

# A data-capture system for post-marketing surveillance of drugs that integrates with hospital electronic health records

Keiichi Yamamoto<sup>1</sup>  
Shigemi Matsumoto<sup>2</sup>  
Kazuhiro Yanagihara<sup>2</sup>  
Satoshi Teramukai<sup>1</sup>  
Masanori Fukushima<sup>1,2,3</sup>

<sup>1</sup>Department of Clinical Trial Design and Management, Translational Research Center, Kyoto University Hospital, Kyoto, Japan; <sup>2</sup>Outpatient Oncology Unit, Kyoto University Hospital, Kyoto, Japan; <sup>3</sup>Translational Research Informatics Center, Foundation for Biomedical Research and Innovation, Kobe, Japan

**Purpose:** In conventional clinical studies, the cost of data management for the purposes of quality control tend to be high and collecting paper-based case report forms (CRFs) can be burdensome, because paper-based CRFs must be developed and filled out for each clinical study protocol. Use of electronic health records (EHRs) for this purpose could reduce costs and improve data quality in clinical studies. Kyoto University Hospital launched an EHR system in January 2005. At the same time, a replicate of that database was established for other purposes. At the Outpatient Oncology Unit of Kyoto University Hospital we developed a data-capture system that includes a cancer clinical database system and a data warehouse for outcomes studies. This system allows us to accumulate data at low cost and apply it to various uses in clinical or outcomes studies. Here we report on the application of this system to the post-marketing surveillance of drugs.

**Methods:** We evaluated the availability of this system and identified problems for future development. With this system investigators can register cases for post-marketing surveillance, and the registered cases are listed on a screen. When CRFs for a particular case are required, data can be extracted from the list and CRFs are produced in PDF format.

**Results and conclusion:** In this study we confirmed the applicability of our new system to post-marketing surveillance in providing prompt and efficient data exchange. We expect it to reduce the cost of data management and analysis and to improve the quality of data in post-marketing surveillance of drugs.

**Keywords:** cancer, electronic data capture, data management, pharmacovigilance

## Introduction

In conventional clinical studies, the cost of data management for the purposes of quality control tend to be high and collecting paper-based case report forms (CRFs) can be burdensome, because paper-based CRFs must be developed and filled out for each clinical study protocol. Additionally, in conventional clinical or outcomes studies the study data are generally the property of the sponsor, and even if the data belong to a researcher's own institution the data cannot be used easily.<sup>1</sup>

Daily clinical practice generates an enormous amount of clinical information, which is recorded in the form of paper medical records. Clinical information for a particular patient is held by that patient's own physician, making the sharing of patient information between physicians problematic. In addition, if clinical records are required elsewhere, such as for a research article, the paper record must first be put into software such as Microsoft Excel to make it usable.

Correspondence: Keiichi Yamamoto  
54 Shogoin Kawahara-cho, Sakyo-ku,  
Kyoto, 606-8507 Japan  
Tel +81 75 751 4717  
Fax +81 75 751 3399  
Email kyamamo@kuhp.kyoto-u.ac.jp

In Japan, medical information technology has recently been promoted as a national project via the e-Japan Strategy, and an electronic health record (EHR) system was launched at Kyoto University Hospital in January 2005. Various types of clinical information are computerized using the EHR. The system was introduced with the primary aim of collecting clinical information, constructing databases, and enabling extraction of data for hospital management, analysis of hospital affairs, and clinical studies. A possible secondary use of EHRs is the storage of data for clinical and outcomes studies, which might yield financial savings and improvement in data quality compared with paper-based records. Direct use of EHR information to complete CRFs may obviate the need to cross-check CRFs against patient records.<sup>2-4</sup>

However, in Japan there is a history of hospital information systems being specialized as computerized order systems for physicians. Thus, these systems were not built with the concepts in mind of long-term data storage or making data accessible for re-use. In addition, EHRs tend to be structured to allow data to be entered freely, in natural language, and various forms of clinical information, including images, are stored together in the same field, irrespective of format. Additionally the structure of the EHRs does not allow for searching on a specific disease, prescription, or other fields across patients, because the original aim of EHRs was to improve the success of medical treatment only in individual patients. For instance, our hospital's EHR format has functions for listing patients by specific disease, test, prescription, and other fields, and it allows the time series of a particular patient's treatment to be displayed. However, it doesn't allow fields such as specific disease, test, or prescription, to be combined from different patients. Further, it can't collect and combine data on a particular patient that have been written under different fields of the EHR, such as background, laboratory data, complications, combination drugs, and adverse events, and so a single CRF cannot be generated. Therefore, the direct use of data from records, as is generally attempted when data are sourced for clinical or outcomes studies, is difficult.

