THE ELEGANT MACHINE: APPLYING TECHNOLOGY TO OPTIMIZE CLINICAL TRIALS

A new model for conducting clinical research programs in the future is beginning to evolve which will emphasize collaboration, interdependency, and close interactive sharing of information among the various stakeholders in the research process: patient, investigator, sponsor, project manager, medical monitor, data management organization, laboratories, biostatisticians, and regulatory authorities. This model will conserve the use of information by minimizing redundancies, transcriptions, and data conversions in a continuous data flow, simplifying the data clarification process, and using the "best fit" technology tool for capturing and accessing clinical information. This view of clinical research will be based on a generic data repository based upon evolving International Conference on Harmonization/Food and Drug Administration (ICH/FDA) data standards which will also support trials management and safety monitoring with respect to the data repository. The current state of alternative data capture technologies such as interactive voice response systems (IVRS), remote data entry (RDE), image recognition, Internet technologies, hand-held computers, voice recognition, and wearable monitoring devices will be reviewed.

Key Words: Data management; Electronic data capture (EDC); Optical imaging; Interactive voice response (IVR); Data repository

INTRODUCTION

IT IS AXIOMATIC TO STATE that the process of clinical research and development can generally be improved through the proper and prudent application of information technology (IT). All too typically, however, technology implementations are costly, time-consuming, and disappointing. Often this disappointment results from inadequate implementation and insufficient process improvement. In far too many instances, the technology disappoints because it is overly complicated and ambitious—it tries to do too many things at once, and does not do them well enough.

In technology—just as in industrial design, fashion, and art—people often find that a simple but powerful solution is the best answer. People admire do-everything, feature-laden software suites, but they are generally more comfortable with a simple, intuitive, and familiar tool that is optimally suited to do the job at hand. In essence, people look for technology solutions that are *elegant*:

Originally presented at the DIA 12th Annual Symposium and Exhibition: "Innovative Technologies and Strategies for the Global Management of Clinical Information," March 16–29, 1997, Philadelphia, Pennsylvania as "The Elegant Machine: How Technology Can Transform Clinical Data Management."

MA 0215.

powerful, streamlined tools optimized for a primary need in a dignified, restrained, and tasteful manner.

People also want technology to provide complete, integrated solutions to their problem. The analogy of a *machine*—an integrated structure consisting of a framework and various parts for doing some particular kind of work—can be used to describe the perfect model. Machines that are elegant feel right to the user—they do just enough and they do it well (1).

The pharmaceutical industry needs such solutions for the process of clinical research and development. To dramatically improve clinical research, elegant machines are needed.

IT AND THE PHARMACEUTICAL INDUSTRY

The pharmaceutical industry has often seemed reluctant to employ new technologies. In various cases, this has been attributed to the absence of competitive pressures to force change, concerns with regulatory acceptance, lack of vision, and insufficient talent to identify and effectively implement new technologies. This, in turn, limited the market for software providers, so that few adequate, high quality software solutions were available for certain needs unique to the industry, and so many pharmaceutical information systems were often antiquated, poorly integrated, and disappointing.

In this respect the pharmaceutical industry was not unusual. A few years ago a typical manufacturing company would have entirely different systems for financial accounting, human resources, materials planning, scheduling, field service, and so forth. The advent of enterprise resource planning (ERP) systems changed this by providing a variety of modules to manage the entire supplier to customer value chain. Implementation of an ERP system would involve reengineering of business processes to eliminate unnecessary handoffs and steps and to more clearly assign responsibility. One of the major benefits of an ERP implementation is the consolidation of data into a central standardized information base, data that can be shared and "mined" to discover key information about customers and suppliers which in turn leads to more opportunities and efficiencies. More recently, there has been a movement toward component-based ERPs that mix and match individual modules using a common, standards-based data store with apparently seamless integration. The end goal is an end-to-end solution that drives the core business processes of an organization. One would expect these same concepts to be applicable to clinical research and development (R&D).

