

*Research Paper* ■

# Human and Automated Coding of Rehabilitation Discharge Summaries According to the International Classification of Functioning, Disability, and Health

RITA KUKAFKA, DRPH, MA, MICHAEL E. BALES, MPH, ANN BURKHARDT, OTD, OTR/L, BCN, CAROL FRIEDMAN, PHD

**Abstract Objective:** The International Classification of Functioning, Disability, and Health (ICF) is designed to provide a common language and framework for describing health and health-related states. The goal of this research was to investigate human and automated coding of functional status information using the ICF framework.

**Design:** The authors extended an existing natural language processing (NLP) system to encode rehabilitation discharge summaries according to the ICF.

**Measurements:** The authors conducted a formal evaluation, comparing the coding performed by expert coders, non-expert coders, and the NLP system.

**Results:** Automated coding can be used to assign codes using the ICF, with results similar to those obtained by human coders, at least for the selection of ICF code and assignment of the performance qualifier. Coders achieved high agreement on ICF code assignment.

**Conclusion:** This research is a key next step in the development of the ICF as a sensitive and universal classification of functional status information. It is worthwhile to continue to investigate automated ICF coding.

■ *J Am Med Inform Assoc.* 2006;13:508–515. DOI 10.1197/jamia.M2107.

## Introduction

The International Classification of Functioning, Disability, and Health (ICF) is designed to provide a common language and framework for describing health and health-related states. The goal of this research is to investigate human and automated coding of functional status information (FSI) using the ICF framework. In the first phase of this process, we attempted to identify redundancies in the codes; determined coding issues pertaining to the qualifiers; and as-

essed the level of domain knowledge required to perform coding. This process allowed for a qualitative assessment of the ICF with respect to a set of desirable qualities for controlled medical vocabularies.<sup>1</sup> In the second phase, which is the focus of this manuscript, we assessed the feasibility of using natural language processing (NLP) to assign ICF codes automatically. We first extended an existing NLP system to encode rehabilitation discharge summaries according to the ICF. We then conducted a formal evaluation in which six subjects each used the ICF to encode FSI in 25 discharge summaries. We compared the coding performed by expert coders, non-expert coders, and the NLP system. The results show that coders can achieve high agreement on ICF code assignment, and suggest that it is worthwhile to continue to investigate automated ICF coding.

---

Affiliations of the authors: Department of Biomedical Informatics (RK, MEB), Columbia University, New York, NY; Mailman School of Public Health (RK, CF), Columbia University, New York, NY; Department of Occupational Therapy (AB), Columbia University Medical Center, New York, NY

Ann Burkhardt is presently at Long Island University, Brooklyn Campus, Brooklyn, NY.

Support for this research was provided by the National Center on Birth Defects and Developmental Disabilities (NCBDDD), Centers for Disease Control and Prevention (CDC), United States Department of Health and Human Services (HHS).

The authors acknowledge the invaluable contributions of our colleagues at the Mayo Clinic; Marcelline Harris, Guergana Savova, and Christopher Chute, without whom this work would not have been possible.

Correspondence and reprints: Rita Kukafka, DRPH, MA, Department of Biomedical Informatics, Vanderbilt Clinic, 5<sup>th</sup> Floor, 622 West 168th Street, New York, NY 10032; e-mail: <rita.kukafka@dbmi.columbia.edu>.

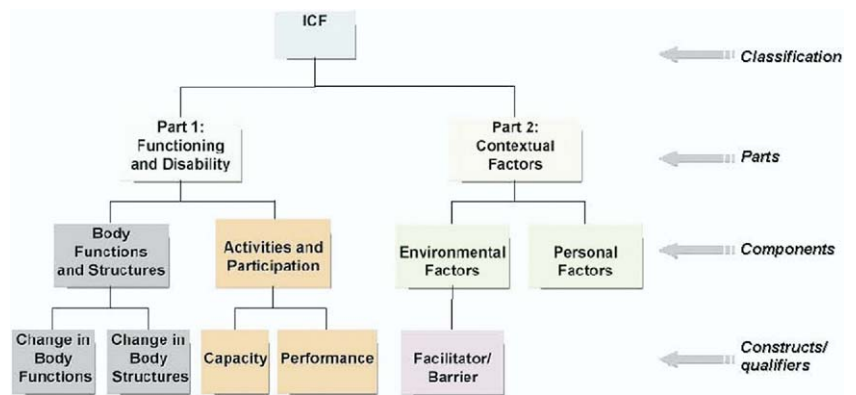
Received for review: 3/20/06; accepted for publication: 6/12/06.