At the Outpatient Oncology Unit of Kyoto University Hospital we developed a data-capture system for observational cancer clinical studies (ie, outcomes studies) that would integrate with an EHR system, and allow patient prognosis, prognostic factors, and outcomes (including drug effectiveness and safety) to be evaluated.<sup>5-7</sup> The purposes of this system are to integrate clinical research with clinical practice, to collect the information that is necessary for a clinical or outcomes study in an effective manner into a

database, and to use the information at the appropriate time to contribute to the improvement of survival rates in cancer. Using this system, data can be accumulated at low cost and put to various uses in clinical or outcomes studies.

The purpose of this study was to develop a system additional to our new data-capture system which can collect and put together patients' data from various areas of the EHR into a single CRF for post-marketing drug surveillance. We also aimed to identify data that could be collected from the current EHR or our new system, and to evaluate the efficiency of our new system.<sup>8</sup>

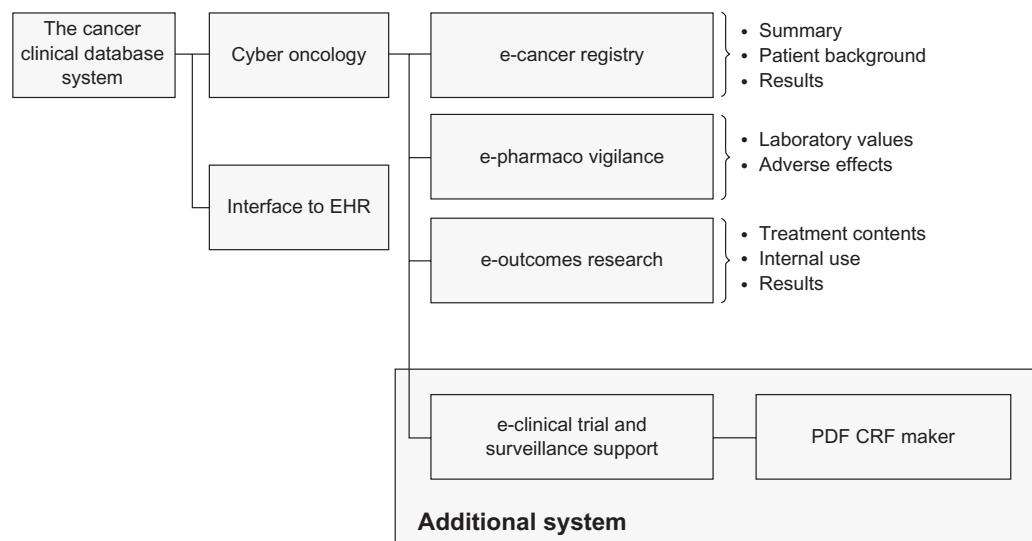
## Methods

### Overview of our system

Kyoto University Hospital launched an EHR system in January 2005. At the same time, a replicate of that database, known as Open DB, was established for other purposes. Using Open DB, we developed a combined cancer clinical database system, called Cyber Oncology, and data warehouse for outcomes studies. The cancer clinical database system permits clinical data to be stored in appropriate data categories, and offers high-level support for everyday clinical oncology practice and online analysis of stored data. The information stored in Cyber Oncology is transferred regularly to a data warehouse, during which process it is anonymized. Data in the warehouse are used for analysis of adverse effects, drug interactions, and evaluation of outcomes. Data for individual studies are extracted from the information on each protocol in the data warehouse and transferred to a data center, at which time data-mart databases for the study are made.

The system comprises four sub-systems: the Cyber Oncology cancer clinical database system, the clinical support system, the daily reports system, and the data-analysis system. Cyber Oncology has four modules, of which the e-Cancer Registry is the core module; the other modules are the drugs safety monitoring module (e-Pharmacovigilance), into which laboratory values and data about adverse events can be entered, the treatment results investigation module (e-Outcomes Research), and a clinical trial and surveillance support module (e-Clinical Trial and Surveillance Support) (Figure 1).

