Refereed papers

Measuring the impact of different brands of computer systems on the clinical consultation: a pilot study

Charlotte Refsum MA
Medical student

Pushpa Kumarapeli BSc (Hons)
PhD student

Aruni Gunaratne MBBS BSc
F2 doctor

Richard Dodds MBBS BSc (Hons)
F2 doctor

Ali Hasan MBBS BSc
F2 doctor

Simon de Lusignan MSc MD FRCGP FBCS CITP
Reader

Primary Care Informatics, Division of Community Health Sciences, St George’s, University of London, London, UK

ABSTRACT

Background UK general practitioners largely conduct computer-mediated consultations. Although historically there were many small general practice (GP) computer suppliers there are now around five widely used electronic patient record (EPR) systems. A new method has been developed for assessing the impact of the computer on doctor–patient interaction through detailed observation of the consultation and computer use.

Objective To pilot the latest version of a method to measure the difference in coding and prescribing times on two different brands of general practice EPR system.

Method We compared two GP EPR systems by observing use in real life consultations. Three video cameras recorded the consultation and screen capture software recorded computer activity. We piloted semi-automated user action recording (UAR) software to record mouse and keyboard use, to overcome limitations in manual measurement. Six trained raters analysed the videos using data capture software to measure the doctor–patient–computer interactions; we used interclass correlation coefficients (ICC) to measure reliability.

Results Raters demonstrated high inter-rater reliability for verbal interactions and prescribing (ICC 0.74 to 0.99), but for measures of computer use they were not reliable. We used UAR to capture computer use and found it more reliable. Coded data entry time varied between the systems: 6.8 compared with 11.5 seconds ($P = 0.006$). However, the EPR with the shortest coding time had a longer prescribing time: 27.5 compared with 23.7 seconds ($P = 0.64$).

Conclusion This methodological development improves the reliability of our method for measuring the impact of different computer systems on the GP consultation. UAR added more objectivity to the observation of doctor–computer interactions. If larger studies were to reproduce the differences between computer systems demonstrated in this pilot it might be possible to make objective comparisons between systems.

Keywords: clinical consultation, computer system, user action recording
Introduction

Computerisation has improved the efficiency of consultation tasks, however, the relationship between the tasks and the computer system features remains poorly investigated. Routine data recording in electronic patient records (EPRs) has become comprehensive. Computerised prescribing reduces errors, saves time in repeat prescribing and improves patient safety. Computer systems provide efficient mechanisms for clinical data recording. Financially incentivised disease management targets have improved completeness of clinical data coding and have optimised the chronic disease management role. Computer systems make maintenance of disease registers, transferring and linking of laboratory test results and ability to perform clinical audits much easier than on manual systems. Investigation into the consultation process has mostly focused on aspects of doctor–patient communication, therapeutic process, prevention and knowledge management aspects or adding new tasks, such as online referral to secondary care. However, using the EPR in the consultation has a time cost; our previous pilot study suggested that general practitioners spend 25% of consultation time using the computer.

Few tools exist for assessing the differences between the various EPR system interfaces. GP computer systems in the UK have developed organically; different systems evolved in isolation with diverse interface features. Early GP clinical systems were developed by enthusiasts; they initially focused on automating the tasks that were labour intensive to carry out with written records, for example repeat prescribing. A combination of the introduction of computer system accreditation and functional requirement demands led to large number of small system developers being replaced by a few major clinical system suppliers. Ninety percent of the UK primary care clinical system supply market is shared between EMIS, InPractice and iSoft. Each supplier has released a number of versions. The result is clinical systems with diverse functional and interface features, with dissimilar levels of impact on consultation tasks.

We carried out this study to pilot a further development in our method for observing the impact of the computer on the consultation. We did this by investigating the differences between the impacts of two brands of GP clinical systems in carrying out two common computer-mediated consultation tasks; coding of a problem title and prescribing.

Method

Introduction

Our method consists of three parts:

1. First we tested the rating method (using an observational data capture (ODC) tool called ObsWin) that we developed in our previous study. Previously we used simulated patients in live consultations for the first time, but we expected real consultations to be much more heterogeneous and needed to know if the rating method remained reliable.

