Health Informatics Journal

Volume 23 Number 1 March 2017

Health Informatics Journal

Article

Syntactic and semantic errors in radiology reports associated with speech recognition software

Health Informatics Journal
2017, Vol. 23(1) 3–13

DOI: 10.1177/1460458215613614
journals.sagepub.com/home/jhi
SAGE

Michael D Ringler, Brian C Goss and Brian J Bartholmai
Mayo Clinic, USA

Abstract

Speech recognition software can increase the frequency of errors in radiology reports, which may affect patient care. We retrieved 213,977 speech recognition software–generated reports from 147 different radiologists and proofread them for errors. Errors were classified as "material" if they were believed to alter interpretation of the report. "Immaterial" errors were subclassified as intrusion/omission or spelling errors. The proportion of errors and error type were compared among individual radiologists, imaging subspecialty, and time periods. In all, 20,759 reports (9.7%) contained errors, of which 3992 (1.9%) were material errors. Among immaterial errors, spelling errors were more common than intrusion/omission errors ($p < .001$). Proportion of errors and fraction of material errors varied significantly among radiologists and between imaging subspecialties ($p < .001$). Errors were more common in cross-sectional reports, reports reinterpreting results of outside examinations, and procedural studies (all $p < .001$). Error rate decreased over time ($p < .001$), which suggests that a quality control program with regular feedback may reduce errors.

Keywords

PowerScribe, quality control, radiology report, report errors, speech recognition

Introduction

Although there are many ways for radiologists to provide the results of an imaging examination to referring clinicians, the signed written report remains the primary and often sole means of communication.[1,2] In many radiology practices, transcribed dictation by a professional transcriptionist has been replaced by real-time speech recognition and self-editing of reports. In general, speech recognition software (SRS) greatly improves turnaround times for reports compared with remote transcription and allows for more immediate control over report editing than traditional paper markup or asynchronous transcription modification.[3–6] However, this often leaves the radiologist as the sole author and editor of the final text that is placed in the radiology information system or electronic medical record.

Most currently available commercial SRS products rely on nearest match for each word transcribed and do not check for logical relevance or perform natural language processing for real-time

Corresponding author:
Michael D Ringler, Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905, USA.
Email: ringler.michael@mayo.edu

recognition and transcription of dictation. Self-editing can be prone to typographic or other errors that may not be noticed by the radiologist. Consequently, persistent undetected report errors that could impede understanding or lead to erroneous conclusions are much more common than seen with expert transcription by a trained professional with excellent typing skills and an understanding of medical language and context. In addition to errors in radiologic diagnosis, simple errors in syntax and grammar that occur during report transcription can have dire consequences for which the signing radiologist is liable.[7] Communication errors are extremely common and one of the top reasons radiologists are sued for medical malpractice.[8] Self-editing errors by a radiologist represent a mitigable threat to appropriate patient care.

The primary reason for the emergence of errors in syntax and semantics in radiology reports is the relatively recent application of SRS and consequent assimilation of transcription duties by the radiologist.[4,6,9] Examples of such errors fall into several categories: (1) omission of appropriate words/phrases, which includes deletions and missing words; (2) intrusion of incorrect words/phrases, which includes interjection, incorrect words, wrong word substitution, insertions, or right–left substitutions; and (3) spelling errors, which includes word truncation, most likely due to the manual editing of text by a radiologist through typing errors or inaccurate selection of text to be removed or edited. Additional errors that do not necessarily fit into the above categories include incorrect dates, image/series numbering errors, measurement scale errors (e.g., cm vs. mm), template errors, and punctuation errors.[10] Different combinations of these error types may lead to nonsense phrases with variable effects on interpretation and comprehension of the final report. Some omissions or intrusions (particularly the word "no" in various contexts) and word substitutions (such as "new" rather than "no") can potentially affect patient care.[1]

At our institution, a portion of all the signed radiology reports generated with SRS are regularly audited by transcriptionists and assessed for syntactic and semantic errors. Any potential errors comprising apparently illogical or inappropriate words/phrases, misspellings, or other errors discovered by the transcriptionist trigger a notification of the staff radiologist with an opportunity to correct the report and notify the referring service. We reviewed the data from these audits to investigate any possible patterns that could help us improve our report quality in the future. Specifically, we investigated four hypotheses, on the basis of our clinical experience: (1) error rate and type vary by radiologist; (2) error rate and type vary by type of imaging examination—we expect more complicated and longer reports, such as cross-sectional (CS) imaging or procedural reports, to contain more errors than shorter plain radiography (CR) reports; (3) the implementation of a quality control program with regular feedback should decrease errors over time; and (4) a Dictaphone hardware upgrade should decrease error rates.

Methods

Our institutional review board approved the study protocol. Informed consent was waived because of the retrospective nature of the study and deidentification of the patients and radiologists involved in the analysis. As part of our department's radiology report quality assurance, an in-house transcriptionist reads every single report dictated using an SRS dictation system (PowerScribe; Nuance Communications, Inc) 2 days per month for every staff radiologist and evaluates these reports for potential errors. If the percentage of reports containing at least one error exceeds 3 percent on either day for a particular radiologist, then all of their reports are similarly scrutinized every subsequent day until their error rate drops below 3 percent. All radiologists with error rates below 3 percent continue to have all their reports audited, on average, 2 days per month. A total of 13 trained medical transcriptionists with experience ranging from 1 to 23 years participated.

Errors are categorized as "material" or "immaterial." *Material errors* are believed by the transcriptionist to potentially alter interpretation of the radiology report. Material errors trigger an email notification of the staff radiologist to allow him or her to correct or revise the report and notify the referring service. *Immaterial errors* are further subcategorized as spelling mistakes or intrusion/omission errors. Reports with multiple errors are only counted once and are classified by the most egregious error type (material > intrusion/omission > spelling). Incorrect date, incorrect measurement, and left–right substitution are all classified as material errors. Punctuation errors are ignored. Errors in radiologic interpretation were not included in this study. The use of macros or standard templates was not recorded.

In addition to type of error and radiology staff member, the date of the report and imaging subspecialty of the examination (ISE) were captured. Imaging methods included computed tomography (CT), magnetic resonance imaging (MRI), plain radiography (CR), nuclear medicine (NM), neuroradiology (NR), and ultrasonography (US). Specific ISEs recorded were "CT Body" (computed tomography of the chest, abdomen, pelvis, or extremities), "CT Neuro" (computed tomography of the head, neck, or spine), "CR," "MR Body" (magnetic resonance imaging of the chest, abdomen, pelvis, or extremities), "MR Neuro" (magnetic resonance imaging of the head, neck, or spine), "NM," "NR" (NR procedures such as lumbar puncture and myelography), "US," "V&I" (vascular or interventional procedures), and "OS" (reinterpretation of any radiology study from another/outside institution of any body part or modality). For analysis purposes, several groups of ISEs were created, including a cross-sectional (CS) imaging group (CT Body, CT Neuro, MR Body, MR Neuro, and US), a procedural group (V&I and NR), and a diagnostic group (CT Body, CT Neuro, CR, MR Body, MR Neuro, US, and NM).

We retrospectively retrieved all reports generated by SRS and signed by 147 different radiologists from 3 January 2011 through 16 April 2014. Mammography reports were excluded because only a small fraction of these examinations are interpreted and reported using SRS at our institution. Similarly, many of our CR examinations and procedures are transcribed without SRS, either by immediate direct transcription in the room, or asynchronously through a digital dictation and remote transcription system. The main reason for the use of direct transcription over SRS with CR examinations is that direct transcription generates a finalized report faster, and CR results are often emergently required. Therefore, the number of CR reports completed with SRS is far less than the total number of these types of examinations reported at our institution. Radiologists were grouped into four categories on the basis of their total error percentage quartiles over the entire time period: group 1 (<5.5% total errors), group 2 (5.5%–7.9% total errors), group 3 (8.0%–10.5% total errors), and group 4 (>10.5% total errors).

Reports were divided into four time periods of exactly 300 days to analyze trends over time: 3 January to 29 October 2011, 30 October 2011 to 24 August 2012, 25 August 2012 to 20 June 2013, and 21 June 2013 to 16 April 2014. A total of 36 radiologists were excluded from the time analysis portion only because of an insufficient number of reports (<100) reviewed by transcription as part of the quality control project in any of the time periods to mitigate confounding of time-based trends by individual radiologists.

Our department updated the dictation microphones from the PowerMic I to PowerMic II Dictaphone (Nuance Communications, Inc) during August and September 2013. To test for differences in error rate as a result of this hardware modification, we also compared reports created in the 6 months immediately before (1 February through 31 July 2013) and immediately after (1 October 2013 through 31 March 2014) the upgrade.

Descriptive categorical data are presented using counts and percentages. Contingency (χ^2) analysis and multiple logistic regression were used as appropriate for comparing nominal data, with calculation of odds ratios (ORs) and 95 percent confidence intervals (CIs). p values <.05 were considered statistically significant. Analyses were performed using JMP version 9.0.3 (SAS Institute, Inc).

Table 1. Percentage of errors per radiologist.

Total errors per radiologist (%)	No. (%) of radiologists ($N = 147$)
0–2.5	3 (2.0)
2.6–5.0	29 (19.7)
5.1–7.5	34 (23.1)
7.6–10.0	38 (25.9)
10.1–12.5	19 (12.9)
12.6–15.0	11 (7.5)
15.1–17.5	6 (4.1)
17.6–20.0	1 (0.7)
20.1–22.5	2 (1.4)
22.6–25.0	2 (1.4)
25.1–27.5	1 (0.7)
27.6–35.0	0 (0)
35.1–37.5	1 (0.7)

Results

Errors by radiologist

A total of 213,977 reports were retrieved. Among these, 20,759 (9.7%) had errors, including 3992 (1.9%) with material errors. The mean (standard deviation (SD)) total error percentage by radiologist was 8.7 percent (5.0%; range, 0.8%–35.1%), and the percentage differed significantly among radiologists ($p < .001$; Table 1). The mean (SD) percentage of material errors per radiologist (out of total errors) was 16.2 percent (7.9%; range, 0.0%–38.7%), which also varied significantly among radiologists ($p < .001$). Among all immaterial errors ($n = 16{,}767$; 80.8% of all errors), spelling errors ($n = 10{,}151$; 60.5%) were more common than intrusion/omission errors ($n = 6616$; 39.5%; $p < .001$).

Errors by exam type

When the data were separated by ISE category, the mean (SD) overall error percentage per category was 11.4 percent (5.2%), material error percentage was 20.8 percent (4.6%) of total errors, and spelling error percentage was 63.1 percent (7.1%) of immaterial errors. These error rates also varied significantly by ISE ($p < .001$). The ISEs NR and MR Body had the most errors, and US and CR had the fewest (Table 2).

Percentages of errors for different types of reports are shown in Table 3. Compared with in-house dictations, reports dictated on outside examinations (OS category) were significantly more likely to result in an error (OR, 1.55; 95% CI, 1.48–1.62) or material error (OR, 1.67; 95% CI, 1.52–1.84; $p < .001$). CS reports were much more likely to contain an error (OR, 3.72; 95% CI, 3.51–3.95) than CR reports ($p < .001$), although material errors were more likely with CR than CS reports (OR, 1.18; 95% CI, 1.02–1.36; $p = .03$). There was no difference in spelling errors between CS and CR reports ($p = .55$). Total errors (OR, 1.91; 95% CI, 1.76–2.07), material errors (OR, 1.69; 95% CI, 1.43–2.00), and spelling errors (OR, 1.79; 95% CI, 1.47–2.17) all were more common in the procedural group than the diagnostic imaging group (all $p < .001$).

Error type trends

The four groups representing quartiles of radiologists, comprised 38 radiologists in group 1 (<5.5% total errors), 36 in group 2 (5.5%–7.9% total errors), 37 in group 3 (8.0%–10.5%), and

Table 2. Errors by ISE category.

ISE	Material errors[a] ($n = 3992$)	Immaterial errors[b]		All errors/no. of reports (%) ($N = 20,759$)
		Omission/intrusion ($n = 6616$)	Spelling ($n = 10,151$)	
NR	96	92	208	396/2010 (19.7)
MR Body	455	870	1221	2546/14,079 (18.1)
OS	669	836	964	2469/17,924 (13.8)
V&I	100	49	185	334/2506 (13.3)
CT Body	1205	2179	3553	6937/55,094 (12.6)
MR Neuro	438	920	1197	2555/25,303 (10.1)
CT Neuro	316	506	823	1645/17,296 (9.5)
NM	187	295	553	1035/12,162 (8.5)
US	279	492	843	1614/28,925 (5.6)
CR	247	377	604	1228/38,678 (3.2)

ISE: imaging subspecialty of the examination; NR: neuroradiology; MR Body: magnetic resonance imaging of the chest, abdomen, pelvis, or extremities; OS: reinterpretation of any radiology study from another/outside institution of any body part or modality; V&I: vascular or interventional procedures; CT Body: computed tomography of the chest, abdomen, pelvis, or extremities; MR Neuro: magnetic resonance imaging of the head, neck, or spine; CT Neuro: computed tomography of the head, neck, or spine; NM: nuclear medicine; US: ultrasonography; CR: plain radiography.
[a]Material errors are those believed by the transcriptionist to potentially alter interpretation of the radiology report.
[b]Immaterial errors are those believed not to alter interpretation of the radiology report (i.e. intrusion/omission errors or spelling mistakes).

Table 3. Comparison of errors by report type.

Report category	Total reports	Errors		
		All[a]	Immaterial[b]	Material[b]
Origin of examination				
In-house	196,053	18,290 (9.3)	14,967 (81.8%)	3323 (18.2%)
OS	17,924	2469 (13.8)	1800 (72.9%)	669 (27.1%)
Imaging type				
CS	140,697	15,297 (10.9)	12,604 (82.4%)	2693 (17.6%)
CR	38,678	1228 (3.2)	981 (79.9%)	247 (20.1%)
Study type				
Procedural	4516	730 (16.2)	534 (73.2%)	196 (26.8%)
Diagnostic	191,537	17,560 (9.2)	14,433 (82.2%)	3127 (17.8%)

OS: reinterpretation of any radiology study from another/outside institution of any body part or modality; CS: cross-sectional imaging group (see the "Methods" section); CR: plain radiography.
[a]Number (%) of reports with an error.
[b]Number (%) of error type per all errors.

36 in group 4 (>10.5% total errors). With the exception of comparisons between adjacent groups 4 and 3 ($p = .15$) and 2 and 1 ($p = .45$), all other comparisons between groups demonstrated significantly increased probability of material error with increasing total error percentage ($p < .001$; Figure 1). The largest difference was between groups 4 and 2 (OR, 1.59; 95% CI, 1.42–1.79) and the smallest was between groups 3 and 1 (OR, 1.41; 95% CI, 1.23–1.62). There was no association between radiologist group and proportion of spelling errors ($p = .12$).

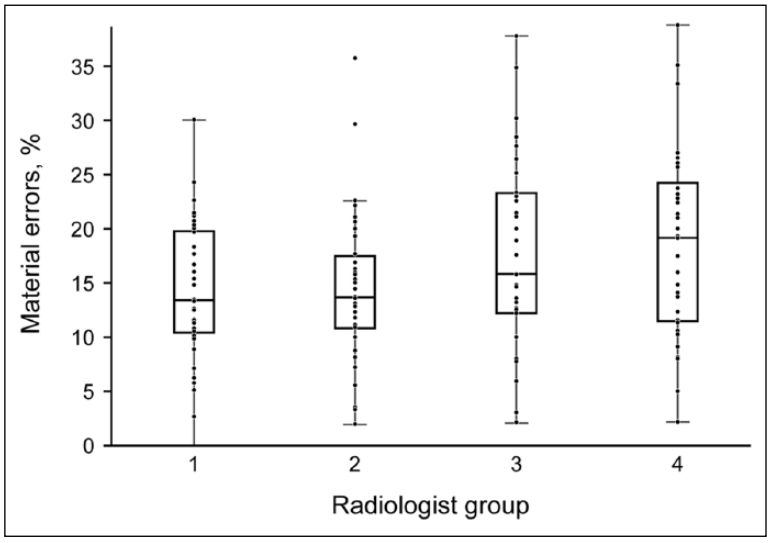

Figure 1. Percentage of material errors by radiologist error quartile ($N = 147$). Boxes show the median, interquartile range, and range of percentage of material errors out of total errors by radiologist for each radiologist group. Radiologist groups with higher total percentage of errors generally also had a higher percentage of those errors being material.

Since all types of errors were more common in the procedural group of reports, and the distribution of radiologists is known to differ between this group and the diagnostic imaging group, we hypothesized that the different radiologist makeup of each group may explain this disparity. Testing for associations between radiologist group and ISE group (χ^2) as potential covariates not surprisingly uncovered a possible relationship between the procedural reports and higher error rate radiologist group (Table 4; $p = .004$). To control for this potential confounding, a multiple logistic regression model was performed demonstrating that dictating procedural reports remains an independent predictor of total error rate ($p < .001$), with an adjusted OR of 1.27 (95% CI, 1.17–1.38) for a procedural report versus a diagnostic report, regardless of radiologist group. All other ISE groups and radiologist groups did not covary.

Error rate over time

The overall error rate decreased significantly over time when comparing either of the first 2 time periods with any later time period ($p < .001$). The largest decrease in total error rate occurred between the first and third periods (OR, 0.68; 95% CI, 0.65–0.71), with an actual mean (SD) error percentage change from 10.1 percent (6.7%) to 7.4 percent (4.3%). The error rate stopped decreasing between the last two time periods, with a significant increase in error rate only among group 4 radiologists ($p = .003$; Figure 2). The total error percentage for the 6-month period after upgrade to the PowerMic II Dictaphone was 9.0 percent, compared with 8.5 percent before the hardware change, which was not significantly different ($p = .06$).

Discussion

SRS-related error rates reported in the literature vary between 4.8 and 38 percent among finalized radiology reports.[1,4,10,11] Higher error rates, in general, have been reported in studies that examined

Table 4. Errors by radiologist group: procedural versus diagnostic study type.

Study type	Errors by radiologist group[a,b]			
	Group 1 ($n=1732$)	Group 2 ($n=2596$)	Group 3 ($n=4492$)	Group 4 ($n=9470$)
Procedural ($n=730$)	11/147 (7.5)	69/854 (8.1)	64/554 (11.6)	586/2961 (19.8)
Diagnostic ($n=17,560$)	1721/47,710 (3.6)	2527/39,166 (6.5)	4428/50,206 (8.8)	8884/54,455 (16.3)

[a]Values are all errors/number of reports (%).
[b]Group 1 (<5.5% total errors), group 2 (5.5%–7.9% total errors), group 3 (8.0%–10.5% total errors), and group 4 (>10.5% total errors).

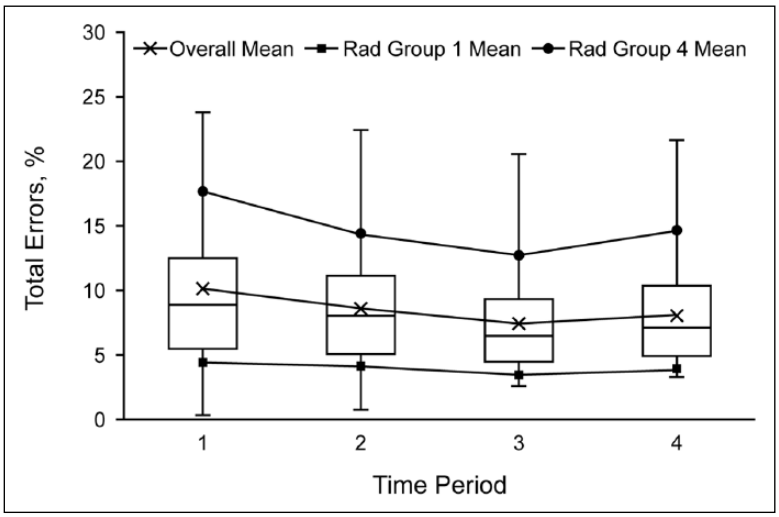

Figure 2. Box plot of total errors by time period ($n=111$). Boxes show the median, interquartile range, and range of error rates for each time period. Lines show overall mean values and mean values of radiologists in the first and fourth error groups. The time-dependent variability is greater among radiologists with the most overall errors (group 4).

only CS modalities. For example, Pezzullo et al.[4] reported a total error rate of 35 percent using SRS to interpret spine MRI, and Quint et al.[10] reported 22 percent total errors in CT of the head, neck, chest, abdomen, and pelvis. However, only a 6 percent error rate was reported by Chang et al.[1] among radiography ("CR group") reports, compared with a 38 percent error rate in their "non-CR" group. A study by McGurk et al.[11] that excluded MRI reports found a 4.8 percent error rate. Our results examining nearly a quarter-million reports confirm this trend, with a lower error rate of 3.2 percent for CR, compared with 11.0 percent for CS (OR, 3.72). Chang et al.[1] calculated a relative risk of error in their "non-CR" group of 3.5 compared with CR. Our reported OR converts to a relative risk of 3.42 (95% CI, 3.23–3.62), showing excellent agreement with Chang et al.[1] despite large differences in error rates for each modality group between the two studies. This is good evidence for a real increase in the probability of SRS-related error in CS imaging reports versus CR reports (3.4- to 3.5-fold increased risk). This effect may persist even when macros are used.[5]

There are many reasons for differences in error rate between various studies in the literature, not the least of which is heterogeneity of particular software vendors, versions, and equipment.[12] Our errors may be at the low end of the spectrum because of a strict quality control policy in place. There is evidence for this in the time-dependent decrease in error rate since the transcriptionist-auditing

program began essentially at the beginning of our study period (Figure 2), with the notable exception that we do not have data predating the quality control program to compare. Of interest, the decrease did not continue throughout, suggesting that there may be a lower limit to the error rate achievable at a large institution. The significant variation in error rate among radiologists may play a role. The group 4 radiologists (with the highest and most variable error rates) actually had increased total error rates in the last time period (Figure 2), which suggests that routine feedback regarding errors does not affect every radiologist in the same way, at least over time. Another possibility may be degradation of speech recognition for some radiologist voice models over time, or perhaps aging hardware/microphones in some areas frequented by these radiologists. Radiologist-dependent variability in error rate has been documented by others, as well,[1] ranging from 0 to 100 percent.[10]

Despite the increased probability of error in CS compared with CR studies, a greater proportion of material errors were present among the CR reports. Similarly, Rana et al.[5] found a greater incidence of "major errors" among their CR reports compared with CS, despite greater total errors in the CS cohort. Although CS reports may theoretically be more likely to contain an error because they contain more words and phrases than CR reports,[5] this does not explain the opposite discrepancy in material errors. In fact, Chang et al.[1] found the opposite, with "very significant" errors found in 8 percent of reports in the "non-CR" group and 0.5 percent in the "CR group." It is possible that the relatively shorter ratio of interpretation time to report dictation/editing time in CR examinations than in CS examinations results in lower awareness of significant intrusions, omissions, or other errors in CR reports. Because of the large number of reports in our data set, we did not quantify length of report, although we would expect that it might correlate with total error frequency.[5]

Despite this notable exception in CR reports, we generally found a greater fraction of material errors to be associated with greater total error rates. This was specifically the case among OS reports, procedural reports, different radiologist groups, individual radiologists, and ISEs. We also found increased error rate in procedural reports (e.g. vascular interventional, lumbar puncture) compared with diagnostic reports (OR, 1.91) and OS compared with in-house examinations (OR, 1.55). We could not find other reports of similar findings in the literature. Although we did not specifically track the use of templates, the CS ISE reports with the fewest errors, US, were frequently made with the use of templates.

