Reliability and validity of the American Hospital Association's national longitudinal survey of health information technology adoption

Jordan Everson,[1] Shoou-Yih D Lee,[1] Charles P Friedman[1,2]

[1]University of Michigan School of Public Health, Ann Arbor, Michigan, USA
[2]University of Michigan School of Information, Ann Arbor, Michigan, USA

Correspondence to
Jordan Everson, University of Michigan, School of Public Health, 1435 Washington Heights, M3149A SPH II, Ann Arbor, MI 48109-2029, USA; jeverson@umich.edu

Received 29 October 2013
Revised 24 February 2014
Accepted 25 February 2014
Published Online First 12 March 2014

ABSTRACT

Objective To evaluate the internal consistency, construct validity, and criterion validity of a battery of items measuring information technology (IT) adoption, included in the American Hospital Association (AHA) IT Supplement Survey.

Methods We analyzed the 2012 release of the AHA IT Supplement Survey. We performed reliability analysis using Cronbach's α and part-whole correlations, construct validity analysis using principal component analysis (PCA), and criterion validity analysis by assessing the items' sensitivity and specificity of predicting attestation to Medicare Meaningful Use (MU).

Results Twenty-eight items of the 31-item instrument and five of six functionality subcategories defined by the AHA all produced reliable scales (α's between 0.833 and 0.958). PCA mostly confirmed the AHA's categorization of functionalities; however, some items loaded only weakly onto the factor most associated with their survey category, and one category loaded onto two separate factors. The battery of items was a valid predictor of attestation to MU, producing a sensitivity of 0.82 and a specificity of 0.72.

Discussion The battery of items performed well on most indices of reliability and validity. However, they lack some components of ideal survey design, leaving open the possibility that respondents are not responding independently to each item in the survey. Despite measuring only a portion of the objectives required for attestation to MU, the items are a moderately sensitive and specific predictor of attestation.

Conclusions The analyzed instrument exhibits satisfactory reliability and validity.

BACKGROUND AND SIGNIFICANCE

Health information technology (HIT) is composed of numerous interdependent functionalities that may be variably configured, implemented, and used. The accelerating rate of HIT adoption has driven interest in measuring the pattern, causes, and consequences of hospitals' use of HIT.[1] [2] Despite this interest, no survey instrument has been formally validated as providing high quality data on hospitals' HIT use. Likely due to this lack of validated sources of data, most HIT studies have focused on single or a few hospital sites, or on a specific technology such as computerized provider order entry (CPOE) and clinical decision support.[3] Those studies that use national data from unvalidated instruments cannot be sure that their results measure real adoption effects, rather than relationships driven by biases or attenuated relationships driven by measurement error. Validation of national

data sources would allow for more confidence in research on many hospitals and a broad range of functionalities over a longer time horizon, producing more generalizable results.

This paper uses classical measurement theory to analyze the psychometric quality—that is, the internal consistency, construct validity, and criterion validity—of a battery of survey items on HIT adoption in the American Hospital Association (AHA) Information Technology (IT) Supplement Survey. Past research has evaluated the quality of other hospital HIT surveys by comparing the consistency of responses to similar questions across multiple instruments.[4] [5] This research has shown that hospitals that were identified as using CPOE by Healthcare Information Management Systems Society (HIMSS) Analytics were not consistently identified as using CPOE by the Leapfrog Group, and that the HIMSS identification of hospital electronic medical record (EMR) use is similarly inconsistent with the AHA Annual Survey of Hospitals' general measure of EMR adoption.

In 2007, The Office of the National Coordinator for Health Information Technology contracted with the AHA to begin measuring the progress of hospital HIT adoption. The AHA has an established record of gathering high quality data on hospitals through their annual survey.[6] Beginning in 2008, the AHA began administering a supplement to its annual survey measuring hospitals' HIT use. The AHA IT Supplement Survey serves an important function in assessing the state of HIT nationally, because it provides comparable year-by-year data for all hospitals regardless of participation in public programs like Meaningful Use (MU). These data have been used in several high profile publications tracking the adoption of HIT.[7–10]

Despite the AHA's track record of measuring overall hospital characteristics, it is not clear that the AHA IT Supplement Survey has resolved the problems identified with other measures of HIT use, particularly because the AHA's prior measure of EMR use, included on the general hospital survey, showed little agreement with HIMSS's measures of adoption. Measurement of HIT adoption by survey is subject to several sources of unreliability and invalidity, including the social desirability bias of having an advanced system and respondents' incomplete knowledge of HIT functionalities throughout the entire hospital or several hospitals in a larger system for which they respond. Despite these concerns, the AHA IT Supplement Survey is likely the best method, short of some type of automated functionality checking,

of accurately measuring and tracking hospital adoption of a broad range of HIT functions.

