Does manual anaesthetic record capture remove clinically important data?

J. M. van Schalkwyk1*, D. Lowes2, C. Frampton3 and A. F. Merry4,5

1 Auckland City Hospital, P/Bag 92024, Auckland 1023, New Zealand
2 University of Auckland, 22 Princes Street, Auckland 1010, New Zealand
3 Department of Medicine, Christchurch School of Medicine and Health Sciences, University of Otago, Dunedin North, New Zealand
4 Department of Anaesthesiology, University of Auckland, Private Bag 92019, Auckland, New Zealand
5 Department of Anaesthesia, Auckland City Hospital, Auckland, New Zealand

* Corresponding author. E-mail: johanvs@adhb.govt.nz

Editor’s key points

- Compared with computerized record-keeping, manual record-keeping results in loss of clinically relevant information.
- The loss of information in handwritten records is sufficient to influence inferences by clinicians.

Background. Numerous studies have shown smoothing and inaccuracies in handwritten anaesthetic records, but the clinical relevance of these findings is unclear. We therefore sought to determine whether the behaviour of anaesthetists differed in assessing anaesthetic records re-synthesized from either handwritten or automated records.

Methods. In a recent New Zealand study (ACTRN12608000068369), both manual and automated records were acquired from the same anaesthetics. Manual records were digitized using digital callipers. Selected data (systolic, diastolic, and mean arterial pressure; heart rate; \(S_{pO_2}\); \(E'_{CO_2}\)) were replayed in a computerized anaesthetic record-keeping system with which the participants were familiar, to present manual and corresponding automated anaesthetic records. Ten anaesthetists, randomly selected from participants in this study, assessed 24 replayed records (a manual and an automated record from each of 10 anaesthetics, with two of each displayed twice). They indicated where and how they would have intervened if administering these anaesthetics. We compared the number of interventions for each pair of anaesthetics and subjective measures of anaesthetic quality.

Results. In our selected sample of unstable anaesthetics, the mean (SD) number of interventions per anaesthetic was 4.0 (2.9) vs 5.2 (3.4) for manual and automated records, respectively (P = 0.013). Subjective measures did not differ significantly between record types. Assessors identified 32 artifacts in six manual records (0.32/record assessment) and 105 artifacts in eight automated records (1.05/record assessment), P = 0.14. Replicability was moderate (COV 39.8%).

Conclusions. In comparison with computerized record-keeping, manual record-keeping resulted in loss of clinically relevant information.

Keywords: computer simulation; medical records systems, computerized; monitoring, intraoperative

Accepted for publication: 16 April 2011

The anaesthetic record has become a key element of good anaesthetic practice: it can facilitate effective monitoring of patients, it is an important medico-legal document, and it may serve as an important tool for research, audit, and quality improvement in anaesthesia. All of these functions depend on the accuracy of the record, and on the relevance and comprehensiveness of the data collected.

Since the pioneering work of Zollinger and colleagues1 in 1976, several studies have demonstrated that in comparison with computer-generated anaesthetic records, in manually constructed records, there is a tendency to smooth data,2–4 omit information, and record incorrect information.5–9 Conversely, the point has been made that automated records have limitations in relation to the indiscriminate recording of artifacts and to insufficient sampling frequencies for some parameters.10 However, the clinical relevance of these various differences remains unclear.

We have therefore undertaken a study to determine whether there are clinically relevant differences between manually and automatically recorded physiological data obtained during the same anaesthetics.

Methods

This is a substudy within the trial, ‘A randomised clinical trial of a multimodal system designed to reduce drug
administration and recording error in anaesthesia’ (Australian New Zealand Clinical Trials Registry registration number ACTRN1260800068369, hereafter referred to as the ‘System Study’), conducted with the approval of the Northern Y Regional Ethics Committee (ref. NTY/07/10/112) and written consent from participating anaesthetists.

