

# Workflow in Clinical Trial Sites & Its Association with Near Miss Events for Data Quality: Ethnographic, Workflow & Systems Simulation

Elias Cesar Araujo de Carvalho<sup>1,2,3</sup>, Adelia Portero Batilana<sup>1</sup>, Wederson Claudino<sup>4</sup>, Luiz Fernando Lima Reis<sup>5</sup>, Rafael A. Schmerling<sup>5</sup>, Jatin Shah<sup>6,9</sup>, Ricardo Pietrobon<sup>7,8,9\*</sup>

**1** Research on Research Group, Department of Surgery, Duke University, Durham, North Carolina, United States of America, **2** Cesumar, University Center of Maringá, Paraná, Brazil, **3** UEM, State University of Maringá, Paraná, Brazil, **4** Oncologistas Associados Cancer Institute, Rio de Janeiro, Brazil, **5** Institute for Teaching and Research, Sirio Libanes Hospital, São Paulo, Brasil, **6** Research on Research Group, Duke-NUS Graduate Medical School, Singapore, Singapore, **7** Department of Surgery, Duke University Health System, Durham, North Carolina, United States of America, **8** Duke - NUS Graduate Medical School, Singapore, Singapore, **9** Research on Research Group, Duke University, Durham, North Carolina, United States of America

## Abstract

**Background:** With the exponential expansion of clinical trials conducted in (Brazil, Russia, India, and China) and VISTA (Vietnam, Indonesia, South Africa, Turkey, and Argentina) countries, corresponding gains in cost and enrolment efficiency quickly outpace the consonant metrics in traditional countries in North America and European Union. However, questions still remain regarding the quality of data being collected in these countries. We used ethnographic, mapping and computer simulation studies to identify/address areas of threat to near miss events for data quality in two cancer trial sites in Brazil.

**Methodology/Principal Findings:** Two sites in Sao Paulo and Rio Janeiro were evaluated using ethnographic observations of workflow during subject enrolment and data collection. Emerging themes related to threats to near miss events for data quality were derived from observations. They were then transformed into workflows using UML-AD and modeled using System Dynamics. 139 tasks were observed and mapped through the ethnographic study. The UML-AD detected four major activities in the workflow evaluation of potential research subjects prior to signature of informed consent, visit to obtain subjects informed consent, regular data collection sessions following study protocol and closure of study protocol for a given project. Field observations pointed to three major emerging themes: (a) lack of standardized process for data registration at source document, (b) multiplicity of data repositories and (c) scarcity of decision support systems at the point of research intervention. Simulation with policy model demonstrates a reduction of the rework problem.

**Conclusions/Significance:** Patterns of threats to data quality at the two sites were similar to the threats reported in the literature for American sites. The clinical trial site managers need to reorganize staff workflow by using information technology more efficiently, establish new standard procedures and manage professionals to reduce near miss events and save time/cost. Clinical trial sponsors should improve relevant support systems.

**Citation:** Araujo de Carvalho EC, Batilana AP, Claudino W, Lima Reis LF, Schmerling RA, et al. (2012) Workflow in Clinical Trial Sites & Its Association with Near Miss Events for Data Quality: Ethnographic, Workflow & Systems Simulation. PLoS ONE 7(6): e39671. doi:10.1371/journal.pone.0039671

**Editor:** David Joseph Diemert, The George Washington University Medical Center, United States of America

**Received:** May 13, 2011; **Accepted:** May 29, 2012; **Published:** June 29, 2012

**Copyright:** © 2012 Carvalho et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The study was funded by Sanofi Aventis. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** This study was funded by Sanofi Aventis, with Mr. Carvalho and Batilana being directly compensated as consultants. Dr. Claudino is currently conducting clinical trials with Sanofi-Aventis. Dr. Reis is Chief Scientific Officer of CRC Sao Paulo. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials, as detailed online in the guide for authors.

\* E-mail: rpietro@duke.edu

## Introduction

While clinical trials have typically been conducted in developed countries such as the United States of America, developing countries have recently emerged as important new locations for clinical research [1]. In particular, the BRIC countries (Brazil, Russia, India and China) and VISTA (Vietnam, Indonesia, South Africa, Turkey, and Argentina), with their large and ethnically diverse populations, have become major players in the globalization of clinical research [2]. At the same time, pharmaceutical companies are differentiating clinical research capabilities across these countries to inform their

investment decisions. Among the most important capabilities is the ability to deliver high quality data. As a result, efforts to improve and ensure the production of high quality data in all countries are essential for enhancing their competitiveness as viable locations for the conduct of clinical trials, especially in light of previous studies that have linked research in developing countries with lower levels of data quality [3].

Even usual errors in data acquisition and transcription during clinical trials can directly affect data quality, due to the additional effort necessary to correct possible errors and ensure an acceptable level of data quality. The typical example is that of a busy clinical research coordinator (CRC) who is not adequately trained to

collect data and who passes the task of data entry to another CRC who might enter data that he/she does not understand.

While previous studies involving clinicians, radiology interpretation, healthcare management systems, have evaluated the relationship between workflow and data quality [4,5,6,7], most have been conducted in a clinical rather than a research environment, frequently using workflow data that had been automatically captured. Although this approach is appealing, it does not apply to a clinical research setting where each step of a Principle Investigator (PI), Clinical research coordinator (CRC), nurse, receptionist and pharmacist is not automatically captured in an electronic system. Moreover, previous studies [5,8,9,10,11,12] investigated methods of standardizing and improving the quality of clinical trial procedures. Between them, only a few workflow studies were conducted in research settings using a pure ethnographic approach [13].

Recent studies have used Unified Modeling Language (UML) [14] - a method that helps assess workplace efficiency, to assess and suggest improvements in clinical research workflow [15,16,17]. UML is composed of a set of graphical notations initially created to help understand software processes. It allows us to model a system and help visualize and understand its operation. It is currently used to model processes in diverse areas such as web applications, business processes and health care [14].

Despite their benefits, UML models are static and hence are limited by their description of a problem in a specific instance in time. Dynamic models can be simulated to understand their behaviour and analyze a range of “what if scenarios” thus providing crucial information before the practical implementation of a system level change/policy. Previous studies have [15,18,19,20] used simulation techniques to represent diverse aspects of the real world in an interactive way. Systems Dynamics (SD) a modeling and simulation technique has been widely used to evaluate the behaviour of systems [21,22,23]. It includes a set of principles for modeling and simulation that are used to conceptualize and evaluate complex systems. Through a SD model, we can represent the relationship between elements, their activities as well as the flow and accumulation of information between them. We can also feed the model with qualitative and quantitative data and then create multiple models representing a variety of “what if” scenarios that help in decision making. Given the complex nature of health systems and clinical research, SD has previously been used to model decision support systems in this area [24,25,26,27].

Although there is a wide variety of publications in each of these areas, to the best of our knowledge, no previous study has evaluated the workflow [28] of clinical trials in depth through a mix of techniques such as ethnographic studies to capture information about work routine, used UML [14] to graphically represent workflow in the form of activity diagrams [29] and SD modeling [4,13,30] to evaluate the resulting behaviour over time. Modeling and simulation performed before initiation of a clinical trial may help trial planners and managers to improve the quality of data and save time.