OBJECTIVES

This study examines the reliability and validity of a battery of items that measure HIT adoption on the AHA IT Supplement Survey. The survey has changed over the 5 years since its initial publication, mostly by the addition of questions as new areas of interest appeared. The battery of 31 questions examined in this study is one of the few parts of the survey that has been asked every year and that can form the basis of a longitudinal dataset for continuous tracking. This battery of questions is divided into six categories and captures functionalities that are considered essential components of a comprehensive HIT system in hospitals. Previous research has used the battery to measure the degree of hospital adoption of HIT by counting the total number of functionalities implemented.[7] [11]

We use classical measurement theory to analyze the properties of response to this battery of survey items with the goal of addressing the following research questions:

1. To what extent is the instrument reliable, as indicated by various indices of internal consistency among the individual items?
2. Regarding construct validity, do the functionalities display a latent factor structure that parallels the survey's categories?
3. Is the survey instrument valid in relation to an external criterion?

METHODS
Data

The AHA IT Supplement Survey was developed through a collaborative and iterative process including input from survey experts, chief information officers, other hospital leaders, health policy experts, and HIT experts using a modified Delphi panel.[7] This study focuses on a battery of questions measuring the use of 31 separate HIT functionalities that appears in every year of the survey. The survey questionnaire divides the measured HIT functionalities into six categories: *electronic clinical documentation*, *results viewing*, *computerized provider order entry*, *decision support*, *bar coding*, and *other functionalities*. Over the years these functionality categories have expanded—for instance the survey now asks about the system's ability to capture gender, ethnicity, and vital statistics—but the categories have remained the same.

Like the AHA's general survey, the IT Supplement Survey is administered to all hospitals that are members of the AHA. The hospital's chief executive officer is asked to assign completion of the survey to the most knowledgeable person in the organization. Non-responding hospitals receive emails and telephone calls to remind them to complete the survey. While data are available for 2008–2012, this analysis will focus on the data gathered in 2011 and released in 2012, because the results of our analysis were similar over multiple years, and because the test of validity through comparison to attestation for MU is most meaningful when the survey and attestation occurred in the same year.[12]

The sample for this analysis is limited to non-federal, general, acute care hospitals in the 50 states. The exclusion of federal, specialty hospitals, and hospitals in US territories is appropriate because those hospitals might adopt HIT in quite different configurations from general hospitals, and their deviation from the general pattern of adoption might bias the psychometric evaluation in this study. A total of 2331 non-federal, general, acute care hospitals, out of 4565 surveyed (51.1%), responded to the

full battery of functionality questions in 2011. Missing data were not imputed in order to avoid inflation of reliability.

Measures

Each of the 31 measured functionalities is coded on a 1–6 scale. In the survey, 'fully implemented across all units' is coded 1, 'fully implemented in at least one unit' is coded 2, 'beginning to implement in at least one unit' is coded 3, 'have resources to implement in next year' is coded 4, 'do not have resources but considering implementing' is coded 5, and 'not in place and not considering implementation' is coded 6. We reverse coded the response so that a higher score corresponds to the functionality being closer to full implementation. Past studies have used various cutoff points to create dichotomous measures of 'implementation' and 'no implementation': some consider a function implemented if it is implemented in at least one unit, and some require it to be implemented in all units.[7] [11] We defined implementation as a functionality being at least 'fully implemented in at least one unit,' reasoning that this division is the most distinguishable and that this is more consistent with the MU objectives of the functions being used on a certain percentage of patients rather than throughout the hospital. We also varied the cutoff points in additional analyses and the results were not sensitive to the choice of the cutoff point.