In the System Study, 509 pairs of records of routine anaesthetics were collected at a large, New Zealand teaching hospital where the use of an automated anaesthetic record keeper (ARK) is now a normal practice. Record pairs consisted of a comprehensive, manually constructed anaesthetic record (‘manual record’), and physiological data from the same anaesthetic, acquired from a Datex A/S3 monitor (Datex-Ohmeda, Helsinki, Finland) by an ARK (SAFERsleep OR, Version 4.9.6, SAFERsleep LLC, Nashville, TN, USA), and stored in an encrypted Extensible Markup Language (XML) format (‘automated record’). In these cases, the ARK record was made in parallel with the manual record, but no information other than the physiological data was collected in the ARK, and the anaesthetist making the manual record was not able to see the ARK. Eighty-nine anaesthetists took part in the study (54 consultants and 35 trainees) and 14 who were eligible declined to participate.

Fifty record pairs from the System Study were selected for the present analysis, as described below. The manual records were digitized using a Vernier digital calliper (ROK International Industry Co.: TestnTools, Auckland, New Zealand; accuracy 0.02 mm, with associated ‘UCB DCMS’ software) and imported into Microsoft Excel 2003 (Microsoft Corporation, Redmond, WA, USA) via the USB port of a Pentium 4 Personal Computer running Windows XP Professional (Microsoft Corporation). The callipers were calibrated before, and repeatedly during the digitization of each record. Software was developed to automate storage and manipulation of the digitized data. All data were recorded by the same observer (D.L.).

In the System Study, the sampling interval was set to 20s. However, the ARK displays data only once per minute, and manual data were usually recorded every 5 min. We therefore used piecewise linear interpolation to generate values once per minute for the manual records, and we used the first valid datum in every 1 min epoch for the automatically collected records. Where the digital record contained data spaced more than 1 min apart, piecewise linear interpolation was used, as for manual data.

Digital data, both automated and manually acquired, were replayed under Windows XP using the ARK, to provide a standardized, familiar view of the data. Computer screen views were captured every minute and stored. Case records that lasted more than 2 h were truncated at 2 h. Where cardiopulmonary bypass (CPB) was initiated, the record was truncated at the initiation of CPB.

A web-based program written by one of the authors (J.M.S.) in Hypertext Preprocessor (PHP), Javascript, and Hypertext Markup Language (HTML) was then used to replay the screen views one-by-one. This program permitted anaesthetists recruited from the Department of Anaesthesia at Auckland City Hospital (ACH) to step through records frame-by-frame and annotate each record at points they deemed appropriate.

Users could add free comments where desired. The interface permitted users to ‘fast-forward’ through repeated clicks/key presses, but the maximum rate at which this could be done was set to one frame per second, a limitation established to prevent users ‘over-running’ a point at which they would have intervened. It was not possible for users to move back to a previous time, thus preventing post hoc modification of responses. ImageMagick software (Version 6.4.8, ImageMagick Studio LLC) was used to crop and subtract Portable Network Graphics (PNG) images from one another, and render the background of images transparent, to lower image bandwidth for web-based display. Anaesthetists’ assessments of displayed anaesthetic data were stored in an SQL database (MySql version 4.1.22, running on a UNIX-based server, version 4.0.27-standard). Full, documented source code of the program is freely available.11

The system was tested by four anaesthetists who, using the web-based interface, assessed 28 record pairs from the start of the System Study, and 22 pairs from the end of the study, presented in a random order. On the basis of this initial assessment, we decided that a process of data enrichment was appropriate, as many anaesthetics were largely uneventful and assessor fatigue was a problem. We also abandoned our initial plan to compare record quality in early and later records in this study. Three of the four testers then each examined the 50 records and ranked them according to the instability (or variability over time) of the physiological data contained in the record. The three ranks thus obtained were summed for each record. The top 20 records (representing the most unstable anaesthetics) were then presented to the study statistician who randomly selected 10 of these records for further use.

