Physician Documentation Quality: Completeness and Readability of EMR versus Paper.

Paul E Manicone, MD[1,3], Zev Waldman, MD[1], Jeremy R Kern, MD[1], Matthew Sharron, MD[2], Sonali Basu, MD[2], Jie M Goh, PhD[4], Guodong Gao, PhD[4], Ritu Agarwal, PhD[4] and Brian R Jacobs, MD[2,3].

[1]Pediatric Hospital Medicine Division, Children's National Medical Center, Washington, DC, [2]Critical Care Medicine, Children's National Medical Center, Washington DC; [3]Center for Pediatric Informatics, Children's National Medical Center, Washington DC; [4]School of Business, University of Maryland, College Park, Maryland

INTRODUCTION

Adoption of the Electronic Health Record has become a national priority. Computerized physician documentation (CPD) has been available for decades, yet recent data suggest that inpatient adoption rates are less than 2%. The overall quality of commercial CPD has been poorly studied. There is no validated, reliable tool to objectively measure quality of clinical documentation in the medical literature. Most studies to date have focused on subjective assessments of implementation, user satisfaction and perceptions of accuracy. Studies with objective measures of quality are lacking. As a result, electronic medical record (EMR) documentation widely varies at the vendor level, institutional level and user level.

We sought to construct a robust scoring tool that included both the essential features of documentation compliance (completeness) and that of documentation quality (readability).

OBJECTIVE

To compare the quality of CPD to paper-based documentation at a pediatric tertiary care medical center.

METHODS

This study had a retrospective cohort design and was performed at a large urban tertiary care pediatric medical center. The study was conducted using an objective 16 point scoring tool developed to allow comparison of computer generated versus paper-based inpatient notes in regards to completeness and readability. Scoring was done from randomly selected paper documents scanned into the medical record in May of 2008 and from Electronic notes in May to October of 2008.

Scoring; 1 point was given for every element present in the *completeness score* (max completeness score=8). Elements of the *readability score* received 0, 1 or 2 points (max readability score=8). Readability scoring included specific definitions and instructions for each element (see appendix item 1). The sum of these two scores formed the *Overall Note Quality Score* (max score=16).

The *Completeness Score* contained 8 variables: Date, time subjective report, vital signs, objective report, assessment, plan and identifiable author .

The *Readability Score* contained 4 variables: Legibility, supportive data, organization, and clinical reasoning.

Physician documents were compared from three time periods:

1. The *paper period* (1 month prior to EMR implementation, May 2008)
2. the *transitional electronic period* (the first 8 weeks after implementation, July 2008)
3. the *established electronic period* (6 months after EMR implementation, October 2008)

Physician documents included in this study were *History & Physicals*, *Progress notes*, and *Consultation notes*. Data was collected to identify the type of note, location, service, the type of physician authoring the note, and degree of clutter present in the note. Each physician scorer was given 75 encounter numbers which linked to a specific patient medical record and a specific inpatient encounter. The scorer collected data on only one note per encounter. Scorers did not know whether they were opening an EMR document or a paper document until they entered this encounter number in the computer. The system then directed them to either the EMR or the digitally scanned paper documentation. Prior to opening an encounter, the scorer was committed to scoring a specific type of note.

Clutter was categorized into three categories: Clutter free (less than 5% of the note's content), Moderate Clutter (5-20% of the note's content) and Severe Clutter (greater than 20% of the note's content).

Inter-rater validation and reliability testing was conducted between 6 physician scorers.

SPSS v13 was used for analysis. The students t-test was used to compare mean completeness and readability scores during each study period. Subgroup analysis was performed to examine differences by location, service, type of note and physician. Significance was defined as p value <0.05. This study received approval by our institutional review board.

RESULTS

Validation testing revealed a 93% agreement. Pre-study reliability testing resulted in a mean Kappa score of 0.71.

A total of 411 notes were scored (144 paper notes, 125 transitional electronic notes, and 142 established electronic notes). Paper and electronic note group characteristics were similarly represented (see table 1).