Accordingly, the objective of this study was to 1. Map the workflow of PIs, CRC's, nurses, receptionists and pharmacists of two oncology clinical sites in Brazil, documenting ethnographic observations using a standard task diagram language [14] and 2. Introduce workflow modeling using UML and propose the use of simulation modeling to understand the underlying behaviour of the workflow model and 3. Conducting computer simulation experiments [31] using a system dynamics approach [21] to estimate work efficiency gains when solutions are applied.

## Methods

### Ethics

We obtained ethics approval from Comitê de Ética em Pesquisa do Hospital Pro Cardíaco, at Rua Dona Mariana, Rio de Janeiro, Brazil and Comitê de Ética em Pesquisa (CEPesq) of Hospital Sírio Libanês, São Paulo, Brazil for conducting the study at the trial sites in Rio de Janeiro and São Paulo respectively.

### Inclusion and Exclusion Criteria

We evaluated two oncology trial sites working with Qualidoc (consultant and training support at all levels of personnel in ISO9001: 2008) recommended by Agência Nacional de Vigilância Sanitária (ANVISA) - a National Agency of Sanitary Surveillance in Brazil. Both sites were located in Brazil: the first being a clinic-based environment in the city of Rio de Janeiro with one PI oncologist, two nurses CRCs, one pharmacist, one clinical research administrator and fourteen studies in phase III and the second being a hospital-based environment in São Paulo, with two PIs oncologists, two nurses CRCs, four pharmacists, one clinical research administrator and three studies in phase III. Both sites present a high volume of trial recruitment. We considered sites with an active phase III or IV trials as high volume sites. We excluded Phase I and II trials since they do not represent the majority of trials conducted in BRIC countries, also presenting a very diverse workflow when compared to phase III/IV trials.

### Operational Definitions

Although some technical terminology might encounter different definitions across clinical trials, workflow, and qualitative research literature, for the purposes of this manuscript, we have standardized the most common concepts to facilitate understanding (Table 1).

### Ethnographic Study

Two researchers from our team [AB, EC] made ethnographic field observations for approximately 40 hours and conducted qualitative interviews at two sites in Rio de Janeiro and São Paulo (Brazil), where they shadowed PIs, CRCs, nurses, receptionists and pharmacists. We recorded information on tasks (S1), interactions between other actors (defined in table 1), communication patterns, information needs, and other aspects related to their workflow. We conducted semi-structured interviews using open-ended questions to complement and fill gaps in our ethnographic observations, being broadly directed to (1) understand the average order of tasks in different stages of each protocol, (2) main causes that affect the average order of tasks, including the most common interruptions; (3) main communication contacts for each actor, how frequently communication occurs, how frequently the workflow is affected by the patterns of communication and what type of noise affects the performance of communication flow; (4) documentation models and migration from paper to computer interface used for documentation; (5) How documentation templates are affected when trial site staff are confronted with any query about events, for example, a subject enrolled sometime ago; (6) information necessary that affect the workflow, causing the CRC to search out information from different sources and need to contact other personnel; (7) How clinical trial workflow has been integrated with information technology and if there are in any area where some disconnection between them exists causing a disruption in the regular workflow. These seven concepts were extracted from the study of Khan et. al. in [13].

Observations were made in a non-participatory manner, in that observers were not directly involved in actively collecting data for

**Table 1.** Terminology standardization.

Term	Description
Protocol	Clinical trial documentation that outlines the objective, methodology, design, analysis plan. It also describes the background and reason for the conduct of the study and puts forth a standard method for the conduct of the clinical trial [22]
Actor	A person who is a part of the trial workflow and interacts with tasks or other actors [21]
Principal Investigator (PI)	An individual who is directly responsible for the conduct and completion of a funded clinical trial. He/She directs the research project and reports study results directly to the sponsors [22]
Sub-Investigator	Physicians designated and supervised by the PI to perform procedures and monitor subjects in the clinical trial [22]
CRC (Clinical Research Coordinator)	Individuals responsible for operational tasks and for providing support to PI/Subinvestigator in a clinical trial study. They could be involved in the process of inclusion, recruitment and still maintain the registry of participants, provide the signature of informed consent forms and schedule procedures and lab tests, ensuring accuracy of source documentation, dispensing study medications and maintaining databases with clinical research data, filling the CRF [22]
Research Subject	Once the patient agrees to be a part of a clinical trial and signed the informed consent document, he/she becomes a research subject [22]
CRF (Case Report Form)	It is a paper-based or electronic record of subject data specifically used in clinical trial research [22]
Source Document	a document where collected data is first recorded for a clinical trial and later entered in the CRF [22]

doi:10.1371/journal.pone.0039671.t001

the clinical trial, therefore attempting to minimize the Hawthorne effect that is characteristic of every observational study [32]. Observers did not have access to protected health information (PHI) or any specific data related to the ongoing clinical trials to ensure confidentiality. Since observers did not have access to raw data for privacy and security reasons, the identification of data quality issues was made following the concept of near miss events (the happening of an unexpected event that did not generate important problems, but had big potential to) [33]. Following is an example of a near miss event: a CRC schedules an appointment in his/her personal electronic calendar, notifying other staff personnel about the same. On the day of appointment, in the event of the CRC's absence, the patient would still be received and assisted by other staff personnel at the site. Yet some important detail/document, available only to the CRC may be missed during this process. Consequently, it may generate a greater problem at the end of the study.

When we observed a task that had a higher potential of near miss events, we highlighted it as a potential problem. Information on tasks was later submitted back to the staff at each of the two sites to ensure that information was not misrepresented, thus ensuring appropriate data triangulation [34].

All data collected through the ethnographic study and interviews were transcribed in a Google Docs Word file. After analysis, we synthesized emerging themes using atlas.ti [35], a software for qualitative data analysis. Each interview was coded by two researchers using Grounded Theory methods [36] where emerging themes were extracted from patterns evolving from the text and by avoiding pre-conceived hypothesis. Ambiguities if any were resolved by discussion. Categories were reduced to major themes through discussion amongst the study team while re-reading the transcripts.

### Observation Categories

Observers documented CRC tasks by selecting tasks descriptions from a list of tasks compiled from the emerging themes obtained from the previously described ethnographic study and by findings from previous studies [13]. A large group of major and minor categories were then used to facilitate the classification of any possible CRC tasks. For the purpose of analysis, this large

group of tasks was summarized into a smaller number of categories.

### Workflow Mapping

Information from the ethnographic study was condensed using a modified version of UML (Unified Modeling Language) Activity Diagrams (UML-AD), version 2.0 [37]. All modeling was conducted using Astah community version 6.1 [38]. The variation of activities between two sites was represented by the workflow (S6) pattern “exclusive choice” [39]. Exclusive choice represents a situation where there are two or more exclusive alternate paths to a workflow and only one of them can be chosen. The terminology for workflow and UML- AD diagram are present in Supporting information S6.

### Preliminary Simulation Model

Next, we formed an external panel with five clinical researchers, affiliated to the Research on Research group (RoR), Duke University, to discuss the findings and summarize it as a simulation model. The simulation was performed by taking the UML activity diagram as the basis for a System Dynamics model [21,22,23]. All modeling was conducted using Vensim DSS [40]. This process involved two steps. First, a baseline model was created to reproduce the current workflow of each clinical trial site along with the reported problems related to near miss events for data quality (Table 2). A system analysis of each major loop in the system identified points where workflow could be potentially improved in terms of a better execution of tasks. Second, the model was presented to each participant for feedback, who suggested modifications. Interventions were then tested in subsequent models, where we simulated an improvement in the system. Finally, the model was fed with qualitative data and simulated using different scenarios.