Ten consultant anaesthetists of the 54 at ACH who had participated in the System Study, excluding the four involved in the initial testing, were randomly selected. All participants were given password-protected access to the web-based program, so that they could evaluate the records at times convenient to themselves. Before they assessed their first record, they were introduced to the user interface and taken through a sample record, during which time they used the web interface to make appropriate annotations under supervision. A paper aide mémoire was also provided. The demonstration record can be viewed interactively on the Internet.11

Each case was associated with a short ‘stem’ that describes minimal salient features of the case (e.g. ‘Older patient for transurethral resection of a bladder tumour. Airway management with a Proseal LMA allows spontaneous and controlled ventilation to occur’). The stem was identical for each record pair. The anaesthetist had time to read this stem while the screens were cached from the Internet database, and the stem was displayed for the duration of the ‘case’. The anaesthetist then moved through the record at his or her own pace, although the program advanced to
the next frame once a minute had elapsed. He or she could record evaluations at any frame.

Cases were presented in a random order, with the inclusion of two duplicated pairs of records to check for consistency; the total number of replayed records assessed by each consultant was thus 24. The randomization was constrained so that the duplicates were not presented consecutively. The time taken to evaluate each case was recorded automatically.

Data presented to the anaesthetist were limited to the following: systolic arterial pressure, diastolic arterial pressure, mean arterial pressure, heart rate, \( E_{CO2} \) and \( S_{pO2} \). As mean arterial pressure was not recorded on the manual records, this was derived from the systolic and diastolic values, using the formula of Razminia and colleagues. Where intra-arterial pressures were available from the automated record, they were used.

The ‘intervention’ options available to the anaesthetist are shown in Figure 1. The display proceeded in a fixed fashion, as it had in the original record, regardless of the interventions chosen by the assessing anaesthetists (i.e. these interventions did not have any impact on the subsequently displayed data). Assessors were informed of this invariance before starting the assessment and reminded of it with every intervention.

Assessors were asked to identify any information on the records considered to be artifactual. Each anaesthetist was also asked to assess the case on several additional grounds, at the end of the case (Fig. 2). Assessors were also asked whether they thought the records had been made manually or by the ARK.

Two endpoints were established a priori for the study. The number of points identified by assessing anaesthetists as warranting clinical intervention was considered to be a measure of clinical response to the presented record. Our null hypothesis was that there would be no difference between the manually and automatically collected records for this endpoint. In addition, we asked the question ‘In similar circumstances, would you be happy for someone close to you to have a similar anaesthetic?’— a more subjective, dichotomous measure of the assessing anaesthetist’s response to the totality of data presented. Other parameters assessed included the number of artifacts identified in the records, the time taken to extract the data from the records, and rating of ‘anaesthetic quality’ from 1 to 5 on a Likert scale.

**Statistical analyses**

The number of points per record warranting intervention was compared between automated and manual records using a general linear model that included terms for anaesthetist and anaesthetic case. The proportion responding ‘No’ as to whether the assessor would be happy for someone close to them to have a similar anaesthetic, and the number of artifacts identified per case was compared between manual and automated records using the Wilcoxon signed-rank test, using the anaesthetic case (\( n=10 \) pairs) as the replicate. The ranking of the anaesthetic quality and the length of time taken to assess each record was also compared between the manual and automated records using the non-parametric Wilcoxon signed-rank test, using the anaesthetic case (\( n=10 \) pairs) as the replicate. For all analyses that compared the automated and manual record results, the first of the duplicates was used. The variation between the duplicates, where the anaesthetist scored the same case with

---

Fig 1 Options for intervention. This screen could be invoked at any time during the course of the ‘anaesthetic’. A sample completion of the form is shown.
the same approach (manual or automatic) twice, is represented as the coefficient of variation. This is calculated as the square root of the variance between these duplicates divided by the mean.

**Results**

Of the 10 anaesthetists randomized to the study, two were women and eight were men. The median duration of practice since qualification was 11 yr (inter-quartile range 7–17 yr). Eight of 10 anaesthetists stated that they were predominantly in public hospital practice, with the other two in mixed private and public hospital practice.

In the selected, ‘unstable’ anaesthetic records presented to the assessing anaesthetists, the mean (SD) number of points per record identified as warranting intervention was 4.0 (2.9) for the manual records and 5.2 (3.4) for the automated records \((P=0.013)\).

The proportion of ‘no’ responses to the question ‘In similar circumstances, would you be happy for someone close to you to have a similar anaesthetic?’ was 0.13 for the manual records and 0.06 for the automated records presented (Wilcoxon’s signed rank, \(P=0.174\)). The median Likert scale rating of ‘anaesthetic quality’ was 3 in both groups \((P=0.72)\). Responses to the question ‘Do you think you have ever encountered this record before?’ were not analysed as this question was considered by several participants to be ambiguous.