Table 1: Description of physician notes scored by time period

		Paper Period	**Transitional EMR period**	**Established EMR period**	**Total**
Type of Note	H&P	53	48	54	155
	Progress Note	45	36	43	124
	Consult Note	43	41	45	129
Type of Service	Medical Service	95	74	91	260
	Surgical Service	37	37	40	114
	ICU Service	12	14	11	37
Type of Physician	House Staff Physician	19	24	26	69
	Attending Physician	26	15	24	65
	House Staff & Attending	67	80	84	231

Group analysis:

Mean Overall Note Quality score was 13 in the Paper Period (PP) and 14.5 in the Electronic Period (EP) ($p<0.001$). (Figure 1)

Mean Completeness score was 6.9 in the PP and 7.6 in the EP ($p<0.001$) (Figure 1)

Mean readability score was 6.1 in the PP and 6.9 in the EP ($p<0.001$) (Figure 1)

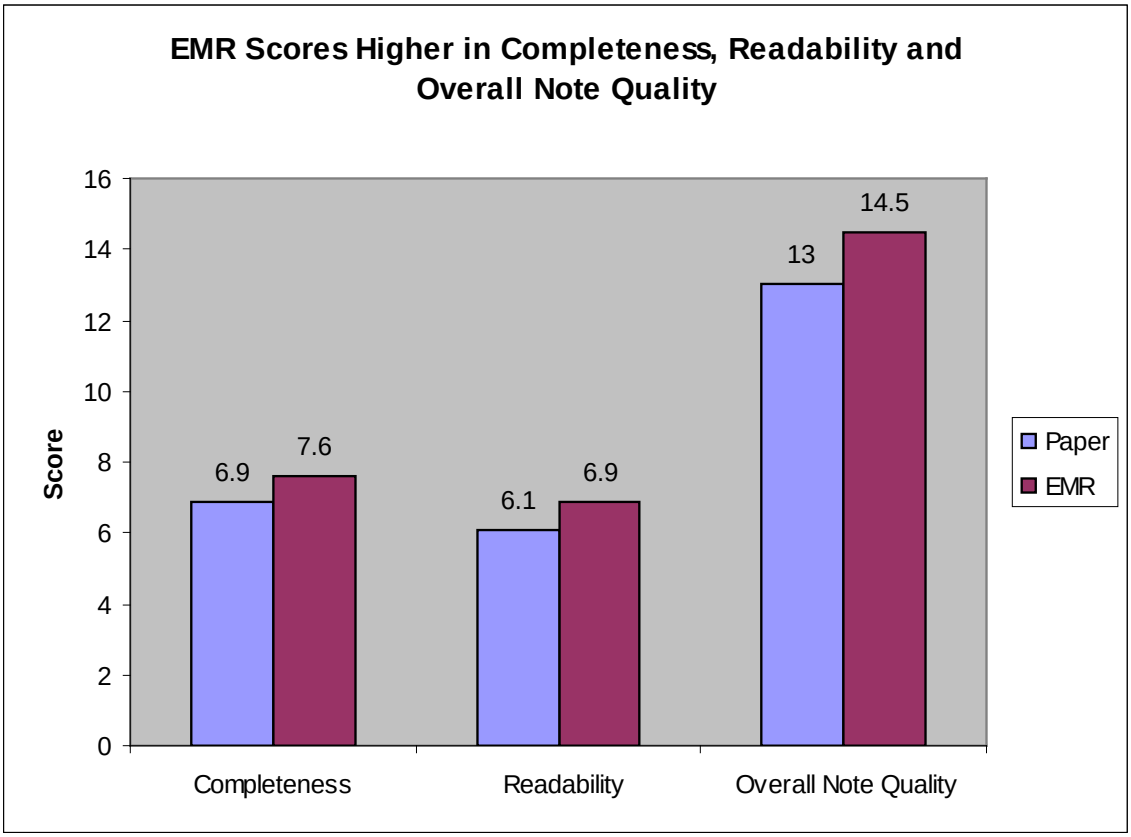


FIGURE 1

Subgroup Analysis:
Overall note quality scores by *Type of Note* were significantly higher in the EP for H&Ps, and Consultation Notes. ($p=0.008$ and $p<0.001$ respectively.). Progress notes were not statistically higher in the paper and EMR ($p=0.083$) (Figure 2.)

