Possible Combinations of Electronic Data Capture and Randomization Systems

Principles and the Realization with RANDI2 and OpenClinica

D. Schrimpf1; M. Haag2; L. R. Pilz3
1Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany;
2Heilbronn University, Heilbronn, Germany;
3University of Heidelberg, Medical Faculty Mannheim, Mannheim, Germany

Keywords
Random allocation, software design, electronic data capture, randomized controlled trial, open source

Summary
Background: Clinical trials (CT) are in a wider sense experiments to prove and establish clinical benefit of treatments. Nowadays electronic data capture systems (EDCS) are used more often bringing a better data management and higher data quality into clinical practice. Also electronic systems for the randomization are used to assign the patients to the treatments.

Objectives: If the mentioned randomization system (RS) and EDCS are used, possibly identical data are collected in both, especially by stratified randomization. This separated data storage may lead to data inconsistency and in general data samples have to be aligned. The article discusses solutions to combine RS and EDCS. In detail one approach is realized and introduced.

Methods: Different possible settings of combination of EDCS and RS are determined and the pros and cons for each solution are worked out. For the combination of two independent applications the necessary interfaces for the communication are defined. Thereby, existing standards are considered. An example realization is implemented with the help of open-source applications and state-of-the-art software development procedures.

Results: Three possibilities of separate usage or combination of EDCS and RS are presented and assessed: i) the complete independent usage of both systems; ii) realization of one system with both functions; and iii) two separate systems, which communicate via defined interfaces. In addition a realization of our preferred approach, the combination of both systems, is introduced using the open source tools RANDI2 and OpenClinica.

Conclusion: The advantage of a flexible independent development of EDCS and RS is shown based on the fact that these tool are very different featured. In our opinion the combination of both systems via defined interfaces fulfills the requirements of randomization and electronic data capture and is feasible in practice. In addition, the use of such a setting can reduce the training costs and the error-prone duplicated data entry.

1. Introduction
Clinical trials (CT) are in a wider sense experiments to prove and establish clinical benefit of treatments. In the last decades governmental and supra-national institutions installed laws and rules to conduct these sensitive research projects. It is believed that the best way to compare two treatments is to randomize the patient allocation guaranteeing treatment by equivalent chance. Thus, a patient population with common characteristics is divided into different groups, getting in most cases an already accepted standard (controls) and new treatments without official approval. With the right sample size based on the primary study hypothesis, resulting demographic differences in the groups will be minimal and makes a comparison possible. Thus it can be taken that randomized clinical trials (RCT) are the method of choice whenever a scientific comparison of treatments should be considered [1].

To perform such a RCT patient’s information has to be collected for the process of randomization and the CT itself. An important postulation of randomization is that it must be performed independently and uninfluenced by the investigators, clinical staff, and patients [2]. It has been good practice for decades to outsource randomization into a separated, independent institution or to use an electronic un-manipulable device. Thus, in both tasks randomization and the following CT procedure data are collected often separately. Typically, some demographic and clinical data acquisition is in parallel, especially in
case of stratified randomization by demographic and clinical aspects (like age, gender, data of physical exams, pretreated patients, etc.) [3]. Additionally, for adaptive randomization processes further information from already treated patients are required to adjust allocation for ending up with a total smaller sample or by treating more patients with the probably better treatment [4]. This information should be available instantaneously for randomization systems (RS). For both tasks a way of good, exact, and ready to hand data management should be mandatory. Nowadays electronic data capture systems (EDCS) are used more often bringing a better data management and higher data quality into clinical practice [5]. Coincidently, the usage of such systems can save costs [6]. There are several strategies in realizing EDCS dependent on the main task to be accomplished [7, 8].

In conducting randomization in a CT identical data are collected in RS and EDCS, respectively. Separated data storage may lead to data inconsistency and anyhow, data samples have to be aligned. The article discusses solutions to combine RS and EDCS. In detail one possible solution is realized and introduced.

2. Material and Methods

To conduct a clinical trial several tasks have to be executed. In planning a study for RS and EDCS the following points are relevant: i) study endpoints, ii) collection of basic data/patient characteristics, iii) randomization method (RM), and iv) collection of longitudinal data. During the study course monitoring is essential to obtain good data quality and to register unexpected events.

2.1 Randomization and RS

Many RM are used depending on the different study situations. The simplest method “complete randomization” assigns the patients equally to treatments avoiding selection bias without the use of preconditions and patient characteristics. The disadvantage of this procedure is the possibility of an imbalance. Other RMs balances the assignment more directly with the possibility to do this also for subgroups, e.g. sorted by gender or age cohorts. To perform such “stratified” randomization it is necessary to collect the considered patient properties during the assignment process [3]. Some new RMs makes use of information of the study course, e.g. to adapt the allocation in favor of the probably better treatment. This can be done by using a binary treatment outcome (success or failure) to adjust the allocation [9]. It is obvious that such procedures need a RS for realization including the possibility to use patient and study information and they are not feasible with classical paper based randomization lists. Further advantages of an electronic RS are the possibility to check the inclusion and exclusion criteria automatically during the randomization to avoid failures and furthermore, the independence of the accessibility of a centralized study office.