Cyber Oncology is available as a front-end component of the EHR system, collecting relevant information (eg, patient background or laboratory data) from the EHRs. The cancer clinical database system permits clinical data to be stored in appropriate data categories. In addition, data that are input directly into Cyber Oncology (eg, adverse-effect data) are converted into EHR information on screen, and can be



**Figure 1** Structure and functions of the cancer clinical database system.  
**Abbreviations:** EHR, electronic health record; CRF, case report form.

copied and pasted into a patient's EHR by medical staff. The system deals with the standard registration form for cancer, certified by the Ministry of Health, Labour and Welfare of Japan. Adverse effects are classified based on the Common Terminology Criteria for Adverse Events V3.0 (CTCAE V3.0). The system is comprehensively incorporated into everyday clinical oncology practice, and all information collected is entered into a database, making it possible to extract high-quality data when necessary. In addition, the cancer registry function of the system is used across the hospital as the standard cancer registry system.

## Method of data extraction for the post-marketing surveillance of drugs

At first an investigator registered cases that were appropriate for surveillance in Cyber Oncology and EHRs, including the computerized physician order entry system, and the registered cases could be viewed on a screen. When CRFs were required, the investigator selected the relevant case from the list on screen. Data were extracted automatically from Cyber Oncology and the EHRs, and CRFs produced in PDF format. The data that could not be extracted automatically by the system were added manually onto the PDF version of the CRF, checked and printed out by the investigator, and sent to the sponsor of the surveillance. The sponsor could enter the data into a data-management system (Figure 2).

## Results

We developed a system that directly accumulates data for post-marketing surveillance of drugs from our hospital's

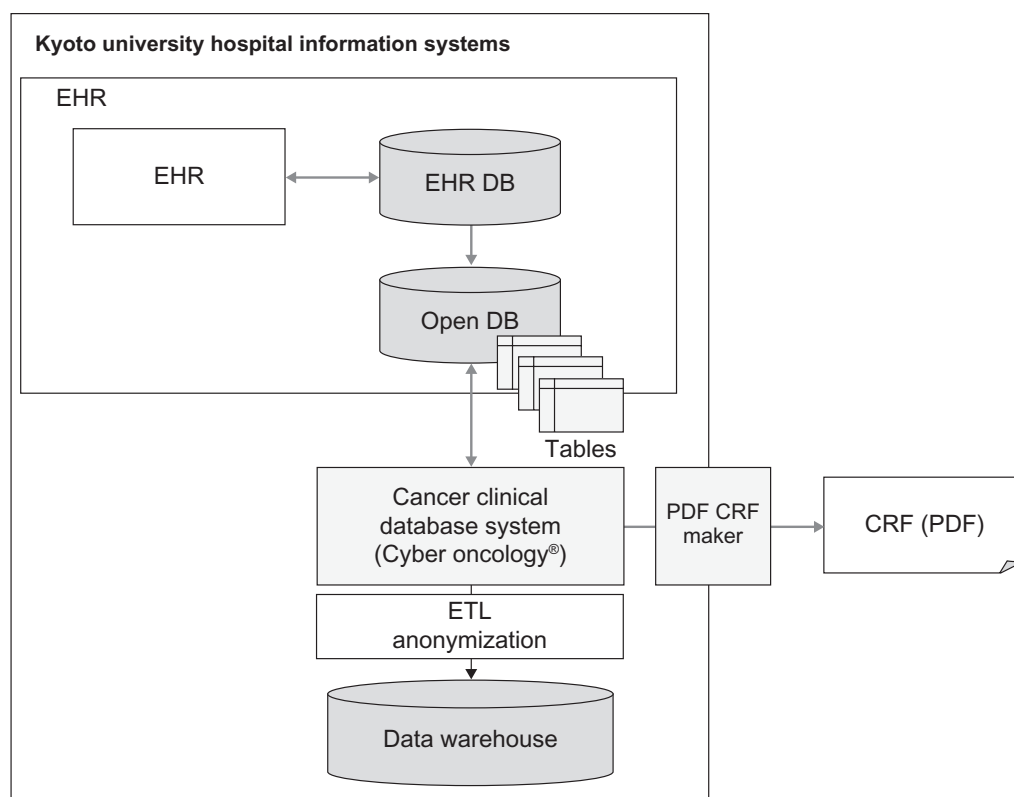
EHR system and our new additional system. We added two screens to our existing system. One was a screen to list cases by study and the other was for printing individual cases, and for converting paper CRFs to PDF format. The cost of making the additional system was about 20 person-days.