### Results

Overall, the two sites evaluated in this study demonstrated consistent data collection system, with only a few discrepancies between them. The inconsistencies were related to specific infrastructure aspects. In addition, a group of areas with higher

**Table 2.** Variables classified as problematic on visited sites.

<b>Visit-EDC asynchrony</b>	The CRC does not have the time to transcribe the content from the paper-based or electronic medical record to the computerized electronic data capture (EDC) in a timely manner.
<b>EMR-EDC non-interoperability</b>	Electronic medical record (EMR) not integrated with EDC, therefore may cause error or lack of information during transcription from one the EMR computer screen to the one for the EDC.
<b>CRC-data entry discontinuity</b>	At the time of the encounter with the subject, investigator takes notes on the EMR, nurse takes note on nurse plan and CRC makes the final data entry into EDC.
<b>Lab-EDC non-interoperability</b>	Laboratory system not integrated with EMR, requiring re-entry with potential for transcription errors.
<b>Decentralized accounting for drug dispensation</b>	System for drug stock control between pharmacy and trials use different systems, with a potential for discrepancy and error.
<b>Lack of decision support systems to assist CRC</b>	CRC frequently has to consult a variety of documents not available at the time of interaction with research subject. This factor is compounded by a single CRC being in charge of multiple studies supported with paper-based documentation, which might lead to confusion and error.

doi:10.1371/journal.pone.0039671.t002

potential for flow quality issues, i.e., following the near-miss event concept, were also identified, mapped and simulated. In total, we observed the activities of four Clinical Research Co-ordinators (CRCs), and identified a total of 139 tasks (Refer to supporting information S1).

### Ethnographic Field Observations and Workflow Mapping using UML Activity Diagrams

**General observations.** The UML-AD detected four major activities in the workflow, namely (a) evaluation of potential research subjects prior to signature of informed consent (S2), (b) visit to obtain subjects informed consent (S3), (c) regular data collection sessions following study protocol (S4), and (d) closure of study protocol for a given project.

Activity A, prior to the signature of the informed consent, the potential research subject underwent a clinical exam as well as additional laboratory exams as per clinical site routine. These are common pathological exams and work as a pre-screening method. The results of these exams help in the evaluation of inclusion and exclusion criteria. During activity B - the visit to obtain subjects informed consent, potential subjects received an explanation about the entire study, described in the informed consent form while being offered the opportunity to ask questions. The PI then requested the subject to sign the informed consent. At this point an individual schedule for the study subject was established. Activity involved a series of encounters representing regular data collection sessions following the variations across different study protocols. Finally, during activity D, the subject either reached completion or was excluded from the study protocol, when study closure was achieved, including the collection of any remaining interventional drug when applicable and also included the plan for any further follow-up when required by protocol. Since each of these activities is distinct, they were represented by different UML-ADs (Refer Supporting information S2, S3, S4).

**Observations of workflow tasks with higher likelihood of quality issues.** Our field observations pointed to three major emerging themes (based on table 1) corresponding to areas where workflow was associated with potential near miss events for data quality issues.

**Lack of standardized process for data registration at source document.** The first emerging theme resulted from the presence of a number of intermediate documents in the transfer of data from research subject to the electronic CRF by CRC.

Examples include: (a) Nurse reports were paper-based and later manually transferred to an electronic data capture system; b) Drug stock control was made on tables of text editor or spreadsheet and then printed; c) Results from laboratory exams were printed from their home pages or received by email and subsequently printed, or printed from laboratory computer system applications. Although all of these sources do not violate any current Good Clinical Practice regulations, as CRFs were filled out based on the original source data, the presence of intermediate documents – even if used only for care purposes - increases the odds of error secondary to unintentional transcription errors. In addition, these intermediate steps represent rework.

**Multiplicity of data repositories.** We observed that the information being transferred was registered in different binary files such as Microsoft Word and Excel files. The files were not linked and later were manually transferred using a copy and paste functionality. In one of the sites, documents directly filled out with subjects information were not directly inserted into the CRF by the CRC, but remained in paper-format (source document) for later transcription into the electronic CRF.

The paper-based documentation trail also results in searchability problems, in that, locating specific data fields for validating systems in the area or when checking for possible errors was time-consuming. Adding to the complexity of this process, research nurses were required to fill out multiple forms in accordance to study protocol, this information being transcribed by investigators and later copied and finally pasted by the CRCs. Other parallel systems also contained different, non inter-operable systems such as pharmacy services using multiple Microsoft Excel and Word documents for stock control, drug validity, and medication disposal.

**Scarcity of decision support systems at the point of research intervention.** Given that a single CRC is concomitantly in charge of multiple studies, it is not unusual that the information about specific details pertaining to each study were sometimes forgotten. This oversight resulted in a major rework loop, in that the correction could only be accomplished through the central coordinating center located in the European Union, not before having to contact the central coordinator for Brazil. The total rework process resulted in a total time loss of approximately two uninterrupted hours. In another instance, the complete period to complete a CRF lasted about one hour, reportedly at least three times more than what that task would

usually require. According to the CRC, the process was slow since, although she had undergone previous training, that training had occurred several months before and most study-related processes were new, requiring constantly re-checking to ensure information accuracy. Since the documentation was long and not easy to search, the whole workflow was significantly delayed, ultimately affecting the CRC's ability to complete it on time. With this delay, some of the CRF completion was left to another person on staff, enhancing the probability of near miss events for data quality issues. Other unexpected tasks for which there is no automation also prevented that the CRC complete the protocol on time, such as accounting issues related to the appointment such as the CRC spending one complete morning solving financial problems related to the study. In addition, study manuals frequently were confusing and lacked appropriate usability as attested by confusion generated while the CRC attempted to interpret them.

**Preliminary Simulation Model**

The preliminary simulation model (Figure 1) envisioned the following three main emerging themes (S5) generating rework: 1) Lack of standardized process for data registration at source document, 2) Multiplicity of data repositories and 3) Scarcity of decision support systems at the point of research intervention. The model has the following features: (a) all emergent themes are represented as stocks (rectangle), (b) a blue arrow represents information input and (c) a red arrow represents some action that will limit the growth of a stock.