Overall note quality scores by *Location* were significantly higher in the EP for Floor and ICU notes. ($p<0.001$ and $p<0.001$ respectively) (Figure 3)

Overall note quality scores by *Author* were significantly higher in the EP for Combination notes (Attending plus house staff). ($p=0.03$). House staff only notes and Attending physician only notes were not significantly improved in the EP ($p=0.299$ and $p=0.294$ respectively) (Figure 4)

Overall note quality scores by *Service* were significantly higher in the EP for Medical services. Surgical services and ICU service notes were not significantly improved ($p=0.12$ and 0.2 respectively). (Figure 5)

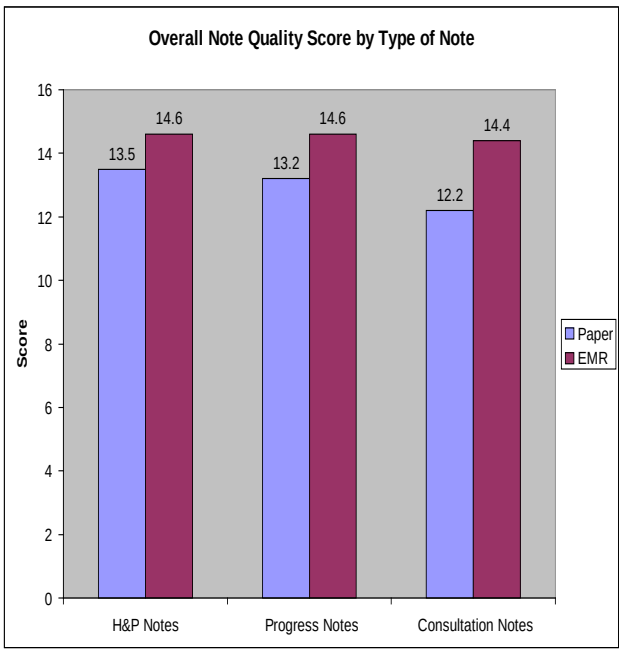


Figure 2

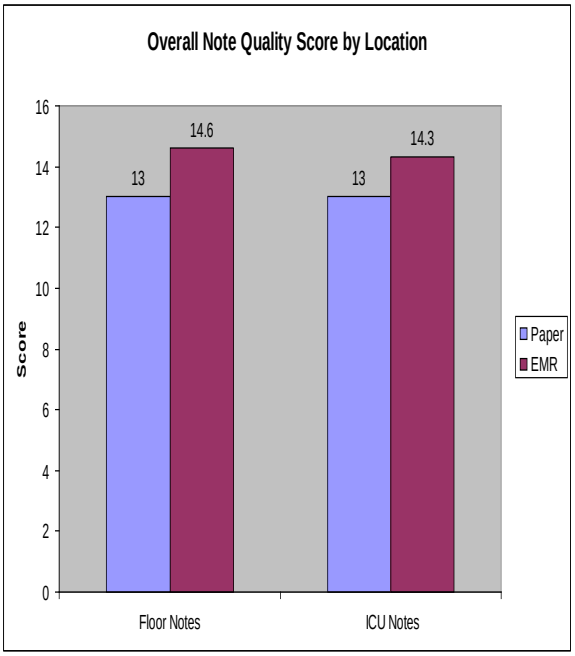


Figure 3

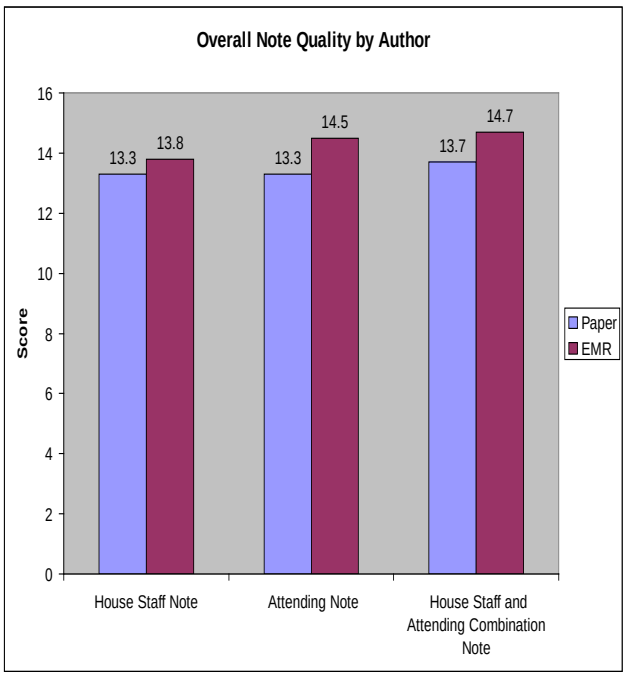


Figure 4

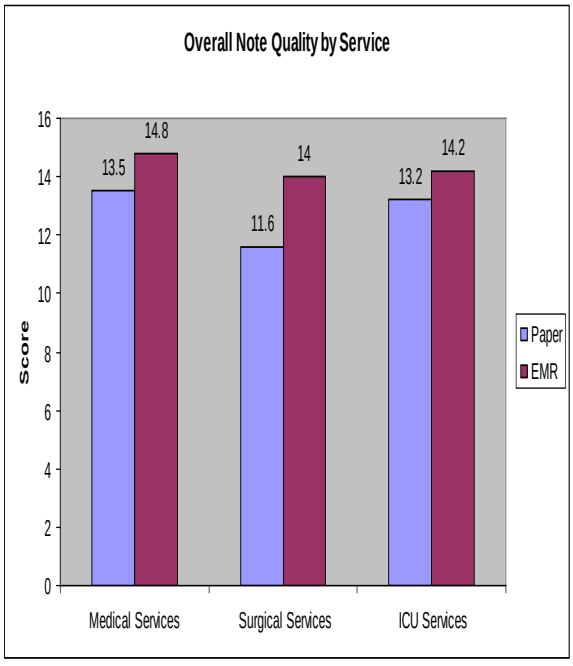


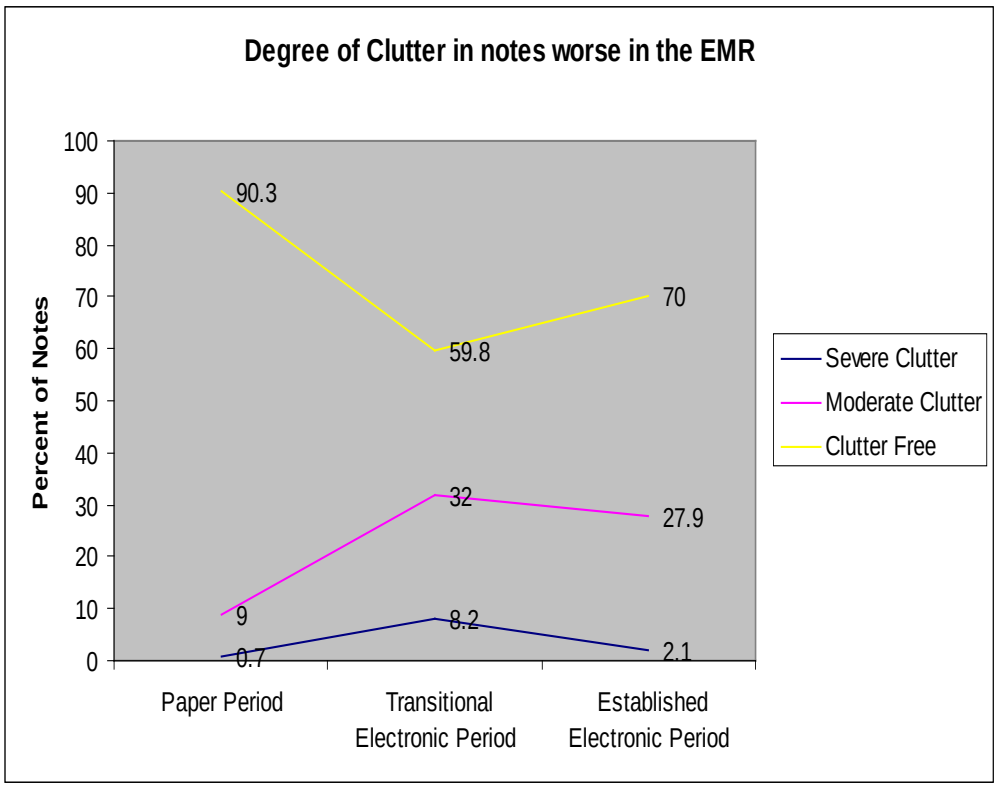
Figure 5

Clutter:

Severe clutter was significantly worse in the EP compared to the PP. Severe clutter was present in 0.7% of Paper notes, 8.2 % of transitional period electronic notes, and 2.1% of established period electronic notes. (Figure 6)

Moderate Clutter was significantly worse in the EP compared to the PP. Moderate clutter was present in 9% of Paper notes, 32% of transitional period electronic notes, and 27.9% of established period electronic notes. (Figure 6)

Clutter Free notes were significantly worse in the EP compared to the PP. Clutter free notes were present in 90.3% of Paper notes, 59.8% of transitional period electronic notes, and 70% of established period electronic notes. (Figure 6)



DISCUSSION