Of the 100 replays of manual records, 38 were assessed as automated; of the 100 replays of automated records, 26 were assessed as manual records.

Examination of the 20 pairs of duplicate assessments, inserted to determine whether assessments were replicable, showed good agreement for determining whether a loved one should have the procedure (18/20 automated, 20/20 manual) and moderate agreement for allocation as automated or manual (90% automated, 60% manual).

The median duration of online case assessment was 6.1 min (range 1.5–22.1). Total time taken per anaesthetist to assess all 24 records ranged from 1.9 to 4.9 h. The median (range) time for assessing automated records was 5.7 min (1.5–19.0), and for manual records, it was 6.5 (1.6–22.1; \(P=0.13\)).

Assessors identified 105 artifacts in eight of the 10 automated records (1.05 per record assessment) and 32 artifacts in six of the 10 manual records (0.32 per record assessment, \(P=0.14\)). All assessors identified artifacts. The number of artifacts reported per assessor ranged from 7 to 26, with a mean of 14 artifacts per assessor: 99 reports related to arterial pressure, 11 to heart rate, five to \(S_pO_2\), and 22 to \(E'_CO_2\).

In addition to the above analyses, we wrote an online query tool that displayed a graphic of colour-coded interventions by case and assessor. Figure 3 shows the interventions considered necessary by one assessor. These interventions are superimposed on graphics of the electronic and manual records that are similar to those presented to the assessors, but slightly compressed vertically. Displaying the data in this way demonstrated considerable variation in the thresholds for and choices of intervention between assessors.

The coefficient of variation representing the variation between the duplicate assessments was 39.8%, indicating a moderate level of agreement when the ratings from the same record were reassessed by the same observer.
**Discussion**

On assessment of replayed manual records, experienced anaesthetists identified fewer points that they felt warranted clinical intervention than was the case with the corresponding records made from the same anaesthetics with an ARK. This finding provides strong evidence that clinically relevant data are lost during manual capture of anaesthetic records. At the same time, three times as many artifacts were identified in automated records as in manual ones. This finding also indicates a degree of selectivity in the data recorded in the manual records. There was no significant difference between record types in responses to the more global assessment question: this finding could be interpreted in various ways, but we think that it indicates an acceptable standard of anaesthesia (i.e. the accurate records did not elicit concern) and of manual record-keeping (again, because these records were also mostly seen as clinically acceptable).

Our findings are consistent with previous studies that have demonstrated smoothing of data recorded manually,\(^1\)–\(^4\) and inaccuracies in manual records,\(^5\)–\(^8\) but go further in demonstrating that this well-established loss of information in handwritten records is sufficient that a clinician assessing the record for the purposes of the law, audit, or research might well draw different clinical inferences from manual vs automated records.

Our study has several limitations. It does not permit evaluation of the advantage or disadvantage of either method in relation to patient care. More interventions do not necessarily imply better management or outcomes, and the anaesthetists caring for the patients of course had direct access to the patient, with the ability to cross correlate physical findings and data that were directly available from the relevant monitors, in addition to any use they may have made of the records in aiding their clinical decisions. Furthermore, we have not examined the influence of the method of record-keeping on actual patient management or outcomes; our data relate only to post hoc, clinically related interpretation of records. The number of participants and records examined was relatively small, and all assessors come from a single institution, so the results obtained might not be generalizable. It is also possible that the standard of manual record-keeping was influenced (positively or negatively) by the fact that these records were made during a research study.

We selected records that showed substantial variability, and the demonstrated effect might have been less marked with unselected records; however, the general principle of data enrichment is an accepted strategy to increase the power of clinical studies, and we did this using a method that was unlikely to be subject to bias. It is possible that assessors’ ability to move rapidly through the displayed record biased their assessment in favour of ‘dramatic'
changes; however, we felt that constraining the assessing anaesthetist to real-time display was impractical and likely to result in fatigue and inattention, in the absence of a real patient to attend to. It might also be argued that web-based presentation of data is not representative of normal evaluation of a record by an anaesthetist (e.g. in the context of providing expert evidence). However, we screen-captured the computer-based interface with which participants were familiar, tested the web interface on commonly used browsers, and our software checked that screen resolution was adequate.