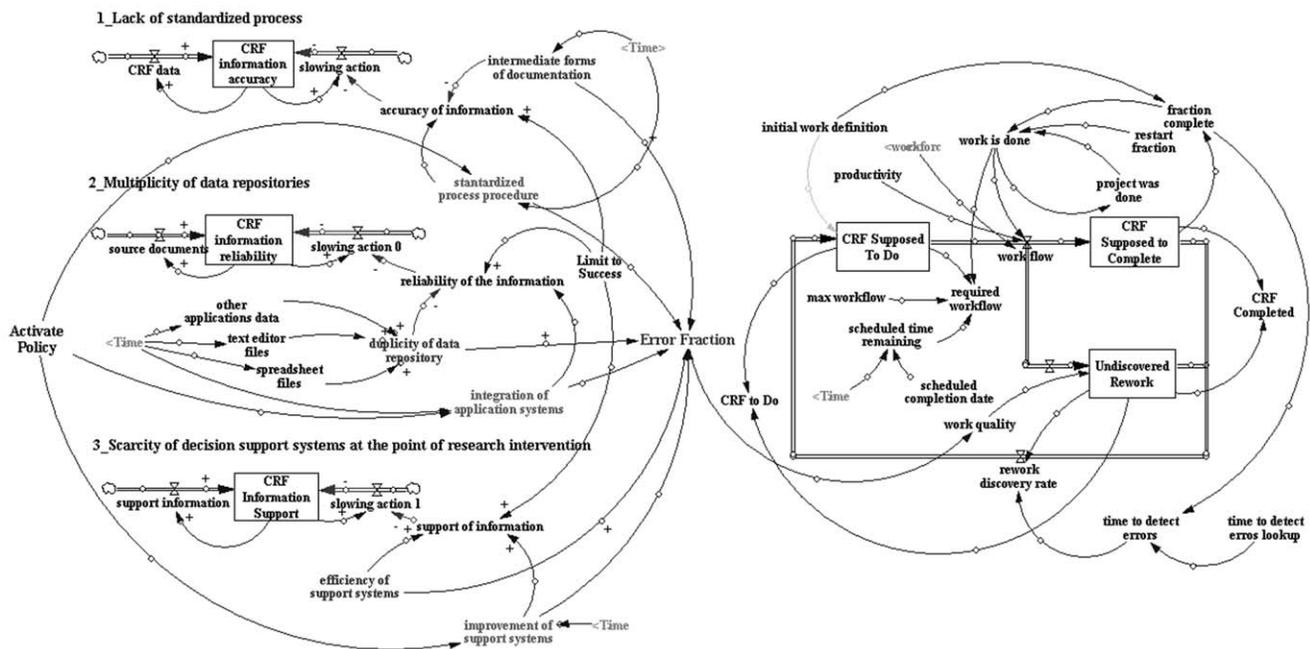
The theme "lack of standardized processes" has as its central feature represented by a stock "CRF information accuracy". Aligned with our ethnographic findings, the variable representing "intermediate forms of documentation" has a negative influence over "accuracy of information". The variable "accuracy of information" will speed up the processing of CRF completion. In other words, the greater the "accuracy information", less delay in successfully completing CRFs we will have.

The second theme "multiplicity of data repositories" is represented by a stock of reliable CRF information. Aligned with our ethnographic observations, it was seen in the form of three types of data repositories of information that would lead to the same information in different formats (Word documents, spreadsheets in Excel and software applications like central pharmacy or laboratories). In our model, we simulated how data multiplicity would affect reliability of CRF information. In our model the variable "data repository multiplicity" is influenced by three distinct variables: "other applications" (laboratories application, central pharmacy application and others), "text editor files" (Word documents), and "spreadsheet files" (Excel documents). The variable "data repository multiplicity" has a negative effect on the variable "information reliability." In other words, the greater the multiplicity of data sources, the less reliable the information and the greater the delay in successful completion of CRFs.

The third theme "scarcity of decision support system at the point of research intervention" indicates that the current workflows for filling out paper and computer-based CRFs are in need of a better decision support system. The model also contains a variable representing the "efficiency of a support system," which positively impacts the speed in completing CRFs.

Additionally, we can observe that the main variables of our model "intermediate forms of documentation", "data repository multiplicity" and "support of information" are influencing another variable called "Error Fraction". The variable "Error Fraction" is used in our model to represent the stocks - 'CRF to Do', 'Rework' and 'CRF Completed'. It receives information from the left part of model, representing the three emergent themes and has a directly proportional relationship with the variable - "Rework" flow. This means that higher instances of 'Error' will generate more rework and as a result lesser CRFs will be completed.

As we did not have access to real data, we made the following assumptions: the values for "intermediate of documents" and "multiplicity of data repositories" was simulated starting on zero and growing each six months until the limit of 20% and 30%



**Figure 1. System dynamics model showing the three major emerging themes identified and rework caused by these themes.** Available for best visualization in <http://goo.gl/XydTd>. doi:10.1371/journal.pone.0039671.g001

respectively (Refer to supporting information S5). We did this to simulate a situation where the staff started working with no problems but faced them increasingly over time. The value of efficiency of support system were simulated with 50% all the time, with this parameter we simulate a system with 50 percent deficiency.

We present two situations based on the outcomes of simulating the model (figure 2). In the first situation labeled as “planned”, we assume the lack of an error in the work done. The simulation results represent CRFs to be done and CRFs completed. In the second situation labeled as “Simulated with Rework” we attempted to simulate of what happened on observed environment. This simulation was based on three emergent themes presented in this manuscript. In this second situation work could not be completed as planned. This was primarily a consequence of rework that was generated because of error fraction caused by emergent themes consequences (rework point on figure 2).

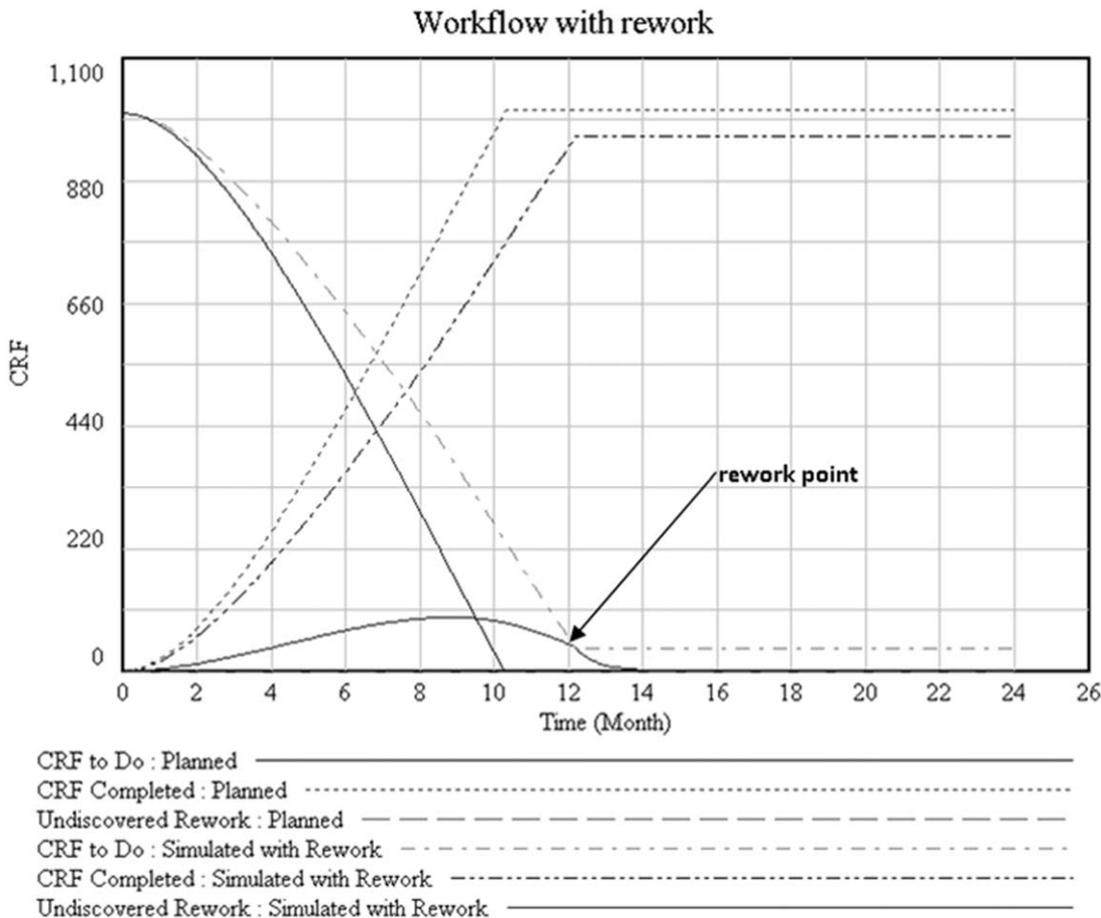
**Policy Model Results**

The policy model results were divided into three categories of policies with an aim to reduce the rework: (a) Standardized process procedure, (b) Integration of application systems, (c) Improvement of support systems (figure 3). In “standardized process procedure” policy, we suggest the clinical trial administrative staff of each site to periodically revise their standard procedures and perform training with the research team when working with CRFs to