2.2 EDCS

The main task of an EDCS is data entry either manually or by import from other electronic systems like laboratory facilities. This digitalized information is easing monitoring and interim analyses. As a direct consequence an EDCS also ends up with a higher data quality, e.g. by checking the data immediately during the submission. If failures during the data collection appear, EDCS can support a query management. This means that it is possible for the monitor to contact the investigator to clarify the contentious issues. Other important functions every EDCS has to support is i) a full audit trail, in which every action is logged for a complete traceability; ii) instantaneous check of input data for plausibility; iii) double data entry (DDE) by two independent users [10]; and iv) export of data for reporting and analyzing.

An EDCS should support common electronic standards: For the highest level of abstraction the definition of clinical trials and the exchange of operational data, the “Operational Data Model” (ODM) [11] is provided by “Clinical Data Interchange Standards Consortium” (CDISC). Advantages using CDISC standards are for example the possibility to exchange the study metadata between different EDCS solutions [12] or to archive the study data [13]. In case of web-based applications technical standards for communication processes should be used, e.g. SOAP (originally Simple Object Access Protocol) [14], recommended by the “World Wide Web Consortium” (W3C) [15]. There are different EDCS available, some of them are complete commercial solutions like the system secu'Trail [16] and others are based on open-source licenses like the system OpenClinica [17] with the possibility to get additional commercial support.

Since the EDCS is the primary point for data entry with high data robustness, it should be the preferable source of information for randomization. In particular by the usage of response adaptive RMs intersections between the RS and EDCS will occur occasionally.

3. Results

There are three possible scenarios for the usage of EDCS and RS. Figure 1 illustrates the variants schematically. i) Variant I assume two complete independent systems; ii) in Variant II both tasks are realized in one system; and iii) in Variant III both systems are separated and communicate via defined interfaces. The advantages and disadvantages of all variants are shown and Variant III is explained as an example realization.

3.1 Variant I: Two Independent Systems

Historically, randomization was separated from the data management resulting that two separated installations exists, either based on electronic devices or on paper based procedures. Since paper based procedures are not feasible for complex randomization methods, they are not considered in detail.

A positive aspect of using two independent systems is the flexibility, since both systems are not linked with the implication of independent development and separation of tasks.

The solution has some disadvantages: i) the need to install and maintain two systems increasing the effort of adminis-
It; ii) users trainings for both systems; iii) the configuration of the trial and data entry separately in both systems. Especially the last point leads to a duplicated data entry which possibly can cause inconsistent databases, which can be moderated with an extensive monitoring and increased costs.

Examples for this variant are independent RS are like the “randomizer” from the Medical University of Graz [18], RANDI2 [19], or MinimPy [20] used in parallel with an independent EDCS like secuTrial [16], OpenClinica [17], or REDCap [21].

3.2 Variant II: Randomization Integrated into EDCS

Another approach is to integrate both systems, which means that data entry and randomization take place in one application. A great advantage is that only one system has to be installed and maintained, and that the users need only training and operation skills in one system. The main advantage in view of data quality is that the independent data entry into two separated systems is no longer required preventing failures. Thus, the integrated randomization component benefits directly from the better data quality of the EDCS. The disadvantages are the different requirements of EDCS and RS, because it might be difficult to find and implement an adequate solution. Especially in case of new adaptive RMs, the RS needs to be flexible. As a consequence, a realization of Variant II is more complex and difficult to adapt for specific settings. An example for a realization of this variant is the mentioned system secuTrial which integrates basic randomization methods directly into the EDCS [16].

3.3 Variant III: Two Systems with Defined Communication Interfaces

Our preferred attempt is to combine both systems via a defined interface, trying to prevent the cons from Variant I and sustaining the pros of Variant II. In such a setting data entry takes place in the EDCS, and are subsequently handed over into the RS. The use of standard technologies is indicated in such an implementation. With the mentioned CDISC ODM and SOAP it is possible to create standardized web services for the data interchange in context of clinical trials.

Thus, the systems supporting standards can be chosen flexible, e.g. it is possible to use a RS for a special randomization method or a RS with a simulation component with the preferred EDCS.