Surprisingly, spelling errors were the most common type of immaterial error and usually did not significantly differ between group comparisons. SRS systems do not make spelling mistakes. This means that many of our radiologists type their reports or make edits when proofreading their reports rather than use the SRS process. The spellcheck function in our software is not automatic and must be manually triggered by the user before signing the report, no doubt further contributing to spelling errors in a busy environment. This strongly contradicts earlier work demonstrating decreased spelling errors with SRS.[13] Ironically, then, a consequence of using a technique that should not result in spelling errors has been a rather large increase in spelling errors in an environment in which spellcheck is not mandatory or visible in real-time as highlighted or otherwise marked text. The human interaction component to technology cannot be overlooked. Unfortunately, we do not know the proportion of spelling errors that contributed to material errors, and therefore the clinical consequences are unknown. Contributors to this phenomenon, as well as other errors, may include cursory report editing due to pressure for quick turnaround time or other failures in the proofreading process, as well as an underestimation of actual error frequency.[10] Busy inpatient working environments and nonnative English-speaking status have also been linked with increased error rates.[11,14] We did not look for associations between error rate and radiologist experience level or presence of trainees, although others have found no such relationships.[5,10,11]

Our study has several limitations. Report errors were not automatically parsed by a computer but were elucidated via human proofreading, with its inherent fallibility and subjectivity. This may result in underestimation of the number of errors, as well as misclassification of types of errors. This limitation is likely to be small, however, given that experienced professional medical transcriptionists were used. Another limitation is the absence of subclassification of material errors. Whereas spelling errors are the most common immaterial error, it is unclear whether this is also true among the material errors, which are more likely to obfuscate report meaning to the extent of complicating or altering patient management. Although such judgments were made by transcriptionists, it was not feasible for a quarter-million reports to be re-reviewed by radiologists or other physicians. A related limitation is the difficulty in quantifying and comparing "very significant," "major," and "material" types of errors. Our use of "material" error, defined as any error that could potentially impede understanding of any part of the report, may then include errors that would not necessarily be categorized as "major" or "very significant" in other publications.

Other potential biases are related to the retrospective nature of the study and data collection. The most significant may be that reports from radiologists with error rates greater than 3 percent were reviewed more frequently than those with lower error rates, possibly skewing the data. However, since only 8 of 147 radiologists had an average total error rate less than 3 percent, this skew is likely mild. We also note that most CR reports at our institution are transcribed directly to a transcriptionist, without the use of SRS, which may bias our results for studies that are performed off-hours or in areas where CR is not routinely reported.

SRS has clear advantages, including ease of integration with the radiology information system and picture archiving and communications systems,[4] decreased report turnaround time[3–5,9,13,15,16] as dictation and transcription processes are combined,[6] and shorter reports.[13] Claims of cost-effectiveness[13,17] can be dubious depending on which costs are included in the analysis and how *effectiveness* is defined.[18] SRS has been shown to contribute to decreased radiologist productivity.[9,19,20] Those who attempt to account for this fact in economic analyses find that SRS results in net increased cost.[4,20] SRS has been demonstrated to decrease productivity in other fields as well, such as among endocrinology and psychiatry secretaries.[21] This may someday be overcome with continued improvements in SRS technology. For example, natural language processing software or use of "send-to-editor" functionality (e.g. report review by transcriptionist after SRS recognition) may potentially increase accuracy and efficiency of the radiologist.[22] Probably the most serious downside to SRS is the higher error rate.[4,5,11,23] The American College of Radiology[24] recommends that radiologists proofread their final reports to minimize these types of semantic and syntactic errors.

Future research

Our results suggest several potential areas of focus for future research. Given the variability in error rates among different radiologists and different types of imaging examinations, departments with limited resources may wish to take a more targeted approach to the problem. Selective auditing of CS and procedural reports, for example, could potentially have a larger impact on overall error rates. Further study of differences in error frequency between reports generated using templates and macros, compared with traditional SRS, is certainly warranted. In addition to our US data results, other studies have found the regular use of macros or templates to be potentially helpful for reducing report errors.[11,25] We are currently expanding our use of templates into other divisions, as are many other departments throughout the world, to determine whether error frequency can be further decreased.

Given the high incidence of spelling errors in our data, we have decided to permanently enable the spellcheck feature of our SRS, so that it is no longer optional. It will be interesting to note any

future changes in error frequency and report turnaround time. As suggested earlier, further research regarding the "send-to-editor" functionality of some SRSs is greatly needed. Although, in theory, the "send-to-editor" function may combine the advantages of SRS and transcription, there are also potential disadvantages, including some potential loss of efficiency and increased turnaround time.[22] As SRS technology progresses, new implementations of hardware and software must be tested in real clinical environments before being widely adopted, particularly given the associated cost. For example, our Dictaphone hardware upgrade was expected to improve speech recognition, presumably based on vendor testing, but in our hands it had no effect on report error rates.

Conclusion

SRS-related errors are more common in CS reports (compared with CR), OS reports (compared with in-house examinations), and procedural studies (compared with diagnostic). When the total error rate increases, the fraction of material errors usually increases as well, except in the case of CR, in which material errors were more common than in CS reports. Error rates are highly variable among radiologists. Spelling errors are the most common type of immaterial error when automatic spelling correction is not mandatory, which suggests that editing radiologists often type rather than use SRS for report editing. A hardware upgrade from PowerMic I to PowerMic II Dictaphones had no effect on error rate. A quality control program with regular feedback can decrease errors over time, but there may be a limit.

For departments that use SRS, we recommend the following actions: (1) regularly audit reports, with feedback, for quality control; (2) focus efforts and resources on reports that are longer and more technical—in radiology departments, these include CS and procedural reports; (3) use automation whenever possible, including templates, macros, and automatic mandatory spellcheck; (4) perform trials of all costly hardware and software upgrades in your environment under your conditions; and (5) regularly retest the system for efficacy after any substantial changes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Chang CA, Strahan R and Jolley D. Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. *J Digit Imaging* 2011; 24(4): 724–728.
2. Kushner DC and Lucey LL; American College of Radiology. Diagnostic radiology reporting and communication: the ACR guideline. *J Am Coll Radiol* 2005; 2(1): 15–21.
3. Prevedello LM, Ledbetter S, Farkas C, et al. Implementation of speech recognition in a community-based radiology practice: effect on report turnaround times. *J Am Coll Radiol* 2014; 11(4): 402–406.
4. Pezzullo JA, Tung GA, Rogg JM, et al. Voice recognition dictation: radiologist as transcriptionist. *J Digit Imaging* 2008; 21(4): 384–389.
5. Rana DS, Hurst G, Shepstone L, et al. Voice recognition for radiology reporting: is it good enough? *Clin Radiol* 2005; 60(11): 1205–1212.
6. Langer S. Radiology speech recognition: workflow, integration, and productivity issues. *Curr Probl Diagn Radiol* 2002; 31(3): 95–104.
7. Smith JJ and Berlin L. Signing a colleague's radiology report. *Am J Roentgenol* 2001; 176(1): 27–30.

8. Physician Insurers Association of America. *PIAA data sharing reports*. Rockville, MD: Physician Insurers Association of America, 1985–2003.
9. Mehta A and McLoud TC. Voice recognition. *J Thorac Imaging* 2003; 18(3): 178–182.
10. Quint LE, Quint DJ and Myles JD. Frequency and spectrum of errors in final radiology reports generated with automatic speech recognition technology. *J Am Coll Radiol* 2008; 5(12): 1196–1199.
11. McGurk S, Brauer K, Macfarlane TV, et al. The effect of voice recognition software on comparative error rates in radiology reports. *Br J Radiol* 2008; 81(970): 767–770.
12. Devine EG, Gaehde SA and Curtis AC. Comparative evaluation of three continuous speech recognition software packages in the generation of medical reports. *J Am Med Inform Assoc* 2000; 7(5): 462–468.
13. Ramaswamy MR, Chaljub G, Esch O, et al. Continuous speech recognition in MR imaging reporting: advantages, disadvantages, and impact. *Am J Roentgenol* 2000; 174(3): 617–622.
14. Kanal KM, Hangiandreou NJ, Sykes AM, et al. Initial evaluation of a continuous speech recognition program for radiology. *J Digit Imaging* 2001; 14(1): 30–37.
15. Lemme PJ and Morin RL. The implementation of speech recognition in an electronic radiology practice. *J Digit Imaging* 2000; 13(2 Suppl. 1): 153–154.
16. Zick RG and Olsen J. Voice recognition software versus a traditional transcription service for physician charting in the ED. *Am J Emerg Med* 2001; 19(4): 295–298.
17. Sferrella SM. Success with voice recognition. *Radiol Manage* 2003; 25(3): 42–49.
18. Reinus WR. Economics of radiology report editing using voice recognition technology. *J Am Coll Radiol* 2007; 4(12): 890–894.
19. Hayt DB and Alexander S. The pros and cons of implementing PACS and speech recognition systems. *J Digit Imaging* 2001; 14(3): 149–157.
20. Issenman RM and Jaffer IH. Use of voice recognition software in an outpatient pediatric specialty practice. *Pediatrics* 2004; 114(3): e290–e293.
21. Mohr DN, Turner DW, Pond GR, et al. Speech recognition as a transcription aid: a randomized comparison with standard transcription. *J Am Med Inform Assoc* 2003; 10(1): 85–93.
22. Williams DR, Kori SK, Williams B, et al. Journal Club: voice recognition dictation: analysis of report volume and use of the send-to-editor function. *Am J Roentgenol* 2013; 201(5): 1069–1074.
23. Al-Aynati MM and Chorneyko KA. Comparison of voice-automated transcription and human transcription in generating pathology reports. *Arch Pathol Lab Med* 2003; 127(6): 721–725 (also published Erratum in: *Arch Pathol Lab Med* 2003; 127(10): 1348).
24. ACR practice parameter for communication of diagnostic imaging findings. Resolution 11, http://www.acr.org/~/media/C5D1443C9EA4424AA12477D1AD1D927D.pdf (accessed 30 September 2014).
25. White KS. Speech recognition implementation in radiology. *Pediatr Radiol* 2005; 35(9): 841–846.

Health Informatics Journal

Article

Identifying an appropriate Content Management System to develop Clinical Practice Guidelines: A perspective

Health Informatics Journal
2017, Vol. 23(1) 14–34

DOI: 10.1177/1460458215616264
journals.sagepub.com/home/jhi
SAGE

Sandeep Reddy, Sally Herring and Allison Gray
Centre for Remote Health, Flinders University, Flinders University, Australia

Abstract

Clinical Practice Guidelines are widely used to inform and improve the quality and consistency of clinical practice. Developing and publishing Clinical Practice Guidelines is a complex task involving multiple components. Electronic Content Management Systems are increasingly employed to make this task more manageable. The Content Management System market offers a variety of options for publishing content on the Internet. However, there are limited products that comprehensively address the requirements of publishing Clinical Practice Guidelines. The authors are involved in publishing guidelines for remote clinical practitioners in Australia and present their perspective about identifying an appropriate Content Management System. Several elements essential to addressing their unique editing needs are defined in this article. Unfortunately, customisation is very expensive and laborious: few Content Management System providers can comprehensively meet the needs of Clinical Practice Guidelines publishing. Being pragmatic about the level of functionality a product can offer to support publication is essential.

Keywords

Clinical Practice Guidelines, Content Management System, editing, publishing, reviewers

Introduction

Clinical Practice Guidelines (CPG) have been defined as 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances'.[1] CPG are widely used to inform and improve the quality of clinical practice by standardising patient care and minimising inappropriate or harmful interventions.[2] A driving force behind their development and uptake is the desire for clinical practice to adopt an evidence-based approach.[3]

Developing CPG is a complex and evolving process involving some phases.[4] These include identifying relevant areas for guideline development, establishing and managing guideline development groups, reviewing the evidence and modifying guidelines accordingly. There is additional

Corresponding author:
Sandeep Reddy, Centre for Remote Health, Flinders University, Flinders University, Corner Simpson and Skinner Street, The Gap, Alice Springs, NT 0870, Australia.
Email: sandeep.reddy@flinders.edu.au

complexity in developing CPG for health practitioners in remote practice settings.[5,6] Aside from developing fit-for-purpose clinical guidelines, one has to consider the setting's context, culture and community perspective. The isolation, extended training requirements and high staff turnover, which impact remote service delivery, also influence CPG development and uptake.

Publication of CPG focusing on remote and Indigenous primary health care in Australia commenced in Alice Springs, Central Australia, in the early 1990s.[7] The first to be published was the pocket-sized *CARPA Standard Treatment Manual* (STM).[8] The practice guidelines have evolved across several editions to incorporate a suite of five Remote Primary Health Care Manuals (RPHCM):[9]

- *CARPA STM*: Clinical guidelines that cover 'what to do' in remote settings for conditions that are common, high risk, unfamiliar and dangerous, have significant public health implications or require coordinated, standardised care.
- *Minymaku Kutju Tjukurpa – Women's Business Manual (WBM)*: Treatment guidelines for women's health issues in the remote context. It is a culturally respectful resource, particularly relevant for female doctors, midwives, nurses and Aboriginal and Torres Strait Islander Health Practitioners and Health Workers.
- *Clinical Procedures Manual (CPM) for remote and rural practice*: A guide to conducting routine and emergency clinical procedures faced by remote health practitioners. Accurate, practical and best-practice clinical direction for procedures that may be less common in mainstream primary health care.
- *Medicines Book for Aboriginal and Torres Strait Islander Health Practitioners and Health Workers*: Visual, plain English guide to medications referred to in the STM and WBM or commonly available in remote clinics. Designed primarily for Aboriginal and Torres Strait Islander Health Practitioners and Health Workers who supply and monitor medicines but may not be able to access or read other medicines reference books.
- *Reference Book for the RPHCM*: The background, evidence and rationales for the protocols and procedures in the RPHCM suite. Designed to enhance and inform practitioners' understanding of clinical decision-making in remote practice.

The RPHCM suite is published both electronically and in hard copy. The manuals are utilised by remote area nurses, doctors, midwives and Aboriginal and Torres Strait Islander Health Practitioners and Health Workers. While the RPHCM suite is designed for the remote Aboriginal and Torres Strait Islander health context, they are also known to be used in other primary health care settings.[10] As the manuals expanded, developing, reviewing and updating content also became more complex. The RPHCM editorial review process, which has evolved over the years, is described below.

RPHCM recommendations are made by expert consensus on the basis of evidence review. Expert opinion and consensus decision on current best practice provide direction where remote-specific research is not yet available. Primary evidence reviews are conducted by content experts, synthesising existing guidelines, evidence summaries, systematic reviews, primary studies and expert opinion, who then make recommendations for change to RPHCM content. Protocols are updated by Editorial Working Groups of clinicians, content and context experts on the basis of these recommendations, the best available evidence and requirements of the remote Indigenous context. Users of the manuals assess relevance, readability and practicality of the updated protocols for the remote primary health care clinic and their input incorporated by the Working Groups wherever appropriate. Final review and endorsement of recommendations occur via a multidisciplinary editorial committee of experienced clinicians and clinical academics with significant remote expertise and experience. In line with the National Health and Medical Research Council

(NHMRC) model, the next review of the RPHCM will incorporate conflict of interest declarations for all decision makers, clear documentation of the evidence base at the recommendation level and wider publication of the development model.

A project team, funded by the Australian Government, manages the development and publication of the RPHCM suite. The authors are part of the RPHCM project team, reporting to a Governance Committee that consists of representatives of partner organisations.

The need to accommodate remote reviewers and the complex editorial review process led to the adoption of an electronic system in 2007, to manage content development and review and facilitate concurrent publication in the paper and electronic formats. These types of electronic systems are described in the publishing world as Content Management Systems (CMS). While the early adoption of a CMS assisted in streamlining and managing the revision and editing of the content, over time it was identified that exporting content to publishable formats remained problematic. This key issue compounded by unmet development expectations, some reviewer's reluctance to adopt the system, access and authorisation issues, difficulties in manipulating content, and need for third party layout and website development software led the team to reconsider their expectations. In response to this, the project team in late 2013 reassessed their CMS requirements, and, while reviewing their current CMS, explored whether alternative CMS could address their needs.

While exploring the literature to guide CMS selection, little to no peer-reviewed content outlining the use of CMS in CPG publishing was identified. No guidance was available to assist developers of CPG in selecting, adapting or utilising a CMS. This article endeavours to address this gap in the literature by describing CMS functional specifications, requirements for CPG publishing and the authors' experience in adapting their CMS to suit their publishing needs. This will potentially benefit clinicians, academics, organisations and professional guideline developers in the efficient publication of CPG.

CMS: general concepts and features

Definition

Various definitions of CMS exist in the public arena. Following review of the literature and based on our user experience, we define a CMS to be *an electronic information system including web-based systems that enable storage, organisation, management and export of content from a single source*.[9,11,12]

General features

A CMS is intended to organise content derived from a variety of sources such that it can be accessed from a single portal.[13] The management of data structure and typesetting in one place enables administrators to have better version control. There is an expectation that a good-quality CMS will deliver the following:[14,15]

For contributors

- *Seamless access*. The CMS should enable a feature-rich environment, which provides content authors with one-stop access to create, publish and update content. There has to be an intuitive environment allowing contributors easy access to the full range of CMS content and features as appropriate.

- *Multiple user accesses*. The CMS should allow access to users outside the administrator/editorial team. These users should be able simultaneously to review and/or author content. A multi-level authorisation process may be necessary for different categories of reviewer/contributor/editor/administrator, and the CMS should be set up to support this process.
- *Simple interface*. The interface should be easy for users to navigate and engage with, without requiring high levels of computer literacy, knowledge of HTML or other advanced technical skill. The interface should resemble a basic word processor that is familiar to users and not require specialised expertise.
- *Division between content and presentation*. There needs to be a clear boundary between content being authored or reviewed and its eventual presentation. It becomes unrealistic to publish in various formats if the content cannot be separated from presentation. The CMS should provide a style-based environment for authoring that provides formatting features during the publication process.

For content management

- *Option for content reuse*. The CMS should not only allow for easy creation of content but also allow the same content to be used across different documents and user groups. This will enable consistent management of content across several platforms. CMS should handle and manage content in logical fragments, creating an ideal environment for the reuse of information. This enhances consistency of content, versioning and storage capacity.
- *Stable cross-referencing*. There has to be the ability to link flexibly documents while maintaining stable connections when restructuring the content. The CMS should be able to generate reports of broken internal links and, ideally, changes to external links.
- *Metadata capture*. Keywords, subject, authors and other metadata must be able to be captured to allow effective coding, searching, indexing and creation of the table of contents. Through a simple interface, the CMS should allow the identification, searching and reporting of arbitrary metadata. These features are essential in managing a large and complex content repository.
- *Workflow*. A workflow model to manage content from initial draft to the final format independent of organisational change is a requirement. Accessible statistics outlining various status updates at the individual user and project levels is critical. Furthermore, the workflow model should provide details about the recent use, content changes and versioning.

As a system

- *Security*. As the integrity of the content is important, there must be satisfactory security features to prevent unauthorised or untracked content change. This means that access control, audit trails and the capacity to roll-back to earlier versions should be built into the CMS. A choice of different categories and permissions should be available, potentially both across and within contributors. This can be offered through an authentication process, controlled by the administrator.
- *Easy integration with associated systems*. The CMS should be able to interact with software already being used to present content. The CMS may be just one of the numerous software packages necessary to present information in published form and must be flexible enough to work with programmes allowing PDF, word processing and design file output. The process

for achieving integration must be fully documented and based where possible on open source and industry standards.

- *Ability to generate reports*. The CMS should be able to generate a range of reports for users and administrators, including user activity and location of the user. It should also report workflow status and currency of pages, and backing for customised reporting.
- *Increased efficiency and value*. Using the CMS there has to be a decrease in costs and time in preparing, presenting and managing content. The CMS should help in eliminating publishing hurdles. The content updates, user registration/access allocation and minor interface adjustments should be able to be done in-house without reliance on external IT support to minimise time and expense.
- *Easy to implement*. Implementation costs are too often overlooked when considering a CMS. The financial outlay, effort and time involved in application and integration of CMS can be considerable, particularly if the CMS has a proprietary standard and will not integrate with currently existing legacy content. Ideally, a CMS that incorporates a cloud solution will decrease implementation and integration costs significantly.
- *Capacity to change the scale*. The CMS should be scalable to adjust to the number and variety of contributors and contributor types, with the potential to be deployed across the organisation. The CMS should be able to support increased user levels and resources required for extended and concurrent usage. Open standard-based CMS should allow for this.
- *Training and support*. There has to be open communication between the vendor and client about training requirements. The vendor must specify support that is available, including materials and training sessions. Good-quality training adapted to the particular context is necessary for an effective CMS. Training can be provided to administrators who can on-train other users. Capacity for re-training as the system is refined and developed further should be considered.
- *Costs*. This can be complex depending on the type of CMS being sought. There could be direct costs including one-off installation costs and ongoing licence and maintenance cost. Also, conversion costs have to be considered. When existing digital content is to be moved into a new/different CMS, conversion has to occur. Further indirect costs such as supporting hardware, software and operating costs have to be considered. Therefore, a CMS vendor has to be transparent about total costs involved while purchasers must plan for the level of uncertainty in customisation time and subsequent costs.

Component CMS

It is important to distinguish between the popular and widely available Web CMS products, and the Component CMS product that is offered only by limited providers for niche users. Typically, the term CMS is assumed to refer to software managing web content delivery.[16] However, content management has broadened to cover management of a considerable range of data, including compound documents, images, email correspondence and digital assets.[17] For complex technical documentation, content has to be organised as ‘components’, so it can be assembled, reassembled and published. When content is organised only at the document or page-level, it cannot be manipulated as flexibly. To cater to such management requirements, a Component CMS is necessary. The Component CMS is relevant to publishing CPG as it assists with storing, accessing, editing and managing protocol or topic-level content and also to managing web content delivery. The remainder of the article refers to this type of CMS.

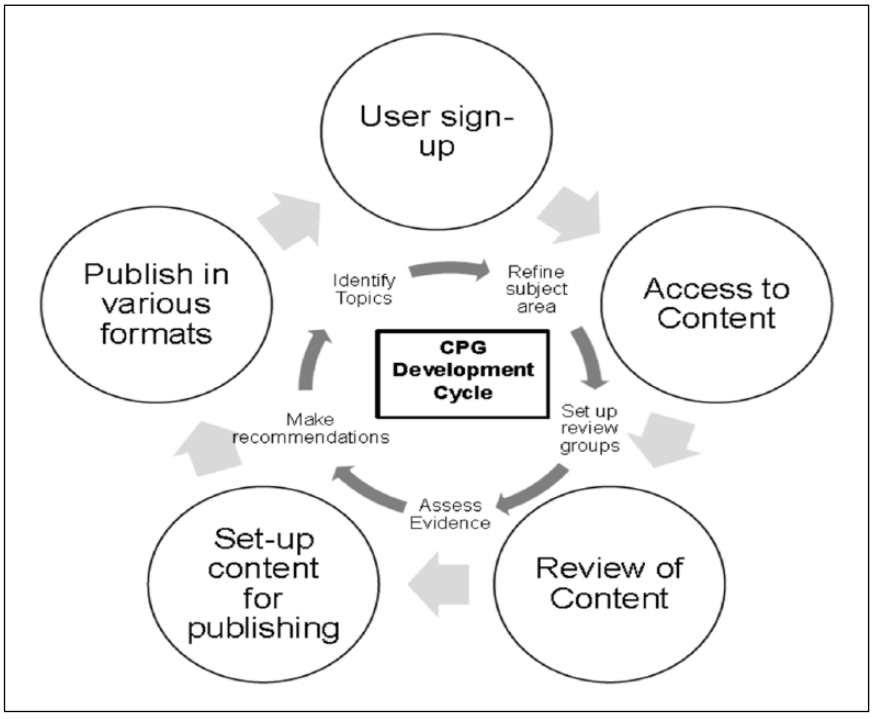

Figure 1. CMS in CPG publishing.