Data from the Centers for Medicare and Medicaid Services (CMS) list of hospitals that attested to Medicare MU in 2011 were acquired from the CMS website.[13] These hospitals were matched to hospitals from the AHA IT survey based on the hospital's National Provider Identifier (NPI) and AHA identification. Because the AHA survey does not have a complete list of hospital NPIs, hospitals were also manually matched based on names and addresses. Seven of the 835 hospitals that attested to MU in 2011 were not matched due to missing data. The sample used for this analysis contained data on 447 (53.5%) of the hospitals that attested and 1884 that did not. Hospitals that responded to the survey were significantly more likely to attest to MU than hospitals that did not; 19.2% of respondents attested, while 16.0% of non-respondents attested (two-sample t test: t=2.8, p=0.003).

Validity and reliability assessment

In classical measurement theory, the two key psychometric properties of a measurement tool are its reliability, defined as the extent to which the tool produces consistent results, and validity, the degree to which the tool measures what it purports to measure, for example the extent to which an IQ test measures intelligence. Reliability is often measured by comparison of similar or identical survey questions under the assumption that answers to questions measuring the same underlying concept should be highly correlated. Validity is usually assessed by comparison to some external criteria.[14] There are many tests of these measurement properties, only some of which are pursued here.

In our psychometric analysis, we assumed that the 31 survey questions on the AHA IT Supplement Survey were representative of distinct HIT functionalities, that they form homogeneous subgroups, and that they contribute to a total measure of HIT adoption. On the basis of these assumptions, the following three steps were taken to assess the psychometric properties of the 31 survey questions. To address the first research question, item analysis was performed to examine the extent to which each functionality was correlated with scores on the total scale and the category it belongs to, and to identify low-correlation items, which may measure a different concept than indicated by

the survey's construction. Each item's relationship with these aggregate scores was assessed using corrected part-whole correlation, which is the correlation between a functionality's implementation status and the total number of HIT functionalities implemented, excluding the measured functionality.[15] In addition, we calculated internal-consistency reliability using Cronbach's α to test the degree to which respondents answered consistently on similar questions. The internal consistency was calculated for the total battery and the six categories.

To address the second research question, we performed principal component analysis (PCA) to examine whether the 31 items were indicators of an underlying, dominant factor and to test whether the survey's grouping of items into six categories is valid. The validity of these categories would be undermined if items that are grouped by the survey loaded onto factors that are more strongly associated with items in other categories, indicating that the errant item is poorly categorized. Because these factors are unlikely to be uncorrelated, oblique rotation using promax rotation was used.

To address the final research question, we examined the criterion validity of the items—the ability of the items to predict an outcome external to the survey itself—by comparing the survey responses to MU attestation. A strong correlation between MU attestation and the implementation of functions necessary to achieve MU would indicate that the items are valid. Attestation under the HITECH Act is a reasonable external criterion against which the validity of the items can be evaluated for three reasons. First, MU attestation is external to the AHA IT survey, which is not used to qualify hospitals for MU. Second, the survey items and MU attestation do not measure an identical phenomenon. Whereas the survey measures the implementation of HIT functions, indicators of actual use of the functions must be reported to achieve MU. Third, because of the incentives around it, attestation is close to a 'gold standard' in the measurement of HIT. To receive the subsidies made available by the federal government under Stage 1 MU, hospitals are required to attest to achieving specific usage milestones for 14 core functionalities and 5 of 10 additional menu objectives.[16] Falsely attesting to the achievement of these criteria is fraud.[17] As such, hospitals would be expected to have a strong interest in ensuring that they were being truthful in reporting their achievement of the MU criteria. The same motivations do not apply to the AHA's voluntary survey.

We matched 11 items included in the 31-item AHA survey instrument to 7 of the 14 core MU objectives, following methodology used by Jha et al[9] to describe achievement towards MU. These 11 matched items were used to create a subscale that was employed to predict MU attestation. Some of the 11 matched functions fulfill the same objective, so only seven core MU objectives are measured. For instance, reporting implementation of one of four decision support functions—clinical guidelines, clinical reminders, drug–laboratory interactions, or drug-dosing support—is considered as fulfilling the core MU objective to 'implement one clinical decision support rule relevant to a high priority hospital condition along with the ability to track compliance with that rule.'[16] Like Jha et al[9] and DesRoches et al[10], we require that hospitals report implementation of the function in at least one unit for the survey item to satisfy the MU objective. We believe this definition is consistent with the MU objectives' requirements that the functions are used on a certain percentage of patients, or in a limited way, rather than being used throughout the hospital.