2. Second, we introduced a new method of semi-automating the measurement of computer use – we called this UAR.

3. Finally, we piloted the new method (incorporating UAR) to compare two common activities in the consultation – recording the problem title and prescribing.

Reliability when rating live consultations

We used series of real life consultations observed using the multi-channel video recording method. This has four channels of recording: three video cameras and computer screen capture software. The first and second cameras focused on the front of the patient’s and doctor’s upper bodies respectively to identify the direction of gaze and the body language. The third camera filmed a wide angle view of the consultation; the conventional single camera setting observing doctor, patient and computer. We used ‘Camtasia’ software to record the doctor’s use of computer system features during the consultation. The final multi-channel footage is produced by combining and synchronizing these video channels into one so that they can be viewed simultaneously.

A group of six raters analysed the consultation videos to measure the doctor–patient interactions and doctor’s use of the computer. We used ‘ObsWin’, a software package for ODC, to view and rate the multichannel videos. Raters could keep track of the occurrence and duration of key events by using designated keys on the keyboard. We used clinical medical student volunteers as raters. Each rater was asked to go through an instruction manual followed by practical training sessions. The raters observed and timed 13 different general practitioner–computer interactions identified in a previous pilot study. Raters watched each consultation four times. The first viewing was for
Measuring the impact of different brands of computer systems on the clinical consultation

Raters to become familiar with the consultation activities without carrying out any rating activity. On the second viewing raters recorded variables related to the interactions between the doctor, the patient and the computer. On the third viewing, more specific aspects of doctor and computer interactions were recorded. The final observation run measured the verbal and non-verbal interactions between the doctor and the patient (Figure 1).

We used two consultations from each system to measure the inter-rater reliability for the selected set of 13 variables. We looked at the maximum and minimum values assigned to each variable by raters in each consultation, together with their median and the inter-quartile range. These statistics indicated how close the raters’ observations were for each consultation. Then the correlation of their duration values was calculated based on the interclass correlation coefficients (ICC). ICC measures the variability among ratings for each variable, compared to the total variation across all ratings and variables.

UAR: a new method for capturing computer use

We developed in-house a new tool to capture computer use which we called the user action recorder (UAR). It records the keyboard key presses indicating character, navigation or control keys. A mouse recorder element creates a log of the mouse clicks and the mouse co-ordination. Both these log file entries have a time-stamp for each row of data, enabling identification of the durations for any chosen segment of activity (Figure 2).

We processed the UAR log files to calculate the computer use times for each doctor–computer

Figure 1 Measuring the multi-channel video footage using ObsWin software and output with duration variables

Figure 2 User action recorder (UAR) and its log files for keyboard and mouse use
interaction. Further processing of log files, targeting the time stamp column using the Excel Macro program, gave the duration between adjoining interactions and the time from the consultation start. These were then categorised based on the time gaps. This indicated possible unbroken interactions or longer intervals between them. We then compared these measurements against the ObsWin outputs to classify purpose of each interaction, e.g. prescribing, free text or coded data entry etc. We used the multi-channel video, especially the clinical system’s screen capture element, to identify a marker representing the start and end of a particular doctor–computer interaction, e.g. mouse click, or pressing a character key (see ‘Action’ column in Figure 3). We could then identify the timing of this marker from the UAR log file, enabling us to get an objective measurement of the duration.

Piloting the new method (ODC plus UAR) on two consultation tasks

We selected two common comparable consultation tasks to study in detail; we chose recording of the problem title and prescribing. These two were chosen as we only wanted to compare activities that would occur regularly and would have easily measured start and end points: these items fitted best with these criteria. Recording the problem title usually involves selecting a relevant code from a picking list. We performed a statistical comparison of these data using the Mann–Whitney U test. This is the appropriate non-parametric test (the type of test used for data which is not distributed normally) when data are taken from independent samples.

Ethical approval

We obtained ethical approval for real consultation recording (REC reference: 06/Q1702/139). We recruited patients after informing them about the research project and obtaining consent forms and offering them the option to reconsider their decision to participate both before and after the consultation session.