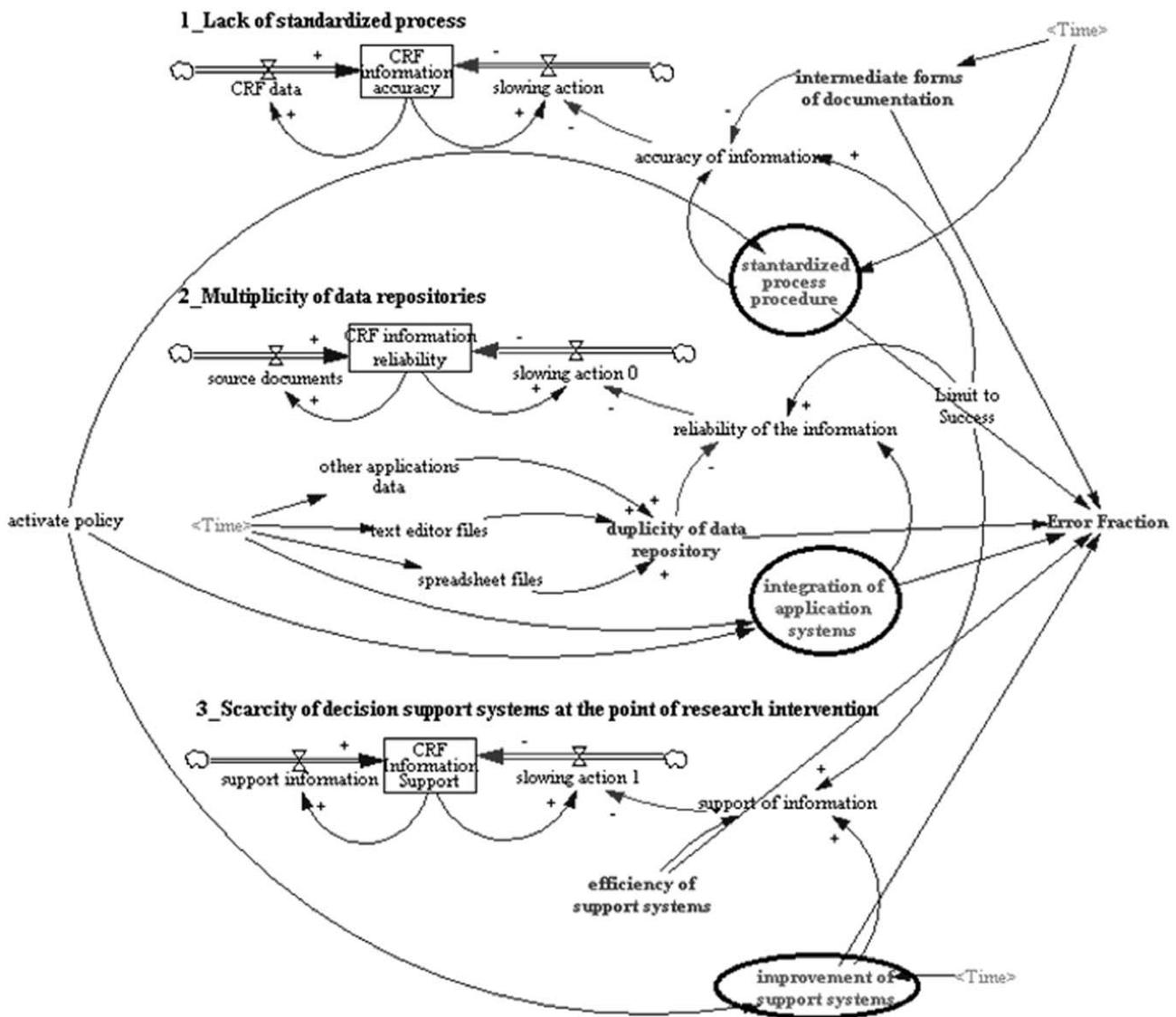
decrease the creation of intermediate documents. For the policy “integration of application systems” the clinical trial administrative staff must work together with IT staff of a site and analyse all the data repositories created in duplicity (as observed during our ethnographic study) and create a solution where the clinical trial staff will insert information in only one application to avoid these duplicities. Finally, in the “improvement of support systems” policy, the clinical trial administrative staff and sponsors need to work more efficiently to obtain support information more fast and friendly when working with CRF. For each one of these policies the PI, CRC’s, nurses, pharmacists and receptionist involved in clinical trial procedures must be trained in new methods and procedures that can facilitate the use of information and resources more efficiently while reducing rework.

As a result of policies applied (figure 4), we observe the latter leads to more CRF’s completed, resulting in less rework. Rework does not disappear altogether but reduces. In course of time, if policies are well implemented, rework should tend towards zero.

Figure 5 gives us an idea of workflow behaviour in our simulation. The dotted line represents workflow planned, which should be complete a thousand of CRFs for ten months. The straight line represents the simulation performed with rework caused by three emerging themes identified. In this situation, the workflow has been increased in order to be able to complete a thousand CRFs which will need twelve months. The dotted line



**Figure 2. CRF’s to do and CRF’s completed in two situations: planned and simulated with rework.**  
doi:10.1371/journal.pone.0039671.g002



**Figure 3. Dynamic model with the policies suggested (circled).** Available for best visualization in <http://goo.gl/XydTd>. doi:10.1371/journal.pone.0039671.g003

represents workflow with the implementation of policies where the time to complete CRF was reduced. The ideal workflow may depend on factors that of not immediate implementation, therefore the implementation of the policies had a small reduction of rework and still kept it.