Measurements of the sensitivity, specificity, and predictive values of the test were generated along with a χ^2 test to determine the likelihood of observing the outcome by chance. The area under the receiver operating curve was also generated as a summary measure of the instrument's success at predicting MU attestation. The survey's success at predicting attestation to MU depends on where the cut point to predict attestation is drawn. Changing the classification rule so that hospitals that report six of the seven measured objectives are still classified as likely to attest to MU increases the sensitivity while decreasing its specificity. In addition, we performed a test of differential validity to show that attestation status was more strongly associated with respondents' answers to the MU-related survey items than their response to non-MU-related items. This provides evidence that MU attesters are not simply responding positively to all aspects of the survey, but rather are more precisely stating that they have implemented the functions associated with the MU program.

All statistical analyses were performed using Stata V.12.[18]

RESULTS
Reliability
Table 1 describes the corrected part-whole correlation of all 31 items as they relate to total HIT adoption and to the category in which they are classified on the survey. The 24 items included in the first four categories, *electronic clinical documentation*, *results viewing*, *computerized provider order entry*, and *decision support*, have previously been defined as the components of a comprehensive electronic health record (EHR) and are all closely associated with the total score of HIT adoption (median correlation: 0.670, range: 0.447–0.725).[7] In addition, the items in the *bar coding* category correlate well with total adoption, although the coefficients are slightly smaller than the first four categories (median: 0.4995, range: 0.434–0.527). Adoption of *other functionalities* is markedly less associated with the total score than the other five groups (median: 0.259). Most, but not all items, were more closely associated with adoption of other items in their category than they were with the adoption of all items.

As table 2 shows, the internal consistency of the entire scale and of most of the categories is quite high, surpassing the threshold of 0.7 commonly used to indicate a consistent scale.[14] Only the *other functionalities* category exhibits low internal reliability (α=0.479). Because α scores are in part dependent on the number of indicators, more items will generally increase α. Despite this characteristic, including *other functionalities* in the total battery decreases its reliability. Additional analysis with prior years of this survey showed that the reliability of the survey is relatively stable over time, but generally increases slightly each year (not shown).

Construct validity
PCA was performed on 28 of the 31 functions, excluding only the items in the *other functionalities* category as they are not highly correlated with the scale or each other. The first extracted component explains a large proportion (43.3%) of the variance in the data, while the other components explain more modest amounts, suggesting that a dominant factor underlies those 28 functions (table 3). Six components were retained because their eigenvalues exceed 1.0.

Factor loadings after promax rotation are presented in table 3. For ease of comprehension and identification of Simple Structure, variables with a factor loading less than 0.10 are left blank.[19] A common rule of thumb for factor analysis is that variables are associated with factors if their loading exceeds 0.3.[20] Using this rule of thumb, the functions in the same category largely load together on the same factor. However, *results*

Table 1 Part-whole correlations of health information technology (HIT) functionalities in the 2011 American Hospital Association IT Supplement Survey