The overall note quality score demonstrated some gains that were expected and some that were not expected. Initially we hypothesized that the EMR would be superior in the area of completeness but inferior in the area of readability. Our observations showed gains in both components of our scoring tool. The EMR's inherent characteristics allow for obvious completeness gains given the automated nature of specific scoring elements such as 'identifiable author' and 'time' and 'date'. The completeness portion of the scoring tool was binomial and subject to less variation between raters. The readability portion of the scoring tool ranged from 0 to 2. This portion of the tool required very specific definitions and scoring instructions in order to achieve an acceptable validity and reliability test. These definitions are provided in the appendix below. These definitions were composed by the multidisciplinary consensus of our authors.

The ONQS was superior in the EMR compared to the paper environment in the areas of both completeness and readability. The readability gains were greater in all 4 elements. As expected, legibility was higher in the EMR given the inherent characteristics of the electronic environment. Clinical reasoning however was also superior in the EMR. Our hypothesis was that this area would be inferior to the paper environment. This unexpected gain may have been in part due to the *composite*

note method of our system (where physician documentation may include contributions from house staff officers and attending physicians), however this only represents a change in the Progress note because H&Ps were consultation notes used the composite method in the paper environment.

Subgroup analysis showed gains in overall note quality in the EMR with wide dissemination when we looked at differences by type of note , type of author, location, and, type of service. Though many subgroups demonstrated gains, those which were not significant were limited by the size of the data set (House staff only notes, and ICU service notes, and Surgical service notes). ICU as a location did demonstrate significant gains which is in part due to a larger data set given that multiple services were included in this data set. The ONQS of Progress notes did not show a statistically significant difference though a trend towards significance was observed (p=0.083). Of note, the largest improvement in mean scores from the paper environment to the EMR were observed in the surgical services and consultation notes (2.4 and 2.2 respectively).

Clutter has been a significant concern in the EMR. Auto-population functions allow for extraneous or less relevant information to not only lengthen a note but at times create reader confusion or disinterest. This was estimated based on the number of apparently extraneous information with respect to the central thrust of the problems stated in the assessment and plan. Severe clutter increased 10 fold and moderate clutter increased more than 3 fold upon implementation of the EMR. 6 months post implementation, severe and moderate clutter improved but still remained elevated as compared the paper environment. This suggests that clutter is an additive problem with contributions from both the user and the system.

Limitations of our study are those inherent to a single center, retrospective cohort study. Efforts to limit bias were achieved through case blinding, pre-selection of note type prior to opening of the chart, standardizing progress note selection as hospital day one, and requiring each scorer to collect data on equal numbers of H&Ps, progress notes, and consultation notes.