The CRFs for the surveillance ran to 24 pages and the total number of items, excluding any repeats, was about 600. The following patient background data could be collected from the EHR: whether in- or outpatient, height, weight, history of allergy, history of adverse events, anamnesis, and complications. From Cyber Oncology we could also collect information on advanced-metastasis cancer, reasons for using the drug, whether at first- or second-line treatment stage, performance status, and site of the cancer. Additionally we could collect laboratory data and pretreatment and combination drugs from EHR, and details on elements of treatment and adverse events from Cyber Oncology. Any data that could not be extracted automatically by the system, such as assessment of physician, progressive disease, and outcome, were entered manually onto the CRF by the investigator.

In the past it took an investigator about 3 or 4 hours to complete a CRF. Using our new system investigators could fill in a CRF within an hour. Furthermore, because adverse events, including whether grade 1 or 2, were being recorded in the Cyber Oncology database, investigators could add these adverse events to CRFs with no further effort.

## Discussion

We identified the following problems in this study. First, EHRs do not have directly available data on adverse events.



**Figure 2** Structure of Kyoto University Hospital's EHR system and the cancer clinical database system.

**Abbreviations:** EHR, electronic health record; CRF, case report form; DB, database; ETL, extract, transform, load.

We could not collect the day of expression, grade, seriousness, outcome, and causality about adverse events from EHRs directly. Instead, we collected such data from Cyber Oncology. Automatic grading of laboratory data from EHRs is possible, but a physician's diagnosis is necessary to decide whether something counts as an adverse event. Because an automatic system for making decisions about adverse events from EHRs has not yet been realized, if investigators from another institution want to extract adverse-event data from EHRs they will need to use an EHR template or develop an additional system to make records about adverse events, akin to the Cyber Oncology database.

It is not sufficient to develop an additional system to collect adverse events. New organizational methods for collecting adverse-event data, which are not currently part of the existing operational processes, might be necessary.

With regard to the collection of adverse-event data and the issue of data quality control in Japan, medical institutions are now required to have a health information manager and a cancer registrar, and their intervention is expected to improve the quality of patient records. When investigators collect necessary data for clinical or outcomes studies from EHRs or our new system, the appropriate intervention of the health

information manager and cancer registrar is necessary to reduce the burden and improve the quality of data entry by physicians. In addition, on a hospital site there will be a Clinical Research Coordinator (CRC) to support each study as it progresses, a Clinical Research Associate (CRA) to monitor whether a study progresses according to the protocol, and a data manager to ensure quality control of data that are collected. Generally, there is no contact between medical staff such as the health information manager, cancer registrar, and clinical trial specialist, and the CRC, CRA, or data manager. When we want to execute clinical or outcomes studies based on EHRs, we must arrange each operation, and must apply a new task as a new quality-control process for clinical or outcomes studies.

When new organizational methods are developed for collecting adverse-event data, then more efficient surveillance of drug safety will be possible. In general, data on adverse effects tend to be collected via voluntary reports from institutions or drug companies, and signals of adverse effects are often not detected promptly. In addition, the information is seldom collected from all treated cases, and frequency estimation is very difficult. There is a large gap between clinical practice and clinical trials for which efficacy and safety are evaluated at the pre-marketing stage. As a rationale