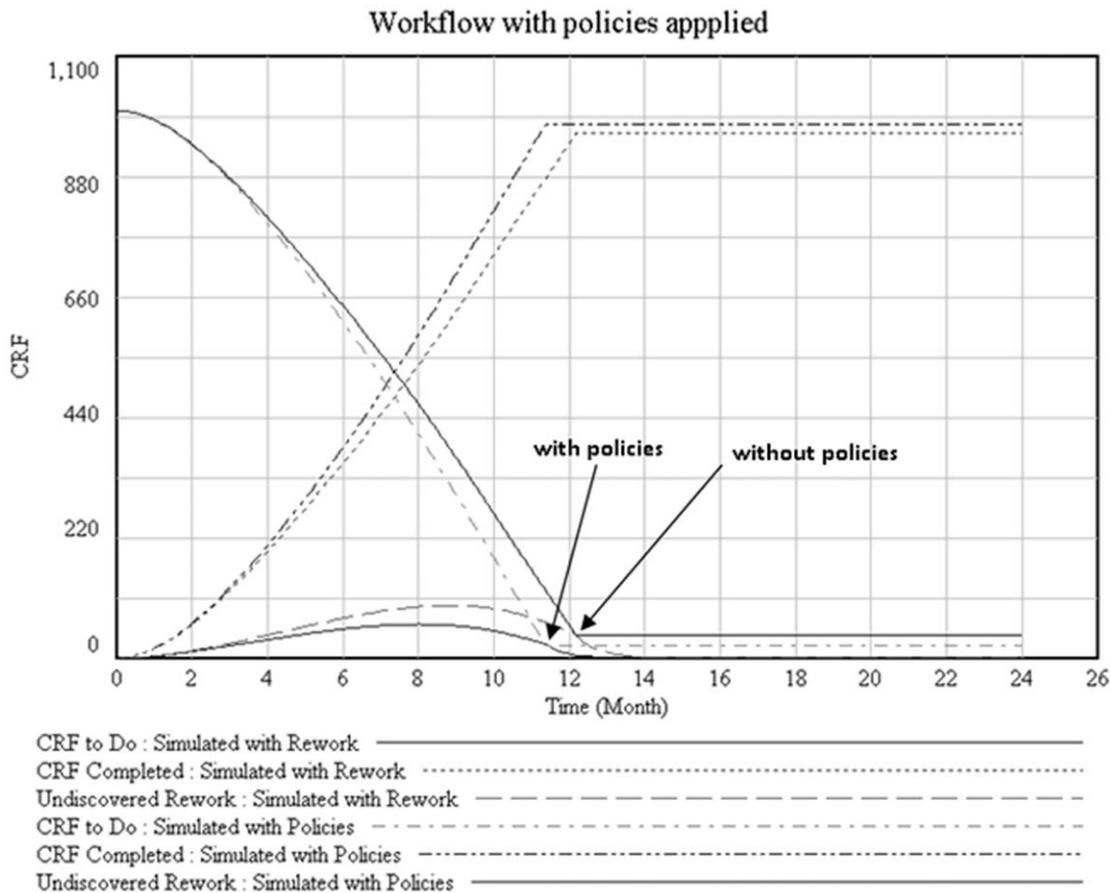
## Discussion

To the best of our knowledge, this is the first study evaluating and mapping workflow at clinical trial sites while also conducting a simulation to represent the potential for improvement once the main issues were addressed.

In our study, several factors affecting workflow pointed towards the lack of interoperability between the electronic medical record and electronic data capture system at the clinical trial site, namely visit-EDC asynchrony, EMR-EDC non-interoperability, lab-EDC non-interoperability, decentralized accounting for drug dispensation). Factors affecting this relation included the presence of standalone and disconnected systems taking care of schedule

management, administration, lab reporting, medical records, and research databases. This results in fragmented and siloed data that are seldom useful for research purposes. In many cases, it has been noted that the data from EMR cannot be directly utilized in a clinical trial environment on account of incompatible data formats and non adherence to clinical research regulations [41]. As a consequence, despite the availability of state of art systems, information is frequently recorded manually on charts followed by separate data entry into the EMR and EDC. The lack of interoperability thus leads to less efficiency, duplication [41], and errors, leading to disruption of the clinical trial pipeline and damages estimated in million dollars/year.

In our study, near miss events for data quality were also influenced by a discontinuity in performance of a specific task. Specifically, rather than having one individual performing the entire task, the process was broken and implemented by more than one individual. Although this practice is aimed at workflow distribution and enhancing productivity and efficiency of different individuals it also serves to increase the possible flow quality



**Figure 4. CRF's to do and CRF's completed in two situations: simulated with rework and simulated with policies suggested.**  
 doi:10.1371/journal.pone.0039671.g004

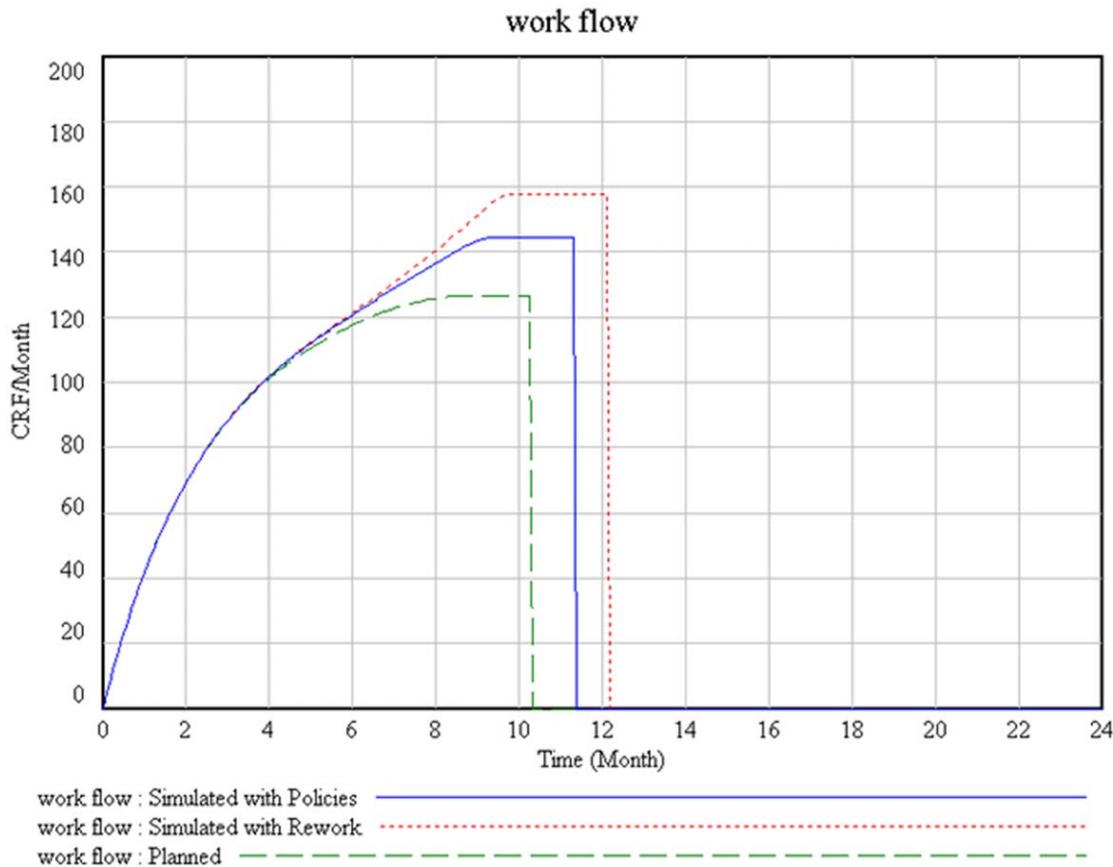
problems resulting from multiple individuals with different levels of information about the subject handling the data. Previous literature suggests that interruption during task implementation may not only have a negative effect on task performance but also lead to serious errors [42]. For example, CRCs may receive unexpected workload and thus forget a critical portion of their task lists or incur in data entry errors [13].

We also noted that an absence of decision support systems for clinical research coordinators had a significant influence on their task performance and resultant near miss events for data quality. Clinical research coordinators and physicians frequently need to answer queries or consult the study material while evaluating a research subject. Lack of readily available guidelines has implications on time and effective subject management in a clinical trial. In our study we identified difficulties encountered by CRC to quickly locate information while handling long paper-based protocols. This difficulty could be alleviated with, for example, electronic documents with better navigability that could be immediately queried at the time when they are needed. Although the benefits of decision support systems in enhancing the safety and quality of clinical practice [43,44,45], reducing errors [46,47] and ensuring adherence to practice guidelines [48,49] has amply been demonstrated in the past, they are equally useful in clinical research settings. In the latter context, decision support tools have been effectively used to determine subject eligibility [50,51]. Based on our results, the implementation of a decision support tool to guide CRC tasks would similarly result in the optimization of clinical trial tasks.