COMPUTER SYSTEMS FOR CLINICAL R&D

In most organizations, the systems used to support clinical R&D activities are classified as clinical or medical systems. These can be classified into at least five major categories:

- *Trials Management systems* to manage patient enrollments, schedules, site interactions, investigator grants, drug supply, and other tasks critical to the performance of a clinical trial,
- *Data Management Systems* which collect clinical data from patient case report forms (CRFs) and prepare these data for statistical analysis,
- *Safety Monitoring Systems* which record serious adverse events and ensure that proper notifications and responses are taken by investigators and sponsors,
- *Statistical Analysis Systems* used to analyze the results of a trial to determine the safety and efficacy of a drug, and
- *Document Systems* used to compose, publish, and manage the supporting information and content of a drug application to a regulatory agency.

Today's clinical trials involve a need for information sharing beyond traditional corporate bounds to an extended team that may include investigators, project managers, central labs, contract research organizations, and others. Sadly, even though these clinical sys-

tems support multiple users, too often this information is collected and distributed inefficiently and its potential is unrealized. Perhaps the most obvious candidate in clinical R&D to benefit from an elegant machine would be clinical data management.

THE INELEGANCE OF TRADITIONAL DATA MANAGEMENT

The classical or traditional data management approach evolved in order to provide a basic level of acceptance regarding the thoroughness and accuracy of the large volume of data required to support a new drug application. This approach involves the collection of vast amounts of patient data on paper CRFs completed by research sites, which is transcribed (typically by study coordinators or administrative personnel) to CRFs from source documents—charts and records completed by the investigating physician. These CRFs are collected by clinical research associates (CRAs) and sent to a data processing center. This center logs receipt and has two operators enter the same data into a database exactly as submitted using a double data entry system—even if obviously wrong! Next, a clinical data reviewer interprets and codes some medical terms, corrects some of the self-evident errors, and sends requests for data clarification (called queries) to the site so the CRFs can be corrected. This iterative, painstaking process is time consuming and inefficient for a number of reasons:

- The data collection organization has no control over data entry quality—anything can be entered on paper and the quality of data depends on such factors as the degree of motivation and learning of the person completing the form,
- The standard process usually involves delays between the time the data are recorded during a patient visit, the time they are transcribed onto the CRF, the time they lie waiting for the CRA to review and pick them up, and the time they are actually entered at the data processing site,
- The reduction in keystroke errors due to double-data entry is questionable and does not necessarily compensate for the time lost entering, correcting, and completing CRFs that are completed with inaccuracies (2),
- The data clarification query process is lengthy, iterative, and very labor-intensive, and
- Variations in protocol and CRF design can contribute to interpretation errors at sites and lengthy custom set-ups for each unique trial.

This typical atmosphere of inefficiency and redundancy can usually be extended to trials management and medical safety monitoring. For instance, trials management systems often require direct input of project events in a separate database—even though much of the key information could readily be gleaned from the CRF database if it were only accurate and available at the right time. And serious adverse events are typically entered into a third system, which results in another time consuming task of reconciling and matching the SAEs reported on CRFs with those reported through the safety monitoring system.

While mature data management organizations have found ways to shorten cycle times and increase efficiencies within the classical data management approach, they are often disinclined to make drastic changes such as eliminating unnecessary steps in the process which are needed to make dramatic improvements in timeliness and quality. And most of today's commercial clinical data systems, while feature-rich, are designed to support the classical approach, rather than reinvent it.

FROM CLASSICAL DATA MANAGEMENT TO ELEGANT MACHINE

To build an elegant machine for data management, one begins by establishing a series of guiding principles. Next the process is examined and broken down into its key functional components. One would then fit together a

suite of technology components to best address the individual steps in the process so that they collectively comprise an integrated, mechanized process streaming a continuous information flow to those who need it. Finally, one would have to ensure that it is properly implemented and keep measuring the results.

For this elegant machine, the following guiding principles would be employed:

- Plan each clinical study well in advance, against the context of the overall program, so that all elements can be effectively designed to facilitate the collection of all essential clean data,
- Standardize data structures, names, and codes in an integrated data repository covering the entire program,
- Push data capture upstream to the site and the patient as close to the point of patient contact as possible; minimize redundancies in data observations, transmission delays, transcriptions, and data transformations wherever possible,
- Assemble a repertoire of data capture tools and choose the "Best Fit" depending on the specific circumstances and objectives of each individual clinical program,
- Provide rapid on-line access to clinical data to give useful information back to all those who provide it, and
- Let technology drive the process wherever possible.