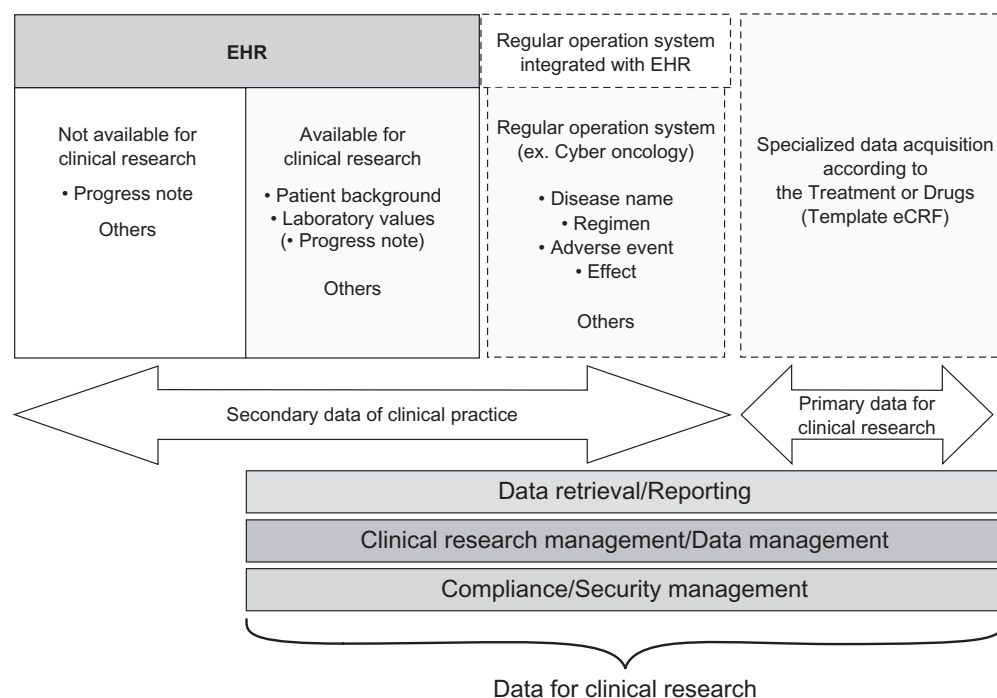
for preventing drug-induced suffering, the post-marketing surveillance of all cases may be effective. Surveillance of all cases has the advantages that it can accumulate detailed data about risk factors and measure correct incidence rates of adverse events, but the disadvantage is that the costs are high. Using the new cancer clinical database system and EHRs it is possible to accumulate information from all cases treated with a given regimen, because it is integrated with the EHRs. It is also very important to accumulate data from everyday clinical oncology practice, and to enable real-time evaluation of the data. We refer to this framework for adverse-effect data collection, which functions using EHRs, as the e-PharmacoVigilance system. Using e-PharmacoVigilance it will be possible to monitor adverse effects of anticancer drugs in large hospitals that have much experience in this field, administer drugs safely to patients, and undertake surveillance of all cases at low cost. It would be useful for both drug companies and regulators.<sup>9</sup>

In current clinical or outcomes studies, when CRFs are filled out, the necessary information is collected from patient records. This means that current clinical or outcomes studies are based on “primary data”, which are collected for a specific purpose. In contrast, data arising from existing records are referred to as “secondary data”. Although primary data represent a new dataset that precisely fits the needs of a study, it is more expensive and time-consuming to collect primary data than secondary data, which are initially collected for

other purposes. To achieve a reduction in the cost of carrying out clinical or outcomes studies, use of secondary, already digitized EHR data seems an appropriate approach.

From this discussion the necessity is clear for an additional system of EHRs to allow the conducting of drug post-marketing surveillance or other clinical research. First, in EHRs there needs to be a section for clinical research, such as progress notes, and currently unavailable details such as patient background and disease, prescriptions, and others. Additionally, there should be a regular operation system which encodes disease name, treatment, adverse effects, and judgments about effectiveness in order to make this information available for secondary use. Cyber Oncology is positioned as an example of such a system. In particular, such as when prospective studies are being conducted, it is necessary to accumulate specialized data from every study, according to the treatment or drugs prescribed. In such cases the flexible input function of the EHR template or others can support every protocol requirement as necessary. Furthermore, high-level reporting and retrieval functions are required, as are online analysis of accumulated data. The system's infrastructure must support protocol control, security management, data management, and compliance with various ethical guidelines and regulations (Figure 3).

When we use our new system, we must convert CRFs from paper to electronic PDF format. Eventually we might



**Figure 3** Diagram showing the clinical research system integrated with the EHR system.

**Abbreviations:** EHR, electronic health record; CRF, case report form.



see all systems produce CRFs that are in a standard format, but at present CRFs are different for every surveillance task. Because of this, CRFs must be redesigned every time that surveillance is requested, and this is a labor-intensive activity. Although the work of producing CRFs on screen is very much easier than with paper-based forms, if a prescribed style of case report was agreed upon, then the job of producing PDF documents would be much quicker and easier. This suggests that it is necessary to standardize the design of CRFs in future, for example for Clinical Data Acquisition Standards Harmonization (CDASH) projects of the Clinical Data Interchange Standards Consortium (CDISC).<sup>10</sup>