Results

Reliability when rating live consultations

We filmed a total of four real consultations with EMIS LV and seven with INPS Vision. The two general practitioners who volunteered to participate in these

![Figure 3](image-url) Measuring the data entry time by processing the UAR log for keyboard use
sessions have been using their respective clinical system for more than five years and can therefore be considered 'expert users'. The multi-channel video provided a highly usable overview of the consultation. The doctors consulted in a familiar environment.

Inter-rater reliability was good in many areas, but poor for computer use, especially when of short duration (Table 1). Raters showed high reliability levels ($r > 0.75$) for doctor–patient verbal interactions, free text data entry and prescribing. Further analysis indicated that, despite the pre-rating training and previewing, raters showed judgemental errors in identifying the start and end point for coded data entry. Having the 'doctor using PC' interaction variable together with other frequently visible interactions might have also caused this variability. Detailed analysis of the UAR log files indicated three seconds as a threshold value to identifying uninterrupted keyboard use; time gaps larger than the threshold value indicate a break of the interaction. There were poor levels of inter-rater reliability ($r < 0.5$) for variables representing coded data entry and 'doctor using PC'.

**User action recording: UAR**

The UAR also captured all keyboard strokes and mouse movements. We therefore decided to use the UAR log files as the primary measurement tool of computer use. We still coded these activities manually using ObsWin, but its prime activity became to label the activity rather than to time it.

**Piloting the new method to compare time taken to record a problem and to prescribe**

We included only the coded data entered using the picking lists. Other methods of structured data entry, such as use of templates, were excluded due to the infrequency of their use; having multiple recording fields also makes it difficult to identify comparable features among systems. Furthermore, we did not observe a sufficient number of repeat prescriptions to make comparisons. Only acute prescribing durations were compared.

Coded data entry in EMIS LV had a mean of 11.5 seconds (median 12.1, inter-quartile range 2.75) compared to 6.8 seconds (median 5.7, inter-quartile range 3.3) in INPS Vision. The mean number of items coded in LV consultations was 1.75, for Vision this was 2. Non-parametric testing with 'Mann–Whitney U' indicated coded data entry in EMIS LV taking significantly longer than INPS Vision ($P = 0.006$). Acute prescribing did not show a statistically significant difference between the two systems ($P = 0.64$). However, LV had a shorter mean time of 23.7 seconds (median 23.8, inter-quartile range 2.1) compared to 27.5 seconds (median 23.6, inter-quartile range 9.0) for Vision (Table 2, Figure 4).

Coded data entry in EMIS LV involved selecting the appropriate consultation heading, entering the search string and then selecting a suitable code from the given list. Navigation was done using the keyboard arrow keys. The picking list for the matching codes was displayed with about 11 items per page. The general practitioner selected the item by pressing the character shown in front of each item. In INPS Vision the general practitioner entered the code directly into the current consultation without linking to a consultation heading. Clicking on the 'Add Medical History' icon launched the data entry window. Entering the search term filled the coded data entry area with the appropriate Read term. If the term offered by the system was not suitable, the doctor would launch the Read dictionary which offers a scrollable picking list. The general practitioners predominantly used the mouse for navigation and selection of items.

Both systems showed similar steps in prescribing; selection of the drug name, dosage and quantity. Vision’s ‘Acute Therapy’ interface presents additional pop-up windows requiring acknowledgement. If the drug name the general practitioner enters is not found, it prompts him or her to use the main drug dictionary. Agreeing to this activates a picking list of matching drug names. Saving the prescription prompts the second pop-up window of ‘drug check results’; a summary of contraindications or interactions. In EMIS LV, the lists of matching drug names are presented without a separate stage of prompting. Other relevant information is shown soon after the drug selection. This is embedded in the same interface requiring no separate acknowledgement (Figure 5).

**Discussion**

**Principle findings**

There is considerable variation in the ways in which computers are used in the clinical consultation. This diversity makes it much more difficult to accurately time events, particularly computer use, in the consultation. We found that we could not reproduce the same level of reliability in the rating of real consultations in comparison with that we achieved in rating a simulated consultation for a blood pressure check.