CPG and CMS: interplay

CPG development and publishing involves distinct and sequential stages.[4] In our experience, each stage is essential for the production of accurate, relevant and up-to-date guidelines. An effective CMS has to fit and support the CPG development process as outlined in Figure 1.

First, the topic areas to be covered are identified and subsequently refined to incorporate only relevant material. CMS access to previous versions of published material for adaptation and updating in subsequent revisions is invaluable. The next stage, establishing individual reviewers and/or working groups to assess current material, can be achieved quickly and easily when the CMS can be used to identify and capture previous reviewers, sign up new reviewers utilising different software platforms simply and easily, establish topic-based user groups for linkages and discussion and manage differing access levels across reviewers and topics depending on need. A CMS can assist in capturing reviewer's expertise across multiple topic areas to differing degrees in a single system, promoting efficiency, reviewer satisfaction and retention, and a sense of the importance of an individual review to the bigger picture. Results of evidence reviews can be posted, viewed, discussed and responded to in topic groups or across content, allowing convenient, transparent and documented discussion and consensus formation. Subsequent changes to and adaptation of content based on the evidence can then be tracked, linked with discussions and other publications, and rolled back if required. The final stage, CPGs publication, is supported by a CMS through an interface mimicking the final published layout, direct publication to electronic mediums and the more complicated hard copy. An effective CMS not only can aid in rapid and widely disseminated publication but is also crucial in maintaining the content integrity and version control.

The RPHCM experience

Role of CMS in RPHCM publishing

Two CMS requirements dominate with CPG publishing: the need to organise the content into 'components' for flexible manipulation of content and the need to 'personalise' CMS interface to suit contributors.[11] The latter is particularly relevant in the RPHCM context, where contributors are volunteers, and although experts in remote and Indigenous health, they may not be as oriented to online reviewing and authoring.

The RPHCM were developed for a specific purpose: to introduce CPG and standardise clinical practice across remote Indigenous primary health care services, beginning with Central Australia then across the Northern Territory. The remote and Indigenous health context requires portable, accessible, visual and easy to read manuals, published both in hard copy and electronically. The manuals are designed for a range of remote practitioners with varying levels of clinical and remote experience, English literacy skills and geographical isolation from specialist services and support. Each of the five RPHCM has a distinct purpose, but all cater to remote area nurses, doctors, midwives and Aboriginal and Torres Strait Islander Health Practitioners and Health Workers. While the manuals cover different content, they are designed to be used as a suite, with content cross-referenced across the individual publications.

User participation is essential in the development of these CPG, and content is created, reviewed and implemented by users. New editions involve reviewing and updating existing content. As content is cross-referenced, occasionally duplicated and consistently formatted across the suite, a Component CMS was initially perceived as the best fit.

The RPHCM project first adopted a Component CMS in 2007. The initial adoption of the CMS occurred in the context of the first publication of a full suite of five manuals, a challenging task involving a considerable volume of content and a large, geographically dispersed volunteer reviewer/contributor pool. However, the transition process, including an installation of the CMS, setting up of production groups, signing up of users with appropriate permissions, and separation of content creation and management from editing/formatting went relatively smoothly. The project has continued to use the same Component CMS for CPG development and publication.

However, despite ongoing refinement driven primarily by user feedback, one major expectation was not immediately addressed. The CMS was unable to export content from the CMS directly to print-ready documents, an essential feature for the mostly hard copy–driven RPHCM publication. The CMS could export content in HTML and Docx formats, which, at the time, were unable to represent accurately and consistently the manuals' formatting, particularly the more image-driven or complex content. External consultants used third party software packages to design appropriate templates for printing, but these continued to be problematic. The difficulty or inability to visualise what the electronic content would look like in hard copy was becoming a major impediment for users of the CMS and the project as a whole. Also, users were experiencing difficulties with the complex registration process resulting in loss of online contributors, and there were delays in the development of an upgraded website.

In combination, these issues prompted the RPHCM project to undertake a review of their CMS against other potential CMS products in late 2013. The key to this process was a stock take of what our requirements were, the features currently available and what may be developed next. This review allowed precise comparison of our current project CMS with comparable CMS products, providing us with an appreciation of the strengths and weaknesses of our CMS and potentially identifying alternative options. Along with the current project CMS, three other reliable Component CMS were identified and reviewed. All three identified systems were Component CMS products.

One CMS had an established reputation in technical publishing and two others, while not widely used, were acquiring a reputation for publishing technical content and had features relevant to the project. The review occurred over approximately 6 months and involved video conferences with the developers, demonstration of products, hands-on access to the products and involvement of independent advisers.

Generic CMS features for CPG development

The RPHCM project first identified publishing and presentation needs that could be provided by any of the CMS that were being reviewed:

Publishing

1. *Ability to support multiple formats*. The CMS had to export to many formats including XML, HTML (web), professional print, PDF and other new standards as they arose. An essential prerequisite was the ability of the CMS to engage with desktop clinical information systems that used XML output with a clinical item index.
2. *Style sheets*. These are used to control the final appearance of each output format. The style sheets would allow the system to be flexible and expandable, by ensuring consistent formatting while permitting subsequent editing of a layout.
3. *Page templates*. Predefined page templates to guide contributors in creating content easily and consistently. The overall page layout, available formatting and heading styles are stipulated in these templates and ideally a non-technical interface is utilised to manage this.
4. *Capacity for extension*. The CMS was to support integration of additional publishing functionality and interface customisation, that is, continual improvement in interface design and incorporation of new features.
5. *Support for print publishing*. The CMS had to decrease or eliminate the need for design time in the preparation of manuals for printing. CMS output also had to be high-quality print-ready.
6. *Website usage statistics*. Statistics such as daily usage, search terms, the source of hits and most popular pages were to be gathered to assess the success of the site and identify any usability issues.
7. *Online-sales enabled*. The CMS was also to facilitate sales of the electronic editions of the manual suite through a shopper-friendly interface on the RPHCM website.

Presentation

1. *Usability and appearance*. Non-technical users with lower levels of computer literacy were to be considered in the design, appearance and usability of the CMS. There was to be an emphasis on familiarity, icons and clarity of display with the colours of the organisations involved in publishing RPHCM.
2. *Accessibility*. The system and associated website had to be user-friendly for people with disabilities, that is, compliant with standards such as W3C Web Accessibility Initiative, including capacity for the use of web readers and so on. Accessibility features can be easily implemented if considered at the development stage of the system.

3. *Multi-browser support*. The CMS pages had to be accessible by all the main web browsers and all standard-based browsers. This aspect was especially important for the RPHCM project as contributors utilised a range of browsers to review and author content.
4. *Client-side functionality*. There were to be limits on technologies required to use the system, and it was expected that users would only need a standard browser for access. If additional software was required to view certain types of outputs, freely available solutions were to be utilised.
5. *Download speed*. Pages were to be of the limited size to allow for acceptable load times for users and to support authoring on low-speed connections. This feature is fundamental to a contributor and user base located mainly in remote Australia with variable connection speeds and Internet capability.
6. *Valid HTML*. The CMS pages were to comply with applicable HTML specifications to guarantee maximum compatibility across different platforms and browsers. This compliance is essential to cater for a varied contributor pool.
7. *Navigation*. Manual contributors had to have access to multiple reliable, direct and consistent navigational aids to promote efficient output.

Customised CMS features for CPG development

Although the generic requirements listed above provided a framework for the initial discussion with CMS product providers, it became apparent that more concrete and detailed requirements were crucial to effective product comparisons. These particular features were identified through a stock take of the current system, incorporation of user and administrator feedback, exploration of trial versions of comparable products and consideration of the guideline development process as a whole. Identification of these specific requirements allowed potential providers to respond to each, at each stage of the process, and was essential to the assessment, selection, development and implementation of our CMS.

Our review and assessment framework covered *Access, Content Management, Editing Tools, Presentation* and *Publishing* features. The findings of the review are outlined in the sections below and are directly compared across CMS products in Tables 1 to 5 in Appendix 1. The CMS utilised by the RPHCM project is referred to as the Project CMS; other CMS reviewed are grouped under comparable CMS products and individually referred to as CMS 1, CMS 2 and CMS 3. Where features were available for the reviewed CMS, it has been marked ✓, and where features were not available or were not clearly distinguished, it has been marked as ×. The narrative has been provided when qualifiers or description of features was required. It must be stated that the comparison in the below tables reflects what the project team identified during the review period and any missing features may have been added to the CMS products since then. Therefore, the tables are to be considered by readers as a checklist of CMS features rather than a decree of the systems reviewed.

The main purpose of this article is to provide a list of desirable CMS features required to publish CPG. It is hoped that the following content will guide CPG developers in making an appropriate determination either while choosing a new CMS or assessing an existing CMS. As this article is not a product review, and does not intend to advise which CMS to select, the CMS products are not named to maintain anonymity for ethical purposes.

Access. These features were especially important as the project relies on unpaid, volunteer reviewers in busy clinical roles. Access to the system was therefore expected to be straightforward while preserving the integrity and security of the system. As the RPHCM utilises multiple levels of

review as outlined in Figure 2, we also required the capacity to assign a varying degree of access to reviewers. A list of 'Access' related features are presented in Table 1 in Appendix 1.

Content management. These features define and separate a Component CMS from a Web CMS. A Component CMS is expected to not only store and manage documents but also allow the publication of documents in multiple formats. A list of 'Content Management' related features is presented in Table 2.

Editing tools. The richer the editing features, the easier will be the process of reviewing and updating CPG content. Customisation of our project CMS over years led to projected and superior editing tools being available. However, not all the editing features listed below may be required by other CPG developers, and the need for Access, Content Management and Presentation features may outweigh the need for comprehensive editing tools. A list of 'Editing' related features is presented in Table 3.

Specific presentation features. These features are largely related to the user interface. While not the highest priority or ranking in our review, we held expectations that the interface would be user-friendly for our clinical reviewers with potentially minimal IT expertise. A list of 'Presentation' related features is presented in Table 4.

Publishing features. The project CMS presented limited capacity for content to be published in multiple formats, and this key issue had not been resolved over a significant period. Third party software was required to support formatting and printing in hard copy and required the transfer of content and some manual layout. Therefore, there was a strong motivation for us to identify an alternative CMS that had in-built ability to export content in multiple formats. However, this appeared an issue across all reviewed products; they needed either third party support, had limited functionality or required extensive customisation. A list of 'Publishing' related features is presented in Table 5.

Outcome of the review

The review, which involved reassessing the project's current CMS and comparing it against other CMS in the market over approximately 6 months, resulted in several educational outcomes. One of the main issues faced in the RPHCM CMS review was a lack of relevant and detailed advice to instruct guideline developers in choosing an appropriate CMS. The RPHCM project involves end-users of its CPG in its review process along with subject experts spread across Australia. The reviewer involvement is entirely voluntary in nature. Any CMS chosen would, therefore, have to present minimal impediments for access and a friendly interface that does not deter less computer-literate users. Our review identified that the alternative CMS products we explored, while providing some enhanced editing, graphics and project management features, largely failed in this area. For the project, retaining our reviewer's confidence and buy-in outweighed any additional editing and export functionalities we would gain by replacing our current CMS. It was also determined that there were editing features present in the current CMS that could not be provided by a replacement CMS, as these features were the result of customisation over years.

For the RPHCM project, the prudent and economical option was to retain our current CMS while working with the provider to address gaps. This result should not be seen as ascertaining the superiority of our project CMS compared to other Component CMS, but a pragmatic decision based on the context and current needs of our project. Other CPG developers who have yet to adopt

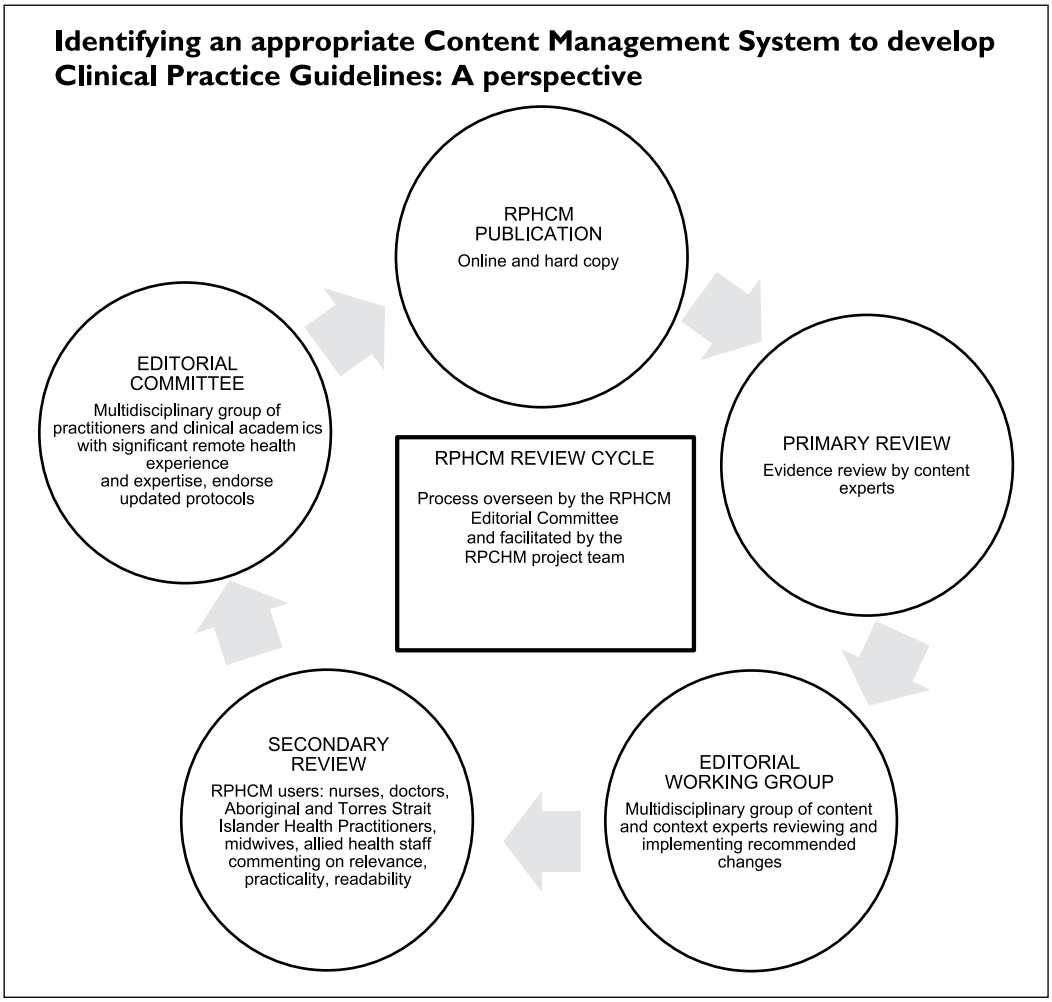

Figure 2. RPHCM editorial review flow chart.

a Component CMS may arrive at a different decision depending on their context and needs and subsequent availability of other CMS products.

Learning that may be useful for other guideline developers is the need for an early and thorough review of project requirements regarding a CMS. This exercise should cover both the generic CMS requirements and requirements unique to the project. As we learnt, specific requirements may tip the balance in choosing an appropriate CMS. Furthermore, considerable time needs to be allowed for the review and identification of a product, including market review, demonstrations and use of trial versions loaded with project content.

It must be noted that, even with an extensive, international market search, there were few products identified that catered to CPG publishing needs. Gaps were noted particularly with generic CMS products, which fulfilled website publishing and management needs but were unable to meet the editing and multi-format publishing requirements of the RPHCM project. Even with niche and component CMS, customisation to address specific client needs is inevitable and

is neither easy nor straightforward. Furthermore, providers may be reluctant to modify the CMS interface significantly, the terminology they employ or other industry standards, to assist less computer-literate users. If the supplier is keen and open to customisation, there remain significant costs involved in the modification. A purchaser must, therefore, be realistic in their expectations. It may not be possible to attain all required CMS features, and one must balance the IT principles of data structure and storage with project content and culture. Additionally, while CPG development occurs over specific and scheduled periods, CMS development timelines may be uncertain.

One significant outcome of the review was realisation of the importance of a strong relationship with the CMS provider. In a CPG publishing environment reliant on an electronic platform, dedicated and long-term support from the provider is critical. All shortlisted vendors in the RPHCM review presented a clear enthusiasm to understand and meet our needs, displaying extraordinary patience in tailoring their presentations and demonstrations to address our concerns. However, it was not clear whether the support and enthusiasm would continue post the sales phase, on a long-term basis, as we did not seek provider references. When the decision was made to retain our current provider, we sought strong assurances that a dedicated support system and contact will be provided for the project, as opposed to our previously generic provider contact. This change by itself vastly improved the provider's ability to react to the project needs, complete deliverables on time and generate project satisfaction. A good relationship with the provider is crucial. It assists in determining the existing system's specifications and the potential for the development and ensures appropriate information sharing with external consultants and future providers as the CPG development needs grow into the future.

Conclusion

CPG development and publication involves challenging and complex management of content. CMS are an ideal tool for improving efficiency, consistency and streamlined publication of guidelines. However, product options are limited, customisation is essential but costly, interfaces must consider less computer-literate users and ongoing development and support is required. Successful selection of an appropriate system requires a detailed and comprehensive understanding of requirements and a pragmatic approach to expectations.

Although unique to their context, the authors hope the outline of their learnings will aid guideline developers in the assessment and selection of a CMS that will promote high-quality CPG development and publication.

Limitations

The RPHCM project has used only one CMS for its CPG publication since 2007 and, therefore, has had prolonged exposure to only one CMS product. However, during their review of alternative CMS products, the authors undertook extensive hands-on testing of trial versions and closely engaged with providers to assess their product features. This focused evaluation has generated a knowledge base regarding CMS selection not previously published or widely available to CPG developers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

References

1. Field M and Lohr KN. Attributes of good practice guidelines. In: Field M and Lohr KN (eds) *Clinical practice guidelines: directions for a new program*. Washington, DC: National Academy Press, 1990, pp. 53–77.
2. Farquhar CM, Kofa EW and Slutsky JR. Clinician's attitude to clinical practice guidelines: a systematic review. *Med J Aust* 2002; 177: 502–506.
3. Keeley PW. Clinical guidelines. *Palliat Med* 2003; 17: 368–374.
4. Shekelle P, Woolf S, Grimshaw JM, et al. Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development. *Implement Sci* 2012; 7: 62.
5. Murphy A and McDonald J. Clinical health practice in a remote setting: the impact of local community relationships. In: *Proceedings of the 2004 annual conference of the Australian Sociological Association*, Beechworth, VIC, Australia, 7–8 December 2004.
6. Struber J. Practical practice guidelines for remote and indigenous health. In: Merrick J (ed.) *Public health yearbook* 2011. New York: Nova Science Publishers, 2012, pp. 437–445.
7. Hampton C and Fallon C. *Evaluation of CARPA standard treatment manual (3rd edition) and GSAT adult chronic disease management guidelines*. Alice Springs, NT, Australia: Central Australian Rural Practitioners Association, 2001.
8. CARPA. *CARPA standard treatment manual*. Alice Springs, NT, Australia: Central Australian Rural Practitioners Association, 1992.
9. Struber J. Partnering with technology. In: *Proceedings of the 11th national rural health conference*, Perth, WA, Australia, 13–16 March 2011.
10. Chaseling M, Mentha R and Davey C. *CARPA standard treatment manual 4th edition evaluation report*. Alice Springs, NT, Australia: Centre for Remote Health, 2008.
11. Hohne WJ, Karge T, Siegmund B, et al. An internet portal for the development of clinical practice guidelines. *Appl Clin Inform* 2010; 1: 430–441.
12. Hauschildt S. *CMS made simple 1.9 beginners guide: LITE*. Birmingham: Packt Publishing, 2011.
13. Baxter S and Vogt L. *Content management system*. US 6,356,903 B1, 2002.
14. Robertson J. How to evaluate a content management system. *Step Two Designs*, http://www.steptwo.com.au/papers/kmc_evaluate/ (2002, accessed 15 October 2014).
15. Merker L. A guide to web content management system evaluation. *OmniUpdate*, http://omniupdate.com/_resources/pdfs/white_papers/cms_evaluation_guide.pdf (2014, accessed 15 October 2014).
16. Samuels J. Types of content management system explained (CMS, DMS, CCMS, ECMS, and others). *TechWhirl*, http://techwhirl.com/types-of-content-management-systems/ (2014, accessed 15 October 2014).
17. Trippe B. *Component content management in practice: meeting the demands of the most complex content applications*. X-Hive Corporation, http://gilbane.com/whitepapers/X-Hive/Xhive_Gilbane_Whitepaper_CCMS6-final.pdf (2005, accessed 15 October 2014).

Appendix 1

Table 1. Comparison of *Access* features between project CMS and comparable CMS products.

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
Registration	Straightforward registration process for users	Need to accept the invitation as a privacy measure Process improved in the latest version	CMS 1 and CMS 2: users receive their accounts by email. There is no need for them to go through the registration process CMS 3: registration process available as an option
	Ability to collect user information as part of the registration process	Can be added	Available in all comparable CMS products
	Ability for team to know when user has accepted invitation or otherwise engaged with the system	Can be configured to send notifications at several different points in time Additional information on a specific need required	Available in all comparable CMS products
	Ability to include in registration or login process acceptance by users of terms of use for CMS and assignment of copyright authorisation for content added/changed	Acceptance of terms could be added on the first login	Available in all comparable CMS products
	Automated username/password recovery for users	✓	Available in all comparable CMS products
	User can update account details, change own password	✓	Available in all comparable CMS products
Assigned levels of user access	Password-protected system access that allows identification of individual users	✓	Available in all comparable CMS products
	Multiple levels of access (user privileges) variously combining the following: • Administrative functions • Reviewing/authoring functions (ability to change documents) • Commenting functions • View only	✓	CMS 1 and CMS 2: ✓ CMS 3:×
	Project team can invite/sign up new users	✓	Available in all comparable CMS products

Table 1. (Continued)

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
	The ability to assign one person different levels of access depending on the role about the materials being worked with, for example, contributor for one topic group, but reviewer only for others	✓	Available in all comparable CMS products
	Capacity for all users to request a view/review level access to any content *OR* All content visible to all users with option to 'turn off' content they don't want to see/aren't required to work with, or to move folders currently being worked on to a 'favourites' area for easy access	All users can be assigned review access as a minimum Other detailed options available	Not available in any of the comparable CMS products

CMS: Content Management Systems.

Table 2. Comparison of *Content Management* features between project CMS and comparable CMS products.

