Using Personal Health Records for Automated Clinical Trials Recruitment: the ePaIRing Model

Adam Wilcox, PhD, Karthik Natarajan, Chunhua Weng, PhD
Department of Biomedical Informatics, Columbia University, New York, NY

Abstract

*We describe the development of a model describing the use of patient information to improve patient recruitment in clinical trials. This model, named ePaIRing (**e**lectronic **Pa**rticipant **I**dentification and **R**ecruitment Model) describes variations in how information flows between stakeholders, and how personal health records can specifically facilitate patient recruitment.*

Introduction

Breakthroughs in the basic biomedical sciences have provided an unprecedented supply of information with the potential to improve human health. Clinical trials research is an important step for translating such information into knowledge that will benefit clinical practice and human health. In recent years, many, including the United States Congress, have expressed their concern about the lack of continuity between the promise of basic science and the delivery of better health[1]. According to recent data, a clinical trial averages $124 million and takes more than a decade to complete per drug candidate[2], with half of this time spent on patient, site, and investigator recruitment[3]. In a recent large, multi-center randomized controlled trial, about 86.8 staff hours and more than $1000 was spent to recruit each randomized participant[4]. Presently, eighty-six percent of all clinical trials are delayed in patient recruitment for 1-6 months and 13% are delayed by more than 6 months[5, 6].

The clinical research environment is increasingly burdened with rising costs, slowly obtained results, and declining participation from both investigators and subjects. The number of subjects needed for an FDA-registry trial has doubled in the past decade, but only 7% of eligible subjects and 3% of eligible cancer patients enroll in clinical trials annually[7]. Although barriers to clinical trial enrollment include patient factors[8] and physician factors[9], a major deterrent to enrollment is the challenge of recruiting eligible patients, which has traditionally been a labor-intensive manual procedure. Studies have shown that up to 60% of the eligible patients can miss being identified[10, 11], which means these patients miss valuable treatment opportunities. Therefore, there is an urgent need for improving the accuracy and efficiency in patient eligibility identification for clinical trials.

The broad deployment of EHR systems has made enormous amounts of information about potential clinical research participants available in EHR systems. In addition, within the past year, Microsoft and Google both released personalized health records (PHR) software. Very soon, patients will be able to enter their health information electronically and share such information online. Mandl and Kohane commented this phenomenon as "Tectonic Shifts in the Health Information Economy."[12]

The advent of PHR and the broad adoption of EHR has created unprecedented opportunities for automating clinical trials recruitment, but also put forth significant regulatory and privacy-related challenges for informatics researchers. "How can we leverage the information in PHR and EHR to improve automated clinical trial recruitment without violating patient privacy" is center to the discussion of this paper. At the end of the discussion, this paper proposes a new informatics model that integrates EHR and PHR for automated clinical trials recruitment.

Model Development

We created our model by following an approach adapted from grounded theory analysis, developing the model inductively from existing data and knowledge. Grounded theory is a qualitative method that starts by collecting data, and then iteratively interpreting that data by coding key points, grouping data and points by concepts, forming categories around the concepts, and eventually creating a theory or model that explains the data. It is a process whereby researchers can gather data and then systematically generate and develop theory directly from the data. We diverged from grounded theory in the data collection, because we generated it directly from experts, but the analysis followed the same qualitative iterations. The analysis was performed by two of the authors (AW, CW), who were experienced with various aspects of informatics and clinical trial recruitment.

For data generation, we first identified various stakeholders who might be relevant to a clinical trial recruitment system. These stakeholders were the

principal investigator, the patient, the physician treating the patient, the study coordinator, and the study sponsor. For each stakeholder, we documented the information resources that were available to the stakeholder, and the information resources that were needed by the stakeholder. We then identified how those needs were currently being met, or might be met in the future.

For the second and third stages of the model development, we coded key points around the stakeholder resources or needs, eventually classifying each resource or need and grouping into themes. This resulted in 7 different themes around the information needed to flow among the stakeholders. These were patient clinical information, patient consent, study inclusion criteria, incentives for participation, identification of physicians with relevant patient populations, communication with patients, and whether a patient fits the study inclusion criteria. The first iteration of coding and grouping was focused on completeness. The second iteration was focused on significance. To this end, we then reviewed both the stakeholders and themes, and identified those that were most significant to the recruitment process. This resulted in 3 significant actors in recruitment (principal investigators, physicians, and patients), and 2 key themes among those actors (criteria matching, and patient consent). Coordinators and sponsors were identified as important stakeholders, but not direct actors in the recruitment process. Criteria matching was defined as the information needed in the process of identifying patients as potential subjects in the study, and patient consent was defined as information indicating patient agreement to participate in the recruitment and study process.

Once the categories and themes were identified, our final stage in model development was to analyze relationships among the categories and themes, to eventually identify theories. We focused on the 3 actors in recruitment, and identified the information flow of the 2 significant themes among these actors. We also coded and classified the methods where the information needs were currently or could be met, as either through paper-based systems or electronic systems. The electronic systems were identified in terms of the main actors use of the systems, and were then classified as either electronic health records or personal health records.

