaker mode change form (Trial 2)	32
Drug administration (Trial 3)	18
Platelet count (Trial 3)	12
Post procedure repeat catheterization (Trial 4)	18
Cardiac markers (Trial 5)	26

¹ Data element is formally defined in ISO/IEC 11179-1.

² A module is a section of a data collection form containing data grouped by topicality, e.g., vital signs, physical exam, lab results. Modules are usually, but not always less than a page.

Clinical global impression (Trial 6)	6
Thyroid function tests (Trial 6)	12
Serum pregnancy test (Trial 6)	6
Medical history (Trial 7)	14
30 Day follow-up (Trial 8)	48
30 Day follow-up (Trial 9)	11
Cardiac enzymes (Trial 9)	9

* Only unique data elements were assessed and counted

Table 1. Characterization of Modules Selected for this Study.

Ten of the analyzed modules reflected different data collection form modules. Five of the analyzed modules were different representations (isomorphs) of the same content (lab results) from different forms. The analysis of multiple instances of similar module content allowed assessment of differences in form elements representation.

Each data element was reviewed by two independent reviewers (informatics graduate students in a health data display class) who were both novices to medical record abstraction. Each reviewer classified the following eight aspects of each data element: Represented scale (nominal, ordinal, interval, ratio), Data collection form representing scale, Task scale, Presence of a search task (yes, no), Presence of external representation for the search task (yes, no), Type of transform tasks, if present (compare, calculate, interpret, other), Type of dimensions required to abstract the data element, Rule representation (internal, external).

Data were collected in a spreadsheet format: one sheet per form, one row per data element. A third person experienced in medical record abstraction adjudicated and reviewed the work of the two independent reviewers; discrepancies were resolved by the adjudicator and final data were reviewed by all three reviewers. Descriptive statistics were then calculated on the final data.

We recognize that the representation in the medical record likely impacts cognition during medical record abstraction. However, we did not assess representation in the medical record because 1) medical record systems should optimize cognitive support for care delivery and clinical documentation rather than secondary data use, and 2) medical record representation differs from institution to institution. The impact of medical record representation on accuracy of abstracted data remains an area for future research.

Results

Of the 250 data elements assessed, 98 (39%) were direct transcription, *i.e.*, once the value was located in the medical record, it could be copied directly onto the data collection form without transformation. For example, a blood pressure value recorded in the medical record in the same units as those required on the data collection form did not need interpretation or calculation if collected as a numeric value. The majority of the data elements, 152 (61%) required transformation of some type. Cognitively, transformation means that a rule is required to change the data value from its source state to the destination state on the data collection form. Collection of age on the data collection form is an example; age would need to be calculated from the date of birth and the date of the screening visit. The types of transformation required include comparison, calculation, interpretation and mapping, shown by percentage in Table 2. In addition, 37 (15%) of the data elements required more than one transformation.

Transformation Type	Percent
Comparison	43%
Mapping (categorization)	29%
Interpretation (also included synthesis)	14%
Calculation	14%

Table 2. Characterization of Transformation

The data collection form representation for each data element was assessed and categorized as either supporting external cognition or not. As expected, external cognition for the 98 direct transcription data elements was supported by the data collection form. For these data elements, the form prompt and field structure made the search and transcription tasks perceptually evident, *i.e.*, no additional cognition on the part of the human abstractor required.

Supporting external cognition for the transformation (rule based) tasks, is more difficult. Unfortunately, the cognitively more complex data elements, *i.e.*, the 152 data elements requiring transformations, were not supported by form-based external cognition artifacts. One hundred and thirteen (74%) of these complex data elements, required internal cognition.

The number of dimensions required for each transformation was also assessed. The mean number of dimensions required for abstracting the data elements that needed a transformation was 2.6, with a range of 1 to 45 dimensions required. Most often, the values for each dimension are held in the abstractor's head prior to and during the transformation.

Therefore, the dimension counts indicate the cognitive load required for the transformation.

Scale mismatch between the represented information, the abstraction task, and the data collection form representation further impacted internal cognitive demands on the abstractor by requiring mental transformations from one scale to another. Each data element was categorized three ways according to Steven's¹⁷ nominal, ordinal, interval, and ratio scales, 1) the scale of the represented information, 2) the scale of the abstraction task, *i.e.*, the transformation, and 3) the scale of the data collection form representation. Table 3 shows the overall shift in scale from the represented information to the data collection form representation.

Data Collection Form Representation

Represented	Nominal	Ordinal	Interval	Ratio
Nominal (138)	138	0	0	0
Ordinal (19)	16	3	0	0
Interval (29)	1	0	28	0
Ratio (64)	23	3	0	38
	178	6	28	38

Table 3. Scale “down shift” from Represented Information to Data Collection Form

Overall, 43 (17%) of the data elements were reduced from the represented information scale to the data collection form representation. This down shift requires transformation, usually in the form of mapping, interpretation, or categorization. Thus, scale mismatch adds to the already significant cognitive load on the human abstractor.