	Frequency	Total part-whole correlation	Category part-whole correlation
Electronic clinical documentation			
Patient demographics	0.918	0.496	0.545
Physician notes	0.475	0.515	0.513
Nursing assessments	0.779	0.683	0.722
Problem lists	0.586	0.619	0.657
Medication lists	0.779	0.697	0.752
Discharge summaries	0.721	0.632	0.712
Advanced directives	0.703	0.630	0.626
Results viewing			
Laboratory reports	0.913	0.538	0.682
Radiology reports	0.910	0.511	0.709
Radiology images	0.915	0.447	0.619
Diagnostic test results	0.766	0.589	0.773
Diagnostic test images	0.696	0.555	0.699
Consultant reports	0.731	0.583	0.641
Computerized provider order entry			
Laboratory tests	0.550	0.706	0.946
Radiology tests	0.549	0.694	0.942
Medications	0.537	0.724	0.925
Consultation requests	0.474	0.714	0.868
Nursing orders	0.551	0.725	0.882
Decision support			
Clinical guidelines	0.477	0.678	0.700
Clinical reminders	0.534	0.695	0.732
Drug allergy alerts	0.766	0.715	0.792
Drug–drug interaction alerts	0.761	0.709	0.797
Drug–laboratory interaction alerts	0.640	0.662	0.776
Drug dosing support	0.614	0.688	0.786
Bar coding			
Laboratory specimens	0.671	0.434	0.510
Tracking pharmaceuticals	0.540	0.494	0.730
Pharmaceutical administration	0.556	0.505	0.717
Patient ID	0.662	0.527	0.698
Other functionality			
Telemedicine	0.428	0.245	0.266
Radiology frequency	0.168	0.259	0.292
Physician personal digital assistant	0.303	0.357	0.354

Table 2 Reliability of the American Hospital Association's health information technology battery and component categories

Scale	Reliability coefficient
Total scale (31)	0.955
Total scale excluding other functionalities (28)	0.958
Total scale excluding bar coding and other functionality (24)	0.949
Electronic clinical documentation (7)	0.865
Results viewing (6)	0.868
Computerized provider order entry (5)	0.970
Decision support (6)	0.915
Bar coding (4)	0.833
Other functionality (3)	0.479

viewing functions load on two separate factors, and the *patient demographics* item loads on a factor (factor 4) different from the expected one—that is, *electronic clinical documentation*. Three functions, *advanced directives*, *clinical guidelines* and *clinical reminders*, do not surpass the threshold to be considered as loading on any one factor.

We repeated the item analysis, PCA, and reliability tests using ordinal (the 1–6 scale) and dichotomous (using different cutoff points) forms of the functionality variables and the results did not change substantively (results are available upon request). We also performed the analyses on four available years of the IT supplement data and found modest improvement in reliability and part-whole correlation over time.

Criterion validity

The survey's success at predicting attestation to MU depends on where the cut point to predict attestation is drawn. As shown in table 4, if reporting implementation of functionality sufficient to fulfill all seven relevant MU objectives is used as the criterion to predict MU, the survey correctly categorizes 83.3% of hospitals that attested (sensitivity), and 71.7% of hospitals that did not attest (specificity). Using the same criterion, 40.9% of hospitals predicted to qualify for MU actually attested (positive predictive value (PPV)). On the other hand, reporting not having fulfilled all measured objectives accurately predicts 94.6% of hospitals that did not attest to MU (negative predictive value (NPV)). If hospitals that omit one measured MU objective are associated with attestation, this improves the sensitivity to 92.8% and the NPV to 96.8%, but reduces the specificity (51.2%) and PPV (31.1%). The area under the receiver operating characteristic curve based on the seven objectives related to the instrument evaluated above is 0.788.

Institutions that attested to MU reported on the survey that they had implemented more MU-related functions than non-MU related functions. The opposite was true for institutions that did not attest. This is an indication of differential validity. As figure 1 shows, hospitals that did not attest to MU reported implementation of a slightly smaller fraction of MU functions than non-MU functions (0.61 and 0.63, respectively; p<0.001). Conversely, hospitals that did attest to MU reported implementation of a larger fraction of MU functions than non-MU functions (0.91 and 0.86, respectively; p<0.001). The interaction between attestation and MU item status was statistically significant (p<0.001).

DISCUSSION

In general, the AHA HIT survey is a reliable and valid measure of hospital HIT use. Of the 31 functions, only the three functions in the *other functionalities* category behaved poorly in terms of part-whole correlation and reliability. All other functions related well to the total scale, demonstrating that the battery measures an underlying concept of hospital HIT adoption.