FUNCTIONAL ELEMENTS OF THE MACHINE

Applying this approach to data management, the process can be broken into five main components (Figure 1). First, the data are captured using forms or interfaces. Next, the data are stored in a central location that organizes them so they can be accurately understood and properly compared to other data. Third, the data are cleaned to correct inaccuracies and omissions. Fourth, the data are worked to compute derived variables and understand their message. Finally, the data are presented so they can be interpreted and acted upon by project managers, physicians, regulators, and other stakeholders. Each of these functional elements will be described in greater detail.

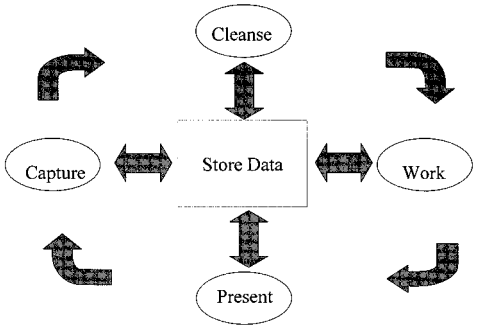

FIGURE 1. Functional elements of an elegant machine for clinical data management.

THE CLINICAL DATA REPOSITORY

Traditionally, data have been stored in formats provided by commercial data management systems, which usually employ different logical data models and thus make data interchange difficult. The data management systems were optimized for double data key entry, not the integration of data from advanced data capture tools. And there was far too much variation in database design and nomenclatures between companies, countries, drugs, and even individual studies for the same drugs. While these factors have made data interchange challenging in the past, evolving FDA/ICH interchange standards should begin to provide a framework that will eventually lead to a universal clinical data interchange format that can further be extended to data collection as well as regulatory submissions.

The idea of a data repository is the heart of the elegant machine. Rather than be considered an intrinsic part of a specific software package, it is a central warehouse where clinical information can be deposited in multiple forms and withdrawn for multiple purposes—a location that ensures that the data of record are properly organized, secure, and unambig-

uous. In many cases, the best starting point for a repository will be the mainline clinical data management system that is already in use augmented with data import/export tools. Most data management systems have the advantage of providing tools to facilitate the cleaning of data even if they are collected by other data capture methods. A second alternative would be to build a data warehouse that only accepts data collected from other systems but is optimized for data analysis purposes and data transfer using a conventional relational database management system. Ideally, this same data repository would also house trials management and serious adverse event reporting information without increasing data redundancies.

It is preferable to decouple the repository from the tools used to capture, process, analyze, and present data, so the best tool can be picked for each specific job. But the repository also establishes a basis for making more extensive use of the wealth of information that has become a by-product of clinical trials if one capitalizes on its potential as a data warehouse. When managed as a warehouse, it ensures that all types of data are consistent, and that multiple redundant copies are eliminated. It also provides information that can potentially be later mined for future analyses after the trial ends.

DATA CAPTURE

Data capture is the process by which clinical data are inserted into a database. The main problem with data capture by key entry under the classical model is that CRF information is often incomplete, inaccurate, and late. These problems can be collected once a "point of sale" approach to data capture is adopted—an approach that ensures the information is correctly captured electronically as soon as it is created. Investigators are often measured, however, more by their ability to find evaluable patients than by their data entry skills; to change this impression, they would need to recognize their purpose as one of providing not just patients, but accurate, timely patient information. Of course, to succeed in this role investigators need study designs that eliminate ambiguity and confusion and tools that help them trap errors at the source. Since this is not typical, the most common practical data collection tool today is still the paper CRF key entered into a database—using either double data entry or single data entry with computer-assisted edit checks and/or 100% verification of key variables. This is particularly true for those trials that have not been designed for other more effective methods. Key entry is thus the least common denominator—it works in almost any case, and sets the baseline for measuring potential improvement.