To reach a connection between the systems with preferable minimal user interactions and without separate data entry, the systems should provide interfaces outlined in the following by showing the complete process of communication between the systems and users. First, the new clinical trial has to be generated in both systems. The main part of this task should happen in the EDCS, since RS normally needs less information. Responsible for this task is the principle investigator, data manager or statistician. To avoid the duplicated entry of the trial structure the EDCS should offer the possibility to export the study metadata and the RS should be able to import and map this information to its internal structure. For reasons of data reduction and data economy only the necessary elements for the randomization process should be used. Selection of elements and configuration of the RM are normally done by a statistician. To ensure secure communication authentication between user and systems, and between systems are required.

After the initial configuration of the trial the phase of patient recruitment begins. Involved in this task are i) investigators for the entry of the patient data, and ii) monitors to control the data quality. These user groups should not necessarily interact with the RS. In an ideal setting the randomization should be triggered automatically after a patient is added in the...
EDCS by an investigator. Following the triggered randomization the RS takes care of the synchronization process and interacts with interfaces of the EDCS. Since in the configuration phase of the study the RS has to use the interfaces of the EDCS, business logic for the communication has been already available. Now RS requests the data of the new patient from the EDCS and performs the randomization. Afterwards, the randomization result should be delivered to the user. This can happen for example with an email notification. In addition for unification of the data sources the RS should transfer the result to the EDCS.

The procedure explained above will be enhanced in case of an adaptive RM. The main additional component is synchronization of patient data during the trial. These data are necessary to adapt the randomization process. There are three different solutions to execute the synchronization of patient data, i) triggered by the complete entry of a case report form (CRF) in the EDCS, ii) immediately before a patient is randomized, or iii) at fixed time intervals.

Table 1 summarizes the necessary functionality an EDCS should provide to an external system.

### Table 1  
Necessary functions an EDCS has to offer to an external RS  
<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentication</td>
<td>Possibility to authenticate via interface</td>
</tr>
<tr>
<td>List of available trials</td>
<td>Access to a list of all available trials</td>
</tr>
<tr>
<td>Study metadata</td>
<td>Availability of metadata of a specific trial offering the possibility to configure the randomization algorithm</td>
</tr>
<tr>
<td>Retrieval of patient information</td>
<td>The access to the documented information of a patient (initial and course data) to execute the randomization</td>
</tr>
<tr>
<td>Entry of patient information</td>
<td>The possibility to write patient data in the EDCS to add the randomization result</td>
</tr>
</tbody>
</table>

#### 3.4 Example Realization

To demonstrate the feasibility of the combination of EDCS and RS with defined interfaces an example realization is implemented. The EDCS OpenClinica [17] and the RS RANDI2 [22] is used.

Both systems are chosen in regard to their range of functions and their open-source nature. Especially, the open-source nature opens the possibility to adapt the applications freely for example realizations.

RANDI2 is a flexible system which allows conducting the randomization by a web interface and was actually developed with our participation. Thereby RANDI2 supports common RMs, like i) complete randomization [23], ii) truncated randomization [23], iii) block randomization (fixed and variable block sizes) [24], iv) Weiß’s Urn design [25], v) minimization [26], and vi) stratified response adaptive randomization [27]. The randomization methods are realized as plug-ins, thereby it is possible to implement new or to adapt existing procedures easily for a specific study setting. In general, RANDI2 has a modularized architecture opening the possibility to extend the application with a module for the communication with an EDCS.

As counterpart of RANDI2 the EDCS OpenClinica in version 3.1.4 is used, since this application is among others i) open source and ii) a web service component is already available for communication with external applications. Also, OpenClinica provides the typical functionality of an EDCS. The web service component of OpenClinica provides a SOAP interface to exchange CDISC ODM messages.

Table 2 lists the available web services with their methods and short descriptions. According to the interfaces requested in Variant III most of the services already exist, only one additional web service needs to be implemented to access the stored patient data. Therefore, the existing “Data” web service was enhanced with a method called “exportDataSet”. This web service requires the identifier of i) the study, and ii) the export dataset (method of OpenClinica to export the stored information in various formats). With this additional web service all necessary EDCS interfaces are completed.

The main part of the realization is done in the RS including the following graphical user interfaces i) list of all available trials (uses “Study” web service); ii) import of a specific trial (uses “Study” web service); iii) configuration of the trial. In addition to the graphical components the following server functions are implemented: i) import of the subject data for the randomization process (uses the new “Data” web ser-
vice); ii) export of the randomization result (uses “Event” and “Data” web service).

In the following, the process from trial generation to randomization in Variant III is shown: i) creation of the study in OpenClinica including event/CRF definitions (and also including the treatment definition); ii) creation of the export dataset including all randomization relevant information. The scope of the export is set to complete data sets, thus only patients with the full randomization information will be exported; iii) configuration of the study in RANDI2 including the selection of the trial with all relevant CRF items and the RM with corresponding parameters and stratification definitions. In principle the amount of stratification factors can be defined freely in RANDI2, but from a statistically point of view the number of stratification factors is limited by the sample size and study design.