The high reliability of the categories measured here is particularly convincing because it is likely an underestimate of true reliability. Cronbach's α is an exact measure of reliability when items are τ-equivalent: when average scores on each item and the items' variances are identical.[14] Because the items measured here do not fulfill either of these criteria, the observed α represents a lower bound of the true consistency of these items. Despite this quality, the 28 items are able to generate highly reliable scales from as few as four variables, indicating that measurement error is not a problem. In fact, the high inter-correlations of some of these measures, such as the CPOE category, raises the question of

Table 3 Principal component analysis of 28 health information technology (HIT) functions in the 2011 American Hospital Association IT Supplement Survey

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
E1. Patient demographics			0.217	0.321		−0.101
E2. Physician notes		−0.106	0.396	−0.116		
E3. Nursing assessments			0.353			
E4. Problem lists			0.454	−0.123		
E5. Medication lists			0.384			
E6. Discharge summaries			0.444			
E7. Advanced directives			0.296			
R1. Laboratory reports				0.496		
R2. Radiology reports				0.504		
R3. Radiology images				0.526		
R4. Diagnostic test results				0.118		0.573
R5. Diagnostic test images						0.633
R6. Consultant reports			0.116			0.416
C1. Laboratory tests	0.470					
C2. Radiology tests	0.477					
C3. Medications	0.436					
C4. Consultation requests	0.399					
C5. Nursing orders	0.410					
D1. Clinical guidelines	0.114	0.262		−0.150		0.158
D2. Clinical reminders		0.288		−0.138		0.142
D3. Drug allergy alerts		0.448		0.104		−0.102
D4. Drug–drug interaction alerts		0.464				
D5. Drug–laboratory interaction alerts		0.476				
D6. Drug dosing support		0.431				
B1. Laboratory specimens					0.376	
B2. Tracking pharmaceuticals					0.557	
B3. Pharmaceutical administration					0.538	
B4. Patient ID					0.494	
Cronbach's α	0.970	0.918	0.849	0.867	0.833	0.856

Principle component analysis was performed only for the 2330 hospitals that had no missing data.
The six extracted components initially explained 43.3%, 10.1%, 6.5%, 5.8%, 4.0%, and 4.0% of the variance in the data. The corresponding eigenvalues were 12.13, 2.83, 1.82, 1.62, 1.13, and 1.13.

whether they are measuring distinct functions that can be meaningfully separated, or whether implementation patterns and vendor packaging of similar functionality makes these survey items nearly redundant. Because CMS requires that certified EHR vendors offer technologies that fulfill all MU objectives, vendors may bundle multiple components together, and hospitals that purchase a single function may simultaneously purchase, but not necessarily implement, these other related functions.[21]

The high consistency of categories also raises the possibility that respondents are inattentive to the individual questions and simply checking all the boxes within each category, creating an artificially high reliability. The battery of items lacks some of the qualities of a good psychometric instrument that would reduce the strength of this effect: similar items are grouped together
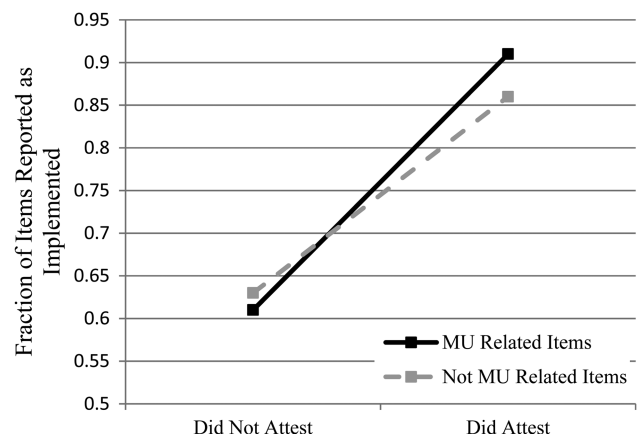
Table 4 Cross-tabulations for fulfilling all measured Meaningful Use (MU) objectives in the 2011 American Hospital Association Information Technology Supplement Survey and attesting to MU

	Attested	Did not attest	PPV and NPV
All 7 MU objectives	370	534	0.409
Not all MU objectives	77	1350	0.946
Sensitivity and specificity	0.833	0.717	2331
Pearson χ² (1)=450.83, p<0.001			
≥6 objectives	415	919	0.311
<6 objectives	32	965	0.968
Sensitivity and specificity	0.928	0.512	2331
Pearson χ² (1)=286.55, p<0.001			

NPV, negative predictive value; PPV, positive predictive value.