Key entry under the classical model, however, is not conducive to most of the principles defined above for the elegant machine. Instead, an elegant machine would provide a repertoire of data capture techniques that would best meet the needs of improving data quality, timeliness, accuracy, and cost effectiveness for the specific needs of a particular clinical trial or a group of investigators. There is already a proven experience base using techniques such as interactive voice response, remote data entry, and image recognition. Other emerging technologies such as Internet data collection systems, hand-held computers, voice recognition, and wearable monitoring devices are also beginning to become practical alternatives for collecting clinical information. All of these technologies have in common a dependence on up front planning to ensure that CRF designs are optimized and that data input quality checks are properly in place when the first data arrive. Each has its own advantages and disadvantages, and is suited for some types of trials more than others.

INTERACTIVE VOICE RESPONSE

Interactive voice response and other voice/telephony technologies provide real-time, interactive database access to users equipped with a touch tone telephone. Its advantage is that it is readily familiar in most developed countries, and can be used by any site with access to a telephone. Voice systems can sup-

port multiple native languages, can operate automatically around the clock for global trials and acute care scenarios, and provide immediate access to a real time transaction processing database without requiring implementation of a complex computer network.

Since most people cannot maintain a dialogue with a computer for more than a few minutes, however, IVR is suited only for applications that require minimal data entry. It is best for a brief series of yes/no, numeric input and multiple choice questions. Some people can get easily "lost" or confused while using an IVR system, and there still is site reluctance in some locations such as Eastern Europe and Asia. And system reliability can be adversely affected by poor telephone quality in some countries. So while this technology is eminently suitable for some applications—such as patient randomization or simple patient diary responses—it is not, at this time, a practical solution for conducting most large-scale clinical trials. Still, it provides a useful supplemental tool to apply to many clinical trials (3).

REMOTE DATA ENTRY

RDE has been promoted as the answer to clinical trials data collection for over a decade, but the supporting data are dubious (4). As a society that has witnessed the implementation of barcodes in supermarket checkout lanes and automated teller machines in banking, people can intuitively recognize the benefits of entering data directly into a database. RDE systems, which are typically run from laptop PCs connected by modem to computer networks, provide strict controls and audit trails for the collection of patient information. And RDE's ability to provide immediate feedback on data plausibility can eliminate up to 80% of entry errors at the source, with resultant savings for data management (The 80% figure has been cited in numerous presentations at DIA, IDC, and Barnett conferences by both customers and vendors, but has not, to the author's knowledge been proven in a documented, controlled study). RDE also allows CRAs to conduct some monitoring tasks remotely—shortening on-site travel time and helping to improve data quality earlier. Since computers are so useful for organizing and finding information, major efficiencies are gained simply from eliminating paper once all the study information is stored in computer format and from making data more immediately accessible to those who need it.

Or at least that would be the case if most RDE systems worked as advertised and if sites actually used them as sponsors would hope. But, in many cases the systems are found to be overly-complicated for the users, who do not feel adequately trained. The cost of a unique PC for each site and difficulties with site support and data transmission reliability can also be a significant obstacle. The need to assign one laptop per sponsor (or sometimes per protocol) can be a logistics nightmare for a site that is conducting several studies for different sponsors, all on different pieces of hardware using different software. RDE systems also are difficult to modify in the field. As a result of these perceived difficulties, many sites are still reluctant to support RDE. In the end, they do not entirely trust computers, and feel a lot more comfortable with paper even if it is less efficient, which, in turn, discourages sponsors away from RDE (5). While there have been many instances of successful experiences using RDE, there have also been many that were not particularly triumphant.

Newer RDE systems based on Internet or intranet technologies promise to ameliorate this reluctance assuming they can overcome the reluctance of sponsors to see multiple trials conducted on the same platform and allay performance and security concerns. If the data are stored on a central server instead of the site, however, the electronic CRF must still be transcribed from separate source documents that are still retained by the sites, which requires close monitoring. Thus, the first generation of Internet RDE products based on a host-centric model probably will not entirely resolve the tendency for delays in entry, transcription errors, and the need for rigorous monitoring that verify that the information was accurately captured from a real patient.

This is the appeal of RDE systems integrated with an electronic medical record system. In theory, if the electronic CRF is just another view of the original electronic source document, transcription errors and much monitoring effort would be eliminated. The integration of the system with direct patient care would support immediacy, accuracy, and consistency of results, and on-site monitoring time would be truly reduced once the EMR-provided CRF data viewed off-site become indistinguishable from the source document itself. Of course, the present high degree of fragmentation among EMR systems and the fact that clinical research design does not directly correspond with primary care processes makes this more vision than reality at this time.