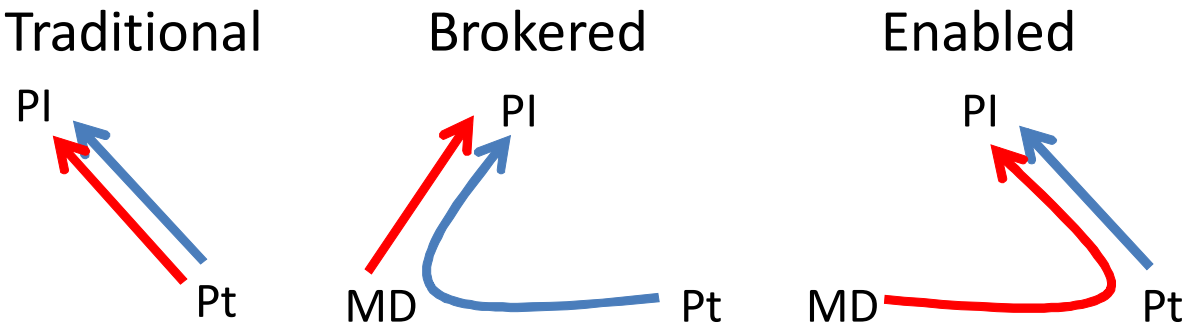

Figure 1: Information flow among actors in the recruitment process. Two types of information– criteria matching (red) and patient consent (blue) flow between the principal investigator (PI), patient (Pt), and physician (MD) in the process. In the traditional model, only the investigator and patient are involved. In the brokered model, the physician is the source of criteria matching information, and the broker for the consent. In the information-enabled model, the patient is the source of consent, and the personal health record, controlled by the patient, acts as a broker of the criteria matching information coming from the physician.

Model Description

We identified the main flow of consent information going between the principal investigator and the patient, with the physician serving as a broker between the other two actors. The criteria matching information also flowed between the principal investigator and the patient, but was less complete when it did so; the more complete data originated with the physician and the medical record. This led to a model containing three variations, or levels, of information flow of consent and criteria matching between the actors in patient recruitment. These levels are traditional, brokered, and information-enabled (Figure 1). In the traditional level, consent and criteria matching flow directly between the principal investigator and the patient. In the brokered level, the physician identifies patients by matching

their clinical information with the selection criteria, and then requests patient consent. In the information-enabled model, clinical information flows from the physician to the patient, where it can be matched with selection criteria, and patient consent is given directly by the patient to the principal investigator.

The traditional level is frequently implemented as recruitment posters, where principal investigators can publish matching criteria in advertisements, and patients can individually match their own data, and give consent and volunteer participation when they consider themselves eligible. This model has two main limitations: the matching criteria must be understood by the patient, and the recruitment must be initiated by the patient. While most traditional implementations use paper-based methods, some use electronic tools to facilitate recruitment. For example, BreastCancerTrials.org allows a patient to enter clinical information, and match multiple clinical trials at once, leading to potentially more visibility of trials, at least to those patients who use the system.

At the brokered level, the physician performs the matching of clinical data to the patient, at the request of the principal investigator. The physician also asks the patient for consent to participate. The physician is the provider of the clinical data, and the broker for patient consent. This is potentially a more efficient mechanism, because the physician is a single point of access to multiple patients. In addition, the physician may have more precise matching information with clinical information in the medical record, allowing for more complicated or more sensitive matching criteria. Another advantage to the brokered level is that patient consent may be more likely than other mechanisms, since physician recommendations for participation in clinical trials is known to be a significant factor influencing patient decision making. The disadvantage of this model is that patient recruitment is dependent on the physician participating as a broker in trial recruitment. Like the traditional level, the brokered level also benefits from electronic data by improving the efficiency of implementation. For example, electronic data can be queried for matching criteria, and then alerts can be delivered to the physician during convenient opportunities for patient recruitment.

At the information-enabled level, the data that would be queried in the brokered level are moved to the personal health record (PHR), where they can be queried with consent of the patient. Patients who have elected to be notified of potential trials can have data queried for matching criteria. These data can include both patient-entered data, and data collected in electronic medical records and then transferred to PHRs. The benefits of the information-enabled level is that the data are as complete as the data in the brokered level, but consent is then obtained directly from the patient as in the traditional level. The main disadvantage is that it depends on broad acceptance and use of PHRs, far above current use. In addition, the enabled level may not be as efficient as the brokered level, because patients are not being recommended directly by their physicians. However, other factors that influence participation, such as timelines for recruitment may be better facilitated at the enabled level.

The actual implementation of While all three levels make use of electronic data to improve the efficiency of an otherwise paper-based process, the brokered and enabled levels also can use data either passively and actively. For example, at the brokered level, passive use of data would include querying the data in the EMR to identify a set of patients who could be recruited, while active use would be alerting physicians to recommend participation during a visit. At the enabled level, passive use of data would include querying to identify available trials that matched an individual patient's data, while active use would alert patients directly of trials they matched, and request participation directly. Figure 2 indicates how this would work in implementation.

