Dear CRF, Please Meet the Trial Protocol:
Intelligent Electronic Data Capture (iEDC) with Expert System Redefines Clinical Data Collection

MITCH SCURTU

Mitch Scurtu *has been a Data Management professional for 26 years. He holds four Master's degrees, and has developed two technologies - iEDC (intelligent EDC) and OGP (Optimal Global Pricing).*

iEDC DEFINITION

iEDCs are EDCs that can take care of the quality of clinical data at no extra cost. The intelligence infrastructure prevents and contains errors at the source. When Data Entry is complete, the clinical databases are virtually FDA and EMA ready.

ABSTRACT

EDC developers / vendors are in a race to move data cleaning up front where errors are happening as being the more cost effective data management (DM) operation. The limitation on how much data cleaning can be done upfront comes from the IT EDC systems that are built to essentially clean data with queries at the backend of the DM operation.

iEDC is relaxing this systemic limitation. 99.9% of typical errors in a clinical database can be prevented or contained from entering the database.

An FDA-approved trial protocol is expected to be followed during a clinical trial. There are several ways to secure the protocol compliance in the DM portion of the clinical trial.

- By process, and there are as many procedural ways to secure protocol compliance as there are DM shops.
- By a systemic approach, which may be the more cost-effective method. The protocol is build into Data Capture, Data Management and Data Extraction. The distinct benefit of the systemic approach is that this will result into a quality controlled environment. Data is clean upon data entry and alleviates or obviates the extra effort needed to clean data at the backend of the DM operation.

INTRODUCTION

Randomized controlled clinical trials constitute the main foundation for the regulatory approval of drugs and devices in the United States. Frequently constituting expensive, complex operational endeavors, clinical trials are not only a crucial element in determining health care practice but also represent the main access channel to a market in excess of billion dollars every year. Despite the significant amount of research, operational cost, and complexity in conducting clinical trials, in the end a trial is represented by its dataset. This dataset is expected to be in full compliance with protocol specifications while also carrying high data quality standards. As we will argue in this manuscript, these assumptions might not always be met.

In the era preceding data entry using electronic data capture (EDC) systems, error rates were primarily associated with the transference of information from paper-based case report forms to electronic records. With the advent of EDC, the evaluation of data quality became more subtle and therefore complex. With the exception of studies making use of dual data entry (ie, two independent data entries with comparisons), the traditional method of comparing data sources is no longer available. Data quality metrics are now mainly restricted to issues such as missing and implausible value rates, inconsistencies across fields, and missing encounters. While these are important sources of error, they are by no means the only types of error and arguably not the most important.

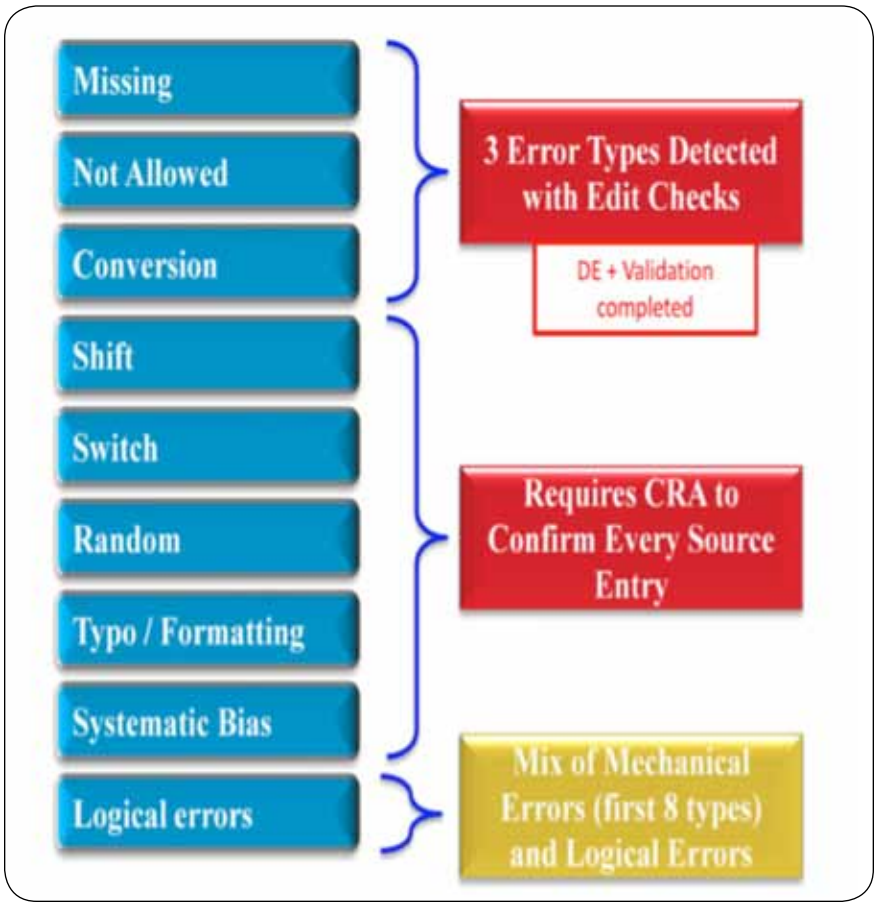
For example, errors originated from a failure to comply with protocol guidelines might not be obvious while entering information into the EDC or, worse, when analyzing the data and generating information that will guide clinical and policy decision making.