Although our report has made a significant contribution in terms of understanding, representing, and simulating strategic workflow decisions to reduce the incidence of near miss data quality issues in clinical trial sites, our study has limitations. As in any observational study, the presence of observers in the clinical trial site might interfere with the way subjects behave, potentially introducing bias into our observations [52]. However, since most of our findings were confirmed by triangulation [53] to be part of the daily routine of the investigated sites, this effect should be considered minimal.

The results obtained from the suggested interventions implemented in our model simulation were not validated through real life implementation, since it would require a substantial investment from the site and sponsor's perspective. Given the highly regulated environment within clinical trials, the level of variability in our UML model is reduced thus enhancing its applicability. Several of the investigators were familiar with the clinical trial process and, therefore, could have unintentionally introduced their opinions into the observational process.

Another issue in comparing workflows across trial sites is the absence of a taxonomy listing variables that should be compared. This taxonomy would ideally focus on areas of dissonance between workflow and data quality and requires a specific article describing this approach. Moreover, although the two sites where we collected data represent active Cancer trial sites in Brazil, our sample size was small and may not be judged representative. A larger sample size could increase the generalizability of our results, but qualitative studies are aimed at depth rather than generaliz-



**Figure 5. The Workflow of all process simulated for three situations: Planned, with reword identified and with policies suggested.**  
doi:10.1371/journal.pone.0039671.g005

ability, the former being achieved by the data collected at each site.

Although we conducted our study in a developing country (Brasil), there is little evidence to suggest that the problems we describe are restricted to developing countries. Moreover, the low granularity of our system level recommendations is justified given its occurrence every time a new approach is applied to a field. Given that, to the best of our knowledge, this is the first study evaluating and mapping workflow at clinical trial sites while also simulating its results using System Dynamics.

Finally, we make a substantial assumption that near-miss events are a proxy for real data quality events. This assumption was made given the qualitative nature of our study, where samples are frequently small and exploratory, therefore unable to reach results that would qualify as definitive. However, we would also like to point that an assumption of association between near miss and “real events” has been made in a wide range of the literature and ratified by entities such as the Institute of Medicine in its report “To err is human” [54].

In conclusion, in order to reduce rework and optimize trial performance, it is essential to improve the efficiency prevailing in trial sites. Clinical trial sites and sponsors should review and improve their standard procedures and manage human resource allocation to better address the workflow issues summarized in our study. Sponsors and site managers can improve support systems for CRCs. Although some interventions such as integration between EMR and EDC, might not seem cost-effective or even technologically possible at this time, strategic decisions should be made to facilitate their future integration. For example, when

planning for an EMR system, hospital administrators should evaluate the ease of integration with external EDCs. Also, future research should integrate systems modeling simulation so that these changes can be constantly monitored and frequently evaluated in the quest for data quality excellence. As a future study we intend to submit a new manuscript about the new taxonomy for what should be compared in workflows across sites focusing on areas of dissonance between workflow and data quality.

## Supporting Information

### Supporting Information S1 List of all task categories found through the field notes.

(DOC)

### Supporting Information S2 UML - Activity Diagram for pre-signature consent form process.

(DOC)

### Supporting Information S3 UML - Activity Diagram for signature consent form and first visit.

(DOC)

### Supporting Information S4 UML - Activity Diagram for next visits.

(DOC)

### Supporting Information S5 Emergent themes main variables in Vensim simulation software.

(DOC)

## Supporting Information S6 Workflow Terminology. (DOC)

## Acknowledgments

Ms. Julie Simkins, RN for assisting with manuscript editing and translation of terms from Portuguese to English. Ms Siderleny Casari Martins, Ms Ingrid Slany, Ms Livia Gomes as research coordinators, Mr Gabriel Mustrangi as clinical research analyst that cooperate with this manuscript. The authors thank the Research on Research Group (<http://>