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
Document management	Ability to store documents for consideration once in single searchable repository, but 'package' them for presentation in a variety of functional groups, for example, by working group topic, by manual	✓	Available in all comparable CMS products
	Ability to store separately or link to resource/reference documents for reviewer's information but not for review or publication as part of the project	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability for multiple users to view and comment on documents simultaneously	Many reviewers can access the document at once, but only one user per topic	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability to lock documents under review to prevent multiple users attempting to make changes at the same time	×	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability to archive old documents	Several methods available	CMS 1 and CMS 2: ✓ CMS 3: ×

(Continued)

Table 2. (Continued)

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
	Ability to track changes made each time a document is saved, including user, time and date stamp. Ability to roll-back content to an earlier version	History available as editing segment	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability to simultaneously update nominated segments of duplicated text appearing in different documents (transclusion/global change)	✓	CMS 1: ✓ CMS 2 and CMS 3: ×
	Find and replace function	Search for and replace function available by editing segment	CMS 1 and CMS 2: ✓ CMS 3: × Global find/replace as an interface available in CMS 2
Administrative functions	Ability to create templates for documents on the CCMS. Templates to include capacity for Text/paragraphs • Bullet-pointed lists – to 3 levels • Numbered lists • Five levels of headings (include document title) • Tables • Bullet-pointed lists within tables • Images • Text boxes – 3 styles to output as 3 different frame colour/width combinations • Flow charts	All of these features available in existing templates	CMS 1 and CMS 2: ✓ CMS 3: × Fixed programme structure/layout in CMS 2 may not suit project need for flexibility
	Ability to create templates for exportable Word or PDF documents	CMS currently creates a template to our specifications. Self-created under development	CMS 1 and CMS 2: ✓ CMS 3: × In CMS 2, function of third party software Fixed programme structure/layout in CMS 1 may not suit project need for flexibility
	Ability to create new documents on the CCMS	✓	Available in all comparable CMS products
	Ability to assign and monitor user tasks	✓	Available in all comparable CMS products

Table 2. (Continued)

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
	Ability to monitor document review activity by user, business type, time and date	✓	Available in all comparable CMS products
Help documents and support	Online plain English help/support documents for users	Being updated with a new version	CMS 1 and CMS 2: ✓ CMS 3: ×
	Online plain English help/support documents for administrators (project team)	In the process of being developed	CMS 1: ✓ CMS 2 and CMS 3: ×
	Ability to provide timely accessible ongoing support as required	At the time of review, a dedicated contact person was not available but a dedicated email address for support/ requests was available	Available in all comparable CMS products
	Process available for requesting ongoing development of the CCMS or leave comments about CCMS access and design	✓	Available in all comparable CMS products
Workflow management	Accessible statistics: percentage of tasks completed, tasks assigned	✓	Available in all comparable CMS products
	Gantt chart (or similar) to track deadlines and task completion	×	CMS 1 and CMS 3: ✓ CMS 2: × Burn down charts are available in CMS 3
	Searchable statistics on reviewers/ users of the CCMS (i.e. working group membership, what protocols have they contributed to)	Information exists but reports cannot be generated	Not available in any of the comparable CMS products
	Searchable statistics/documentation: versions/changes by reviewers versus project administrators	×	Not available in any of the comparable CMS products
Data recovery	Provision of a detailed data management, backup and recovery strategy	✓	Available in all comparable CMS products

CMS: Content Management Systems; CCMS: Component Content Management Systems.

Table 3. Comparison of *Editing Tools* between project CMS and comparable CMS products.

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
Commenting on document content	Contributor date and time logged for all comments and recorded with the remark Preference for contributor name rather than email address to appear for privacy reasons	✓	CMS 1 and CMS 2: ✓ CMS 3: × Feature available in CMS 1 and CMS 2 with contributor name In CMS 2, users are identified by email. This could be changed to name
	Comment recorded close to the section being discussed on or relevant part otherwise identified	Available for editing segment or the whole document	CMS 1 and CMS 2: ✓ CMS 3: × Available in CMS 1 and CMS 2 for editing section or the whole document CMS 2 utilises pointer to identify exact text reference
	Email alert or content of comment sent out to other registered members of the topic group. Users are able to control the frequency or turn off email alerts	Currently working on a new notification system	Not available in any of the comparable CMS products
	Users able to 'reply ' to the comment establishing a discussion thread	✓	CMS 1 and CMS 3: ✓ CMS 2: ×
	Comments are searchable	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
	Comments can be archived	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
Reviewing/ authoring tools	Spell checker with capacity for user-defined dictionary	Not available	In the development process in CMS 2 and CMS 3
	Intuitive search function that allows searching of various content, for example, whole CMS, documents, images and comments	✓	CMS 2: ✓ CMS 1 and CMS 3: ×
	Ability to tag index terms in the CCMS and transfer this data to the hard copy publication software for automated compilation of the index	Transfers word documents	CMS 1: ✓ CMS 2 and CMS 3: ×
	Ability for users to log a 'change category' each time changed content is saved, with a rationale required for significant changes	✓	Limited features available in CMS 1, CMS 2 and CMS 3

Table 3. (Continued)

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
	Ability for users to export Word or PDF version of the document for reviewing when no Internet	Very limited features available	CMS 1: ✓ CMS 2 and CMS 3: ×
Images/digital assets	Single searchable repository of images	Images will be searchable in new search interface, by whole repository or specific folders	CMS 1 and CMS 2: ✓ CMS 3: ×
	System of linking images to documents that allow then to be stored once but used multiple times as needed	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability to search for and import images from within an open document	✓	CMS 1: ✓ CMS 2 and CMS 3: ×
	Reverse list of image links allowing us to know where a given image has been used	×	CMS 1: ✓ CMS 2 and CMS 3: ×
	Ability to store high-resolution images on the CCMS which can be Exported directly for hard copy publication Automatically downsized by integrated programme to low-resolution images for CCMS and web viewing	×	Not available or limited functionality in the comparable CMS products
	Ability to archive old versions of images	✓	Available in all comparable CMS products Old versions are tracked through the history system
Cross-referencing	Functional hyperlinks within and between documents in the CCMS	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability to link to a particular section of a document, not just the start of the document	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability to search for content and create hyperlinks from within an open document	✓	CMS 1 and CMS 2: ✓ CMS 3: ×

(Continued)

Table 3. (Continued)

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
	List of reverse links to document or sections of the document (i.e. which documents contain a link to the open document) to minimise unintentional breaks to links	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
	Hyperlinks remain functional when exported to HTML (web) content	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability to create active links to external electronic content (ex to other websites)	✓	CMS 1 and CMS 2: ✓ CMS 3: ×
	Ability to recognise hyperlinked text in exported printed documents to allow for page referencing (i.e. see page xxx)	Visible in Word export	CMS 1: ✓ CMS 2 and CMS 3: ×
Versioning	Ability to create a point in time versions of document content	✓	Available in all comparable CMS products
	Ability to compare versions of documents	✓	Available in all comparable CMS products
	Ability to work with different versions concurrently, for example, version of the content (v1) currently on the website can have changes/corrections made without having to upload content under review (v2)	✓	CMS 1 and CMS 2: ✓ CMS 3: ×

CMS: Content Management Systems; CCMS: Component Content Management Systems.

Table 4. Comparison of *Presentation* features between project CMS and comparable CMS products.

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
User interface	Plain English terms/instructions on pages used by reviewers making them suitable for users with varied levels of computer literacy	✓	These terms were configurable in all the CMS that were reviewed
	Ability to change current terms/language that may not be appropriate for our users to terms/language of our choosing	✓	These terms were configurable in all the CMS that were reviewed
	Minimal 'clicks' and intuitive sequencing for movement between screens/functions	✓	Available in all comparable CMS products
	Capacity to have more than one document open at once	✓	Available in all comparable CMS products

Table 4. (Continued)

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
	Capacity to return 'back' to the document being originally worked on	Back button enabled Documents able to be opened in a new window	Available in all comparable CMS products
	Reviewing tools that resemble common word processor functions, including standard keyboard shortcuts	✓	Available in all products with some products offering additional options
	Intuitive use with minimal need for training for people with basic computer literacy	✓	CMS 3: ✓ CMS 1 and CMS 2: ×

CMS: Content Management Systems.

Table 5. Comparison of *Publishing features* between project CMS and comparable CMS products.

Specification	Project requirement	Feature in project CMS	Features in comparable CMS products
Multi-channel publishing	Ability to publish CCMS XML content to HTML website	✓	CMS 1 and CMS 2: ✓ CMS 3: × Available in CMS 1 and CMS 2; however, will need customisation to fill requirements
	Ability to publish to website one document at time, to allow individual updates	In development	CMS 1: ✓ CMS 2 and CMS 3: × Potentially available in CMS 2, but depended on how the connection to the website was designed
	Ability to export CCMS XML to a programme that facilitates hard copy publication	Integration with InDesign templates for output: require further refinement Possibility of using third party software for automated output	CMS 1: ✓ CMS 2 and CMS 3: × Available in CMS 1; in CMS 2 only through third party support
	Ability to publish CCMS XML to other electronic modalities, for example, e-book, mobile or PDA apps	Beta version of CMS for phone available	CMS 1: ✓ CMS 2 and CMS 3: × Available in CMS 1, but will require customisation

CMS: Content Management Systems; CCMS: Component Content Management Systems; PDA: personal digital assistant.

Health Informatics Journal

Article

Development and evaluation of a record linkage protocol for Utah's Controlled Substance Database

Health Informatics Journal
2017, Vol. 23(1) 35–43

DOI: 10.1177/1460458215618520
journals.sagepub.com/home/jhi
SAGE

April F Mohanty
George E. Wahlen Department of Veterans Affairs Medical Center, USA

Jacob Crook
Utah Department of Health, USA

Christina A Porucznik
University of Utah School of Medicine, USA

Erin M Johnson and Robert T Rolfs
Utah Department of Health, USA

Brian C Sauer
George E. Wahlen Department of Veterans Affairs Medical Center, USA

Abstract

Utah's Controlled Substance Database prescription registry does not include master identifiers to link records for individual patients. We describe and evaluate a linkage protocol for Utah's Controlled Substance Database. Prescriptions (N = 22,401,506) dated 2005–2009 were linked using The Link King software and patient identifiers (e.g. names, dates of birth) for 2,232,725 patients. Review of 998 randomly selected record pairs classified 46 percent as definitely correct links and 54 percent as probably correct links. A correct link could not be confirmed for <1 percent. None were classified as probably incorrect links or definitely incorrect links. Record set reviews (N = 100 patients/set for 10 set sizes, randomly selected) classified 27–49 percent as definitely correct links and 39–63 percent as probably correct links. Fewer had too little information to confirm a link (5%–22%) or were probably incorrect (0%–6%). None were definitely incorrect. Overall, results suggest that Utah's Controlled Substance Database records were correctly linked. These data may be useful for cross-sectional and longitudinal studies of patient-controlled substance prescription histories.

Corresponding author:
April F Mohanty, IDEAS Center, George E. Wahlen Department of Veterans Affairs Medical Center, 500 Foothill Dr., Salt Lake City, UT 84148, USA.
Email: april.mohanty@va.gov

Keywords
controlled substance, prescription registry, The Link King

Introduction

Unintentional fatalities due to prescription medications are a well-established problem in Utah[1] and the United States.[2,3] Since 2003, the leading cause of injury death in Utah has been poisoning from prescription drugs with opioid-related poisonings as the primary offender.[1] During the years 1999–2007, deaths related to prescription pain medications increased sixfold in Utah.[4,5] The increase was largely due to deaths from prescription opioid pain medications. The number of emergency department encounters stemming from prescription drug overdose has also continued to increase in Utah from 2001 to present.[6,7]

To support identification of potential cases of prescription medication misuse, inappropriate prescribing, and related adverse outcomes throughout the state, Utah's Division of Occupational of Professional Licensing within the Utah Department of Commerce maintains a Prescription Drug Monitoring Program which tracks all outpatient prescriptions for Schedule II–IV drugs dispensed in Utah.[6] Legislative mandate (Utah Code Section 58-37f-101), established in 1995, requires all outpatient pharmacies to submit controlled substance dispensing records to Utah's Controlled Substance Database (CSD) The CSD has been a valuable tool for prescribers to look up a patient's access to controlled substances; however, due to the lack of a master patient index/identifier, the CSD has been a difficult resource to use for research and population analytics.

Our objectives are to describe and evaluate a linkage protocol for Utah's CSD, an outpatient prescription medication registry, which does not include unique master patient identifiers to link patient records.

Methods

Data sources

Utah's CSD contained over 22 million prescription records from approximately 500 pharmacies dated 2005−2009. On a monthly basis, using the standard of American Society for Automation in Pharmacy, Version 2, pharmacies report the following variables with each dispensing record: patient name, street address, zip code, date of birth and sex, pharmacy identification number, prescriber identification number (Drug Enforcement Agency (DEA) number), National Drug Code Number, prescription number, date written, date filled, new/refill number, metric quantity of drug, days supply of drug, and number of authorized refills. However, the CSD does not include a patient master identifier to link multiple dispensing records for each patient. Also, information submitted to the CSD does not follow a standardized, structured format, leading to considerable variability in submitted patient identifier information, which may include a mix of Social Security Number (SSN), phone numbers, names, driver's license, or other unspecified identification numbers or text strings. It should be noted that the greatest potential for variation in patient identifiers occurs across multiple pharmacies and tends to be more consistent for a patient within a given pharmacy. Extensive record preprocessing and cleaning to standardize the linking variables were required before implementing the linkage protocol including removal of prescriptions by veterinarians by DEA number, removal of implausible patient names or birthdates, and standardization of sex coding. We also attempted to decode 9-digit identifiers that were a mix of SSN, phone numbers, and driver's license numbers to determine whether

Table 1. Availability of patient identifiers for the linkage of Utah's CSD records.

Patient identifier	% Missing in raw CSD	% Missing in cleaned CSD
Required by the Link King		
First name[a]	0.02	0.08
Last name[a]	0.00	0.06
9-digit ID[b]	37.97	66.91
Date of birth[a]	0.02	0.08
Recommended by The Link King		
Middle name/initial[a,b]	100.00	91.54 (imputed in standardization)
Maiden name[b]	100.00	100.00
Gender[a]	0.01	0.03
Race/ethnicity[b]	100.00	100.00
Zipcode[c] or city name or county	2.67 (zip code)	2.50 (zip code)

CSD: Controlled Substance Database.
[a]Patient identifiers used to link Utah's CSD records.
[b]The 9-digit ID was a mix of patients' driver's license, phone number, and/or social security number and was frequently not present or systematically captured in Utah's CSD.
[c]Flex variable used to link Utah's CSD records.

including them as a matching variable influenced our results. Our evaluation (of a subset of records examined in this study)[8] revealed that including the 9-digit IDs did slightly improve the match likelihood but did not change our overall matching decisions; thus, we chose not to include it in our final linkage methods. The proportion of missing identifiers required for the link was low (<0.3%; Table 1).

Linkage methods

The Link King. We used a freeware data linkage SAS application, The Link King (available from www.the-link-king.com/).[9] The Link King allows for integration of restrictive deterministic matching algorithms with more flexible probabilistic algorithms. The Link King was originally developed to de-duplicate the client database of Washington State's Division of Alcohol and Substance Abuse.[9] Probabilistic algorithms used in The Link King were adapted from those developed by MEDSTAT for the Substance Abuse and Mental Health Administration's integrated database.[10] The Link King's probabilistic and deterministic algorithms compare record pairs that may be within a block of records that match on certain user-selected criteria. For each of the record pair comparisons, the algorithms derive a cut-point ranging from Levels 1–4 to classify the likelihood of a correct match. The highest certainty of a correctly linked record pair corresponds to Level 1 with each succeeding level corresponding to slightly lower certainty of a correct link. There are two other levels, Level 6 (probabilistic probable twins) and Level 7 (probabilistic maybe), which are considered to be possible links by the probabilistic but not the deterministic algorithms; there is no Level 5. These certainty levels allow the user to decide which record pairs to accept, reject, or manually review as correctly linked records. The probabilistic statistical algorithms adjust the weights for a given record pair based on the uniqueness of the patient identifier value. For example, a common last name (e.g. Johnson) is assigned a lower weight compared to a less common last name (e.g. Eisenhower). The program also allows the user to include a "flex variable" such as a patient's address, date or place of birth, or zip code. Users can adjust the weight of the flex variable to positively or negatively impact the resultant probabilistic score. Authors of prior studies who

Table 2. Patients with one or more name variations.

No. of variations	No. of patients (%)
1[a]	1,934,295 (86.6)
2	255,132 (11.4)
3	35,130 (1.6)
4	6240 (0.3)
5	1365 (0.1)
6	380 (<0.1)
7 or more	183 (<0.1)
Total	2,232,725 (100.0)

[a]Patients with one or more variations in patient identifiers also include those who were considered unmatchable (i.e. all blank values for all patient identifiers).

have used The Link King have reported positive predictive values and sensitivity values above 90 percent depending on The Link King selected certainty score.[9,11]

Protocol settings

To link Utah's CSD records, we included patient's first and last name and date of birth since SSNs were not routinely collected (Table 1). We also included the following Link King recommended variables, when available: patient's middle initial and gender. We used patient zip codes as the flex variable, and based on The Link King protocol recommendations, we set the program to restrict zip code from having a negative weight on the probabilistic score since we anticipated true changes in patient addresses to be frequent. Therefore, when a zip code of a given record pair was present, the probabilistic score of a correct record link was only positively impacted. Overall, we followed The Link King's default settings which are designed to provide users with "conservative" results, consolidating only records where there is a high certainty of the linkage.[12,13] Although we note that record linkage is a complex task and optimal linkage results may require the user to customize the settings. The Link King has numerous options for customizing linkage procedures and supporting documentation to help users understand how the various settings impact linkage certainty. The default settings we used included the blocking level, probabilistic weight settings, and the name rarity cut-point of 0.3. The Link King uses the "rarity/commonness" of a name as part of decision-making criteria in the deterministic protocol. Matching criteria for the data elements are relaxed when a patient's name is considered to be "rare." Given the number of records linked in this study and the cost and time required for manual review, we retained links classified as Levels 1–3 and excluded Levels 4, 6, and 7.

Manual review

Utah's CSD records linked from years 2005 to 2009 which contained 22,401,506 prescription records for 2,232,725 patients; therefore, 90 percent of prescription records were aggregated for the same patient, longitudinally. Few patients (13%) had two or more name variations (Table 2). Among these 298,430 patients, where The Link King protocol classified records as true links (Levels 1–3) but where the patient names, sex, or date of birth were not an exact match for two or more linked records, a subset of record pairs and record sets were manually reviewed by two researchers and together adjudicated.

Table 3. Selection of record sets for manual review.

Set size (records/set)	No. of patients (%)[a]	Sets sampled
2	38,130 (12.8)	100
3	23,005 (7.7)	100
4	24,558 (8.2)	100
5–6	33,616 (11.3)	100
7–8	21,757 (7.3)	100
9–13	32,147 (10.8)	100
14–21	27,258 (9.1)	100
22–37	28,379 (9.5)	100
38–74	30,361 (10.2)	100
75+	30,219 (10.1)	100
Total	298,430 (100.0)	1000

[a]Number of patients with one or more variations in identifiers within a set of records consisting of two or more records/set.

Record pair review

For the record pair review, we randomly selected a subset of pairs (N = 998) for manual review. Patients with more linked records in Utah's CSD had a higher likelihood of being selected for review, and multiple record pairs for a given patient could be selected. Reviewers manually examined all linkage variables and classified each record pair as one of the following five "linkage certainty" categories: (1) definitely not the same person, (2) probably not the same person, (3) there is not enough information to determine whether or not they are the same person, (4) probably the same person, or (5) definitely the same person.

Record set review

For the record set review, we categorized patients into 10 different sampling set sizes based on the number of records linked by The Link King protocol (Table 3). Ranging from set sizes of 2 to 75+ records/patient, we randomly selected 100 sets for manual review for each of the 10 sizes or 1000 total sets. Reviewers manually examined the linkage variables across all records corresponding to each set, and each set was classified according to the same linkage certainty categories described for the record pair review.

This work was reviewed and approved by the Institutional Review Boards of the University of Utah and the Utah Department of Health.

Results

Reviewers of the record pairs (N = 998 pairs) classified 46 percent as definitely correct links and 54 percent as probably correct links (i.e. the same person). Few (<1%) had too little information to confirm a correct link. None were classified as probably incorrect links or definitely incorrect links.

Reviewers of the 10 record sets classified 27–49 percent as definitely correct links and 39–63 percent as probably correct links (Figure 1). Fewer sets were classified as having too little information to confirm a link (5%–22%) or probably incorrect links (0%–6%). None were classified as definitely incorrect links. The proportion of sets classified as either lacking enough information to confirm a link or probably an incorrect link was slightly lower for set sizes of 2–8 records

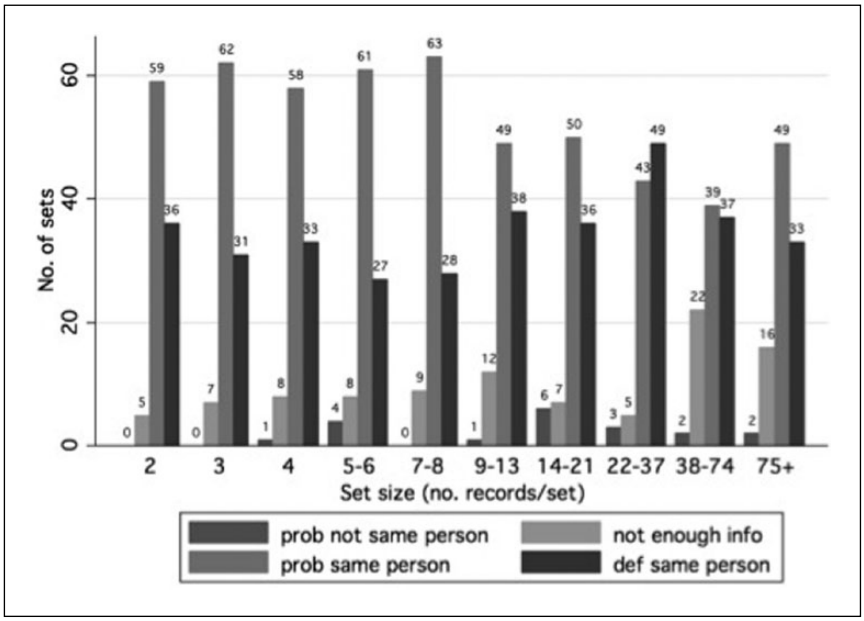

Figure 1. Results of manual review: for each set size, number of sets by reviewers' confidence of a correct match.

compared to set sizes of 9–75+ records, 8 percent versus 15 percent, respectively. However, contrary to our expectation, there was no evidence to suggest that the proportion of incorrect links increased with higher set sizes.

Discussion

Using The Link King's probabilistic and deterministic algorithms, we successfully implemented a linkage protocol among 22 million prescription records from Utah's CSD. Overall, our manual review suggested that the pairs and sets of records corresponding to specific patients were correctly linked. Therefore, the linked records of Utah's CSD may be valuable for cross-sectional and longitudinal studies.

The Link King has been used to link a variety of administrative data such as from Medicaid files,[11] surveillance data including Colorado, Connecticut, and Oregon's Hepatitis C and HIV/AIDS data,[14] and Washington State's Division of Alcohol and Substance Abuse data.[9,15] Prior studies have demonstrated the utility of The Link King for linking administrative records. For example, Beil et al.[11] attempted to link 80,414 records in the North Carolina Medicaid files to public health surveillance files using The Link King. Authors reported validity estimates based on manual review of two randomly selected subsets of record pairs: those that were and were not successfully linked. For The Link King certainty Levels of 1, 2, and 3 (those considered to be correct links in our study), Beil et al. reported sensitivities (95% confidence intervals) of 83.7 percent (78.5%−89.2%), 89.2 percent (83.8%−94.9%), and 89.2 percent (83.8%−95.0%), respectively. Specificities (95% confidence intervals) for Levels 1, 2, and 3 were 89.3 percent (81.6%−97.6%), 88.2 percent (80.7%−96.5%), and 88.2 percent (80.7%−96.5%), respectively. Had we considered our manual review as the reference standard and that true links were those identified as definitely or probably correct links, we would have reported a positive predictive

value (95% confidence interval) of 99.1 percent (98.5−99.7%) for the linked (Levels 1–3) record pairs. Since we did not manually review records that were not linked, we cannot compare the sensitivity or specificity of our link to Beil et al. As in our study, Beil et al. chose to exclude Levels 4 and 5, which had lower specificities, since their priority was to have a correct link rather than capturing all possible matches.

We are unaware of prior studies of The Link King that included manual reviews of record sets (greater than 2) corresponding to a single patient; previously, the focus has been on record pairs. While our manual review revealed that record set sizes of 9 or greater had a slightly higher proportion of incorrect links or links that lacked enough information to confirm a link compared to set sizes of 2–8, there was no evidence to suggest that the proportion of incorrect links increased as the number of corresponding patient records increased. It may be important to examine the certainty of links across larger set sizes for future studies that rely on linked, aggregated data for longitudinal analyses. In the particular case of controlled substance data where individuals might be actively attempting to evade detection in order to obtain more medications, this limitation may be even more critical when interpreting results.