This study was also limited by the scoring of digitally scanned paper documents rather than the actual hard copy. Accessing the archives of hard copies would have made the data collection impractical and would have negatively impacted the bias of scoring a paper document.

Finally an important limitation of this study was the absence of *accuracy* in our scoring tool. We acknowledge that accuracy is a fundamental goal of physician documentation and accordingly, it should be included in a scoring tool for quality. A single center, retrospective study however, is not able to approach this particular element in a sensible way. A prospective study with interviews between unbiased physician scorers and the physician authors themselves is required. In addition, a study of this nature would require a multi-center effort where the paper environment was in use at one location and the EMR was in use at another.

CONCLUSIONS

In this study we demonstrated that measuring physician documentation for quality can be reliably performed using the *Overall Note Quality Scoring Tool*. In addition, we observed significant gains of the EMR over the paper medical record in *overall note quality, completeness and readability.* The highest note quality scores are achieved when notes are *co-authored* by house staff and attending physicians. Lastly we reported the problem of clutter. Although the EMR contains more *clutter* than paper notes, significant improvements were made in this area as users became more proficient with the EMR.

We believe these results warrant further study with a more granular scope into the strengths and weaknesses of EMR documentation. It is clear that the EMR adoption will increase. What is less clear is how evidence based *best practices* for physician documentation can be established and adopted by commercial vendors and health care systems to ensure that physician documentation is complete, readable and accurate.

APPENDIX

ITEM 1: Scoring definitions for Readability.

Readability Score: 0, 1 or 2 points for each of the following (max score = 8)

Legibility:

0: Mostly Illegible: More than 20% words illegible per page
1: In Between: >5 and < 20% words illegible per page
2: Mostly Legible: Less than 5% illegible words per page

Data Presentation:

Paper Note:

0: No data (labs, meds, problems, imaging reports, etc.) included in the note to support the diagnoses
1: Data is present, but no attention to any particular elements
2: Data is presented in summarized fashion (e.g.: conventional lab work grids) and/or attention is brought to relevant elements. Score 2 if presented in the corresponding system or problem-based section of the note (i.e. electrolyte data in the FEN section)

Electronic Note:

0: No data (labs, meds, problems, imaging, etc.) included in the note to support the diagnosis.
1: Auto-populated data (labs, meds, problems, imaging reports, etc.) **and** no attention to relevant elements.
2: Presented in summarized fashion (e.g.: conventional lab grids or truncated auto-populated data) and/or attention is brought to relevant elements. Score 2 if relevant data is in the corresponding system or problem-based section of the note (i.e. electrolyte data in the FEN section)

Organization:

0: Information is out of conventional order
1: Generally in order with small deviations from convention
2: Orderly SOAP note or template used and followed.

Clinical Reasoning:

0: No written assessment **or** explanation of plan. Score 0 if only assessment is template based statements (i.e. 'progressing as expected).
1: Written assessment **or** explanation of plan is present, but not both
2: Both Written assessment and explanation of plan present.

REFERENCES:

1. "Implementation of a computerized clinical documentation system (CCDS) in a tertiary care pediatric medical center; a longitudinal survey of users". Paul Manicone, MD et al. 2009 PAS abstract citation #754045

2. Comparison of handheld computer assisted and conventional paper chart documentation of medical records: a randomized, controlled trial. Stengel, D MD; Bauwens, K MD, Walter, M MD, Kopfer, T MD, Ekkernkamp, A MD PhD, Journal of Bone & Joint Surgery - American Volume. 86-A(3):553-560, March 2004.

3. An Observational Study of Anesthesia Record Completeness Using an Anesthesia Information Management System. Driscoll, William D. MA; Columbia, Mary Ann RN; Peterfreund, Robert A. MD, PhD Anesthesia & Analgesia. 104(6):1454-1461, June 2007.

4. Effects of Billing and Documentation Requirements on the Quantity and Quality of Teaching by Attending Physicians. McConville, John F. MD; Rubin, David T. MD; Humphrey, Holly MD; Carson, Shannon S. MD Academic Medicine. 76(11):1144-1147, November 2001.