Error Research into clinical databases and DM processes yielded the Error Profile of clinical databases. Part of the Error Profile is the error classification below.

Inferences from this error classification:

The EDC IT solutions to capture clinical trial data

- May not rigorously observe **protocol requirements** for data quality as the cost of the manual data cleaning may exceed allocated budgets
- Skills of highly qualified clinical personnel may be underutilized while cleaning **mechanical errors**



Edit Checks in EDCs detect typically two - three types of mechanical errors. Remaining five types of mechanical errors and logical errors will have to be cleaned manually. Two distinct sub-processes have emerged in an EDC operation: Data Capture and Data Cleaning. As per Deming and TQM theory, this may not be the most cost effective DM operation.

- Adaptive Clinical Trials capabilities may be limited as EDC databases will be clean late in the clinical trial process
- Delays in both locking databases and FDA review times may lead to **missed medical and market opportunities**.

"Experts say EDC software isn't ready to support the clinical trial needs of large pharmaceutical companies and they are right" as per Jashua Walker and Michael Barrett (*Business View Brief*, June 13, 2010). The promises that came with the emergence of EDC technologies of higher quality clinical data, lower DM cost, and shorter DM processing times are yet to be delivered. Data quality problems in EDC operations is contemporary with a 600% increase in recalled drugs. The complexity of the EDC IT solutions can render the DM operations unmanageable, which in turn is further impacting both protocol compliance and quality/cost of clinical databases.

iEDC SYSTEM DESCRIPTION

The theoretical background of iEDCs are the precepts of Total Quality Management (TQM) as an integrative philosophy of management for continuously improving the quality of products, processes, and services. These precepts include Quality Systems, Quality Processes, and Quality Clinical Data as necessary conditions for quality of services. For the sufficient conditions for quality services and as a part of the Error Profile of clinical databases, Error Research has identified and quantified the error sources in the clinical DM operation. Error types have been identified, quantified, and matched to error sources. The intelligence infrastructure on four levels – field, form, application and global – has been developed to prevent errors entering the clinical database or contain error sources. At the field level (outer shell), 57 field attributes will define the quality of data points as per protocol. At the form level (second shell from the outside), the intelligence infrastructure will prevent errors occurring between eCRF forms, visits as per protocol. At the application level (third shell from the outside), the Automated Navigator and Intelligent Cursor are implemented together with the remaining protocol provisions. At the global level (first shell from core), the previous three intelligence levels are integrated to attain the congruity of the system. The core represents TQM.

NECESSARY CONDITION

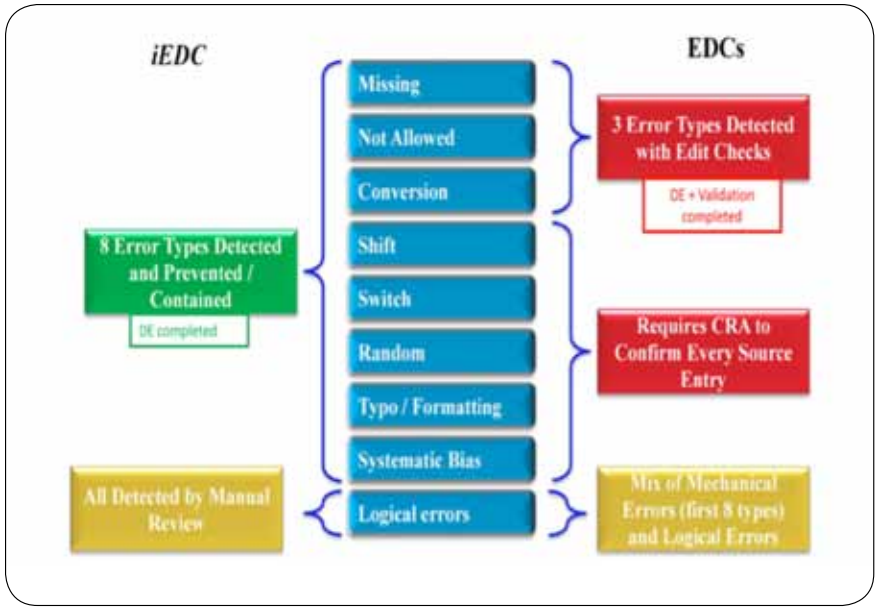
The iEDC system underwent a ten years continued development / improvement cycle.