Limitations

The results of our study should be interpreted in view of several limitations. First, since we relied on manual reviews in our evaluation of the linkage, and manual reviews are vulnerable to human error, we did not report the validity of the linkage. In cases where there was limited information to determine the correctness of a link, one could argue that human judgment is less precise and possibly less valid than The Link King's decision-making, such as for record links categorized by our manual review as "not enough information to determine whether or not they are the same person." Our manual review of record pairs revealed that The Link King's decision-making appears valid, and for the record sets unlike the manual reviewers, The Link King is able to make a decision regarding sets in the context of all record pairs. Second, we did not review unlinked records and were thus unable to comment on the proportion of these that could have been deemed by our manual review as missed links. A third limitation is that we used strict criteria in our manual review of record sets. For example, a set size of 5 records or a set size of 75 records would both be classified as "probably not the same person" if only 1 record had a different last name than the other 4 or 74 records, respectively. In other words, we did not down-weight the impact of incorrect/questionable records as the set size increased in our manual review. Nonetheless, we show that the error rate is consistently low, regardless of the set size. Our conservative criteria may have led to lower sensitivity. Finally, since we did not retain The Link King assigned levels of certainty, we were unable to examine the concordance of certainty assignments given by The Link King and the manual review for the record pairs.

Strengths

Strengths of our study included use of Utah's CSD, a large database that had a low proportion of records missing the required linking variables. Also, we included a large number of record pairs and record sets in our manual review, and to date, no study has reviewed match certainties for record sets of three or more.

Conclusion

The Link King software is accessible and efficient, and it allowed us to develop a protocol for successful and cost-effective linkage of prescription records in Utah's CSD. Contrary to our

expectation, the proportion of incorrect links remained stable, even as the number of records in a set increased. Overall, our manual review suggested a high proportion of correct links, including for patients with 75 or more records. With the growth of many record linkage methods, future studies are needed to compare results and efficiency across methods that include reports of sensitivity, specificity, precision, and recall. Extensions of our work include the application of The Link King to merge Utah's CSD to external data sources, such as from Medicaid, Medical Examiner, or Emergency Department databases, thus broadening the scope of public health and epidemiologic research that can be accomplished with this resource.

Acknowledgements

This manuscript is the result of work supported with resources and the use of facilities at the VA SLC IDEAS Center, Salt Lake City, UT. J.C., C.A.P., E.M.J., R.T.R., and B.C.S. designed the study and wrote the protocol. J.C. carried out the data linkage protocol. A.F.M., J.C., and B.C.S. undertook the statistical analysis. A.F.M., J.C., C.A.P., and B.C.S. assisted with interpretation. A.F.M. wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Utah Department of Health; the Centers of Disease Control and Prevention Grant (5 R21 CE001612); the Department of Veterans Affairs; and the VA Advanced Fellowship Program in Medical Informatics of the Office of Academic Affiliations, Department of Veterans Affairs. Utah Department of Health personnel were involved in the study and in the authorship of this article.

References

1. Rolfs R. *Prescription medication overdose deaths in Utah*. Utah Health Status Update, 2008, http://health.utah.gov/opha/publications/hsu/08Nov_RxDrugDth.pdf
2. Paulozzi LJ, Budnitz DS and Xi Y. Increasing deaths from opioid analgesics in the United States. *Pharmacoepidem Dr S* 2006; 15: 618–627.
3. Paulozzi LJ and Ryan GW. Opioid analgesics and rates of fatal drug poisoning in the United States. *Am J Prev Med* 2006; 31: 506–511.
4. Centers for Disease Control and Prevention. Increase in poisoning deaths caused by non-illicit drugs—Utah, 1991–2003. *MMWR Morbid Mortal Week Rep* 2005; 54: 33–36.
5. Rolfs RT, Johnson E, Williams NJ, et al. Utah clinical guidelines on prescribing opioids for treatment of pain. *J Pain Palliat Care Pharmacother* 2010; 24: 219–235.
6. Rolfs RT, Porucznik C, Sauer B, et al. *Utah pain medication management and education program: interim report on the implementation of health bill 137*. Utah Department of Health, 2007, http://health.utah.gov/prescription/pdf/report_to_legislature2007.pdf
7. Sims SA, Snow LA and Porucznik CA. Surveillance of methadone-related adverse drug events using multiple public health data sources. *J Biomed Inform* 2007; 40: 382–389.
8. Yoneoka Y, Xu W, Liu Z, et al. *Turning junk data into value, using 9-digit mixed identifiers to enhance linkage results for Utah controlled substance database*. Utah: University of Utah, 2009.
9. Campbell KM, Deck D and Krupski A. Record linkage software in the public domain: a comparison of Link Plus, The Link King, and a "basic" deterministic algorithm. *Health Informatics J* 2008; 14: 5–15.

10. Campbell K, Deck D, Cox C, et al. The Link King. *User Manual (online)*, www.the-linkking.com\user_manual.zip (2005, accessed 3 November 2006).
11. Beil H, Preisser JS and Rozier RG. Accuracy of record linkage software in merging dental administrative data sets. *J Public Health Dent* 2013; 73: 89–93.
12. Campbell KM. Rule your data with the Link King. In: *Pacific Northwest SAS users group conference*, Portland, Organ, September 10–11, 2004, pp. 20–30. SAS Conference Proceeding.
13. Campbell K, Deck D, Cox A, et al. *The Link King user manual: version 5.2*. Olympia, WA: Division of Alcohol and Substance Abuse, 2006, http://www.the-link-king.com/user_manual.zip
14. Speers S, Klevens RM, Vonderwahl C, et al. Electronic matching of HIV/AIDS and hepatitis C surveillance registries in three states. *Public Health Rep* 2011; 126: 344–348.
15. Campbell KM. Impact of record-linkage methodology on performance indicators and multivariate relationships. *J Subst Abuse Treat* 2009; 36: 110–117.

Health Informatics Journal

Article

Health Informatics Journal
2017, Vol. 23(1) 44–55

DOI: 10.1177/1460458215623077
journals.sagepub.com/home/jhi
SAGE

Using smartphone-based applications (apps) in workplace health promotion: The opinion of German and Austrian leaders

Anita Dunkl and Paul Jiménez
University of Graz, Austria

Abstract

Reaching the actual target group for a web-based health promotion project turns out to be a difficult task. In this article, individual and organizational factors which can influence the decision of using apps in workplace health promotion are analyzed. Furthermore, we analyzed the opinion about feedback possibilities of apps in workplace health promotion. A study with 438 leaders was conducted, as leaders can be seen as a key factor in the success of health promotion projects. The results showed that younger leaders and leaders with a more positive attitude toward workplace health promotion are more likely to use an app. Furthermore, leaders with a positive attitude are more interested in expert-feedback than in instant feedback received from an app.

Keywords

application, feedback, leadership, target group, workplace health promotion

Introduction

Smartphones have, little by little, invaded our modern everyday life and are already an established means to keep in touch with our family and friends all around the globe. But the pocket-sized supercomputers are not only an aid regarding communication but have also gained importance as a reliable source of information via Internet connection. Furthermore, there is a variety of smartphone applications (apps) to choose, which range from simple video games to pass the time to reliable assistants to help us face and conquer everyday tasks and challenges.

In line with this increased availability and usage of smartphones, apps for health promotion practices and health outcomes are advancing. Despite the increased number of developed health apps which can be downloaded at the app store, little scientific research has been done about apps for health promotion practices and health outcomes and there is even less research about psychological interventions in the organizational context.[1,2] Some notable publications in the

Corresponding author:
Anita Dunkl, Institute of Psychology, University of Graz, Universitätsplatz 2, 8010 Graz, Austria.
Email: anita.dunkl@uni-graz.at

development and analysis of apps in the organizational context include apps for individual stress prevention and leadership development,[3–6] but there is a lack of holistic workplace health promotion (WHP) apps which can be implemented in the organizational process.

Reaching the right target group and raising participation rates is one major challenge in WHP. This challenge is the same for app-based projects in WHP. Both individual and organizational differences are able to influence participation, such as gender, age, health status, experience, or size of organization.[7–10] Therefore, the aims of this article are to (1) identify individual and organizational aspects which might influence the decision of using an app in WHP and (2) analyze the opinion about feedback possibilities of apps in WHP. More specifically, we investigate the opinion of leaders, as leaders can be seen as a major key factor in the success of WHP projects.[11,12]

Smartphone apps and health promotion

Apps in health promotion have multiple benefits, such as being relatively inexpensive, reaching a large number of individuals even in remote places, supporting highly stigmatized groups without exposing them, sharing health information on social media, 24-h accessibility, providing updated content to the app any time, implementing visual materials (e.g. graphs, videos), and repeating lectures/trainings as often as wanted.[3,4,6,13–16]

Another major advantage is delivering instant feedback and support.[4,15] Instant feedback helps participants monitoring their personal progress and can motivate participants to proceed with health apps. In their guidelines for Internet interventions, Proudfoot et al.[17] stated that professional support should be included to assist an online intervention. This professional support can be delivered frequently by qualified health professionals using e-mail, forums, audio chat, webcam, or even face-to-face support. Ritterband et al.[15] also suggested that instant feedback and support can be provided in many ways, for example, personalized e-mails, instant messaging communication, or even phone and face-to-face meetings.

However, with external support via phone or face-to-face meetings, the reach of apps will be reduced, the costs increase, and the app will be more difficult to organize.[18] On the other hand, a totally unguided program has more drop-out rates and even can raise ethical issues, such as failing treatments, negative side effects, or misinterpretation of the instant feedback on the screen.[16,18] Furthermore, a program without any experts behind is lacking quality and will be less effective.[19]

A major challenge is quality assurance, as currently nearly everybody is allowed to develop an app.[20] Another potential risk is data security and privacy.[1,21] The user's lack of technical skills or knowledge related to the use of new media is an additional barrier that can prevent persons to try app-based health promotion programs.[20]

Reaching the right target group is another critical aspect that has to be noted. Health apps are mainly used by those individuals already supporting a healthy lifestyle,[22,23] a point which can be found also in other health promotion activities.[10,24] This is critical, especially when it comes to WHP, where the organization puts time, money, and effort into efficient, successful health promotion activities.

Individual and organizational differences in WHP programs

As previously mentioned, healthier people have higher participation levels in WHP programs. Additionally, women are shown to participate more often in health intervention programs than men,[7,10,25] but there is also evidence that with offering the right activities, men and women are participating alike.[9] Next to gender, there seems to be a trend where younger employees have the highest and elder employees have the lowest participation level in WHP programs.[9] However,

when it comes to long-term participation, older participants are more likely to participate in follow-up programs.[24,26]

Furthermore, WHP programs are usually conducted in larger enterprises. In smaller organizations, the owner usually is responsible for all management functions and additionally managing health programs therefore is less interesting. Moreover, the benefits of WHP activities are less obvious in the short term, as problems such as health issues or accidents rarely occur in small organizations.[8,27] However, there is evidence that although larger companies usually organize more health promotion activities, the participation levels in smaller organizations are higher.[9]

Several methods have been suggested to increase participation rates, such as individually tailored interventions, personalized mails, or giving instant individual feedback.[2,25,28,29] For health promotion at the workplace, recruitment via company counselors or in particular leaders should be considered to raise participation.[11,12,22] Successful WHP projects are presented by leaders encouraging their employees to adopt a healthy lifestyle and give information about health-relevant issues.[30] Furthermore, leaders who illustrate the benefits of a WHP program are able to influence the participation level positively.[31] Therefore, leaders are the promoters of WHP and must be seen as the major target group when implementing app- or web-based health promotion programs at the workplace.

Research questions

In this article, we want to identify individual and organizational aspects which might influence the decision of using an app in WHP and analyze the opinion about feedback possibilities of apps in WHP. For this purpose, a study with leaders is conducted, as leaders can be seen as a major key factor in the success of WHP projects.[11,12] Referring to the current status of research, following hypotheses were stated:

1. Female leaders have a more positive opinion of using apps in WHP and their feedback possibilities.
2. Younger leaders have a more positive opinion of using apps in WHP and their feedback possibilities.
3. Leadership experience is positively related to acceptance of apps and their feedback options in WHP, as experienced leaders have the information and skills to conduct and promote health promotion projects at the workplace.
4. Similar to the findings that healthier individuals more often participate in health promotion activities, we propose that leaders with a more positive attitude toward health promotion are more interested to use an app and have a more positive opinion for app-related feedback.
5. Working in a larger organization is positively related to a higher acceptance of apps in WHP and its feedback possibilities.
6. Having a higher number of direct subordinates is positively related to a higher acceptance of apps in WHP and its feedback possibilities.

Materials and methods

Recruitment and selection

Austrian and German leaders were invited to participate in an online study. In this study, we defined leaders as persons who assume responsibility for one or more employees in a lower hierarchy.

In cooperation with a German market research company, the participants were invited by sending out e-mails. The participants had to fulfill the requirement of currently having a leading position; otherwise they were excluded at the beginning of the survey. On the whole, 884 participants declared their interest to be part of this study. Out of this sample, 418 participants had to be excluded as they did not have a leading position; 28 participants cancelled their participation after answering only a few questions. In the end, a sample of 438 leaders who filled-in all questionnaires could be used for this study.

Participants

In this sample of 438 leaders, 64.2 percent were working in Germany and 35.8 percent were working in Austria; 29 percent were females and 71 percent were males. In all, 42.5 percent were 40 years or younger, 34.9 percent were between 41 and 50 years old, and 22.6 percent were older than 51 years. They worked in different business sectors, working mostly in the business sectors service/consulting (28.3%), manufacturing (12.3%), commerce (12.3%), health care (6.6%), and insurance (6.4%). Company sizes ranged from one to nine employees (26.2%), 10 to 49 employees (17.1%), 50 to 199 employees (16.8%), and 200 employees and more (39.9%).

The leaders mostly worked in middle or first-line management (68.7%); 31.3 percent were chief executive officers (CEOs). Most of the leaders (38.2%) had their leading position for longer than 10 years, 28.8 percent had their leading position for 5–10 years and 22.7 percent for 2–5 years. The number of direct subordinates varied, ranging from only one subordinate (18.7%), two to five subordinates (29.5%), six to ten subordinates (19.2%), and more than 10 subordinates (32.6%).

Measurements

Next to the sociodemographic data (gender, age, leadership experience, organization size, and number of direct subordinates), the acceptance of using apps in WHP and feedback possibilities of apps were measured with scales previously developed for this study. All items are listed in Table 1.

The *attitude toward WHP in general* was measured with six items. These six items can be combined into one scale "positive attitude toward WHP." The questions could be answered on a 5-point Likert scale from 1 (I disagree) to 5 (I agree).

The *acceptance of using apps in WHP* was measured with two items (Table 1). These two items could be answered on a 5-point Likert scale from 1 (I disagree) to 5 (I agree). These two items are combined into one scale "acceptance of using apps in WHP."

Attitude toward feedback was measured with five items. These items can be categorized in two scales: (1) expert-feedback (three items) and (2) feedback from new media (two items). Answer scale was a 5-point Likert scale ranging from 1 (I disagree) to 5 (I agree).

Feedback possibilities from an app were measured referring to three possible WHP aspects: physical fitness, mental fitness, and nutrition. In all, 12 items could be answered on a 5-point Likert scale ranging from 1 (I disagree) to 5 (I agree). There were four similar answers for each of the three aspects physical fitness, mental fitness, and healthy nutrition (Table 1). The 12 items and can be categorized in four scales: (1) individual feedback, (2) analyses of results, (3) advice/suggestions, and (4) sharing results.

Results

Descriptive statistics

Means, standard deviations, internal consistencies, and correlations of the scales can be found in Table 2. Internal consistencies of all study variables are high, ranging between .80 and .95, with the exception of feedback from new media (.51).

Table 1. Questions used in the study.

No.	Item	Scale
1	Workplace health promotion projects could improve employee health in the long run and can make a positive contribution to the companies' performance	Positive attitude toward WHP
2	I would wish that health-related problems of employees could be detected in an early phase so that interventions can be initiated	
3	In my opinion, interventions in workplace health promotion can contribute to less sickness absences	
4	An organization has to take responsibility for the health of its employees	
5	There is no point in workplace health promotion, as employees usually are not enthusiastic about workplace health promotion activities (inv)	
6	In our organization, workplace health promotion is not necessary (inv)	
7	I would like to use a new technology (e.g. a smartphone app) to support workplace health promotion projects	Acceptance of using apps in WHP
8	With technical help (e.g. a smartphone app) health-relevant aspects can be measured more efficiently	
9	For an adequate feedback, experts are absolutely necessary	Expert-feedback
10	To give professional feedback, trainings would be helpful	
11	Especially for critical feedback, it is important that experts give it	
12	It is an interesting option to give feedback with new media like smartphone apps (e.g. via chat)	Feedback from new media
13	Giving feedback indirectly via new technologies is impersonal (inv)	
14	Individual feedback from a smartphone app/a web portal […] would motivate […] for physical fitness	Individual feedback
15	Individual feedback from a smartphone app/a web portal […] would motivate […] for mental fitness	
16	Individual feedback from a smartphone app/a web portal […] would motivate […] to proceed with healthy nutrition	
17	Analyses of results from a smartphone app/a web portal […] would motivate […] for physical fitness	Analyses of results
18	Analyses of results from a smartphone app/a web portal […] would motivate […] for mental fitness	
19	Analyses of results from a smartphone app/a web portal […] would motivate […] to proceed with healthy nutrition	
20	Advices and suggestions from a smartphone app/a web portal […] would motivate […] for physical fitness	Advice/ suggestions
21	Advices and suggestions from a smartphone app/a web portal […] would motivate […] for mental fitness	
22	Advices and suggestions from a smartphone app/a web portal […] would motivate […] to proceed with healthy nutrition	
23	Letting others see my physical training results on a social platform (e.g. Twitter, Facebook) […] would motivate […] for physical fitness	Sharing results
24	Letting others see my mental training results on a social platform (e.g. Twitter, Facebook) […] would motivate […] for mental fitness	
25	Letting others see me eating healthy food on a social platform (e.g. Twitter, Facebook) […] would motivate […] to proceed with healthy nutrition	

As for the individual aspects, significant correlates can be found for age and positive attitude toward WHP. Age is positively correlated with expert-feedback (.16) and negatively correlated with feedback from new media (−.13), individual feedback (−.12), statistical results (−.11), and

Table 2. Means, standard deviations, internal consistencies (Cronbach's α), and correlations between all study scales.

No.	Dimension	Mean	SD	α	1	2	3	4	5	6	7	8	9	10	11	12
1	Gender	–	–	–												
2	Age	–	–	–	.00											
3	Leadership experience	–	–	–	.07	.57*										
4	Number of direct subordinates	–	–	–	.02	−.04	.06									
5	Size of organization	–	–	–	.09	−.09	−.02	.66*								
6	Positive attitude toward WHP	3.83	.81	.84	−.02	.15*	.06	.05	.16*							
7	Positive acceptance of apps in WHP	3.11	1.08	.80	.06	−.09	−.02	.05	.11**	.26*						
8	Expert-feedback	3.93	.88	.85	−.08	.16*	.09	−.06	.03	.39*	.19*					
9	Feedback from new media	2.34	.96	.51	.05	−.13**	−.09	.10**	.10**	−.16*	.30*	−.20*				
10	Individual feedback	2.71	1.25	.95	.09	−.12**	−.03	.08	.10**	.04	.67*	.08	.44*			
11	Analyses of results	2.77	1.25	.95	.09	−.11**	−.03	.07	.11**	.07	.67*	.10**	.43*	.93*		
12	Advice/suggestions	2.81	1.23	.95	.07	−.09	−.03	.09	.11**	.13*	.65*	.13*	.38*	.87*	.90*	
13	Sharing results	2.26	1.19	.95	.05	−.16*	−.09	.12	.07	−.07	.41*	−.09	.45*	.64*	.60*	.60*

Note: N = 438; Spearman's Rho was calculated for ordinal variables.
*Correlation significant ($p < .01$); **correlation significant ($p < .05$).

sharing results (−.16). Positive attitude toward WHP shows a strong positive correlation with expert-feedback (.39) and moderate to low correlations with acceptance of apps in WHP (.26), feedback from new media (−.16), and advice/suggestions (.13).

For the organizational aspects, only low to zero correlations with the outcome variables can be found, showing the highest correlation between size of organization and positive attitude toward WHP (.16).

Individual and organizational differences and acceptance of apps in WHP/ feedback possibilities

A hierarchical multiple linear regression analysis was conducted separately for all outcome criteria. Table 3 presents the two steps of the analysis. The individual predictors such as gender, age, leadership experience, and positive attitude toward WHP (hypothesis 1, 2, 3, and 4) were stepped into the equation first. This step was significant for all outcomes: acceptance of apps in WHP ($F(4,435)=10.453$, $p<.0001$), expert-feedback ($F(4,435)=23.335$, $p<.0001$), feedback from new media ($F(4,435)=23.335$, $p<.0001$), individual feedback ($F(4,436)=2.472$, $p<.05$), analyses of results ($F(4,436)=3.016$, $p<.05$), advice/suggestions ($F(4,436)=3.735$, $p<.01$), and sharing results ($F(4,436)=3.365$, $p<.01$). The second step of the model included the organizational predictors such as size of organization (hypothesis 5) and number of direct subordinates (hypothesis 6). For all outcome variables, this second step did not significantly add to the explained variance.

Age and positive attitude toward WHP in general were the most important predictors. In detail, gender (hypothesis 1) does not show any significant result, which was not expected in our hypotheses. As assumed in hypothesis 2, age is negatively associated with acceptance of apps in WHP and feedback possibilities of an app. In contrast to our hypothesis, leadership experience did not show any significant association with acceptance of apps or feedback options (hypothesis 3). In line with hypothesis 4, leaders with a more positive attitude toward WHP are more interested to use an app. Hypothesis 5 and 6 also could not be supported, as the size of the organization and the number of direct subordinates did not show a significant relation with acceptance of apps and feedback options.

In Table 3, the standardized regression coefficients as well as p values and adjusted R^2 from the last step are presented.

Discussion

In this article, we aimed to (1) identify individual and organizational aspects which might influence the decision of using an app in WHP and (2) analyze the opinion about feedback possibilities of apps in WHP.

Our first hypothesis, female leaders are more likely to use an app and their feedback possibilities, was not supported. In our findings, gender does not show any significant effect. This non-significant result might be influenced by a common gender stereotype, where men are more open using new media and technical devices. Another explanation could be the fact that women are underrepresented in leading positions. In this study, only 29 percent of the leaders were females. So, this non-significant result could be an effect of an inhomogeneous group.

The second hypothesis, younger leaders would have a more positive opinion of using apps and their feedback possibilities in WHP, was supported. The analysis shows that younger leaders would rather accept an app in WHP and use more of its feedback possibilities (individual feedback, analyses of results, advice/suggestions, sharing results). As younger individuals are shown to participate more often in WHP projects,[9] this result was expected. Additionally, younger individuals are more

Table 3. Results of the multiple regression analyses (standardized regression coefficients from the last step).