RDE systems, however, are almost always appropriate for a Phase I environment and ideal for critical Phase II and complex Phase III trials (in developed countries) where rapid access to data is most important and when each extra day's delay counts against the bottom line. They are less likely to pay off for large scale, long-term global trials at this time, although this will improve as Internet/intranet systems mature and eventually become the tool of choice.

OPTICAL IMAGING DATA CAPTURE

Imaging technology is already commonly used among sponsors for preserving scanned or faxed electronic copies of CRFs and other paper documents. Surprisingly, these may actually cause increased costs and delays in data capture when the images are keyed into a data management system such as paper (6). Today's advanced image data recognition engines can now be applied to the scanned images of paper CRFs to extract data automatically into a database with dramatic gains in terms of cost and timeliness (7).

Optical imaging is especially appealing to data management staff and to sites which are already comfortable with paper CRFs because it does not require drastic changes to standard workflows. This means that much of the strain of implementing a new system can be circumvented. If the forms are faxed or mailed soon after the patient visit (rather than waiting to be picked up at the next CRA visit), it can still expedite the flow of information and payments, saving both time and money. Electronic images can be archived and stored on-line and are immediately ready for incorporation into an electronic regulatory submission.

The main drawback with image recognition is that it is inherently inefficient—it still promotes the transcription of data from source document to CRF to database in multiple steps, only faster. Since it cannot trap errors at the point of entry, it will not improve data quality as much as a good RDE system. These paper-based systems can lead to version control problems, since the data of record are still a paper form which must be updated and corrected, rescanned, and so forth. Excellent form planning and design is a critical prerequisite for ensuring quality in data capture. When implemented properly, optical imaging systems can yield many advantages.

Optical image recognition is useful on almost any trial, especially one that incorporates good CRF design techniques that use liberal spacing, preprinted index fields, and drop-out color shading. It is particularly appropriate for large scale simple Phase III and Phase IV studies where the economies of scale are most realizable. Optical imaging is best used for forms with tick boxes and constrained handprint boxes for text and numeric fields and numeric fields; free text must still be key entered in most cases. Optical image recognition systems may also miss detection of some unexpected peripheral comments, although these will be available for viewing on the source image (which is also likely to be used for the NDA submission).

OTHER EMERGING TECHNOLOGIES FOR DATA CAPTURE

There are many other emerging technologies which are promising candidates for inclusion in a repertoire of tools. For example, computerized voice recognition would be exciting

to apply to patient examination scenarios where the investigator could read out findings to a voice recognition computer that immediately populates a database. The data would then be later reviewed against the voice recording and approved by the investigator's electronic signature.

Another alternative with more immediate benefit is the hand-held data collection device (8). Some hand-held devices have been in use for several years with generally positive results when properly implemented, although widespread use has often been limited because of high per unit costs. The continued advances in the 3Com PalmPilot® and the hand-held PCs based on Microsoft's Windows CE® operating system make full-featured data collection a reality with data synchronization easily managed via serial cable, modem, or infrared. These tools are particularly interesting as take home patient diaries or survey registration tools.

Finally, there are several real-time measuring devices that are becoming available to track vital signs over time, or to record when medicine bottles are open and closed. While these can provide very useful data in many instances, they are generally cost prohibitive at this time for most clinical trials. As such devices become more widespread for general healthcare purposes, however, their suitability for clinical trials will increase.

OTHER FUNCTIONAL ELEMENTS

An elegant machine can provide other efficiencies in the areas of cleaning, working, and presenting the data collected during clinical trials. The standardized data repository creates the opportunity to use standard validation and data cleaning tools. Such tools are already provided by the mainstream data management systems and could conceivably be used when alternative systems are used for data capture. Some companies have created their own proprietary systems for cleaning data to augment the shortcomings of some of the commercial systems, and will find that these can be retrofitted to a repository with excellent results. In still other cases (such as RDE systems), the cleaning process may be managed by the data capture tool itself before the data enter the repository. An effective and well-integrated method of data clarification is an essential prerequisite to assembling a repertoire of data capture tools.