- Quality System is intelligence based rather than query based
- Quality Processes were mapped to match the Quality System. This Process Map enabled development of a Project Management Plan Template to provide for
  - Planning the study
  - Organizing the study
  - Managing the study
    - > Task management
    - > Timeline management
    - > Resource management
    - > Budget management
  - Continuous feedback from DM to all horizontally related functions; ie, clinical, medical, statistics, regulatory, etc.
- Quality Clinical Data: The data manager is in charge of the quality of the clinical database while running the DM operation.

Data Manager is communicating effectively and continuously with all related functional groups enabling **Fact Based Decisions**.

SUFFICIENT CONDITION

- Error Research identified and quantified all error sources of the intelligent DM Operation
- Error Research over a ten years time span identified and quantified error types and matched them to the error sources
- Continuously develop the Intelligence Infrastructure consisting of procedures, algorithms, methods to prevent errors and / or contain error sources at the Field, Form, Application, and Global levels



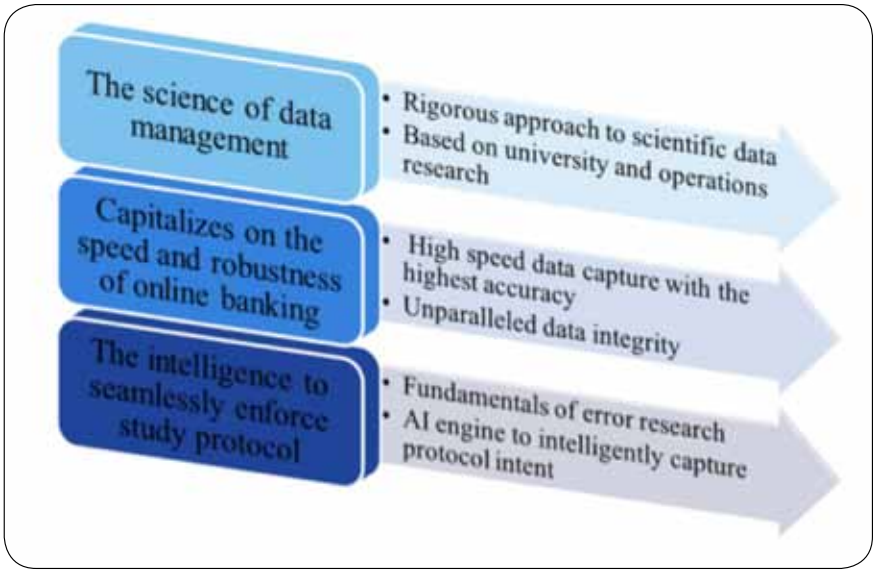
The iEDC system can prevent and/or contain eight types of mechanical errors shortening the data processing time and also the time the FDA requires to review clinical databases. When clinical data enters the iEDC database it is on average 0.1% error rate clean. Clinical trial setup time, data processing time, and database lock in iEDC, live up to the claims of EDCs.

The development effort of iEDCs has yielded Three Breakthroughs:

- iEDCs employ a rigorous approach to scientific clinical data, based on research from universities and the precepts of TQM / Operations Research.
- iEDCs operate on a high speed data capture system with high accuracy and data integrity as the Study Protocol is leveraged in all data capture/management/ extraction

AND last, but not least

- iEDCs rely on the Fundamentals of Error Research with an Artificial Intelligence Engine (Expert System) to capture the Study Protocol intent.



CONCLUSIONS

As the clinical data management community is expecting more efficient tools, EDC developers/ vendors are incorporating progressively more intelligent features into their technologies converging infinitely and asymptotically to the status of an iEDC.

iEDCs are positioned as:

- The next generation of EDCs and
- The key to integrating EHRs with clinical trials. ●

References/citations from this article are available upon request.