	Acceptance of apps in WHP		Expert-feedback		Feedback new media		Individual feedback		Analyses of results		Advice/ suggestions		Sharing results	
	β	p value	β	p value	β	p value	β	p value	β	p value	β	p value	β	p value
Step 1: individual aspects														
Gender (male)	.06	.207	−.07	.116	.04	.433	.08	.086	.09	.068	.07	.155	.04	.372
Age	−.14**	.003	.12*	.010	−.09	.072	−.11*	.022	−.11*	.024	−.10*	.035	−.14*	.003
Leadership experience	.00	.948	.04	.466	−.03	.539	−04	.511	−.02	.676	.02	.737	−.02	.683
Positive attitude toward WHP	.27**	<.001	.38**	<.001	−.17**	<.000	.05	.356	.07	.129	.14**	.005	−.05	.309
Step 2: organizational aspects														
Number of direct subordinates	.03	.667	−.08	.143	.04	.547	.06	.366	.03	.638	.06	.353	.12	.056
Size of organization	.03	.612	.07	.300	.07	.291	.06	.419	.07	.283	.04	.563	.00	.954
Model R^2 adjusted	.08		.17		.04		.02		.02		.02		.02	

Note: **correlation significant ($p < .01$); *correlation significant ($p < .05$).

familiar with new media such as smartphones and apps, as they are usually used by the younger generation.

As the third hypothesis states, leaders with a more positive attitude toward WHP are more interested to use an app but feedback is expected to be given by experts. Feedback from experts is also demanded by older leaders. This refers to quality issues that instant feedback is confronted with. A totally unguided program without human involvement can lead to high drop-out rates and smaller effects as employees are not provided with enough information and knowledge about health-relevant issues.[19] Additionally, ethical issues are raised as feedback can be misinterpreted or might fail its purpose.[16,18] As the statements of expert-feedback do not exclude instant feedback, this issue of quality has to be considered strongly. It can be concluded that the expert's knowledge and skills are essential. This can be also seen in the scale "advice/suggestions from an app," as leaders with a more positive attitude toward WHP would wish this kind of feedback.

In line with past research,[8,27] we expected an effect for organization size and number of direct subordinates, more specifically, that working in a larger organization and having a higher number of direct subordinates is positively related to a higher acceptance of apps in WHP. The findings could not support these assumptions. Larger companies are more willing to invest time and money in health promotion projects,[8,27] but we did not find this result for app-based health promotion, as organization size and number of direct subordinates showed non-significant results. However, very large companies (5001+ employees) have usually fully adopted new technologies, using the Internet for multiple purposes.[32] Therefore, larger companies could be seen as a potential market for apps in WHP. This assumption has to be investigated more closely in future research. Furthermore, leadership experience also shows no significant relation with acceptance of apps and their feedback possibilities in WHP.

Looking at the results, only small effects could be found, having low—although significant—adjusted R^2. As we were able to collect a large sample of leaders in this study, we can benefit from this advantage and see also these small effects. However, more studies including other influencing factors such as daily use of smartphones and apps or affinity for technology and mobile communications should be conducted to investigate possible effects.

Limitations

Data were collected online, which have the advantage of reaching a large number of participants. With this method, it was possible to collect a set of complete data without missing values from 438 leaders, allowing us to find even small effects. With this method, a selection bias could occur as our sample consists of leaders being familiar with the Internet. Therefore, it seems possible that with this recruitment method we asked leaders that already show a high affinity for technology. Generalizing the results to the total population might be difficult.

The participants were recruited with an online panel; therefore, it is possible that the sample in this study has more affinity for technology and mobile communications. However, this conclusion must be done with caution as we did not ask the participants about their frequency of using apps or other new media. Another important aspect which is not considered in this study is the style of leadership. Leaders that engage in a health-promoting leadership style are more likely to invest time and effort for health promotion practices. These aspects should be included in future research.

Practical implications and conclusion

We could find that the most important aspect that leads to a higher acceptance of apps in WHP is a positive attitude toward WHP. Of course, a positive attitude means being aware of the need for prevention of risks at the workplace. Leaders that already engage in WHP activities can use apps

to support health interventions. Apps could support in many ways, for example, information about specific interventions (date, content) can be made available right away with the use of an app, or short questionnaires about the team's stress or motivation level can be used to get a feedback about the team's psychological state. Next studies should include possible features of an app in WHP to investigate the leaders' wishes and expectations.

The findings imply that using feedback possibilities of smartphone apps could be beneficial for WHP. The correlation analyses imply that leaders who would accept an app in WHP are also more interested in its feedback possibilities. The findings further suggest that feedback from experts could be an addition to automatic feedback. Especially when it comes to mental aspects, an automatic feedback without talking to "real" people can be critical, as automatic answers tend to be rather brief with less explanation. In line with this assumption, sharing these short automatic answers on social platforms may not be desirable, as they might lead to stigmatization. This is much more critical in a work-context, where colleagues and leaders might see the automatically generated feedback. In this sense, a web-based intervention with automatic feedback—either on the personal computer or on the smart phone—must be programmed in highest possible quality. Apps can help to sensitize people but in every case we need experts for every area, for example, physical activities, mental fitness, and nutrition. These experts can be involved on many levels, such as planning the content of the program, developing understandable, non-stigmatizing feedback texts, or being available for life-chats.

In the guidelines for Internet intervention research of Proudfoot et al.,[17] professional and other support was already discussed. We should proceed to the next step also for Internet interventions directly. Including "experts" in the process is an important quality criterion for every Internet/software-based application starting from the concept to the implementation. This is even more important when feedback possibilities are included, as the support of "real" people is needed especially when feedback of critical states are displayed. This will further raise the acceptance of using apps or web-related programs in WHP.

Acknowledgement

The authors want to thank the company Mikropis Holding, d.o.o, a partner in the EU Project.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was sponsored by the company Mikropis Holding, d.o.o, a partner in the EU Project, 1646-002 A-LIFE.

References

1. Kratzke C and Cox C. Smartphone technology and apps: rapidly changing health promotion. *Int Electron J Health Educ* 2012; 15: 72–82.
2. Ly KH, Asplund K and Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. *Internet Interv* 2014; 1: 95–101. DOI: 10.1016/j.invent.2014.06.003.
3. Billings DW, Cook RF, Hendrickson A, et al. A web-based approach to managing stress and mood disorders in the workforce. *J Occup Environ Med* 2008; 50: 960–968. DOI: 10.1097/JOM.0b013e31816c435b.

4. Hasson H, Brown C and Hasson D. Factors associated with high use of a workplace web-based stress management program in a randomized controlled intervention study. *Health Educ Res* 2010; 25(4): 596–607. DOI: 10.1093/her/cyq005.
5. Ruwaard J, Lange A, Bouwan M, et al. E-mailed standardized cognitive behavioural treatment of work-related stress: a randomized controlled trial. *Cogn Behav Ther* 2007; 36: 179–192. DOI: 10.1080/16506070701381863.
6. Kawakami N, Takao S, Kobayashi Y, et al. Effects of web-based supervisor training on job stressors and psychological distress among workers: a workplace-based randomized controlled trial. *J Occup Health* 2006; 48: 28–34.
7. Bradley KL, Robinson LM and Brannen CL. Adolescent help-seeking for psychological distress, depression, and anxiety using an Internet program. *Int J Ment Health Promot* 2014; 14: 23–34. DOI: 10.1080/14623730.2012.665337.
8. Champoux D and Brun JP. Occupational health and safety management in small size enterprises: an overview of the situation and avenues for intervention and research. *Saf Sci* 2003; 41: 301–318. DOI: 10.1016/S0925-7535(02)00043-7.
9. Robroek SJ, van Lenthe FJ, van Empelen P, et al. Determinants of participation in worksite health promotion programmes: a systematic review. *Int J Behav Nutr Phys Act* 2009; 6: 26. DOI: 10.1186/1479-5868-6-26.
10. Van Verheijden MW, Jans MP, Hildebrandt VH, et al. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. *J Med Internet Res* 2007; 9: e1. DOI: 10.2196/jmir.9.1.e1.
11. Dellve L, Skagert K and Vilhelmsson R. Leadership in workplace health promotion projects: 1- and 2-year effects on long-term work attendance. *Eur J Public Health* 2007; 17: 471–476. DOI: 10.1093/eurpub/ckm004.
12. Whitehead D. Workplace health promotion: the role and responsibility of health care managers. *J Nurs Manag* 2006; 14: 59–68. DOI: 10.1111/j.1365-2934.2005.00599.x.
13. Griffith F, Lindenmeyer A, Powell J, et al. Why are health care interventions delivered over the internet? A systematic review of the published literature. *J Med Internet Res* 2006; 8: e10. DOI: 10.2196/jmir.8.2.e10.
14. Househ M, Borycki E and Kushniruk A. Empowering patients through social media: the benefits and challenges. *Health Informat J* 2014; 20: 50–58.
15. Ritterband LM, Thorndike FP, Cox DJ, et al. A behavior change model for Internet interventions. *Ann Behav Med* 2009; 38: 18–27. DOI: 10.1007/s12160-009-9133-4.
16. Rozental A, Andersson G, Boettcher J, et al. Consensus statement on defining and measuring negative effects of Internet interventions. *Internet Interv* 2014; 1: 12–19. DOI: 10.1016/j.invent.2014.02.001.
17. Proudfoot J, Klein B, Barak A, et al. Establishing guidelines for executing and reporting Internet intervention research. *Cogn Behav Ther* 2011; 40: 82–97. DOI: 10.1080/16506073.2011.573807.
18. Donker T, van Straten A, Riper H, et al. Implementation of Internet-based preventive interventions for depression and anxiety: role of support? The design of a randomized controlled trial. *Trials* 2009; 10: 59. DOI: 10.1186/1745-6215-10-59.
19. Shimazu A, Kawakami N, Irimajiri H, et al. Effects of web-based psychoeducation on self-efficacy, problem solving behavior, stress responses and job satisfaction among workers: a controlled clinical trial. *J Occup Health* 2005; 47: 405–413. DOI: 10.1539/joh.47.405.
20. Ballew P, Castro S, Claus J, et al. Developing web-based training for public health practitioners: what can we learn from a review of five disciplines? *Health Educ Res* 2013; 28: 276–287. DOI: 10.1093/her/cys098.
21. Miller G. The smartphone psychology manifesto. *Perspect Psychol Sci* 2012; 7: 221–237. DOI: 10.1177/1745691612441215.
22. Geraedts AS, Kleiboer AM, Wiezer NM, et al. Feasibility of a worker-directed web-based intervention for employees with depressive symptoms. *Internet Interv* 2014; 1: 132–140. DOI: 10.1016/j.invent.2014.07.001.

23. Heber E, Ebert DD, Lehr D, et al. Efficacy and cost-effectiveness of a web-based and mobile stress-management intervention for employees: design of a randomized controlled trial. *BMC Public Health* 2013; 13: 655. DOI: 10.1186/1471-2458-13-655.
24. Glasgow RE, Vogt TM and Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999; 89: 1322–1327. DOI: 10.2105/AJPH.89.9.1322.
25. Strecher VJ, McClure J, Alexander GL, et al. Web-based smoking-cessation programs: results of a randomized trial. *Am J Prev Med* 2008; 34: 373–381. DOI: 10.1016/j.amepre.2007.12.024.
26. Glasgow RE, Nelson CC, Kearney KA, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. *J Med Internet Res* 2007; 9: e11. DOI: 10.2196/jmir.9.2.e11.
27. Gardner D, Carlopio J, Fonteyn PN, et al. Mechanical equipment injuries in small manufacturing businesses. Knowledge, behavioral, and management issues. *Int J Occup Saf Ergon* 1999; 5: 59–71.
28. Carlbring P, Maurin L, Törngren C, et al. Individually-tailored, Internet-based treatment for anxiety disorders: a randomized controlled trial. *Behav Res Ther* 2011; 49: 18–24. DOI: 10.1016/j.brat.2010.10.002.
29. Highet N and Culjak G. Are mailouts an effective health promotion strategy? If so, for whom? *Int J Ment Health Promot* 2014; 14: 13–22. DOI: 10.1080/14623730.2012.665340.
30. Gurt J, Schwennen C and Elke G. Health-specific leadership: is there an association between leader consideration for the health of employees and their strain and well-being? *Work Stress* 2011; 25: 108–127. DOI: 10.1080/02678373.2011.595947.
31. Nöhammer E, Eitzinger C, Schaffenraht-Resi M, et al. Zielgruppenorientierung und betriebliche Gesundheitsförderung [Target group orientation and workplace health promotion]. *Präv Gesundh* 2009; 4: 77–82. DOI: 10.1007/s11553-008-0154-9.
32. Papastathopoulou P and Avlonitis G. Classifying enterprises on the basis of WWW use: a behavioral approach. *Internet Res* 2009; 19: 332–347. DOI: 10.1108/10662240910965388.

Health Informatics Journal

Article

Knowledge-guided mutation in classification rules for autism treatment efficacy

Health Informatics Journal
2017, Vol. 23(1) 56–68

DOI: 10.1177/1460458215627187
journals.sagepub.com/home/jhi
SAGE

Kelley Engle
Harrisburg Area Community College (HACC), USA

Roy Rada
University of Maryland, Baltimore County (UMBC), USA

Abstract

Data mining methods in biomedical research might benefit by combining genetic algorithms with domain-specific knowledge. The objective of this research is to show how the evolution of treatment rules for autism might be guided. The semantic distance between two concepts in the taxonomy is measured by the number of relationships separating the concepts in the taxonomy. The hypothesis is that replacing a concept in a treatment rule will change the accuracy of the rule in direct proportion to the semantic distance between the concepts. The method uses a patient database and autism taxonomies. Treatment rules are developed with an algorithm that exploits the taxonomies. The results support the hypothesis. This research should both advance the understanding of autism data mining in particular and of knowledge-guided evolutionary search in biomedicine in general.

Keywords

autism, classification rules, genetic algorithms, medical domain knowledge

Introduction

Genetic algorithms have been applied to rules from decision trees to improve diagnosis in health care.[1] This research extends that work by adding domain knowledge to the genetic algorithm.

Genetic algorithms that incorporate domain knowledge may be called memetic algorithms,[2] and this research contributes to the study of memetic algorithms. Specifically, the discovery of new rules will be facilitated through the use of domain knowledge to guide the choice of concepts to incorporate in a rule.

Corresponding author:
Roy Rada, University of Maryland, Baltimore County (UMBC), Baltimore, MD 21250, USA.
Email: rada@umbc.edu

The machine learning research here is applied to the domain of autism. Autism is a neurodevelopmental disorder that first appears during childhood and generally follows a steady course without remission. Broadened to include "autism spectrum disorders," the disease affects 11 of 1000 children in the United States.[3] Decision-support tools can support management of such widespread, pediatric disorders.[4]

The knowledge about autism that is used for this work comes from the Interactive Autism Network (IAN) dataset. IAN is an online, research registry that connects family members of autistic people with researchers in an effort to help solve the many problems associated with autism.[5]

The No Free Lunch Theorem states that a universal optimization strategy is not possible.[6] Therefore, one method can only outperform another method if it is designed to solve a particular problem or somehow structured to be specialized. Evolutionary algorithms that incorporate heuristic-based knowledge can outperform ones without such knowledge.[7] The domain knowledge can be injected at any phase of the genetic algorithm—including initialization, representation, selection, crossover, and mutation phases.[8] The idea of incorporating domain knowledge into an evolutionary process has also been applied to neural network applications in medicine.[9]

The role of domain knowledge for the medical domain has been investigated in a number of studies including the use of evolving sub-ontologies for traditional Chinese medicine.[10] Verb selection patterns were used within a genetic algorithm to classify newly recognized biomedical terms co-occurring with domain-specific verbs.[11] Of course, other problem domains, such as finance, have been approached with memetic algorithms.[12]

In the work to be presented here, the memetic algorithm is used to produce classification rules for autism treatment efficacy. Within the medical domain, classification rules have frequently aided clinicians and medical researchers. For example, neuro-fuzzy rule-based classifiers provided linguistically interpretable rules for one medical field.[13] Classification rules have been extracted from trained neural networks for breast cancer diagnosis.[14] A genetic programming algorithm was developed for discovering classification rules in breast cancer, dermatology, and pediatric adrenocortical tumors.[15]

Several studies have incorporated artificial intelligence methods for the autism domain. Text mining was applied to biomedical literature for the construction of an ontology which identified rare relations in autism,[16] which led to a literature mining method for uncovering hidden relations from a set of articles in a given domain.[17] Self-organizing maps were used to model attention shift impairment and familiarity preference,[18] both hallmarks of autistic behavior.

Artificial intelligence and genetic databases have been combined in applications to autism. Decision trees[19] were created to predict the severity of autism based on single nucleotide polymorphisms.[20] Genetic and environmental factors were examined using combinatorial fusion analysis and association rule mining to determine associations between autism prevalence and the exposure to mercury and lead during critical stages of a child development.[21] Another genetics study found that association rules were able to successfully predict autism susceptibility genes.[22]

Support vector machines were able to categorize infants in high- and low-risk groups for autism via an analysis of electroencephalogram (EEG) data.[23] An expert system was developed as a screener for autism.[24] Heart rate patterns were compared with common autism behavioral problems, such as self-injury and aggression.[25] The performance of various machine learning algorithms in a healthcare application has been compared.[26]

One parameter of a knowledge-guided, evolutionary search algorithm is the size of conceptual changes to a rule.[27] This article compares the effects of knowledge-guided mutation to the traditional method of random mutation. The algorithm operates on a population of classification rules (for autism treatment efficacy) created from the IAN dataset. The fitness measure is the

classification rule accuracy. The hypothesis is that mutations that implement a small conceptual change will result in small changes in rule accuracy. The results of these empirical tests will assist in the determination of how to best incorporate domain knowledge for both classification rules and genetic algorithms.

The remainder of this article is structured as follows: theory, methodology, results, discussion, and conclusion. The "Theory" section introduces and defines the concepts of domain knowledge, semantic distance, classification rule, and accuracy. The "Methodology" section describes the databases, taxonomies, and algorithms used in the experiments. In the "Results" section, the results of experiments and the analysis of those results are presented. Finally, the "Discussion" and "Conclusion" sections put those results in perspective.

Theory

The fundamental goal was to determine how domain knowledge can be usefully incorporated in knowledge-guided mutation in order to constrain the search and thereby constrain the associated fitness. Mutations involving random changes often lead to extreme fluctuations in the associated fitness measure to the detriment of the overall goal of discovering the optimal solution. The empirical results will help illuminate the relationship that exists between exploration and exploitation within the solution space. The hypothesis is that for a given set of classification rules, a systematic, incremental change in semantic distance will result in parallel change in accuracy.

Next, important terms are defined as follows:

Domain knowledge

Domain knowledge can be conceptualized as meta-data or data regarding the data that pertain to a particular domain. The domain data itself should be specific and relevant to the problem being solved.

Systematic

A *systematic* change is implemented based on the information contained in the IAN Semantic Diagram (presented in the "Methodology" section). Four categories of knowledge guidance (KG) control the amount of change that may be applied to a classification rule for autism treatment efficacy:

KG_1: Knowledge-guided mutation level 1—Involves the minimum amount of change to the classification rule. This method will only allow a medication to be replaced with another medication from the same category.

KG_2: Knowledge-guided mutation level 2—Allows a greater change to the classification rule than KG_1. This method will allow the medication to be replaced with any other medication.

KG_3: Knowledge-guided mutation level 3—Allows a greater change to the classification rule than KG_2. This method will allow the medication to be replaced with any other treatment.

KG_4: Knowledge-guided mutation level 4—Allows the greatest amount of change to the classification rule. This method will allow the medication to be replaced with non-treatment data attributes (such as "patient diagnosis" and "parent expectation of outcome").

Semantic distance

The semantics for the IAN data are formally presented in the drug taxonomy and KG levels (KG_1–KG_4). Semantic distance relates to the relative change between one KG level to another (i.e. KG_1, …, KG_2).

Parallel

Parallel refers to the relationship between the change in a rule antecedent and the resulting change in that rule's accuracy. For each degree of change to a classification rule (i.e. KG_1–KG_4), a similar change is hypothesized to occur in the resulting accuracy. In other words, the smallest degree of change (i.e. KG_1) will result in the smallest (relative to KG_2 through KG_4) amount of change in accuracy.

Accuracy

Accuracy can be defined as follows

$$Accuracy(R) = \frac{n_{totcorrect}}{n_{totrecords}}$$

where R represents a single classification rule, $n_{totcorrect}$ represents the total number of records correctly classified by R, and $n_{totrecords}$ represents the total number of records including those not correctly classified. For any given experiment, where P represents a population of initial classification rules, an individual change in accuracy is computed and an absolute change in accuracy is captured as a_1 to a_n where n represents the total number of records in P. For each rule, this individual accuracy measure is computed using 10-fold cross-validation. The overall average measure, *AvgAcc*, can be formally defined as follows

$$\overline{AvgAcc} = \{a_1, a_2, \ldots . a_n\}$$

Classification rule

A classification rule consists of an *antecedent(s)* and a *consequent* which predict the class of instances covered by that rule. It is represented in IF–THEN form where multiple *antecedents* together form a logical conjunction.

Methodology

The research involved accessing a large database, developing a learning algorithm, and analyzing the results, as will explained in this section. The database comes from IAN and the taxonomies from IAN and the National Library of Medicine. The algorithm involves initializing the population of classification rules and then generating new rules in various ways. Finally, a working example is provided of *knowledge-guided mutation*.

Data

The IAN Project collects data from families with autism including information on demographics, parent medical data, sibling information, as well as a vast quantity of data on the autistic

child including treatment information. Although the primary reason for the creation of IAN is to link researchers in the field of autism to potential subjects, the non-identifiable data stored on these families are accessible to any researchers that have proper Institutional Review Board approval.

The IAN data for this experiment covered approximately 9800 children between the ages of 0 and 18 years who have been diagnosed with some form of autism. The dataset has over 60,000 individual treatment records and thus approximately 6.3 treatments per autistic child.

In the data release that is being used for this research (date of version: 19 March 2012), there are 334 attributes. Some are categorical, some are binary, and some are numeric. Understanding the data requires some intimacy with the phenomenon of autism. For example, the categorical attribute of patient diagnosis included these entries: autism, Asperger's syndrome, childhood disintegrative disorder, pervasive development disorders, and autism syndrome disorder. These different diagnoses reflect the variations of autism as identified by healthcare professionals. Across the database, these different diagnostic categories appear with a wide, relatively even distribution which makes that attribute a useful one for classification purposes.

For the purposes of data mining, identifying the appropriate outcome measure is an important pre-processing step prior to model building. There are a number of outcome measures for the autistic child present in the IAN data that can be derived from IAN's Social Communication Questionnaire, its Social Responsiveness Scaling, or its treatment outcomes.

The Social Communication Questionnaire is utilized as a tool to screen for autism spectrum disorders. It consists of 40 "Yes/No" questions to be answered by the parent and will give a resulting score. The questionnaire is designed to classify a child as autistic or not and does not allow for gradations of severity. Due to the fact that this tool cannot indicate improvement or worsening in a child (a movement in the raw score is not an accurate measure), it will not be used as an outcome measure in any of the analyses.

Social Responsiveness Scaling could be used for diagnosing the severity of the autism disorder in children. It categorizes autistic children as non-autistic, mildly autistic, or severely autistic. However, in the dataset, the breakdown for the autistic children is heavily skewed toward severe autism.

Treatment outcome is a complex outcome measure that tracks the parent's perception of treatment efficacy over time (horizontally) and by treatment (vertically). A parent rates each treatment that the autistic child receives at three distinct time periods: (1) when the treatment is first entered into the system; (2) every year when the parents are requested to update their child's treatments; and (3) when the treatment is stopped. The treatment efficacy measure is provided by the parent using a 9-point Likert scale with four ratings for worsening, 1 for *no change* and 4 for *improvement*. Given the limitations of the other outcomes and despite the challenges of using subjective data,[28,29] the treatment outcomes proved to be the appropriate outcome measure for these experiments.

A 9-point Likert scale was mapped to a categorical measure clinical outcome. Outcome measures with high cardinality typically suffer from over-fitting when used in various data mining activities. Classification rules are particularly susceptible to this phenomenon. Over-fitting would lead to classification rules that will perform well on training data, but when tested on separate data perform very poorly.