## Background

### ICF Framework

The World Health Organization defines health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.”<sup>2</sup> To achieve goals of healthy people and healthy populations, concepts associated with functioning and disabilities need to be characterized in a consistent and comparable way. The ICF was designed to be a common language to describe these concepts. It can be used to describe how functional problems can result in difficulties carrying out tasks, and how these problems are manifested in a person’s environment. Rather

**Figure 1.** The ICF framework is arranged in a hierarchical structure. At the uppermost levels it is divided into two parts, and then into four components: body functions and structures, activities and participation, environmental factors, and personal factors. Adapted from World Health Organization (Switzerland), 2006.<sup>4</sup>



than focusing on the consequences of disease, the ICF addresses components of health.<sup>3</sup>

The ICF framework is based on a biopsychosocial model of functioning and disability, and the codes express FSI, which includes behavioral, social, and environmental aspects that influence health status. FSI is important partly because diagnostic information in medical records, such as that expressed using ICD-9 codes, conveys little detail about the effect of a patient's health status on the individual's daily activities. Two individuals may have the same medical diagnosis, but very different levels of functioning. FSI relates to an individual's capacity to carry out a set of tasks or actions, and to changes in body structures or functions arising from a health condition. It can be used to assess the use of resources and the need for services.

The aims of the ICF are to provide a scientific basis and common language for describing health; to improve communication between different users of the framework, such as health care workers, researchers, policymakers, and members of the public (including people with disabilities); to allow for comparison of data across countries, health care disciplines, and time; and to provide a systematic coding scheme for health information systems. It is intended, therefore, to be a meaningful and practical system that can be used worldwide for development of health policy, for quality assurance, and for outcome evaluation.

The ICF framework is arranged in a hierarchical structure. At the uppermost levels it is divided into two parts, and then into four components: body functions and structures, activities and participation, environmental factors, and personal factors (Figure 1). Each ICF code is assigned one or more numeric qualifiers to indicate the level of a certain parameter, such as the level of facilitation or impairment conferred by a health condition. The meanings of these qualifiers vary depending on the ICF component.

Several recent papers have focused on various aspects of the development and possible uses of the ICF framework, providing overviews of the framework,<sup>5-7</sup> investigating the role of environmental factors in functional status,<sup>8</sup> describing public health opportunities afforded by the development of ICF,<sup>9</sup> linking health status measures to the ICF,<sup>10, 11</sup> and assessing its potential use as a conceptual framework for nursing.<sup>12</sup> ICF-related literature from the fields of psychology, health and human services, and rehabilitation have been summarized by Bruyère and colleagues.<sup>13</sup>

Recently, various studies have begun to investigate manual coding of clinical data using the ICF. In a study<sup>14</sup> in which the ICF was applied to 40 children with various types and degrees of disability, the framework was judged to be applicable, reliable, and strongly correlated with well-established measures of functioning. The authors recommended specific adaptations to improve the framework's ability to reflect the developmental nature of many abilities of children. In a test of inter-rater reliability,<sup>15</sup> independent raters decided which ICF codes to assign to children with disabilities based on the results of standard pediatric functional assessment measures. The authors found high levels of agreement when assessment test items were structured developmentally and corresponded to a single ICF code. In a study<sup>16</sup> in which ICF codes were used to describe the functional status of patients in geriatric care, test-retest reliability kappa values were found to be relatively low. The authors suggest specific improvements to the framework for use in the geriatric care setting. The ICF has also been applied to describe experiences of living with muscular dystrophy<sup>17</sup> and in a proof of concept study<sup>18</sup> to assess the feasibility of using individual items in a generic health measure to create coded functional status indicators.

### Natural Language Processing

Assigning codes to clinical concepts manually, using any framework, can be costly and time consuming. It has been estimated that it may take an hour or more to assign all the relevant ICF codes for a particular patient.<sup>11</sup> However, capturing and encoding a patient's FSI is important, and the merits of FSI in clinical care and public health practice are well established. At the time of this project's inception there was little published on the process of assigning ICF codes automatically.

Because of the difficulties involved in performing manual coding of records, we have examined the feasibility of using computers to perform the coding automatically, by way of natural language processing, or NLP. In this process, a file containing text is input into a software program. The program then extracts the information expressed in natural language and outputs data in coded format. Once the data are in coded format, they can be used for a number of purposes, such as for billing or for identifying individuals meeting a certain diagnostic criterion.

## MedLEE

The NLP tool used for this research was MedLEE, an existing medical language processing program. MedLEE, the Medical Language Extraction and Encoding system, is a rule-based system<sup>19</sup> originally intended to be used for alerting. The system has been in routine use at the Columbia University Medical Center (New York, NY) since 1995, for encoding information in chest x-ray and mammography reports. MedLEE was later expanded to other areas, and it has been adapted for use in other institutions. In several evaluations, the performance of the software has been shown to be similar to that of expert coders such as physicians. MedLEE's knowledge components consist of a lexicon database (a semantic lexicon for clinical text), grammar rules, a compositional mapping tool for multi-word phrases, and an encoding table to map clinical terms to controlled vocabularies. The system's processing components include a preprocessor, a parser, an encoder, an Extensible Markup Language (XML) translator, and an error handler.