After the finalized configuration the recruitment process can start. The investigators can add new patients in OpenClinica and the randomization takes place automatically by a click at an additional link in the subjects’ view of OpenClinica. If RANDI2 is triggered the patient information is synchronized. Thereby, the data defined in the OpenClinica dataset are transferred to RANDI2 in a CDISC ODM format. Since OpenClinica only exports patients data with complete randomization information (defined in the data set), just these patients will be randomized. To inform the investigator and the monitors about the randomization result RANDI2 send a notification email and add the result to the configured treatment event/CRF.

The example realization is tested with data of the completed GemTax III study. For the test the block randomization of block size 8 with the stratification factors ‘trial site’ and ‘gender’ is used.

4. Discussion

Electronic system can support the execution of clinical studies. In this concern one important application is the EDCS, which handles the data entry and offers methods to increase the data quality, e.g. with immediately checks of the entered data or the support for the DDE method. Using such a system can increase the data quality and decrease costs of data management.

Another important tool is a RS, opening the opportunity to perform the randomization independently from a centralized study office. Additionally a RS can check the inclusion and exclusion criteria automatically to prevent failed randomizations. So far we know there is no empirical study available, which compares the error rates of a manual and electronic randomization.

The article shows three possibilities of separate usage or combination of EDCS and RS. i) In Variant I it was assumed that the two systems are complete independent; ii) in Variant II both tasks were realized in one system; and iii) in Variant III both systems were separated and communicate via defined interfaces.

Variant II has the advantage that only one system needs to be maintained and trained; on the other hand a direct integration increases the system complexity. Especially one has to obey that both tasks – documentation and randomization – are intrinsic very different. Variant II for example is realized by the commercial solution secuTrial which integrates common randomization methods like block randomization directly.

In our opinion Variant III is preferable, since this solution preserves the flexibility of a complete independent approach of both, documentation and randomization, and contains the reuse of the entered patient data like in Variant II. Furthermore, the pure handling for investigators of Variant II and III is identical during the study.

To show the feasibility of Variant III for multi-center trials with electronic data capture and randomization the combination of OpenClinica and RANDI2 is realized. The main advantage in such a setting in comparison to Variant I is that the data have to be entered only once. This solution necessitates less effort in training and data monitoring as in a complete independent solutions. Particular advantages of Variant III are given in case of adaptive RMs, since with such procedures the necessary amount of data which needs to be synchronized and checked will increase. This approach is still to be tested in a clinical study.

Besides the chosen applications in our solution other systems can be used as basis for a realization. If a flexible RS is used, it is relatively easy to implement new RMs, assumed that the necessary skills are available. Since in Variant III the whole communication between the EDCS and RS take place by defined interfaces, it is possible to change the RS for a new study without an effect on processes investigators already are involved. Often commercial offers of EDCS and RS are restricted in the possibility to adapt the application freely. This was the reason we used only open source software for our example realization. In consequence of it the internal structure of such applications is public available and the implementation of additional components is possible without restrictions.

Besides the organization form of the applications, some of them are web-based and others are desktop solutions. Web-based approaches are favorable in a multi-center setting through their flexible accessibility by an Internet browser. This means that it isn’t necessary to install special software at study centers, such that web-based applications should be preferred in our eyes.

RANDI2 and OpenClinica fit together with their open source properties and in addition, OpenClinica offers a good connectivity to other applications in a clinical data setting. The thereby used techniques are based on CDISC and W3C standards, which are widely used. CDISC provides the common standards whenever clinical studies are involved, forced by the US Federal Drug Administration (FDA). By the fact that the realization is based on open-source applications, the solution is freely available. Furthermore, it is intended to integrate the additional implemented OpenClinica web service component in its main project.

Generally, once entered data should be used in multiple contexts to avoid the error-prone duplicated entry of data. Such a multiple use of information can lead to higher data quality, faster availability, and lower costs. Especially these systems are useful if special methods and tools are implemented to check and increase the quality of the data entry. Having in mind these aims and easy handling of the system
for the user a combination of both systems seems to be favorable what we have shown in two approaches, Variant II and III. Since the EDCS and RS are very different featured a flexible independent development of those systems is of advantage. In our opinion Variant III fulfills these requirements and is manageable in practice.

References

17. OpenClinica (Internet). OpenClinica, LLC; c2013 (cited December 02). OpenClinica 3.1 User Documentation; Available from: https://docs.openclinica.com/3.1/
18. Randomizer (Internet). Medical University of Graz; c2013 (cited December 02). Web-Based Patient Randomization (Randomisation) Service for Multi-Center Clinical Trials. Available from: http://www.randomizer.at