## References

- Normile D (2008) The promise and pitfalls of clinical trials overseas. *Science* 322: 214–216. doi: 10.1126/science.322.5899.214.
- Thiers FA, Sinskey AJ, Berndt ER (2008) Trends in the globalization of clinical trials. *Nature Reviews Drug Discovery* 7: 13–14. doi: 10.1038/nrd2441.
- Ali M, Park J-K, von Seidlein L, Acosta CJ, Deen JL, et al. (2006) Organizational aspects and implementation of data systems in large-scale epidemiological studies in less developed countries. *BMC Public Health* 6: 86. doi: 10.1186/1471-2458-6-86.
- Borycki EM, Kushniruk AW, Kuwata S, Kannry J (2006) Use of Simulation Approaches in the Study of Clinician Workflow. *AMIA Annu Symp Proc* 2006: 61–65.
- Dwivedi A, Bali RK, Wickramasinghe N, Naguib RNG (2007) How workflow management systems enable the achievement of value driven healthcare delivery. *Int J Electron Healthc* 3: 382–393.
- Lang M, Kirpekar N, Bürkle T, Laumann S, Prokosch H-U (2007) Results from data mining in a radiology department: the relevance of data quality. *Stud Health Technol Inform* 129: 576–580.
- Noumeir R (2006) Radiology interpretation process modeling. *J Biomed Inform* 39: 103–114. doi: 10.1016/j.jbi.2005.07.001.
- Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, et al. (2007) Recruitment to randomised trials: strategies for trial enrollment and participation study. *The STEPS study. Health Technol Assess* 11: iii, ix–105.
- Fuhrer MJ (2005) Conducting multiple-site clinical trials in medical rehabilitation research. *Am J Phys Med Rehabil* 84: 823–831.
- Heller B, Herre H, Lippoldt K, Loeffler M (2004) Standardized terminology for clinical trial protocols based on top-level ontological categories. *Stud Health Technol Inform* 101: 46–60.
- Merzweiler A, Weber R, Garde S, Haux R, Knaup-Gregori P (2005) TERMTrial—terminology-based documentation systems for cooperative clinical trials. *Comput Methods Programs Biomed* 78: 11–24. doi: 10.1016/j.cmpb.2004.11.001.
- Willenberg KM (2004) Managing clinical trials—frustration or bliss? *J Oncol Manag* 13: 24–26.
- Khan SA, Kukafka R, Payne PRO, Bigger JT, Johnson SB (2007) A day in the life of a clinical research coordinator: observations from community practice settings. *Stud Health Technol Inform* 129: 247–251.
- Fowler M, Scott K (1999) UML Distilled: A Brief Guide to the Standard Object Modeling Language. 2nd ed. Addison-Wesley Pub (Sd). 224 p.
- de Carvalho ECA, Batilana AP, Simkins J, Martins H, Shah J, et al. (2010) Application description and policy model in collaborative environment for sharing of information on epidemiological and clinical research data sets. *PLoS ONE* 5: e9314. doi: 10.1371/journal.pone.0009314.
- Kumarapeli P, De Lusignan S, Ellis T, Jones B (2007) Using Unified Modelling Language (UML) as a process-modelling technique for clinical-research process improvement. *Med Inform Internet Med* 32: 51–64. doi: 10.1080/14639230601097705.
- Borlowsky T, LaFountain J, Petty L, Saltz JH, Payne PRO (2008) Leveraging an existing data warehouse to annotate workflow models for operations research and optimization. *AMIA Annu Symp Proc*: 885.
- Anderson JG (2002) Evaluation in health informatics: computer simulation. *Comput Biol Med* 32: 151–164.
- Despont-Gros C, Mueller H, Lovis C (2005) Evaluating user interactions with clinical information systems: a model based on human-computer interaction models. *J Biomed Inform* 38: 244–255. doi: 10.1016/j.jbi.2004.12.004.
- Borycki EM, Kushniruk AW, Kuwata S, Kannry J (2006) Use of simulation in the study of clinician workflow. *AMIA Annu Symp Proc*: 61–65.
- Bossel H (2004) Systeme, Dynamik, Simulation: Modellbildung, Analyse und Simulation komplexer Systeme. 1st ed. Books on Demand GmbH. 400 p.
- Sterman J, Sterman JD (2000) *Business Dynamics: Systems Thinking and Modeling for a Complex World with CD-ROM*. McGraw-Hill/Irwin. 1008 p.
- Lynceis JM, Ford DN (2007) System dynamics applied to project management: a survey, assessment, and directions for future research. *System Dynamics Review* 23: 157–189. doi: 10.1002/sdr.377.
- van Ackere A, Smith PC (1999) AID-SDR171>3.0.CO;2-P.
- Homer J, Ritchie-Dunham J, Rabbino H, Puente LM, Jorgensen J, et al. (2000) Toward a dynamic theory of antibiotic resistance. *System Dynamics Review* 16: 287–319. doi: 10.1002/sdr.200.
- researchonresearch.duhs.duke.edu/) for the templates for writing introduction and discussion sections of the manuscript [55] as well as templates for Literature matrix, Duke University Health System [56].
- McDaniel RR, Driebe DJ (2001) Complexity science and health care management. *Advances in Health Care Management*: 11–36. doi: 10.1016/S1474-8231(01)02021-3.
- Cloutier LM, Boelhe M, Université du Québec à Montréal. Centre de recherche en gestion (2001) Innovation management under uncertainty: a system dynamics model of R&D investments in biotechnology. Montréal: Université du Québec à Montréal, Centre de recherche en gestion]. p.
- Marjamaa RA, Torkki PM, Hirvensalo EJ, Kirvelä OA (2009) What is the best workflow for an operating room? A simulation study of five scenarios. *Health Care Manag Sci* 12: 142–146.
- Dumas M, Hofstede AHM ter (2001) UML Activity Diagrams as a Workflow Specification Language Springer Verlag. 76–90.
- Payne PRO, Mendonca EA, Starren JB (2007) Modeling Participant-Related Clinical Research Events Using Conceptual Knowledge Acquisition Techniques. *AMIA Annu Symp Proc* 2007: 593–597.
- Fishman GS (1978) *Principles of Discrete Event Simulation*. New York, NY, USA: John Wiley & Sons, Inc. p.
- Landsberger HA (1958) *Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry*. Available:<http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED024106>. Accessed: 15 April 2011.
- Barach P, Small SD (2000) Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. *BMJ* 320: 759–763.
- Silverman D (2004) *Qualitative Research: Theory, Method and Practice*. Second ed. Sage Publications Ltd. 390 p.
- ATLAS.ti website (2011) Available: [www.atlas.ti.com](http://www.atlas.ti.com).
- Strauss AC, Corbin JM (1997) *Grounded Theory in Practice*. 1st ed. Sage Publications, Inc. 288 p.
- de Carvalho ECA, Jayanti MK, Batilana AP, Kozan AMO, Rodrigues MJ, et al. (2010) Standardizing clinical trials workflow representation in UML for international site comparison. *PLoS ONE* 5: e13893. doi: 10.1371/journal.pone.0013893.
- Astah Community website (2011) Available: <http://astah.change-vision.com/>.
- White S (2004) Process Modeling Notations and Workflow Patterns. In L. Fischer, editor, *Workflow Handbook 2004*, pages 265–294. Future Strategies Inc. Lighthouse Point, FL, USA.
- SD modeling software. Available from <http://www.vensim.com/>.
- Grossman M, Bates S (2008) Knowledge capture within the biopharmaceutical clinical trials environment. *VINE* 38: 118–132. doi: 10.1108/03055720810870932.
- Brixey JJ, Robinson DJ, Johnson CW, Johnson TR, Turley JP, et al. (2007) A concept analysis of the phenomenon interruption. *ANS Adv Nurs Sci* 30: E26–42.
- Bennett JW, Glasziou PP (2003) Computerised reminders and feedback in medication management: a systematic review of randomised controlled trials. *Med J Aust* 178: 217–222.
- Walton R, Dovey S, Harvey E, Freemantle N (1999) Computer support for determining drug dose: systematic review and meta-analysis. *BMJ* 318: 984–990. doi: 10.1136/bmj.318.7189.984.
- Walton RT, Harvey E, Dovey S, Freemantle N (2001) Computerised advice on drug dosage to improve prescribing practice. *Cochrane Database Syst Rev*: CD002894. doi: 10.1002/14651858.CD002894.
- Kaushal R, Shojania KG, Bates DW (2003) Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch Intern Med* 163: 1409–1416. doi: 10.1001/archinte.163.12.1409.
- Bates DW, Teich JM, Lee J, Seger D, Kuperman GJ, et al. (1999) The Impact of Computerized Physician Order Entry on Medication Error Prevention. *J Am Med Inform Assoc* 6: 313–321.
- Hunt DL, Haynes RB, Hanna SE, Smith K (1998) Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 280: 1339–1346.
- Shiffman RN, Liaw Y, Brandt CA, Corb GJ (1999) Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. *J Am Med Inform Assoc* 6: 104–114.
- Breitfeld PP, Weisburd M, Overhage JM, Sledge G Jr, Tierney WM (1999) Pilot study of a point-of-use decision support tool for cancer clinical trials eligibility. *J Am Med Inform Assoc* 6: 466–477.

51. Ohno-Machado L, Wang SJ, Mar P, Boxwala AA (1999) Decision support for clinical trial eligibility determination in breast cancer. *Proc AMIA Symp*: 340–344.
52. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, et al. (2007) The Hawthorne Effect: a randomised, controlled trial. *BMC Medical Research Methodology* 7: 30. doi: 10.1186/1471-2288-7-30.
53. Holden JD (2008) Hawthorne effects and research into professional practice. *Journal of Evaluation in Clinical Practice* 7: 65–70. doi: 10.1046/j.1365-2753.2001.00280.x.
54. Institute of Medicine (IOM) (2000) *To Err Is Human: Building a Safer Health System* (n.d.). Available: <http://www.nap.edu/openbook.php?isbn=0309068371>. Accessed: 1 June 2011.
55. Shah J, Shah A, Pietrobon R (2009) Scientific writing of novice researchers: what difficulties and encouragements do they encounter? *Acad Med* 84: 511–516. doi: 10.1097/ACM.0b013e31819a8c3c.
56. Pietrobon R, Guller U, Martins H, Menezes AP, Higgins LD, et al. (2004) A suite of web applications to streamline the interdisciplinary collaboration in secondary data analyses. *BMC Med Res Methodol* 4: 29. doi: 10.1186/1471-2288-4-29.