Initialization

The semantic net for this research (Figure 1) relates to the data structure in IAN. The circle at the top level represents all the IAN data, whereas at the second level the circles represent the subentities present in the IAN dataset:

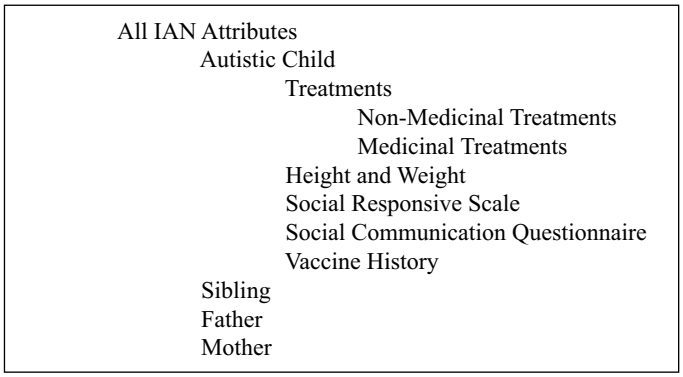

Figure 1. IAN semantic net.
This hierarchical listing depicts a partial taxonomy of top-level IAN attribute. An indentation indicates a descendant relation. It omits the children of the medicinal treatments that are presented separately.

- Autistic child;
- Mother;
- Father;
- Sibling.

The circles at the third level represent the sub-entities associated with the main entity. The autistic child has a number of sub-entities including the following:

- Social Communication Questionnaire;
- Vaccine history;
- Social Responsive Scale;
- Height and weight history;
- Treatment information.

Treatment information includes two sub-entities: medicinal treatments and non-medicinal treatments.

The Medical Subject Headings (MeSH) is a massive thesaurus from the National Library of Medicine.[30] Drug information was obtained from MeSH that was associated with the treatments in the IAN database. The 56 distinct medicinal treatments were then classified into nine categories based on MeSH (Table 1).

The experiments used a classification rule with two predicates in the IF-part and one in the THEN-part. The first IF predicate was the First Autism Diagnosis Category (DiagnosisPredicate). The second IF predicate was treatment. The outcome predicate was one of three values: improvement, no change, or worsening. By combining all the variations of predicates, 1027 distinct rules were created.

Experiment parameters

Knowledge may be exploited to constrain evolutionary search, and this experiment will use domain knowledge to constrain the mutation operation.[31] The genetic algorithm implementation isolated the effects of mutation and did not use the crossover operator. The software to run the experiments

Table 1. Drug taxonomy.

ADHD	Allergy	Anti-anxiety	Seizure	Anti-depressant	Anti-psychotic
Adderall	Benadryl	Buspar	Carbamazepine	Celexa	Abilify
Catapres	Claritin	Ativan	Clonazepam	Citalopram	Geodon
Daytrana	Singulair	Buspirone	Depakote	Effexor	Invega
Dexedrine	Zyrtec	Fluvoxamine	Keppra	Lexapro	Lithium
Focalin		Lorazepam	Lamictal	Paxil	Risperdal
Guanfacine		Luvox	Tegretol	Prozac	Risperidone
Intuniv		Trazodone	Topamax	Remeron	Seroquel
Meta-date			Trileptal	Sertraline	Zyprexa
Methylphenidate				Wellbutrin	
Ritalin				Zoloft	
Tenex					
Vyvanse					

ADHD: attention deficit hyperactivity disorder.
This table shows the drug taxonomy for the six categories of medication with the most drugs. This omits the three categories of anti-fungal, acid reflux, and laxative which had a total of seven drugs among them.

Table 2. Experimental details.

Experiment	Population	Fitness	Hypothesis test
KG_1 versus KG_2	1027 two-predicate rules	Accuracy	One-way ANOVA: Δ in accuracy
KG_1 versus KG_3	1027 two-predicate rules	Accuracy	One-way ANOVA: Δ in accuracy
KG_1 versus KG_4	1027 two-predicate rules	Accuracy	One-way ANOVA: Δ in accuracy

ANOVA: analysis of variance.
This table shows the three experiments to have the same initial population of rules, the same fitness measure, and the same statistical test of the hypothesis.

was coded in Microsoft Access Visual Basic for Applications. The experiment parameters are presented in Table 2.

For each knowledge level of guidance, mutations were performed on the second predicate. The pseudo-code in Figure 2 describes the mutation process and the fitness measure. Each time a rule was mutated, the offspring rule was tested for accuracy. Information about the rule, its mutation, and the performance of the offspring rule was recorded for every mutation and rule. A one-way analysis of variance (ANOVA) test was then conducted in SPSS to determine whether the variability between accuracies at the different knowledge levels was statistically significant.

Working example

A simple example illustrates the method presented in Figure 2. First, the starting classification rule with a baseline accuracy of 80 percent is

$$\text{IF DiagnosisPredicate} = 3 \text{ AND TreatmentName} = \text{Adderall THEN Improvement}$$

This classification rule is mutated under the guidance of KG_1 which is the smallest amount of change that can be applied. KG_1 dictates that the TreatmentName attribute should be changed to

(a)

Definitions

CR = base population of 2-predicate classification rules
$CR_{1\ldots.n}$ = individual classification rules consisting of P1 and P2 where n = 1,027
$CR_{1acc\ldots.nacc}$ = initial accuracy calculated for each 2-predicate classification rule
$CRM_{1\ldots.n}$ = New individual classification rule consisting of P1 and PM2
$CRM_{1acc\ldots.nacc}$ = new accuracy calculated for each 2-predicate classification rule
P_1 = first predicate
P_2 = second predicate
P_{2M} = second predicate after mutation
KGx= The level of knowledge guidance selected (i.e, KG_1, KG_2, KG_3, KG_4)
$Accabs_{1\ldots n}$ = represents the change in accuracy
$\overline{AvgAcc}$ = captures the mean of the absolute change in accuracy

(b)

Pseudo-code

```
y=1
For CR1….CRn
        //The second predicate is mutated according to the selected knowledge level
        P2M = KGx(P2)

        //The new rule consists of the first predicate and mutated second predicate
        CRMy = P1 + P2M

        //The accuracy is calculated for the mutated rule
        CRMyacc = accuracy (CRMy)

        //The absolute change in accuracy is calculated
        Accabsy = ABS (CRyacc - CRMyacc)

        y++

Next Record
End Loop
```

$\overline{AvgAcc} = \frac{CRM_{1acc} + CRM_{2acc} \ldots.. CRM_{nacc}}{n}$ //The average (absolute) change in accuracy is computed

Figure 2. (a) Definitions for pseudo-code: These acronyms appear in the pseudo-code that describes the memetic algorithm used in this experiment. The acronyms are defined in this figure. (b) Pseudo-code: This pseudo-code describes the memetic algorithm of this experiment. The algorithm is essential one "Do Loop." The lines beginning with "//" are comments. "y++" means to add one to y. "ABS" means absolute value. The other acronyms are defined in (a).

another medication in the same family as Adderall. Since Adderall belongs to the category attention deficit hyperactivity disorder, another medication from this same category will be randomly selected—in this case Intuniv. The new classification rule with the calculated accuracy of 82 percent is

IF DiagnosisPredicate = 3 AND TreatmentName = Intuniv THEN Improvement

Table 3. Summary of results.

Experiment	Change in accuracy (%)	Change in accuracy (absolute, %)
KG_1	.25	13
KG_2	−.67	16
KG_3	18	23
KG_4	11	21

The leftmost column indicates the experiment. The second column is the average change in accuracy, while the third column is the average change in absolute accuracy. In each case, the accuracy is being compared.

The absolute change between the two rules is 2 percent. A similar methodology has been used in the financial domain where rules were mutated under constraints of domain-specific taxonomy, but the domain was finance.[32]

Results

In this section, the results of the experiment comparing the different knowledge-guided mutation levels (KG_1 through KG_4) will be presented. The experiments support the hypothesis. Additionally, an analysis of the issue of evolving poor rules is discussed.

In Table 3 and Figure 3, a summary of the experimental results is presented. Since the sign would alter the average significantly, the absolute value is also reported. The results indicate that there is an increase in (absolute) change in accuracy across all experiments.

To establish statistical significance, a one-way ANOVA was conducted in SPSS for the experimental results for KG_1, KG_2, KG_3, and KG_4. The data contained in four separate data files captured the absolute change for each. In Table 4, the descriptive statistics are presented.

Table 5 reports the results of the pairwise comparison (t-test). The results report the mean difference, standard error, and statistical significance. For each of the comparisons, the p value is below .05, thereby confirming statistical significance of the variances. The results of the one-way ANOVA indicate that the variances in absolute change that occurred in the four experiments (KG_1–KG_4) could not have been due to random fluctuation.

Many of the beginning rules in the dataset as well as number of the mutated rules suffered from extremely low accuracy. These rules can be described as *poor rules*. Poor can be defined in two ways: (1) rules that suffer from extremely low accuracy (i.e. 0%) or (2) rules that suffer from extremely low support (.001 or less). In the initial dataset of 1027 rules, 261 suffer from extremely low support. Subsequently, there are 427 rules with an accuracy of 0 percent. The two reasons for including these poor rules in the experiments are presented next.

First, the initial population in a genetic algorithm should exhibit diversity to provide sufficient variation in the mutation phase. In other words, starting with a population of all good parents (i.e. classification rules) who exhibit both high accuracy and high support is not desirable since this will overly constrain the search space. Ideally, the beginning population of classification rules should cover a wide spectrum of the desired search space.

Second, due to the multidimensional nature of the classification rules (i.e. two predicates), it cannot be assumed that poor support or accuracy is due to the first predicate. It is possible that by mutating the second predicate, a rule may be generated that is significantly more robust. Discarding rules with poor support or accuracy could constrain the search space unnecessarily.

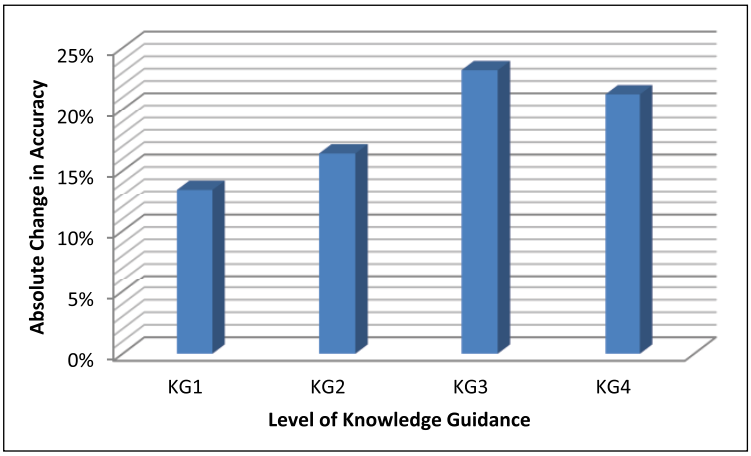

Figure 3. Change in accuracy by level of knowledge guidance.
This figure shows the absolute change in accuracy on the y-axis by level of knowledge guidance on the x-axis.

Table 4. Descriptive statistics.

Experiment	Mean	SD
KG_1	.13	.23
KG_2	.16	.24
KG_3	.23	.32
KG_4	.21	.30

SD: standard deviation.
One-way ANOVA descriptive statistics. The experiment is in the first column, and the mean and SDs are in columns 2 and 3. The number of rules in the population was 1027 for each of the four experiments.

Table 5. Pairwise comparison.

(I) Algorithm	(J) Algorithm	Mean difference (I−J)	SD	Sig.
KG_1	KG_2	−.030	.008	.000
	KG_3	−.097	.011	.000
	KG_4	−.078	.010	.000
KG_2	KG_1	.030	.008	.000
	KG_3	−.068	.011	.000
	KG_4	−.048	.010	.000
KG_3	KG_1	.097	.011	.000
	KG_2	.068	.011	.000
	KG_4	.020	.009	.037
KG_4	KG_1	.078	.010	.000
	KG_2	.048	.010	.000
	KG_3	−.020	.009	.037

Results of pairwise comparisons for KG_1, KG_2, KG_3, and KG_4. The column Sig. is the p value where anything less than .05 is statistically significant.

Discussion

Limitations to this research could affect its generalizability but could be readily addressed in future experiments. One logical next step would be to expand the experimental results to include different attributes in the first predicate. Another step would be to expand the generations of mutation by one of two methods:

- The first would include creating a set of multiple distinct populations (i.e. 100 datasets consisting of 100 rules). KG_1, …, and KG_4 would be separately applied to each of these datasets. The measures would be collected as they were in the current experiment, and the one-way ANOVA test is applied to determine statistical significance.
- The second method to incorporate multiple generations would begin with the initial population (i.e. the dataset of 1026 rules) and apply KG_1, …, KG_4 as was done in this experiment. The results from each of these experiments will become the new populations for further experiments.

The experimental results follow a clear trend for KG_1, KG_2, and KG_3. However, the trend was broken for KG_4, and future experiments could explore this anomaly. From the data and knowledge perspective, a limitation of the experiment was the extent of domain knowledge exploited; in other words, incorporating new sources of domain knowledge might lead to new insights.

Conclusion

This article's section "Discussion" emphasized the limitations of the study, while this section will highlight the generalizability of the results. This article presents the results of an experiment that uses *knowledge-guided mutation* on classification rules for an autism database. Domain knowledge is shown to constrain and guide the mutation operator.

The step-size of the mutation operator tended to correspond to the step-size in the change of performance of the rule that was mutated. A mutation that is restricted to make a small semantic difference in the classification rule tends to lead to a small change in accuracy of the offspring rule. In other words, when the semantic change is allowed to be slightly greater, then the change in accuracy is slightly greater (13% for KG_1, 16% for KG_2, and 23% for KG_3).

The domain knowledge refers to the drug taxonomy or semantic net that was created for the medications in the IAN database. This domain knowledge could be utilized by other autism researchers looking to augment other artificial intelligence methods, such as Naïve Bayes, decision trees, or clustering. This work contributes to the growing body of science about incorporating domain knowledge in machine learning.

Although this research has focused on how to apply the memetic algorithm to the autism domain, the method presented here could be extended to other medical domains, such as cancer and aging. The process of creating a drug taxonomy using MeSH could be applied to other medical domains where the data include medications. This research might help health informatics researchers build prediction or classification models that combine evolutionary algorithms and domain knowledge.

Acknowledgements

The authors would like to thank Dr Paul Law, Founding Director of the Interactive Autism Network and Professor of Pediatrics at Johns Hopkins University, for his help with accessing and using the data.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Vidrighin C and Potolea R. ProICET: a cost-sensitive system for prostate cancer data. *Health Info J* 2008; 14(4): 297–307.
2. Krasnogor N and Smith J. Memetic algorithms: the polynomial local search complexity theory perspective. *J Math Model Algorithm* 2008; 7(1): 3–24.
3. Duchan E and Patel DR. Epidemiology of autism spectrum disorders. *Pediatr Clin N Am* 2012; 59(1): 27–43.
4. Shegog R, Bartholomew LK, Sockrider MM, et al. Computer-based decision support for pediatric asthma management: description and feasibility of the Stop Asthma Clinical System. *Health Info J* 2006; 12(4): 259–273.
5. Daniels AM, Rosenberg RE, Anderson C, et al. Verification of parent-report of child autism spectrum disorder diagnosis to a web-based autism registry. *J Autism Dev Disord* 2012; 42(2): 257–265.
6. Wolpert DH and Macready WG. No free lunch theorems for optimization. *IEEE T Evolut Comput* 1997; 1(1): 67–82.
7. He J, Yao X and Li J. A comparative study of three evolutionary algorithms incorporating different amounts of domain knowledge for node covering problem. *IEEE T Syst Man Cy C* 2005; 35(2): 266–271.
8. Giraldez R, Aguilar-Ruiz JS and Riquelme JC. Knowledge-based fast evaluation for evolutionary learning. Systems. *IEEE T Syst Man Cy C* 2005; 35(2): 254–261.
9. Abbass H. An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artif Intell Med* 2002; 25(3): 265–281.
10. Mao Y, Wu Z, Tian W, et al. Dynamic sub-ontology evolution for traditional Chinese medicine web ontology. *J Biomed Inform* 2008; 41(5): 790–805.
11. Spasic I and Ananiadou S. Using automatically learnt verb selectional preferences for classification of biomedical terms. *J Biomed Inform* 2004; 37(6): 483–497.
12. Aranha C and Iba H. The memetic tree-based genetic algorithm and its application to portfolio optimization. *Memetic Comput* 2009; 1(2): 139–151.
13. Nauck D and Kruse R. Obtaining interpretable fuzzy classification rules from medical data. *Artif Intell Med* 1999; 16(2): 149–169.
14. Setiono R. Generating concise and accurate classification rules for breast cancer diagnosis. *Artif Intell Med* 2000; 18(3): 205–219.
15. Bojarczuk CC, Lopesa H, Freitas A, et al. A constrained-syntax genetic programming systems. *Artif Intell Med* 2004; 30(1): 27–48.
16. Petric T, Urbancic B and Cestnik B. Discovering hidden knowledge from biomedical literature. *Inform Slovenia* 2007; 31(1): 15–20.
17. Petric I, Urbancic T, Cestnik B, et al. Literature mining method RaJoLink for uncovering relations between biomedical concepts. *J Biomed Inform* 2009; 42(2): 219–227.
18. Paplinkski A and Gustafsson L (eds). An attempt in modelling early intervention in autism using neural networks. In: *Proceedings of the joint conference on neural networks*, Budapest, 25–29 July 2004. New York: IEEE.
19. Schetinin V, Jakaite L, Jakaitis J, et al. Bayesian Decision Trees for predicting survival of patients: a study on the US National Trauma Data Bank. *Comput Methods Programs Biomed* 2013; 111(3): 602–612.
20. Jiao Y, Chen R, Ke X, et al. Single nucleotide polymorphisms predict symptom severity of autism spectrum disorder. *J Autism Dev Disord* 2012; 42(6): 971–983.

21. Schweikert C, Li Y, Dayya D, et al. Analysis of autism prevalence and neurotoxins using combinatorial fusion and association rule mining. In: *Proceedings of the 2009 ninth IEEE international conference on bioinformatics and bioengineering*, Taichung, Taiwan, 22–24 June 2009, pp. 400–404. New York: IEEE.
22. Gong L, Yan Y, Xie J, et al. Prediction of autism susceptibility genes based on association rules. *J Neurosci Res* 2012; 90(6): 1119–1125.
23. Stahl D, Pickles A, Elsabbagh M, et al. Novel machine learning methods for ERP analysis: a validation from research on infants at risk for autism. *Dev Neuropsychol* 2012; 37(3): 274–298.
24. Veeraraghavan S and Srinivasan K. Exploration of autism expert systems. In: *Proceedings of the 4th international conference on information technology: new generations*, Las Vegas, NV, 2–4 April 2007. New York: IEEE.
25. Freeman R and Grzymala-Busse J. Functional behavioral assessment using the LERS data mining system: strategies for understanding complex physiological and behavioral patterns. *J Intell Inf Syst* 2003; 21(2): 173–181.
26. Amaral JLM, Lopes AJ, Jansen JM, et al. Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease. *Comput Methods Programs Biomed* 2012; 105: 183–193.
27. Rada R. Gradualness facilitates knowledge refinement. *IEEE Trans Pattern Anal Mach Intell* 1985; 7(5): 523–530.
28. Gan Y, Kakiashvili T and Koczkodaj W. A note on relevance of diagnostic classification and rating scales used in psychiatry. *Comput Methods Programs Biomed* 2013; 112(1): 16–21.
29. Kakiashvili T, Koczkodaj W and Woodbury-Smith M. Improving the medical scale predictability by the pairwise comparisons method: evidence from a clinical data study. *Comput Methods Programs Biomed* 2012; 105(3): 210–216.
30. National Library of Medicine (NLM). *Medical subject headings*. Bethesda, MD: National Library of Medicine (NLM), 2013.
31. Wimmer H and Rada R. Good versus bad knowledge: ontology guided evolutionary algorithms. *Expert Syst Appl* 2015; 42(21): 8039–8051.
32. Du J and Rada R. Knowledge in memetic algorithms for stock classification. *Int J Artif Life Res* 2014; 4(1): 13–29.

Health Informatics Journal

Article

Health Informatics Journal
2017, Vol. 23(1) 69–79

DOI: 10.1177/1460458215627188
journals.sagepub.com/home/jhi
SAGE

An innovative capstone health care informatics clinical residency: Interprofessional team collaboration

Laura M Custis and Shelley Y Hawkins
University of San Diego, USA

Tanna R Thomason
Sharp Memorial Hospital, USA

Abstract

Integrated information systems and wireless technology have been increasingly incorporated into health care organizations with the premise that information technology will promote safe, high-quality, cost-effective patient care. With the advancement of technology, the level of expertise necessary to assume health care information technology roles has escalated. The purpose of this article is to describe a clinical residency project whereby students in a graduate degree health care informatics program successfully fulfilled program competencies through a faculty-lead research project focused on the use of home telehealth with a group of heart failure patients. Through the use of Donabedian's framework of structure, process, and outcomes, the health care informatics students completed essential learning activities deemed essential for transition into the role of an informatics specialist. Health care informatics educational leaders are encouraged to adapt this template of *applied learning* into their practices.

Keywords

health care informatics, health care informatics capstone, health care informatics student, remote patient monitoring, telehealth

Introduction

More than a decade ago, integrated information systems began to be incorporated into health care organizations creating a fundamental change in the health care industry. The premise was that information technology (IT) would promote safe, high-quality, cost-effective patient care.

Corresponding author:
Shelley Y Hawkins, University of San Diego, 5998 Alcala Park, San Diego, CA 92110-2492, USA.
Email: shawkins@sandiego.edu

Globally, health care systems have been experiencing unprecedented expansion and restructuring, in part, to successfully fulfill initiatives such as integrating patient medical records that yield certain incentives and reimbursements for compliant health care systems. For organizations to effectively and efficiently accomplish technological priorities and mandates, IT departments have initially responded by training employees needed to fill these roles. However, employers will not be able to keep pace with the demand for informatics employees since growth is projected to increase somewhere between 27 and 38 percent during the next decade.[1] Furthermore, as technology has become increasingly sophisticated, the level of expertise necessary to assume responsibility for effective integration and operationalization of these information systems has escalated necessitating greater scrutiny as to the existing alternatives for preparing a work force skilled in informatics.

Health care informatics education

In response, a growing number of both nursing and non-nursing graduate degree informatics programs have been developed in an effort to prepare nurses and non-nurses who can confidently and competently transition into health care informatics (HCI) employment.[2] While there is overlap among the fields of HCI, biomedical informatics, clinical informatics, and health information management, the focus of this article is on HCI (sometimes referred to as health or medical informatics). The knowledge base and skill set acquired through an HCI program prepares graduates to apply computer science and IT to health care data. Expertise is developed in managing and communicating data in order to make decisions integral to patient care.[3] Typically, courses such as electronic health records, health care information security, and financial management in health care are required in HCI programs (see Table 1). An essential component of the educational experience requires active learning experiences whereby HCI students are exposed to "real-world" learning through the incorporation of concepts into practice and thereby integration of knowledge, skills, and attitudes toward achieving the essential course competencies. Typically, the HCI program culminates with a "capstone" clinical residency (e.g. HCIN (Health Care Informatics Nursing) 545 in Table 1), that is, an integrative field experience designed to fulfill requirements for completing a graduate level IT health care system project. This application of knowledge is believed to be critical to achieve a higher level learning.