The combination of UAR and ODC through multi-channel video recording provides a mechanism to measure the impact of different features of EPR systems on the consultation. UAR objectively captures
<table>
<thead>
<tr>
<th>Variable and activity</th>
<th>EMIS LV 1</th>
<th>EMIS LV 2</th>
<th>INPS Vision 1</th>
<th>INPS Vision 2</th>
<th>Mean</th>
<th>Median</th>
<th>IQR</th>
<th>Min</th>
<th>Max</th>
<th>ICC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>X – Doctor using PC</td>
<td>44.5</td>
<td>33.3</td>
<td>37.8</td>
<td>44.5</td>
<td>40.0</td>
<td>41.2</td>
<td>7.8</td>
<td>33.3</td>
<td>44.5</td>
<td>0.46</td>
</tr>
<tr>
<td>C – Doctor speaking to patient</td>
<td>56.2</td>
<td>40.3</td>
<td>29.1</td>
<td>41.3</td>
<td>41.7</td>
<td>40.8</td>
<td>7.5</td>
<td>29.1</td>
<td>56.2</td>
<td>0.85</td>
</tr>
<tr>
<td>V – Patient speaking to doctor</td>
<td>18.0</td>
<td>31.8</td>
<td>8.9</td>
<td>21.6</td>
<td>20.1</td>
<td>19.8</td>
<td>8.4</td>
<td>8.9</td>
<td>31.8</td>
<td>0.78</td>
</tr>
<tr>
<td>B – Patient in room</td>
<td>86.8</td>
<td>89.6</td>
<td>99.7</td>
<td>98.1</td>
<td>93.6</td>
<td>93.9</td>
<td>9.6</td>
<td>86.8</td>
<td>99.7</td>
<td>0.99</td>
</tr>
<tr>
<td>N – Third party (interruption)</td>
<td>0.0</td>
<td>0.0</td>
<td>7.2</td>
<td>5.8</td>
<td>3.3</td>
<td>2.9</td>
<td>6.1</td>
<td>0.0</td>
<td>7.2</td>
<td>0.92</td>
</tr>
<tr>
<td>A – Referral</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.03</td>
</tr>
<tr>
<td>S – Coded data (non-QOF)</td>
<td>0.0</td>
<td>0.6</td>
<td>0.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td>0.0</td>
<td>0.9</td>
<td>-0.10</td>
</tr>
<tr>
<td>D – Coded data (QOF)</td>
<td>1.8</td>
<td>1.4</td>
<td>2.3</td>
<td>1.2</td>
<td>1.7</td>
<td>1.6</td>
<td>0.5</td>
<td>1.2</td>
<td>2.3</td>
<td>-0.12</td>
</tr>
<tr>
<td>F – Entry of free text</td>
<td>8.8</td>
<td>6.8</td>
<td>7.9</td>
<td>7.3</td>
<td>7.7</td>
<td>7.6</td>
<td>0.9</td>
<td>6.8</td>
<td>8.8</td>
<td>0.89</td>
</tr>
<tr>
<td>G – Prescribing (non-QOF related)</td>
<td>0.0</td>
<td>0.0</td>
<td>8.1</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>8.1</td>
<td>0.74</td>
</tr>
<tr>
<td>H – Prescribing (QOF related)</td>
<td>3.3</td>
<td>2.4</td>
<td>3.3</td>
<td>11.0</td>
<td>5.0</td>
<td>3.3</td>
<td>2.2</td>
<td>2.4</td>
<td>11.0</td>
<td>0.02</td>
</tr>
<tr>
<td>J – Prompt from PC</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.03</td>
</tr>
<tr>
<td>Q – Eye contact</td>
<td>30.8</td>
<td>38.0</td>
<td>26.0</td>
<td>22.7</td>
<td>29.4</td>
<td>28.4</td>
<td>7.4</td>
<td>22.7</td>
<td>38.0</td>
<td>0.53</td>
</tr>
<tr>
<td>W – Dr speaking and using PC</td>
<td>25.2</td>
<td>11.1</td>
<td>8.2</td>
<td>18.9</td>
<td>15.9</td>
<td>15.0</td>
<td>10.1</td>
<td>8.2</td>
<td>25.2</td>
<td>0.89</td>
</tr>
<tr>
<td>E – Pt speaking and Dr using PC</td>
<td>7.1</td>
<td>1.4</td>
<td>2.6</td>
<td>5.7</td>
<td>4.2</td>
<td>4.1</td>
<td>3.8</td>
<td>1.4</td>
<td>7.1</td>
<td>0.58</td>
</tr>
<tr>
<td>R – Doctor examing patient</td>
<td>0.0</td>
<td>8.4</td>
<td>13.3</td>
<td>0.0</td>
<td>5.4</td>
<td>4.2</td>
<td>9.6</td>
<td>0.0</td>
<td>13.3</td>
<td>0.66</td>
</tr>
<tr>
<td>T – Silence</td>
<td>13.7</td>
<td>17.7</td>
<td>31.9</td>
<td>14.2</td>
<td>19.4</td>
<td>15.9</td>
<td>7.2</td>
<td>13.7</td>
<td>31.9</td>
<td>0.83</td>
</tr>
</tbody>
</table>