In addition, the electronic data capture (EDC) system, which accumulates data electronically from the medical institute, and the data-management system, used for surveillance, don't have an electronic interface, so CRFs are produced in PDF format, printed onto paper, and data then entered manually into the data-management system. In addition, there are concerns about the security and reliability of the data-management system and our system. Technically, we can output data electronically, using forms established by the CDISC, for example. In conclusion, if the EDC system and data-management system had an electronic interface of standard data format, with an appropriate guarantee of security and reliability, data capture for clinical studies could be performed entirely electronically.

We plan to promote the use of our cancer clinical database system in other institutions, and to collate data from as many institutions as possible in the data warehouse, with the aim of establishing a large-scale database for multi-center clinical oncology study data in Japan. At that time, it will not be sufficient to simply install our new system in other institutions. We will also have to standardize the level of quality control, including the use of standard terminology for drug and disease names among participating institutions. We expect this system to become a popular standard for clinical research, such as for the CDISC.

## Conclusion

In this study it has been possible to use EHR data and our new system's data in post-marketing surveillance and we

have shown how efficiency can be improved in carrying out post-marketing surveillance of drugs. This system can be applied not only to post-marketing surveillance but also to various clinical studies or trials. We aim to deal adequately with the problems identified in the course of this study and to increase the examples of its application in order to build a better system in the future. We expect that adoption of the new system will result in lower costs and improved data quality for outcomes studies.

## Acknowledgments/disclosure

The authors report no conflicts of interest in this work. We acknowledge the staff of the Department of Medical Informatics, Kyoto University Hospital, for their kind support.

## References

1. Daniels K, R. *eClinical Trials: Planning and Implementation*. Boston: CenterWatch Inc; 2003.
2. Leveraging the CDISC Standards to Facilitate the use of Electronic Source Data within Clinical Trials (eSDI). Available from: <http://www.cdisc.org/stuff/contentmgr/files/0/2f6eca8f0df7caac5bbd4fadfd76d575/miscdocs/esdi.pdf>. November 20, 2006. Accessed March 29, 2011.
3. Cancer Biomedical Informatics Grid (caBIG). Available from: <https://cabig.nci.nih.gov/>. Accessed March 29, 2011.
4. BRIDG Model. Available from: <http://www.bridgmodel.org/>. Accessed March 29, 2011.
5. Yamamoto K, Matsumoto S, Tada H, et al. A data capture system for outcomes studies that integrates with electronic health records: development and potential uses. *J Med Syst*. 2008;32(5):423–427.
6. Matsumoto S, Nishimura T, Kanai M, et al. Development of a novel information technology (IT) system using the electronic medical record (EMR) in daily clinical practice. ASCO Annual Meeting Proc. Part I. *J Clin Oncol*. 2007;25(18S) (Supplement):17066.
7. Yamamoto K, Matsumoto S, Matsuba H, et al. Development of data capture system for clinical study by the secondary use of electronic health records. *Proc Medinfo 2007 Proceedings of the 12th World Congress on Health (Medical) Informatics – Building Sustainable Health Systems*, August 20–24, 2007, Brisbane, Australia 2007.
8. Yamamoto K, Matsumoto S, Tada H, et al. A data capture system for post marketing surveillance of drugs that integrates with a hospital electronic health record. Society For Clinical Trials 29th Annual Meeting; 2008:72.
9. ICH E2E: Pharmacovigilance Planning. Available from: <http://www.fda.gov/RegulatoryInformation/Guidances/ucm129411.htm>. April 2005. Accessed March 29, 2011.
10. CDISC: Clinical Data Acquisition Standards Harmonization (CDASH) Projects. Available from: <http://www.cdisc.org/cdash>. Accessed March 29, 2011.

### Open Access Journal of Clinical Trials

### Publish your work in this journal

The Open Access Journal of Clinical Trials is an international, peer-reviewed, open access journal publishing original research, reports, editorials, reviews and commentaries on all aspects of clinical trial design, management, legal, ethical and regulatory issues, case record form design, data collection, quality assurance and data auditing

Submit your manuscript here: <http://www.dovepress.com/open-access-journal-of-clinical-trials-journal>

### Dovepress

methodologies. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.