The purpose of this article is to describe a clinical residency project completed by five HCI students in a graduate degree HCI program in southern California, using Donabedian's model as a basis to develop, implement, and evaluate the educational experience. Through this 200-h residency, each student was assigned to a clinical mentor/preceptor in a health care system and was required to apply the knowledge, skills, and attitudes attained through previous courses in order to successfully demonstrate essential program competencies. The evaluation of successful completion of the HCI program competencies was guided by the five learning activities (Table 2) deemed essential for the role of an informatics specialist. These learning activities included (1) educating patients, peers, and/or colleagues about the technology; (2) demonstrating the ability to operate and/or execute the technology device(s) and/or system(s); (3) serving as a consultant regarding the project technology; (4) demonstrating expertise through participation in using the technology; and (5) demonstrating leadership attributes through presentations and/or educational offerings developed in collaboration with other team members. At the culmination of the clinical residency, student evaluations were completed by the clinical mentor, HCI program coordinator, and primary investigator of the research team. Additionally, the student conducted a self-evaluation along with an evaluation of the clinical mentor.

Table 1. Graduate health care informatics program curriculum.

Course name	MS HCI—33 units	MSN HCI 33 units	HCI Certificate 15 to 18 units	Course description
HCIN 540: Introduction to Health Care Information Management	X	X	X	Emphasizes computer hardware, network architecture, clinical application of EHR's, and other software applications, along with regulatory, patient privacy, security and reimbursement issues.
HCIN 541: Introduction to Health Care Delivery Systems*	X		X	Provides overview of health care delivery system, professional roles, care delivery models, and relevant regulatory environment.
HCIN 542: Systems Analysis and Design for Health Care Informatics	X	X	X	Focuses on systems development, project management skills, requirement analysis and specification, feasibility and cost-benefit, logical and physical design, prototyping, and system validation.
MSNC 511: Evidence Based Practice: Role of Theory and Research		X		Explores and critiques the theoretical foundations of nursing science as a basis for the development of research.
ENLC 500: Health Care Leadership, Values and Social Justice	X			Examines leadership theories, corporate ethics, values-focused strategies, and principles of social and health care justice.
HCIN 543: Database Design and Knowledge Management	X	X	X	Applies skills in database design, data structure, modeling, and development within database management systems.
MSNC 512: Influencing the Health Care Environment: Policy and Systems		X		Provides an understanding of nursing's leadership role in health care policy, organization, and finance.
ENLC 557: Strategic Planning and Management of Health Systems	X	X		Emphasizes strategic planning and management as requisite to health care system growth and survival. Includes language, processes, tools and techniques, and marketing.
ENLC 553: Financial Management in Health Systems	X	X		Exploration and evaluation of the financial environment of the health care industry, and how it specifically affects the role of nurses and nurse entrepreneurs.

(Continued)

Table 1. (Continued)

Course name	MS HCI—33 units	MSN HCI 33 units	HCI Certificate 15 to 18 units	Course description
ENLC 556: Management of Health Care System Quality Outcomes and Patient Safety	X	X		Focuses on process of health care delivery from a systems perspective, and emphasizes continuous process improvement as crucial to achieving high-quality outcomes.
MSNC 507: Statistics	X	X		Emphasizes basis of probability concepts and distributions, and inferential statistical methods used in answering research questions. Incorporates SPSS statistical software.
HCIN 544: Advanced Health Care Information Management	X	X	X	Emphasizes EHR and clinical decision support system design, implementation and evaluation. Addresses regulatory, reimbursement, ethical issues and emerging technology.
HCIN 545: Residency in Health Care Informatics Capstone	X	X	X	Provides an integrative field experience to synthesize and apply knowledge attained in the HCIN core courses.

MS HCI: Master of Science in Health Care Informatics; MSN HCI: Master of Science in Nursing—Health Care Informatics; HCIN: Health Care Informatics; MSNC: Master of Science in Nursing Course; ENLC: Executive Nurse Leadership Course; EHR: electronic health record.
Each course is 3.0 academic units.
*Required for individuals having less than 2 years experience in hospital or clinical setting.

Table 2. Structure, process and outcomes of health care informatics (HCI) capstone project.

Structure *Completion of HCI Residency Course Components*	Process	Outcomes

Educate patients, peers, and/or colleagues about technology	Developed a home installation toolkit TH device
	• Loan device agreement
	• Written instructions for use
	• Contact information
	• Extra supplies:
	  • Batteries
	  • Splitters
	  • Phone jacks
	  • Phone cords/cables
	Implemented patient education
	Follow-up phone calls/re-educate on technology use as necessary
	Repeat home visit if unable to resolve problems on phone
		HCI student phoned patient within 24–48 h to schedule date/time for device(s) installation
		Implemented home-based installation and ongoing communications
		Developed trust and rapport between HCI students and participants

Table 2. (Continued)

Structure *Completion of HCI Residency Course Components*	Process	Outcomes
Demonstrate ability to operate and/or execute technology device(s) and/or system(s)	Installed in-home TH device(s) Instructed participant and family member(s)/caregiver(s) with return demonstration	Developed interface between TH device(s) and central processing site (registration site)
Serve as consultant regarding project technology	Provided daily monitoring of patient data with follow-up communication as necessary Participated in weekly 30–60-min team phone meetings focused on enrollment status, biometric data/trends, and troubleshooting issues Provided support with intermittent challenges with battery failure and user issues Revisited home to troubleshoot or replace batteries/device—45 percent of participants needed a second home visit	Evaluated the (devices) technology for: Participant satisfaction Improvements: provided manufacturer with feedback on redesign suggestions (e.g. blood pressure cuff Velcro excess sticky and floor scale extra sensitive to tactile stimulation)
Demonstrate expertise through participation in using technology	Developed quality monitoring process to ensure accuracy of patient data (correct patient with correct data) Monitored daily biometric transmissions Operated TH device and downloading biometric data	Optimized use of resources (devices and personnel) Improved efficiency of TH installation with repetition and practice Maintained integrity and confidentiality of participant biometric data
Demonstrate leadership attributes (e.g. presentations and/or educational offerings) collaboratively developed with other team members	Developed team communication structure with weekly team teleconferences Established secure website for team communication of patient data Facilitated collection of post data (in person or over phone) using study instruments	Participated in the development of a project plan Disseminated structures, process, and outcomes in an educational poster at the university's Scholarly Day (April, 2015)

TH: telehealth; HCI: health care informatics.

HCI residency: telehealth (TH) research study

Specific to this capstone clinical residency project, five HCI students joined a faculty-lead research project to gain an appreciation for conducting research as well as fulfill HCI program competencies. Through activities related to the installation, management, and evaluation of TH services for the study patients inclusive of telemedicine regulations, HCI students were afforded an experiential learning process to determine their readiness for transition to an informatics specialist role. The focus of the research study was to determine the feasibility of using a multimodal technology self-care educational intervention to promote heart failure (HF) patient self-care and reduce hospital readmissions. To meet the requirements of the Institutional Review Board, the HCI students along

with all other team members, completed the National Institute of Health "Protecting Human Research Participants"[4] module. With this educational module, the HCI students gained a fundamental comprehension of the importance of protection of human subjects.

Research study team and mentors

The research team consisted of several academicians, clinicians, and five HCI graduate students. A PhD nurse practitioner was the principal investigator and a lead university faculty for the research study functioned as the primary project mentor for the HCI students with emphasis on the research process. The five bachelor's prepared HCI students had varying professional backgrounds. Two of the HCI students were registered nurses (RNs) with extensive clinical experience, while the other students had backgrounds in business and/or pharmaceuticals. This variety of student educational and experiential backgrounds enriched the overall learning experience for each student frequently resulting in one of the students assuming a mentor role for the other students, one of the many strengths afforded through graduate education.

Donabedian's structure, process, and outcome model

The Donabedian's[5,6] framework of structure, process and outcomes was used to guide this collaborative project and research study. Introduced in 1966, this framework guides understanding and allows for the monitoring of progress throughout the continuum of a project.[7,8] This model was utilized based on the required level of detail needed to consistently and safely implement the home TH intervention.

Structure

Structure refers to the infrastructure and resources needed to achieve HCI residency requirements. Key structural elements for this educational experience were based on utilizing the five "capstone" project learning activities (Table 2). Since several of the HCI students did not have a health care background, it was important that the team review health care compliance and regulations learned in the HCI foundation courses. A secure password-protected website was created for storing the participants' daily biometric data consisting of blood pressure (BP) and weight readings. The monitoring devices used in this study, electronic scales and BP cuffs, were purchased by the researchers and were loaned to the study participant. A telephone modem was required for transmission of data. HCI students learned how to assemble, calibrate, and install devices through instruction from the HCI university professor, reading the handbook, and obtaining telephone tech support provided by the manufacturer's phone assistance staff. Prior to home installation, the HCI students completed pre-testing of the equipment for accuracy and reliability and practiced their teaching skills using researcher-designed scripts. Additionally, the HCI students designed a tool kit for each in-home installation that consisted of documents pertaining to the device loan agreement, installation instructions for both the HCI team and patient, monitoring equipment (i.e. BP cuff, weight scale), modem, and extra supplies (e.g. batteries).

Process

The home TH patient care process has the potential to be complex especially when the device users are older adults. The concept of "process" focuses on what activities are actually completed and a comparison to what was planned.[6] Positive patient and team member outcomes necessitated

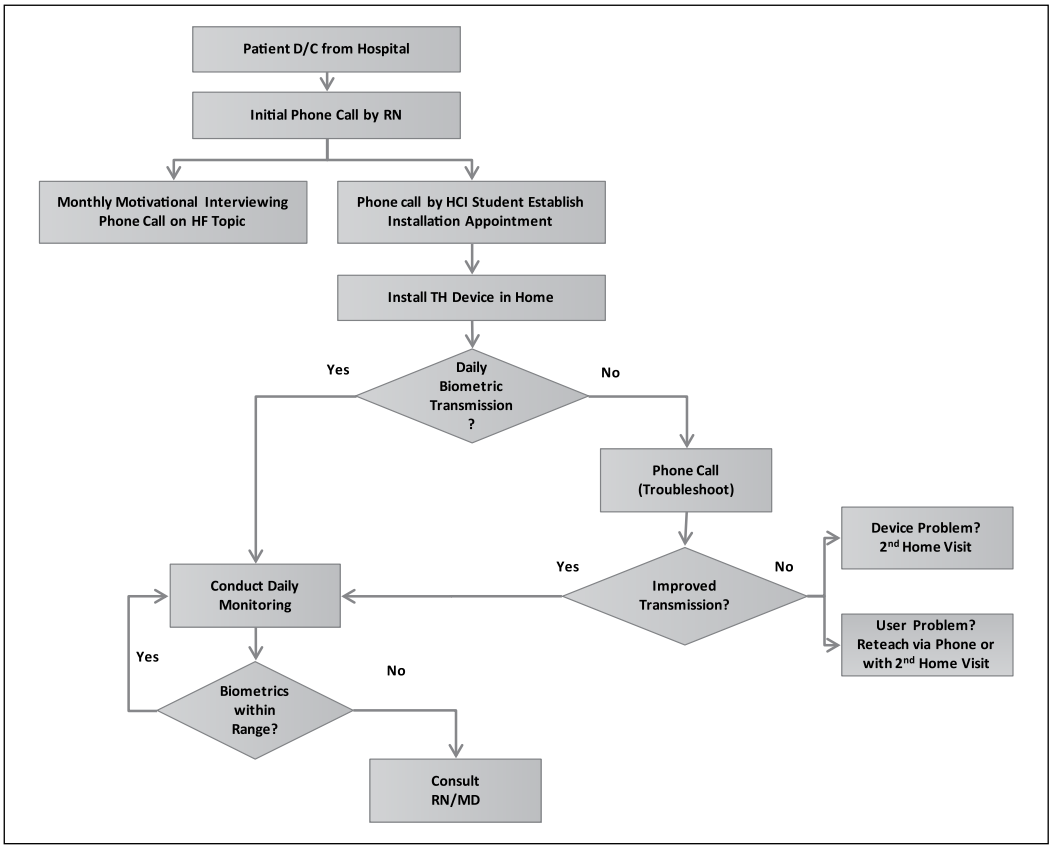

Figure 1. Process for home telehealth device installation and daily monitoring.

coordinated team communication. To achieve high device utilization rates, the installation and patient education are needed to be structured and systematic yet customized for each individual patient/learner. The process implementation phase of this project included a 24-h post-hospital discharge call followed by a HCI student call to schedule home installation of the devices, and monthly scheduled HF self-care nurse coaching sessions by an advanced practice registered nurse (APRN) interventionist. Additionally, the HCI students monitored daily biometric data collection using an American Heart Association algorithm to guide decision-making that was supported by their HCI APRN program coordinator while participating in weekly conference calls with the research team (Figure 1).

Home TH device installation. The HCI students used a researcher-designed protocol to ensure all activities were completed. This process consisted of four phases/steps: (1) identify patients for home TH monitoring, (2) initiate patient contact via telephone, (3) install remote monitoring equipment in patient's home, and (4) troubleshoot, support, and maintain all monitoring equipment and manage data via website (Figure 2).

A team of two HCI students, ideally one RN and one non-RN student completed the in-home installation of the TH device. Initially, the HCI RN's assumed a leadership role with patient interaction while the non-RN HCI team member initiated the equipment set-up. Each subsequent

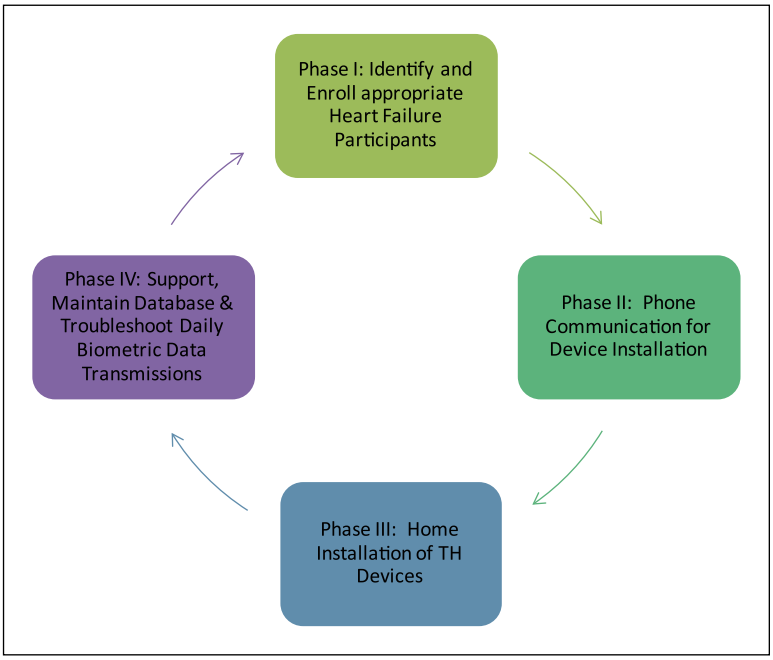

Figure 2. Home teleheath device installation and maintenance.

installation afforded each of the students to become comfortable with either of the two in-home installation roles. Follow-up phone calls and occasional re-visits to the home primarily due to data transmission issues were necessary. For example, one patient experienced functional issues with one of the devices that necessitated one of the HCI students to contact the device manufacturer and eventually secure a replacement. Ultimately, it was important that all HCI student members were comfortable and confident with engaging in participant communication, TH device installation, troubleshooting the equipment, and utilizing a teach-back method of educating participants on use of equipment.

Team leadership and communication. As participants were enrolled in the home TH device group of the study, the HCI student team registered each patient's demographics into the data collection website. One of the two HCI students who installed the home device took a "lead" role in assessing daily biometric data and providing updates to the team (Figure 1). On a weekly basis, 30- to 60-min team conference calls were conducted and continued throughout the 11 months of the study. Conference topics included information on recruitment updates, participant adherence to daily monitoring, data transmission issues, biometric data trends, along with summaries of communication between research team members and the participants. These meetings were highly valuable to the successful structure, process, and outcomes of the experience.

Outcomes

As a result of this collaborative research project, the HCI students completed multiple "process" activities that yielded "outcomes" demonstrating fulfillment of the five "structure"-learning activities. Through weekly prerequisite teleconferences, the mentors and the students debriefed and

monitored current processes and made necessary adjustments throughout the duration of the study. As a result of these ongoing weekly debriefings, the five HCI students consistently demonstrated progress toward fulfilling the five "structure"-learning activities. In addition to completing capstone and course competency requirements, the HCI students learned how to foster trust and develop a close rapport with study personnel and HF participants. The students developed a further appreciation of the research process, the importance of teamwork, and use of effective communication skills. Over time and with repetition, the HCI students increased their confidence and competence with home installations and troubleshooting malfunctions. The HCI students also recognized and engaged in maintaining the integrity and confidentiality of participant biometric data through the secure web site.

Implications for future HCI collaborative learning. Based on evaluations from faculty, mentors, and students, there was consensus that the outcomes of this educational experience supported a positive learning experience for the HCI students. As with all teaching/learning methodologies, especially those that are innovative and novel, there are valuable "lessons" learned from the educational process. The following topics are a detailed description of observations and insights shared by the HCI students and research team colleagues during the HCI residency project:

- *Participant engagement with timely installation.* Timely installation of the home TH device was an important variable to ensure that the participants maintained an interest in the program. Phone calls were made within 24–48 h after hospital discharge. Occasional delays in communication resulted from participants who did not move home immediately (e.g. lived with a family member or in a skilled nursing facility) prior to living at home. Delays with installation resulted in some participants dropping out of the study.
- *Team member role clarity.* With numerous team members participating in the study it was important to establish clarity of roles specific to device installation, data tracking, and ongoing communication with the study participant. The HCI student who installed the home device was assigned to be the "lead" HCI person for the participant. This role consisted of making the 24-h follow-up call and addressing all user or technical issues that occurred during the participant's enrollment in the study.
- *Family/caregiver involvement.* Family dynamics, inclusive of caregiver involvement varied within each household. Some participants refused family engagement, others recognized and agreed to family support with the TH devices, and there were several participants who openly communicated that their preference was to have one or more family members assume total responsibility for the equipment. To ensure consistency in the use of the devices, it was essential that family members involved in the patient's care be present during installation of the TH equipment. In addition, it was necessary to determine if the point of contact for communication was the patient or the family member.
- *Home environment challenges.* By entering the patients' homes, the HCI students experienced the patients' living conditions first hand. The home environment varied among all participants and each one provided unique installation challenges. For example, the location of the phone jack bluetooth modem installation site was sometimes distant from the necessary location of the monitoring devices resulting in connection challenges. The use of extra long phone cords placing the modem closer to the scale remedied this problem. In many cases, the phone jacks in the home were not functioning. This was resolved by using a splitter on the main phone line. If the participant, or a family member, used the landline telephone extension while the monitor was interfacing with the modem, the data would not transmit to the website. This necessitated that the participant repeat their BP and weight

readings. Participants living in mobile home communities often experienced interference with other nearby bluetooth devices. In homes where modem Internet connections were used, transmission of data was better than with the telephone jack. However, in one home, strict firewall security interfered with transmission of data and adjustments were made to the patient's firewall settings.

- *Equipment challenges.* Various equipment challenges were encountered; many were due to the overall design and functionality of the specific device. For example, participants had a difficult time reading the liquid crystal display's (LCD) blue font on the weight scale, and the numbers were too small to read for some participants. The "sensitivity" of the equipment was a common problem. The scale was very "sensitive" to bumping or tactile stimulation while the BP monitor was very "sensitive" to movement/motion. If the participant moved during the inflation phase, the cuff would re-inflate resulting in discomfort to the participant. At times, the cuff re-inflated several times during one reading. If this occurred the participant was instructed to turn the monitor off and repeat the reading. The BP cuff also had extremely strong Velcro creating application and removal difficulty. Sizing the cuff to the arm and then sliding the cuff on and off the arm for each reading resolved this issue. Communication with the manufacturer was provided with subsequent improvements made in the design of similar future devices.
- *Vendor website challenges.* As each TH participant was enrolled in the study, an HCI student registered the patient's demographics (patient initials and contact information) and device serial numbers into the vendor website. The device serial numbers interfaced with each patient's bluetooth modem enabling data to upload into the website. On a couple occasions, data interface challenges occurred when a device was deactivated and used for a subsequent participant. Existing data from the previous participant did not automatically clear resulting in confusion and an overlap of two participant data sets. Equipment serial numbers needed to be erased and then re-entered before using the device on another participant. A second challenge faced was the consistency of the vendor's web site. On occasion, the manufacturer's web site unexpectedly went off-line, resulting in the inability to view the participant's data for a day or two. In each of these occurrences, the vendor was readily available to troubleshoot and correct any issues that were encountered.
- *Working with older adults.* The average age of the HF participant was 79 years. Some participants were quite frail, lived alone or with an elderly spouse. Working with this population enlightened the HCI students with an increased awareness of the daily challenges faced by elderly patients and their families in managing a chronic disease. Establishing a trusting rapport with each participant by incorporating patience and compassion was crucial to ensure success of this relationship and study. Sometimes, this population is reluctant to use new technology and their perception of technology is quite varied. Participants often needed to be retrained and required additional support with simple tasks such as changing batteries. Some of the participants were intimidated by the equipment and needed positive feedback to continue. The 24-h follow-up call after home installation was deemed to be crucial to reiterate the use of the monitoring equipment and provide needed encouragement. Even though each participant demonstrated understanding of the use of the monitors during the onsite installation, the participant was challenged the following morning when confronted with using the equipment alone since essential information for using the devices had been forgotten. To provide support, each participant was given a contact phone number and was reassured that they could call the technology team at any time with questions or concerns about the equipment.

Conclusion

This capstone clinical residency project provided a unique opportunity for HCI students to gain real-time experience in the use of home TH with a group of HF patients while fulfilling course requirements necessary for graduation. Using Donabedian's[5,6] framework of structure, process, and outcomes, the HCI students utilized knowledge and skills acquired throughout the HCI program to engage in activities demonstrating competency as an informatics specialist. Furthermore, the HCI students gained insight and understanding of conducting research including recruitment, the informed consent process, home installation of the TH devices, monitoring of biometric data, and the ongoing communication with the TH participants over the duration of this 4-month home TH intervention designed to promote a positive patient care experience. Ultimately, the students reaped the rewards and satisfaction of developing greater self-confidence related to the roles of informatics specialist, researcher, and patient care manager. HCI educational leaders are encouraged to replicate or modify this template of *applied learning* to facilitate the application of curriculum education (i.e. theory) into practice.

Acknowledgements

The authors wish to thank the following people: Carl Fritz Steen, MBA, BSN, CDE, CCP; Alex Boreen, MS, BA; Leila Cavazos, BA; Erika Turner, MSN, RN; Lyndhell Vivo, MS; Jonathan Mack, PhD, ANP-BC.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: University of San Diego School of Nursing Faculty Research Incentive Grants.

References

1. Department of Health and Human Services and Centers for Medicare and Medicaid Services. Medicaid and EHR incentive program, https://www.federalregister.gov/articles/2012/09/04/2012-21050/medicare-and-medicaid-programs-electronic-health-record-incentive-program-stage-2 (4 September 2012, accessed 4 May 2015).
2. United States Department of Labor. Bureau of Labor Statistics: employment projections, 2015, http://www.bls.gov/emp/ep_table_102.htm (accessed 16 June 2015).
3. American Nurses Association (ANA). *Scope and standards of nursing informatics practice*. 2nd ed. Washington, DC: ANA Publishing, 2014.
4. National Institute of Health—Protecting Human Research Participants, http://phrp.nihtraining.com (2008, accessed 1 June 2015).
5. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966; 44(Suppl.): 166–206.
6. Donabedian A. Methods for deriving criteria for assessing the quality of medical care. *Med Care Rev* 1980; 37(7): 653–698.
7. Gardner G, Gardner A and O'Connell J. Using the Donabedian framework to examine the quality and safety of nursing service innovation. *J Clin Nurs* 2014; 23: 145–155.
8. Ing-Long W and Pi-Jung H. Understanding hospital innovation enabled customer-perceived quality of structure, process, and outcome care. *Total Qual Manag Bus* 2011; 22(2): 227–241.