Other medical language processing systems have been developed for a variety of tasks, such as extracting assertions about molecular binding relationships from biomedical text,<sup>20</sup> detecting acute bacterial pneumonia from radiology reports,<sup>21</sup> and mapping biomedical text to concepts in the UMLS Metathesaurus.<sup>22</sup>

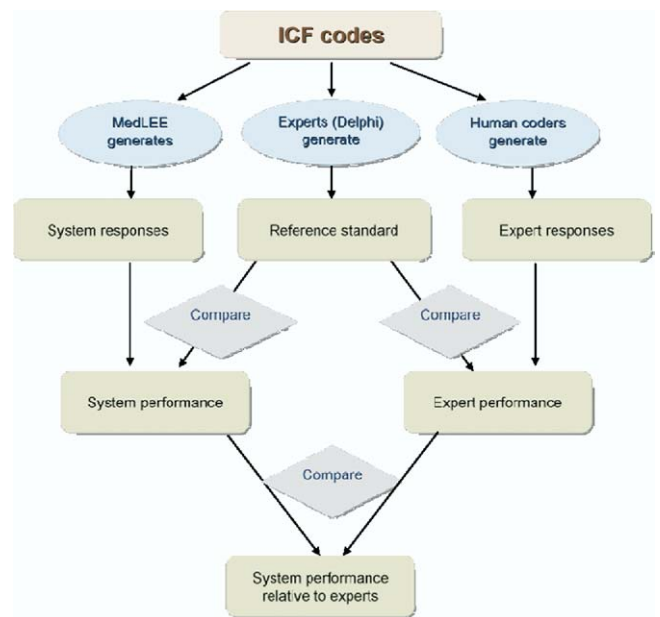
## Methods

An overview of the evaluation appears in Figure 2. To summarize, we first selected ten ICD-9 codes which are related to conditions known to be associated with changes in functional status. We then obtained a set of 251 inpatient rehabilitation discharge summaries and divided them randomly into a training set ( $n = 147$ ) and a test set ( $n = 104$ ). To prepare the NLP system, we used the training set to extend it for use in the FSI domain by adding entries to its lexicon and encoding table.

We also trained three rehabilitation experts on using the ICF. These experts developed a reference standard of coding assignments for the test set. Finally, a second set of human coders (three rehabilitation experts and three non-rehabilitation experts), as well as the NLP system, assigned ICF codes to the discharge summaries in the test set. The reference standard was then compared with the coding assignments of the NLP system, expert coders, and non-expert coders. The methods are explained below in greater detail.

### Modification of MedLEE NLP System

The NLP system was modified for processing rehabilitation records and assigning selected ICF codes automatically. To make this possible, the system had to be extended for use in the FSI domain. This was done iteratively. To begin, the system was given some text to process. Based on the results, entries were added to the lexicon and encoding table. (Table 1 contains selected entries from the lexicon. Entries from the encoding table appear in Table 2.) The same text was then fed back into the system and more lexical entries were added as needed. Each piece of functional status language was assigned a target form in the lexicon, in collaboration with the rehabilitation expert. For example, the phrase "a little trouble" was assigned the target form "minimal assis-



**Figure 2.** Evaluation design. We first trained three rehabilitation experts on using the ICF. These experts then developed a reference standard of coding assignments for a group of 75 rehabilitation discharge summaries. Next, we extended the NLP system to the FSI domain by adding entries to its lexicon and encoding table. Finally, a second set of human coders (three rehabilitation experts and three non-rehabilitation experts), as well as the NLP system, assigned ICF codes to the 75 discharge summaries. The reference standard was then compared with the coding assignments of the NLP system, expert coders, and non-expert coders. Adapted from Hripcsak et al (2002).<sup>23</sup>

tance." Likewise, "ability" was given the target form of "independent", and so on. Qualifiers were assigned based on the highest level of assistance required. For example, the text "stand by to maximal assist" was assigned the same target form as "maximal assist."

Finally, the table data was examined and expanded through the addition of various combinations of codes and qualifiers—combinations that had not yet been added because the particular permutation of codes and qualifiers were not

**Table 1** ■ Functional status information and corresponding target forms. Excerpt from table of functional status language and target forms, used to extend the MedLEE lexicon. The first entry alphabetically is "a little trouble." By collaborating with a domain expert, we assigned this text the target form "minimal assistance." The second entry was "ability," which was given the target form of "independent," and so on.

Language	Target
able to feed herself after set up	independent after set up
able to feed himself after set up	independent after set up
abstract reasoning	cognition
adaptive equipment	adaptive equipment
alert	consciousness functions
ambulate	transferring oneself



**Table 2** ■ This table contains examples of lexical target form combinations which were used to generate the NLP system's encoding table. The first row indicates that when any word or phrase having the target form of "eating" appears near any word or phrase having the target form "requires encouragement," the ICF code *d550.30* should be assigned. Other code combinations were "dressing" and "adaptive equipment for dressing," which were assigned the ICF code *d5400.10*, and "judgment" and "unrealistic," which were assigned the code *b117.3*.