*a Intra-class correlation (ICC) for six raters, italics indicate poor correlation
*b IQR = Inter-quartile range
Measuring the impact of different brands of computer systems on the clinical consultation

Details about doctors’ computer use; it is more reliable in live consultations than direct observation. Detailed screen capture and three video footages provide a comprehensive overview. ODC using ObsWin measures multiple aspects of computer-mediated consultation interactions.

Implications

There is the potential to use this method to compare computer system features and their impact on the clinical consultation. This method also has applications in clinical system testing, simulation, evaluation, and investigations into general practitioners’ computer use styles.

Limitations of the study

This is a pilot study with smaller number of consultations using only one general practitioner for each system. A study with a larger sample might provide more detail about the impact of a variety of interfacing features. Though both coded data entry and prescribing are done in a structured manner, individual doctors’ computer use style will affect the measurements. Having a number of users for each system would help to standardise the findings.

Comparison with literature

Existing methods for analysing consultation activities are more subjective. Some qualitative studies have looked into the behaviours associated with information exchange between the doctor and computer, using a single channel approach followed by tagging of the video. Previous attempts to look into the multi-tasking in consultation have combined video analysis and conversation analysis methods. A cognitive-based observational approach to analyse data entry by clinicians in an outpatient setting used a much more complex set-up; a portable usability laboratory with a video converter, a microphone for conversation and keyboard sound recording. Investigation methods in human–computer interaction use multiple observation methods, but these are often synthesised into a single visual data stream. They lack easily quantifiable measurements flexible enough to code the variety of consultation–system interactions.

Call for further research

Applying this method to a larger sample of consultations could enable researchers to quantify the impact of varying system designs.

Conclusion

This pilot demonstrates how automating the capture of computer use using UAR overcomes the problems of reliability found in rating live consultations. Despite the information management advantages introduced by clinical computer systems, doctor–computer interactions interfere with the social aspect of the consultation. This further development to our method for observing the clinical consultation should contribute to our ability to assess the functioning of computer system features within the consultation. If these results are repeated in a larger study we are closer to developing a method to enable us to develop less intrusive but equally effective clinical systems.

Table 2 Time taken for coded data entry and acute prescribing in EMIS LV and INPS Vision

<table>
<thead>
<tr>
<th></th>
<th>Coded data entry</th>
<th>Prescribing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EMIS LV*</td>
<td>INPS Vision*</td>
</tr>
<tr>
<td>No. of items</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>11.5 (3.0)</td>
<td>6.8 (2.9)</td>
</tr>
<tr>
<td>95% CI</td>
<td>8.7–14.2</td>
<td>5.1–8.5</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>12.1 (2.8)</td>
<td>5.7 (3.3)</td>
</tr>
<tr>
<td>Min</td>
<td>5.7</td>
<td>3.6</td>
</tr>
<tr>
<td>Max</td>
<td>14.4</td>
<td>12.5</td>
</tr>
</tbody>
</table>

* No. of consultations: for EMIS LV, N = 4 and INPS Vision, N = 7
Figure 4 Box-whisker plots comparing coding and prescribing times for EMIS LV and INPS Vision

Figure 5 Coded data entry and prescribing screens in EMIS LV and INPS Vision
REFERENCES


CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Primary Care Informatics
Division of Community Health Sciences,
St George’s – University of London
SW17 0RE
UK
Tel: +44 (0)208 725 5661
Fax: +44 (0)208 725 3584
Email: slusigna@sgul.ac.uk

Accepted March 2008