Figure 1 Differential validity: fraction of Meaningful Use (MU) and non-MU items implemented by attestation status.

under an explicit heading rather than spread across the instrument, no attention-check questions are included specifically to account for respondents' attentiveness, and no questions are reverse coded. While some of these features would be impractical given the subject matter, it is difficult to assess how thoughtfully respondents are answering individual questions on the survey.

The PCA affirms that these functionalities form both a single scale of total HIT adoption and intuitively meaningful categories. The very large amount of variance explained by the first factor prior to rotation indicates that all of the functions can be associated with a single dimension. For the most part, functionalities load onto factors in a pattern parallel to the survey instrument's categorization, demonstrating the validity of the categories. However, some variables loaded on their category's factor much less strongly than others. This may be because the variable is weakly related to the category—for instance adoption of *advanced directives* appears to be only weakly associated with adoption of more central clinical documentation tools—or because of the different frequencies of adoption of the functions. As an example, over 90% of hospitals have adopted the *patient demographics* functionality, while less than 40% have adopted *physician notes*. This discrepancy in commonness or difficulty reduces the amount of covariance these functionalities share.[22] The effect of different frequency levels is most apparent in factors 4 and 6, which divide the *results viewing* component of the survey into two parts, the more commonly adopted and less commonly adopted functionalities. The *patient demographics* item loads on the common *results viewing* factor (factor 4), further indicating the role of frequency on covariance. Additional analysis using item response theory might better describe the relationship between functions that differ in frequency of adoption.[23]

Judging by the high internal consistency of the overall battery and the factor loading patterns, the 28 functionalities, excluding the *other functionalities* category, form an overall scale that collectively measures the underlying capability of HIT adoption. This affirms the approach of previous studies that have used these items additively. In past work, both the *bar coding* and *other functionalities* categories have been excluded from the definition of a comprehensive EHR.[7] Because the items in the *other functionalities* category do not correlate well with the whole and reduce the reliability of the all-item scale, their exclusion from the definition of a comprehensive EHR is supported by empirical assessment. However, the exclusion of *bar coding* functions from the measurement of a comprehensive EHR is less clear from an empirical perspective; they correlate reasonably well with total adoption and form a coherent group unto themselves. The *electronic clinical documentation* category is the most nebulous, since some of the items have relatively low correlations with the category and are weakly associated with the factor (factor 3) that is most representative of the category. Interestingly, this is the category that has been expanded the most over the years, and might demand refinement.

Responses to the survey were moderately successful predictors of MU attestation. Varying the cut point illustrates a clear trade-off between specificity and sensitivity. While these two qualities always form a trade-off, so that increasing specificity by adjusting the cut point results in decreased sensitivity, a more valid overall measurement instrument would allow for the possibility of high sensitivity and specificity.

Regardless of cut point selection, the survey performed least well in terms of PPV—that is, correctly predicting attestation based on fulfilling all objectives as measured by the survey.

There are several ways to explain a low correlation between a self-report of implementation of all MU-relevant functions and MU attestation. First, hospitals may be over-reporting the state of their HIT adoption, perhaps in line with a social desirability bias. However, reported implementation of MU-related items was more closely related to attestation than implementation of other items, showing that the survey responses were somewhat precisely associated with the external gold standard, contrary to an across-the-board desirability effect. Second, not all of the MU core and menu objectives are measured in the AHA IT Supplement Survey, so that hospitals might not have the necessary complement of functions to meet MU despite having all measured functions. Third, the questions included on the survey do not exactly match the MU objectives and measure. For instance, MU requires that at least 30% of patients' medication orders are entered electronically. The survey only measures whether medication CPOE is implemented in one or all hospital units. This represents a different and likely lower bar for achieving MU than attestation, especially for those hospitals that may have only implemented the system in a single unit. Fourth, hospitals were able to attest to MU during any quarter of 2011, with the final attestation period occurring in December, whereas the survey data collection ended earlier than the window for attestation. Some hospitals may have implemented necessary functions between the survey and their attestation date. Even with these limitations, the survey's moderate success at predicting attestation is evidence that it is usefully measuring adoption of HIT. Future studies should look to use the categories of similar items as scales in order to increase overall reliability and reduce the potential effects of attenuation bias caused by using only a single survey item as the key variable of interest.