The process of working the data is still primarily controlled by statistical analysis tools, but much of the computation of derived variables, for example, can be merged with data capture in an EDC environment or in the repository itself. The matter of presenting the data will certainly continue to be combined with statistical analysis, but will begin to capitalize on graphical, web-based data analysis tools which will be applied against the repository from the earliest stage of the trial through regulatory submission and beyond. Through the universal medium of secure Internet access timely information will become more rapidly available to the investigator, study coordinator, medical sponsor, project manager, medical monitor, data manager, laboratory, biostatistician, regulatory authority, and even the patient who will have new opportunities to view, understand, and act as a result.

CHOOSING THE BEST FIT TOOL

Once the advantages of alternative data capture methods are better understood and applied in a standardized manner, it is possible to select the best fit tool or tools to meet the needs of a particular trial. For example, a large Phase III study may use IVR for randomization, image recognition for CRF processing, hand-held or palmtop computers for collecting patient diary data, and a traditional data management system for data cleaning.

While putting together such a portfolio can be quite a challenge today, tomorrow's elegant machines will simplify this technology integration process, enabling sponsors to offer investigators the option of specifying their own best fit tool for a study, irrespective of what the other sites choose. Table 1 shows one approach for defining criteria to choose

TABLE 1
Criteria for Choosing the Best Fit Data Capture Tool

Criteria	Classic	PC RDE	Image recognition	IVRS
Best phase	II–III	I–III	III–IV	II–IV
Number of sites	No limit	1–60	No limit	>1-Unlimited
Lead time required	1–2 weeks	8+ weeks	2–4 weeks	4+ weeks
User computer experience	None	Basic	None	None
Patient/site ratio	Low to high	High	Low to high	Low to high
Cost savings	Neutral (Baseline)	Neutral to low	Low to high	Neutral to middle
Study duration	No limit	<2 years	No limit	No limit
Time to database lock	8–12 weeks	1–4 weeks	1–4 weeks	Neutral

Note: Actual figures may vary depending on specific study parameters.

the best fit data capture tool for certain kinds of trials.

CONCLUSIONS

It would be convenient to conclude by leaving the impression that a repertoire of data captures tools tied together is a simple and manageable task. After all, elegant machines could all be engineered by following a simple recipe:

- Select a standard data repository,
- Standardize data forms and structures,
- Standardize validation systems,
- Assemble a repertoire of data capture tools, and
- Implement them well and continuously improve the process.

Once again, however, real-world experience with technology is a reminder that such efforts are very demanding, fraught with pitfalls, and totally dependent upon the talents and commitment of many hard working, fast learning, multiskilled people. The best of the best fit tools are still more of a glimmer in their creator's minds than reality today, but technology, as everyone knows, is rapidly ever improving. Yet, the fundamental elements discussed above are already readily available, and the principles, goals, and potential of this approach offer a realistic means to improve the R&D process now. When the process and the technology are improved with measurable results, everyone wins, since new, improved drug therapies can be delivered sooner (9). Thus the pharmaceutical industry—developers, vendors, regulators, and customers—must all work together to make the way research and development is conducted hum like an elegant machine.

REFERENCES

1. Gelernter D. *Machine Beauty: Elegance and the Heart of Technology.* New York: Basic Books; 1998.
2. Gibson H, et al. Is double data entry necessary? The CHART trials. *Controll Clin Trials.* 1994;15: 482–488.
3. Schoichet, Creasey, Kasay. Systems Infrastructure for Sponsor/CRO Collaboration on International Clinical Trials. *Drug Inf J.* 1998;32:155–161.
4. Coats, Connolly, et al. Adventures in computerized remote data entry. *The Monitor.* Spring, 1997.
5. Regalado A. Re-engineering drug development, II: Clinical data collection & management. *Start-up*. Feb. 1998;27.
6. Lawson, Brown, Shostak. Results of a data imaging pilot project. *Drug Inf J.* 1997;31:194–195.
7. Schantz HF. Optical character recognition: The mature technology with the brilliant future. *Inform*. July 1996;8–14.
8. Tiplady C, et al. The use of electronic diaries in respiratory studies. *Drug Inf J.* 1997;31:759–764.
9. Lightfoot, Getz, et al. Faster time to market, ACRP's White Paper on future trends. *Appl Clin Trials.* April 1998;66.