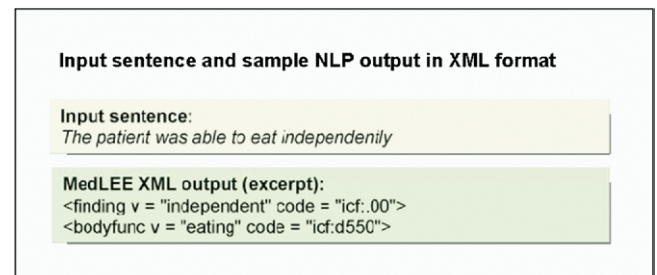
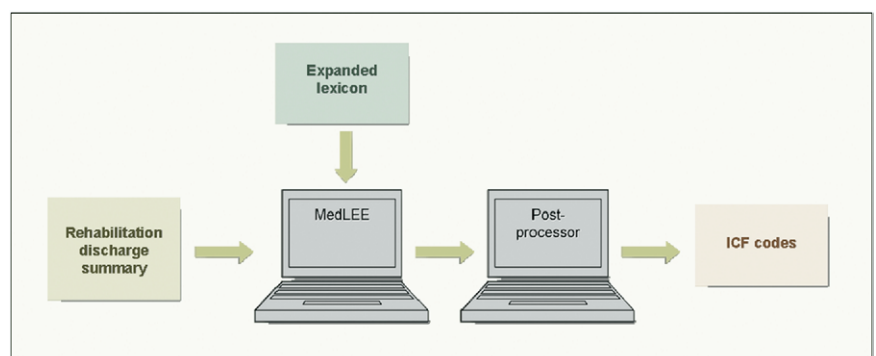
Activity	Qualifier	ICF Code
eating	requires encouragement	d550.30
dressing	adaptive equipment for dressing	d5400.10
orientation functions	confusion	b117.4
dressing	with effort	d5400.10
cognition	inconsistent	b117.3
judgment	unrealistic	b117.3

encountered in the training language. For example, if "maximal assistance for opening containers" and "minimal assistance for opening containers" were both coded, but "moderate assistance for opening containers" was not yet coded, then this combination was added to the lexicon, along with the appropriate qualifier.

To illustrate how the system was used in this research, consider the sample text "X is able to toilet independently with assist of a toilet riser with arms." In this example, the system would recognize three terms from the lexicon: "toilet," "independently," and "toilet riser." Each of these terms has a target form. The target form for "toilet" is "toileting;" the target form for "independently" is "independent;" and the target form for "toilet riser" is "adaptive equipment for toileting." These target forms also appear in the encoding table. The target forms "independent" and "toileting," when they appear together, would be assigned the ICF code *d530.00*. The target form "adaptive equipment for toileting" is in the coding table alone (i.e., without a qualifier), and is associated with the code *d530.10*. For this example, then, the two ICF codes assigned would be *d530.00* and *d530.10*.

To match the responses provided by the human coders, the goal was to assign just one set of qualifiers for each main code. The final step, therefore, was to run a postprocessing algorithm. For each discharge summary the postprocessor

**Figure 3.** Overview of natural language processing steps. The lexicon and encoding table of the NLP system were expanded for the FSI domain. Rehabilitation discharge summaries were processed by the system and ICF codes were generated. The resulting lists of codes were then input into a postprocessor which generated the final ICF code assignments (from zero to five of the selected ICF codes for each discharge summary.)



**Figure 4.** A sentence expressed in natural language, along with the coded form after the sentence has been processed by MedLEE. MedLEE identified the concept "independent" and the associated ICF qualifiers of .00. It also identified the concept of eating, which corresponds to the ICF code *d550*. In this case, the MedLEE output is in XML format.

was supplied with all main code/qualifier combinations from the MedLEE output. The postprocessor algorithm was designed to select the code with the highest numeric value for the first (performance) qualifier and the code with the lowest numeric value for the second (capacity) qualifier. Using these rules, for every main code for which more than one code/qualifier combination was assigned, the algorithm selected just one code/qualifier combination.

To illustrate using the example above, the codes assigned are *d530.00* and *d530.10*. The code with the highest numeric value for the performance qualifier, (pertaining to the lowest level of functional status) is 1. Both capacity qualifiers are 0. The final code selected here would be *d530.10*. An overview of the natural language processing steps is in Figure 3. Figure 4 contains a sample input sentence and MedLEE output in XML format.

### Recruitment and Training of Research Subjects

While preparation of the NLP system was underway, we recruited nine research subjects, and gave them training on the ICF and on using the ICF framework to encode functional status information. Six of the research subjects were rehabilitation experts and three were physicians who were not rehabilitation experts.

### Selection of ICF Codes and Procurement of Rehabilitation Discharge Summaries

Five ICF codes were selected as the focus of the research: *b117* (intellectual functions), *d420* (transferring oneself), *d530* (toileting), *d550* (eating), and *d5400* (putting on clothes). These were selected based on the importance of the corre-

sponding activities in daily living. Next, 251 inpatient rehabilitation discharge summaries in electronic format were obtained through a research collaboration with the Mayo Clinic (Rochester, MN). These discharge summaries were selected on the basis of ICD-9 codes pertaining to the following diagnoses: *spinal cord injury, head injury, above knee/below knee amputation, brain tumor neoplasm, debility, stroke, hip replacement, knee replacement, arthritis, and fractures*. We divided them randomly into a training set ( $N = 147$ ) and a test set ( $N = 104$ ).