Our study only addresses the quality of the battery of HIT functionalities in the AHA survey within one recent year of survey data. While similar results were replicated using other years of the survey, we cannot draw conclusions about the inter-year reliability of the survey.

CONCLUSION

The AHA IT Supplement Survey provides a national source of reliable and valid measures of hospitals' adoption of HIT. The internal consistency of the overall scale and its subscales indicates high reliability suitable for most analyses. The usefulness of the HIT function items in the survey was demonstrated by showing that hospitals' attestation to MU could be predicted reasonably well by identifying their responses to the survey, indicating that the amount of total error in the data is modest. The relatively high number of hospitals who reported having all measured functionalities but not attesting to MU indicates that responses may be biased towards over-reporting implementation of functionalities.

REFERENCES

1. Zhang NJ, Seblega B, Wan T, et al. Health information technology adoption in U.S. acute care hospitals. *J Med Syst* 2013;37:9907.
2. Baird A, Furukawa MF, Rahman B, et al. Corporate governance and the adoption of health information technology within integrated delivery systems. *Health Care Manage Rev*; Published Online First: 25 April 2013. doi:10.1097/HMR.0b013e318294e5e6

3   Wu S, Chaudhry B, Wang J, et al. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006;144:742–52.

4   Diana ML, Kazley AS, Menachemi N. An assessment of Health Care Information and Management Systems Society and Leapfrog data on computerized provider order entry. *Health Serv Res* 2011;46:1575–91.

5   Kazley AS, Diana ML, Menachemi N. The agreement and internal consistency of national hospital EMR measures. *Health Care Manag Sci* 2011;14:307–13.

6   AHA Annual Survey Database™ Fiscal Year 2011. http://www.ahadataviewer.com/book-cd-products/AHA-Survey/ (accessed 30 Sep 2013).

7   Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. *N Engl J Med* 2009;360:1628–38.

8   Jha AK, DesRoches CM, Kralovec PD, et al. A progress report on electronic health records in U.S. hospitals. *Health Affairs* 2010;29:1951–7.

9   Jha AK, Burke MF, DesRoches CM, et al. Progress toward meaningful use: hospitals' adoption of electronic health records. *Am J Manag Care* 2011;17: SP117–24.

10 DesRoches CM, Charles D, Furukawa MF, et al. Adoption of electronic health records grows rapidly, but fewer than half of US hospitals had at least a basic system in 2012. *Health Affairs* 2013;32:1478–85.

11 Blavin FE, Buntin MJB, Friedman CP. Alternative measures of electronic health record adoption among hospitals. *Am J Manag Care* 2010;16:e293–301.

12 American Hospital Association. Hospital EHR adoption database 2011 release: a supplement to the annual survey database, 2011.

13 Center for Medicare and Medicaid Services. *Data and program reports.* http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/DataAndReports.html (accessed 30 Sep 2013).

14 Allen MJ, Yen WM. *Introduction to measurement theory.* Monterey, CA: Brooks/Cole, 1979.

15 Friedman CP, Wyatt JC. *Evaluation methods in biomedical informatics*. New York: Springer, 2006.

16 Center for Medicare and Medicaid Services. Meaningful Use. What are the Requirements for Stage 1 of Meaningful Use? http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Meaningful_Use.html (accessed 30 Sep 2013).

17 Center for Medicare and Medicaid Services. Federal Fraud and Abuse Laws: Remaining in Compliance while Attesting to Meaningful Use. http://www.hitecla.org/sites/default/files/Compliance%20Presentation.pdf (accessed 30 Sep 2013).

18 StataCorp. 2011. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.

19 Gorsuch RL. *Factor analysis*. 2nd edn. Hillsdale, NJ: LEA, 1983.

20 Tabachnick BG, Fidell LS. *Using multivariate statistics*. Upper Saddle River, NJ: Pearson, 2012.

21 Medicare and Medicaid Programs. Electronic health record incentive program. Final rule. *Fed Regist* 2010;75:44314–518.

22 Carroll JB. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika* 1945;10:1–19.

23 Schuur WHV. Mokken scale analysis: between the Guttman scale and parametric item response theory. *Political Anal* 2003;11:139–63.