5. Health Information on the Internet: Accessibility, Quality, and Readability in English and Spanish. Berland, G MD; Elliott, M N. PhD; Morales, L S. MD, PhD; Algazy, J I. MD, MPH; Kravitz, R L. MD, MSPH; Broder, M S. MD; Kanouse, D E. PhD; Munoz, J A. MS, ; Puyol, Juan-Antonio; L, Marielena MD, MPH; Watkins, K E. MD, MSHS; Yang, H; McGlynn, E A. PhD JAMA. 285(20):2612-2621, May 23/30, 2001.

6. Two formulas for determining the readability of subject consent forms. Grundner, T. M. American Psychologist. 33(8):773-775, August 1978

AUTHOR BIOGRAPHIES:

Paul E. Manicone, MD is the associate chief in the Division of Pediatric Hospital Medicine at Children's National Medical Center and Assistant Professor of Pediatrics at the George Washington University Medical Center. He has been a clinical investigator with the Center for Pediatric Informatics at Children's National for 3 years with a primary interest in the transformation to an Electronic Medical Record. He has presented his work at national conferences in the areas of EMR implementation, satisfaction, workflow efficiency, and documentation quality.

Ritu Agarwal is Professor and the Robert H. Smith Dean's Chair of Information Systems at the Robert H. Smith School of Business, University of Maryland, College Park. She is also the founder and Director of the Center for Health Information and Decision Systems. Dr. Agarwal studies the performance impacts of health IT, business models for health information exchange, and the challenges of health IT adoption and implementation. She has published over 100 scientific papers and made presentations at a variety of national and international conferences. She is currently Editor-in-Chief of Information Systems Research.

Sonali Basu, MD is an attending physician in the Division of Pediatric Critical Care at Children's National Medical Center and Assistant Professor of Pediatrics at the George Washington University Medical Center. Her research interests include echocardiographic analysis of myocardial dysfunction in septic shock and resident education development in the pediatric intensive care unit

Goudong (Gordon) Gao, PhD is an Assistant Professor at the R.H. Smith School, University of Maryland, College Park. Professor Gao received his Ph.D. in Operations and Information Management from the Wharton School of the University of Pennsylvania. His research interests include business value of IT and information asymmetry problems in healthcare services. His research has been published or accepted in leading journals like *Management Sciences, Information Systems Research*, *M&SOM*, and *Journal of Management Information Systems.*

Jie Mein Goh, PhD is an Assistant Professor of Information Systems at the IE Business School in Madrid, Spain. Her primary research interest is in the area of technology enabled impacts in health care. She has published her work in various journals such as Information Systems Research and has presented her research in many leading conferences.

Brian Jacobs, MD is the chief medical information officer at Children's National Medical Center, is an attending physician in the Division of Pediatric Critical Care Medicine at Children's National Medical Center, and is a professor of pediatrics at the George Washington University Medical Center. He has published extensively, presented nationally and internationally, and has led the transformation of Children's National's EMR from 2005 to present.

Jeremy R. Kern, MD is a hospitalist in the Division of Pediatric Hospital Medicine at Children's National Medical Center and Assistant Professor of Pediatrics at the George Washington University Medical Center. His current research interests include examing the processes involved in STAT medication orders and the effect on patient outcomes.

Matthew Sharron, MD is an attending physician in the Division of Critical Care Medicine at Children's National Medical Center and Assistant Professor of Pediatrics at the George Washington University Medical Center. He is currently an investigator in the Children's Research Institute at Children's National Medical Center examining the role of platelets and anti-platelet compounds in sepsis and multi-organ dysfunction.

Zev Waldman, MD, MPP, is the Chief Medical Officer of the HSC Pediatric Center in Washington, DC, and Assistant Professor of Pediatrics at the George Washington University Medical Center. He has also been a member of the Division of Pediatric Hospital Medicine at Children's National Medical Center since 2006 and completed the American Medical Informatics Association "10x10" program in 2009. His relevant clinical and research interests include diagnostic decision support systems, electronic documentation quality, and the effect of the electronic health record on medical education.