### Establishment of Reference Standard Using Delphi Method

To develop a reference standard, three of the trained rehabilitation experts achieved consensus on their code selections using the Delphi method. This method was first developed in a series of experiments conducted by the RAND Corporation in the 1950s.<sup>24</sup> The Delphi method can be used to allow coders to reach a consensus through an iterative process. Several variations of the method have been proposed; however, the process generally involves several iterations in which coders individually report their code assignments. These assignments are collected by the investigators and are assessed collectively to determine the level of agreement between the coders. Individuals are then given an opportunity to revise their code selections until a satisfactory level of agreement is reached. Some degree of anonymity is normally provided to individuals when they are recording their responses.

In our research, the three rehabilitation experts worked independently to encode the FSI in 75 of the discharge summaries into an electronic spreadsheet. We then assessed the level of agreement among the coders, and created revised versions of the spreadsheet in which coders were able to see their code selections alongside the code selections of their colleagues. They were then given the opportunity to revise their code selections, if desired, based on the knowledge of their colleagues' responses. The third round of coding took place in a 90-minute long face-to-face meeting at which the three expert coders were able to discuss their code selections and determine final code assignments. These three rehabilitation experts did not participate in the remaining phases of the study.

### Coding of Rehabilitation Discharge Summaries

The remaining coding steps were carried out by the remaining six trained human coders and by the NLP system. We used a test set of 100 rehabilitation discharge summaries which were independent from those used in the training set. These were sorted by file size and divided into two groups. We randomly selected three-quarters of 75 ( $n = 56$ ) from the larger summaries, and one-quarter of 75 ( $n = 19$ ) from the smaller summaries, for a total of 75 summaries. We then sorted these summaries into random order and proceeded sequentially through them, assigning each one randomly to one of the three coders from the group of experts, and to one of the three coders from the non-experts. After any coder was assigned 25 summaries, that coder was not assigned any additional summaries. We then distributed copies of the 25 selected discharge summaries to each of the six coders, along with an electronic spreadsheet and written instructions.

For each discharge summary, participants could assign from zero to five of the selected ICF codes. For each ICF code assigned, participants were instructed to assign one or more numeric qualifiers. In accordance with the ICF specification,<sup>3</sup> items coded as *b117* (intellectual functions) were assigned one qualifier, while items coded using the remaining four main codes each were assigned two qualifiers. The NLP system was allowed to assign an unrestricted number of codes to each of the 75 discharge summaries. To obtain the system's final code assignments, this initial list of codes was run through a postprocessing program. The format of the system's final output matched that of the human coders; for each discharge summary, the system assigned from zero to five of the selected ICF codes.

### Data Analysis

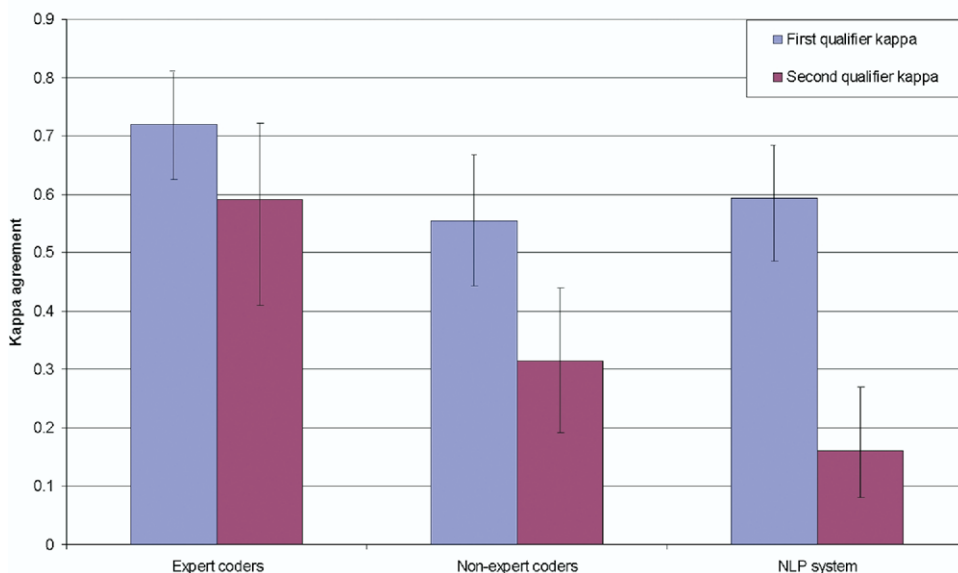
We quantified performance by measuring agreement between all independent coders (the six human coders and the NLP system) and the reference standard. Because the ICF qualifiers are ordinal values that express varying levels of impairment, we used a quadratically weighted version of Cohen's kappa statistic<sup>25</sup> as a chance-corrected measure of agreement to measure inter-rater reliability. To assess the likelihood that the reported kappa values were due to chance, statistical significance was assessed using p-values calculated from kappa variance and z-scores.<sup>26</sup> Using thresholds corresponding to the permissible ordinal values of the ICF qualifiers, receiver operating characteristic (ROC) curves were plotted to depict agreement between coders (grouped and individually) and the reference standard.