A further potential limitation is that selected physiological data were presented, rather than a detailed clinical record. It might be thought that the other information (related to drugs, position, fluids, surgical interventions, and the patients themselves) would compensate for any inadequacies in the physiological data. There were two reasons for the approach we used: first to focus assessment on physiological data, and second to limit extraneous factors (extra degrees of freedom) that might have necessitated assessment of many more cases. In addition, our choice of variables is in keeping with those represented in the literature and comprises variables that are also commonly represented on most anaesthetic charts. We acknowledge that there are limitations to the approach we have taken.

Piecewise linear interpolation was used to provide minute-by-minute data where intervening data were missing. This approach was necessary to mitigate the problem of assessors being confronted by long periods of ‘no data’ and follows the common practice of linear interpolation seen where, for example, plotted heart rate values appropriately become the basis for legal proceedings after an adverse patient outcome. The implications of our findings are substantial, particularly if our results can be duplicated by others. It is possible to take two views of the merits of the differences we have demonstrated. On the one hand, it can be argued that the manual anaesthetic record reflects a filtering of the information by the anaesthetist and that vigilance may be facilitated by the process of making a manual record. In addition, it can be argued that post hoc interpretation of faithfully recorded data (including artifacts) might inappropriately become the basis for legal proceedings after an adverse patient outcome.

The converse argument (to which we subscribe) is that it is difficult to endorse the use of a process, here manual record-keeping, that results in loss and distortion of information, particularly where this loss of information results in changes to subsequent clinical assessment of the record. This point of view is independent of the actual ARK used and does not address other relevant questions such as whether automated record-keeping facilitates or hinders anaesthetists in providing safe anaesthesia, notably in relation to effects on vigilance. We are concerned here only with the merits of accurate record-keeping and the loss of data encountered with manual record-keeping. We believe that it is desirable for ARK users to be able to annotate the record appropriately, without altering primary data. Given such a facility, the ability to retrieve an unbiased chronicle of events during an anaesthetic incident must surely be more credible in defence than the use of a manual technique that introduces smoothing, allows the removal of unwanted data points, contains comparatively fewer data points, and which we have now shown to result in clinically different interpretations from a contemporaneously acquired, automated record.

Although we used a proprietary record-keeping system, all of the software written specifically for this study is based on ‘freeware’ components, and we have made the complete
source code and documentation available under the GNU public licence,\(^\text{15}\) so that others might extend our preliminary work using other ARKs.

The code for data acquisition is at present rather crude and tied to use of digital callipers. It is conceivable that the code might be modified to accept other input devices, for example, appropriately calibrated, pressure-sensitive pads. Other refinements might allow introduction of additional physiological data, even other data such as those related to administration of drugs, or other interventions. There are several potential applications for this software, including impartial assessment of a given anaesthetic record, assessment of multiple anaesthetists’ responses to a given scenario (even with the establishment of ‘reference’ behaviour), the ability to determine how alteration of a particular component of a record alters behaviour, and use in continuing professional development. We anticipate using similar techniques in future research.

In summary, we have shown that manual record-keeping results in clinically relevant loss of information.

### Acknowledgements

The authors would like to thank Drs C.N. Bradfield, L.A. Hopley, and S. Laurent for testing of software and valuable comments that contributed to the development of the study. They would also like to thank all the anaesthetists who participated in this study, and J. Hannam, K.-E. Edwards, and A. Jardim who collected the data analysed here in the System Study.

### Conflict of interest

A.F.M. has financial interests in SAFERsleep LLC, being a major shareholder, and is a director of this company. It is conceivable that our preliminary demonstration of clinically relevant loss of information during manual record-keeping might support the case for use of ARKs like SAFERsleep, resulting in some benefit to makers of ARKs.

### Funding

Funding for this project was through a University of Auckland summer studentship (D.L.).

### References