Discussion

Although from only a limited evaluation in a small sample of data collection forms, the results reported here document the significant cognitive demands in medical record abstraction. Based on our results, a given transform task will likely require more than one transformation, internalizing the rule for each transform, as well as an average of 2.6 dimensions each. Moreover, each of the values involved may also require a scale shift. A human can hold on average from 5-7 chunks of information in working memory¹⁸. Our results show that on average, the cognitive demands bump up against the limits of human cognition. Further, the 9% of data elements requiring four or more dimensions, clearly exceed working memory limits. Moreover, the data collection forms analyzed had little to no external cognition artifacts to

support the most cognitively demanding data elements.

Many authors have cited requiring “abstractor judgment” or “interpretation” as a cause of errors in medical record abstraction^{6, 8, 9, 10}. However, none have suggested why these errors occur or what their relationship is to other types of data error in medical record abstraction. Likewise, the literature does not suggest concrete methods of mitigating or preventing the resulting data errors. Our results contribute a possible explanation and mechanism for a portion of the data accuracy problem that now exists in medical record abstraction. In addition, the theory of distributed representation and the associated representational analysis used here can be applied to analyze data element representation on data collection forms and abstraction tasks to prevent cognitive limit related abstraction errors. Confirming these results in a larger and more diverse sample, and evaluation of data accuracy from data collection form isomorphs are key next steps in this area of inquiry.

Conclusion

The cognitive load required for abstraction of 61% of the data elements in our sample was both high and unsupported with external cognition artifacts on the data collection forms, exceedingly so for 9% of the data elements. The high working memory demands are a possible explanation for the association of data errors in medical record abstraction with data elements that require abstractor interpretation, comparison, mapping or calculation. Existing methods of representational analysis can be applied to identify data elements with high cognitive demands. Further, representational analysis provides a tool to analyze form isomorphs and identify those with the lowest cognitive demands.

References

1. Gardiner RC. Quality considerations in medical records abstracting systems. *J Med Syst.* 1978;2(1):31-43.
2. Herrmann N, Cayten CG, Senior J, Staroscik R, Walsh S, Woll M. Interobserver and intraobserver reliability in the collection of emergency medical services data. *Health Serv Res.* 1980;15(2):127-143.
3. Robinson L, Hughes P. Use a streamlined approach for medical record abstraction. *Qual Lett Healthc Lead.* 1998;10(6):14-15.
4. Kerr EA, Smith DM, Hogan MM, Krein SL, Pogach L, Hofer TP, et al. Comparing clinical automated, medical record, and hybrid data sources for diabetes quality measures. *Jt Comm J Qual Improv.* 2002;28(10):555-565.
5. Nahm M, Johnson CM, Johnson T, Fendt K, Zhang J. Clinical research data quality literature review and systematic analysis. *Clin Trials J.* [submitted].
6. Allison JJ, Wall TC, Spettell CM, Calhoun J, Fargason CA Jr, Kobylinski RW, et al. The art and science of chart review. *Jt Comm J Qual Improv.* 2000;26(3):115-136.
7. Findley TW, Daum MC. Research in physical medicine and rehabilitation. III. The chart review or how to use clinical data for exploratory retrospective studies. *Am J Phys Med Rehabil.* 1991;70(1):S23-S30.
8. Banks NJ. Designing medical record abstraction forms. *Int J Qual Health Care.* 1998;10(2):163-167.
9. Beard CM, Yunginger JW, Reed CE, O’Connell EJ, Silverstein MD. Interobserver variability in medical record review: an epidemiological study of asthma. *J Clin Epidemiol.* 1992;45(9):1013-1020.
10. Feinstein AR, Pritchett JA, Schimpff CR. The epidemiology of cancer therapy. IV. The extraction of data from medical records. *Arch Intern Med.* 1969;123(5):571-590.
11. Gong, Y. The interaction between internal and external information on relational data search. PhD [dissertation]. Houston (TX): University of Texas School of Health Information Sciences; 2006.
12. Zhang J, Norman DA. Representations in distributed cognitive tasks. *Cog Sci.* 1994;81(1):87-122.
13. Zhang J. A representational analysis of relational information displays. *Int J of Human-Computer Studies.* 1996;45:59-74.
14. Bang M, Timpka T. Cognitive tools in medical teamwork: the spatial arrangement of patient records. *Methods Inf Med.* 2003;42(4):331-336.
15. Spiker B, Schoenfelder J. Data collection forms in clinical trials. New York: Raven Press; 1991.
16. Society for Clinical Data Management. Good clinical data management practices [document on the Internet]. Milwaukee: SCDM; 2009 [cited 2009 Nov 4]. Available from: <http://www.scdm.org>.
17. Stevens, SS. On the theory of scales and measurement. *Science.* 1946;103(2684):677-680.
18. Miller GA. The magical number seven, plus or minus two. *The Psychological Review.* 1956;63(2):81-97