### Results

The expert coders performed better than either the non-expert coders or the NLP system, and all three coding groups performed better when assigning the first qualifier than the second. For the expert coders, the kappa value for the first qualifier was 0.719 (95% C.I. 0.626–0.809,  $p < 0.0005$ ) and for the second qualifier 0.591 (95% C.I. 0.410–0.722,  $p < 0.0005$ ). For the non-expert coders, the kappa value for the first qualifier was 0.555 (95% C.I. 0.443–0.667,  $p < 0.0005$ ) and for the second qualifier, 0.315 (95% C.I. 0.191–0.441,  $p < 0.0005$ ). For the NLP system, the kappa value for the first qualifier was 0.593 (95% C.I. 0.485–0.682,  $p < 0.0005$ ) and for the second qualifier 0.160 (95% C.I. 0.082–0.270,  $p < 0.0005$ ). The three coding groups are compared in Figures 5 and 6. The ROC curves (Figures 7 and 8) are graphical plots of the level of agreement between the coding groups, for the first (Figure 7) and second (Figure 8) qualifiers. The points on each curve are thresholds corresponding to the permissible ordinal values for the ICF numeric qualifiers (zero through four).

To summarize, the expert coders coded more consistently with the reference standard than did the non-expert coders. Based on the kappa measures for the performance qualifier, the NLP system performed better than the non-expert coders, but not as well as the expert coders. The NLP system's coding assignments for the second qualifier agreed with the reference standard more often than they disagreed, but only slightly more than would be expected through chance alone (Figure 8).

Kappa agreement between ICF coding groups and reference standard



**Figure 5.** Comparison of the three ICF coding groups and the reference standard. The kappa value indicates the level of agreement between the coding group and the reference standard.

**Discussion**

Preliminary research indicates that NLP software can be used to encode patient information using the ICF with results, at least for the performance qualifier, similar to those of human coders.

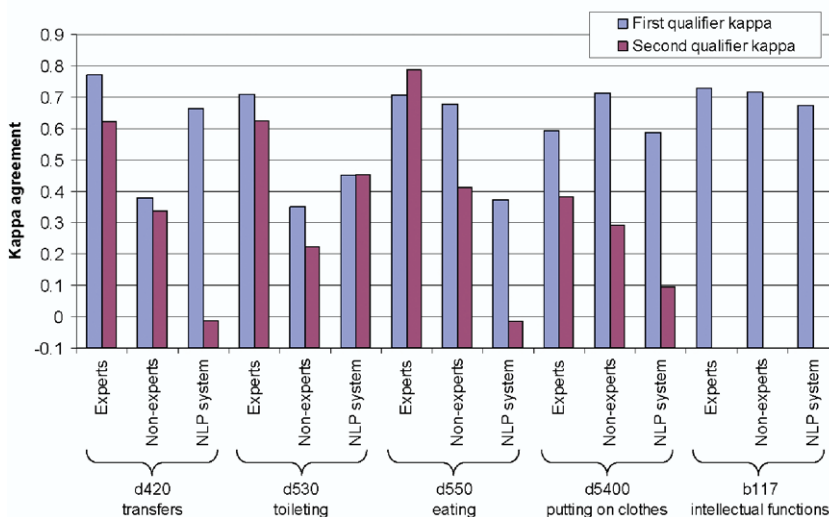
Levels of agreement on code assignments were generally higher for the first qualifier (performance) and lower for the second qualifier (capacity). This may be because assignment of a capacity qualifier is highly subjective; it involves making predictions about an individual’s future capacity. Future capacity, in turn, can be determined by any of a large number of patient-specific factors.

For the NLP system in particular, code assignments for the capacity qualifier did not show strong agreement with the reference standard. While medical language processing systems are effective at handling comorbidities (such as diabetes and chronic obstructive pulmonary disease) they may be less effective at encoding concepts such as level of social

support. Both can affect a patient’s capacity to return to a previous level of functional status. Coding tasks involving complex reasoning, such as those in which disparate pieces of information must be connected, are a difficult challenge for current NLP systems.

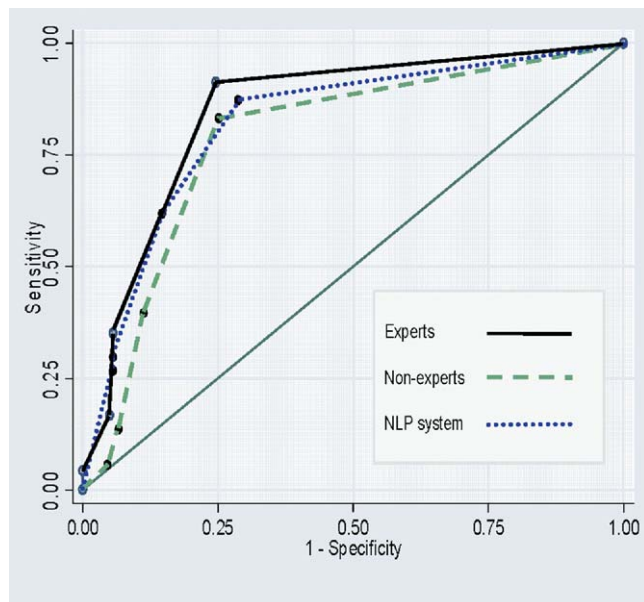
When the NLP system did not agree with the reference standard coding assignments, there were two main types of errors. In some cases, the system assigned an ICF code that was not assigned by the coders who developed the reference standard. For example, one of the discharge summaries contained the text “X will eat with set up assistance only.” In this case, the system assigned the code *d550.20*, corresponding to the ICF representation *eating, moderate difficulty* (performance), *no difficulty* (capacity). The reference standard coders determined there was not enough information to assign qualifiers for this ICF code and assigned *d550.88*, corresponding to the ICF representation *eating, not specified* (performance), *not specified* (capacity).

Kappa agreement between ICF coding groups and reference standard, by code



**Figure 6.** Comparison of kappa coefficients for the three coding groups, by code. Similar levels of agreement were achieved for all five of the selected codes. All kappa values were statistically significant at the  $p < 0.05$  level, with the exception of those for the second qualifier, for the NLP system, for codes *d420* (transferring oneself) and *d550* (eating), which were not statistically significant.



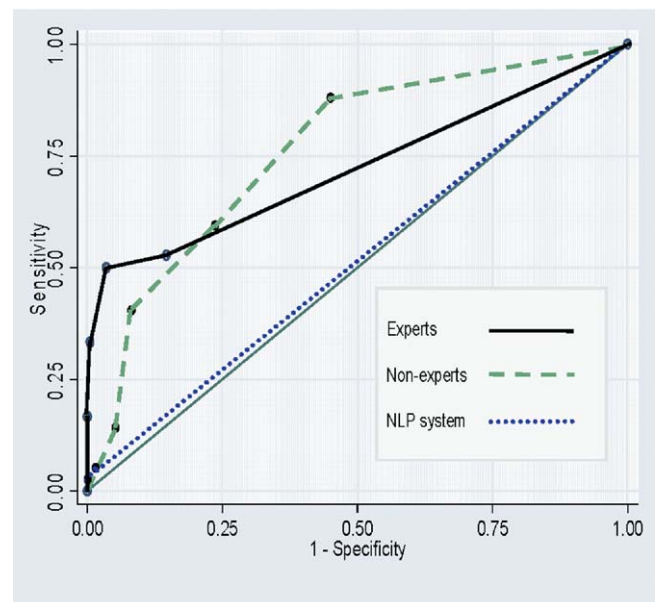


**Figure 7.** ROC curves for first (performance) qualifiers. The areas under the ROC curves for the expert, non-experts, and NLP system were 0.8483, 0.7886, and 0.8229, respectively. To summarize, the expert coders coded more consistently with the reference standard than did the non-expert coders. Based on the kappa measures for the performance qualifier, the NLP system performed better than the non-expert coders, but not as well as the expert coders.

In other cases, a code was assigned by the reference standard coders, but the NLP system did not assign the code. There were at least two reasons for this type of error. The first was that there was functional status language that was not in the lexicon. As an example, in one case the text was: "X is able to dress U or E with Minimal A." In this case, "Minimal A" was short for "minimal assistance". This abbreviation was not encountered in the training sets, and was not in the lexicon. The NLP system only recognized "able to dress." As a result, although the reference standard code for this patient was *d5400.21*, the NLP system assigned *d5400.00*.

A second reason for false negative errors was related to the fact that accurate assignment of ICF codes sometimes requires reasoning based on contextual information. For example, the text from one of the discharge summaries in the test set included: "disposition planning for traumatic brain injury and polytrauma associated with a head on car crash" followed by "X is still not able to communicate." Although the expert coders assigned the code *b117.4*, the NLP system did not assign a code related to cognition. This is partly due to the fact that although the word "communication" was in the extended lexicon, the word "communicate" was not added to the lexicon and it was not in the code table. The assignment of code *b117.4* involved inferred knowledge that the patient who had had a traumatic brain injury and was unable to communicate would also have cognitive impairment.

Whether ICF codes are assigned manually, automatically, or in a semi-automated way, functional status information can be used to convey the health status of an individual or an



**Figure 8.** ROC curves for second (capacity) qualifiers. The areas under the ROC curves for the expert, non-experts, and NLP system were 0.7230, 0.7636, and 0.5156, respectively. The graph shows that the NLP system's coding assignments for the second qualifier agreed with the reference standard more often than they disagreed, but only slightly more than would be expected through chance alone. This may be partly because assignment of a capacity qualifier is highly subjective; it involves making predictions about an individual's future capacity. Future capacity, in turn, can be determined by any of a large number of patient-specific factors.

entire population. Using the framework can help clinicians with specific tasks like predicting a person's ability to live in the community following a change in health status. When the codes are assigned to a population of patients, they can provide a quantitative measure of how functional status is changing over time. Because electronic health records and population databases often contain codes from several biomedical terminologies, ICF might be most expressive when used in conjunction with other terminological systems.

This study is subject to several limitations. First, because this was preliminary research it was necessary to select a subset of just five codes. As a result, certain branches of the ICF framework were not evaluated. Second, the discharge summaries that were used in this research were from a different institution. One of our coders pointed out regional differences in language use. For example, the phrase "stand by assistance," which appeared throughout the corpus, is analogous to "close supervision" which was more familiar to the coders involved in this experiment. These differences may have made it more challenging for the coders to assign codes. Third, since this work was designed to explore the feasibility of using NLP for ICF coding, there were only minimal changes to the NLP system. New entries were added to the lexicon and coding table; however, other parts of the system were not modified. Because of the relatively small size of the training set, the lexical entries that were added were discovered manually. It is possible that the results could be improved given a larger training set and further work to adapt the NLP system.

## Conclusion

This paper describes a methodological evaluation of assigning ICF codes to FSI in rehabilitation discharge summaries. The results show that automated coding can be used to assign codes using the ICF, with results similar to those obtained by human coders, at least for the selection of ICF code and assignment of the performance qualifier. This research is a key next step in the development of the ICF as a sensitive and universal classification for functional status information.

## References ■

- Bales ME, Friedman C, Burkhardt A, Kukafka R. Qualitative assessment of the International Classification of Functioning, Disability, and Health with respect to the desiderata for controlled medical vocabularies. *Int J Med Inform.* 2006;75(5):384–95.
- World Health Organization (Switzerland). Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.
- World Health Organization (Switzerland). International Classification of Functioning, Disability, and Health. Geneva: The Organization; 2001.
- World Health Organization (Switzerland). ICF as the new member in the WHO family of international classifications. 2006. Available at <http://www3.who.int/icf/training/icf.ppt>. Accessed 24 Feb 24, 2006.
- Dahl TH. International classification of functioning, disability and health: an introduction and discussion of its potential impact on rehabilitation services and research. *J Rehabil Med.* 2002;34(5):201–4.
- Ustun TB, Chatterji S, Kostansjek N, Bickenbach J. WHO's ICF and functional status information in health records. *Health Care Financ R.* 2003;24(3):77–88.
- Lollar DJ, Crews JE. Redefining the role of public health in disability. *Annu Rev Publ Health.* 2003;24:195–208.
- Schneidert M, Hurst R, Miller J, Ustun B. The role of environment in the International Classification of Functioning, Disability and Health (ICF). *Disabil Rehabil.* 2003;25(11–12):588–95.
- Lollar DJ. Public health and disability: emerging opportunities. *Public Health Rep.* 2002;117(2):131–6.
- Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the International Classification of Functioning, Disability and Health. *J Rehabil Med.* 2002;34(5):205–10.
- Stucki G, Ewert T, Cieza A. Value and application of the ICF in rehabilitation medicine. *Disabil Rehabil.* 2003;25(11–12):628–34.
- Kearney PM, Pryor J. The International Classification of Functioning, Disability, and Health (ICF) and nursing. *J Adv Nurs.* 2004;46(2):162–70.
- Bruyere SM, Van Looy SA, Peterson DB. The International Classification of Functioning, Disability and Health: Contemporary Literature Overview. *Rehabil Psychol.* 2005;50(2):113–21.
- Battaglia M, Russo E, Bolla A, et al. International Classification of Functioning, Disability and Health in a cohort of children with cognitive, motor, and complex disabilities. *Dev Med Child Neurol.* 2004;46(2):98–106.
- Ogonowski J, Kronk R, Rice C, Feldman H. Inter-rater reliability in assigning ICF codes to children with disabilities. *Disabil Rehabil.* 2004;26(6):353–61.
- Okochi J, Utsunomiya S, Takahashi T. Health measurement using the ICF: Test-retest reliability study of ICF codes and qualifiers in geriatric care. *Health Qual Life Outcomes.* 2005; 3(46).
- Boström K, Ahlström G. Living with a chronic deteriorating disease: the trajectory with muscular dystrophy over ten years. *Disabil Rehabil.* 2004;26(23):1388–98.
- Mayo NE, Poissant L, Ahmed S, et al. Incorporating the International Classification of Functioning, Disability, and Health (ICF) into an electronic health record to create indicators of function: proof of concept using the SF-12. *J Am Med Inform Assoc.* 2004;11(6):514–22.
- Friedman C, Alderson PO, Austin J, Cimino JJ, B. JS. A general natural language text processor for clinical radiology. *J Med Inform Assoc.* 1994;1(2):161–74.
- Rindflesch T, Rajah C, Jayant V. Extracting molecular binding relationships from biomedical text. *Proc 6th Appl Nat Lang Proc Conf; 2000; San Francisco: Association for Computational Linguistics; 2000:188–95.*
- Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc.* 2000;7:593–604.
- Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Annu Symp.* 2001:17–21.
- Hripcsak G, Wilcox A. Reference standards, judges and comparison subjects. *J Am Med Inform Assoc.* 2002;9(1):1–15.
- Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Inform Manage.* 2004;42(1):15–29.
- Cohen J. Weighted kappa: nominal scale disagreement with provision for scaled agreement or partial credit. *Psychol Bull.* 1968;(70):213–20.
- Viera A, Garrett J. Understanding interobserver agreement: The kappa statistic. *Fam Med.* 2005;37(5):360–3.