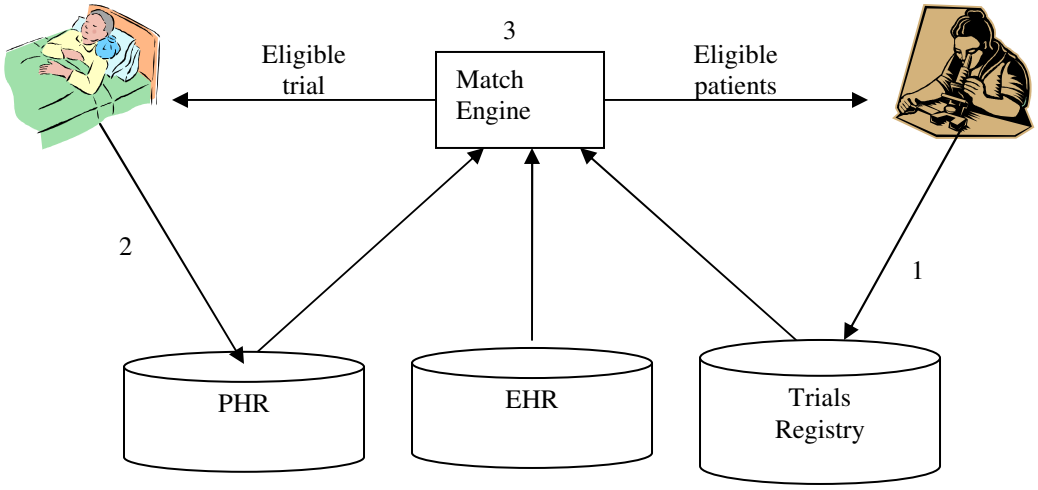

Figure 2: 1) A researcher submits a study protocol into a clinical trials registry. Then, the researcher gets a study population size returned via the match engine. 2) Patients enter their PHR and the match engine provides a list of the clinical studies they are eligible for.

Discussion

Our model of patient recruitment identifies two important requirements: the patient must be matched to the appropriate trial, and the patient must give consent to be recruited. While the patient and principal investigator are required stakeholders, we have also identified the physician as an important stakeholder, who likely has the most complete and detailed information available for matching patients to clinical trials, and can broker between the patient and principal investigator in obtaining consent for participation. Based on the relationships between the requirements and the significant stakeholders, we identified three variations or levels of interaction among the stakeholders: traditional, brokered and information-enabled. At each level, the requirements are facilitated by the use of electronic data.

Our model is significant because it identifies how personal health records specifically can be used, and how the implementation of the model changes with PHR use. Even the more advanced models of recruitment using electronic health records have only addressed mechanisms for automated identification of potential subjects through the electronic medical records of physicians [13]. In addition, our model informs how data can be passive, by allowing access, or active, by evoking alerts. Finally, the model describes information flow independent of the format of that data (paper-based or electronic), and identifies the specific benefits of electronic data.

There are significant issues that still need to be resolved with this model, relating to legal, privacy and institutional review board requirements. Current use has allowed the physician to serve as a broker for consent, but also to access the medical record data directly. This is possible because the physician already has access o the data, and therefore it does not expand the scope of who looks at the data. In the information enabled model, consent is obtained from the patient for participation, but it is not clear what level of consent is needed to identify potential matches to studies. Since the access to the data is by a machine, rather than an individual, such direct access to data may be allowed. In this way, the matching would follow a policy similar to the privacy policy in use at Google Mail [14]. The acceptance of Google and Gmail may indicate the potential of computing matches rather than querying directly. Further model development and research may be necessary to clearly specify the legal and privacy issues.

While this model was successful at identifying the potential use of PHRs in recruitment, it is speculative, and its realization depends directly on the maturity and use of PHRs in the future. Currently, the use of PHRs is still too low to test the utility of the model. However, the model does have content validity. In addition, the value of the model is that it is speculative, that it predicts what the clinical trials recruitment approach may become after widespread adoption of PHRs. The model is also useful in identifying requirements of PHRs, and in directing PHR architecture, while PHR models are still being debated [15]. Finally, the model informs considerations for consent law based on PHRs and patient recruitment.

There are multiple limitations to the model development we have employed. First, our data source from which we derived stakeholders and significant requirements was limited, based on the experience of two experts in recruitment. While we have been involved in multiple levels of patient recruitment for clinical trials, other experts may have identified different stakeholders, that if considered, would be seen as more important than those included in our model. We developed the model using a methodology adapted from grounded theory analysis. Typically, grounded theory is based on field work data, rather than expert-generated topics. In addition, while our analysis did proceed somewhat iteratively, it is not clear that we specifically reached a point of theoretical saturation. However, it was clear that we were not biased in our analysis approach and in our results, as we were surprised by the outcome of the process. Finally, we were limited by the size of the study, and did not perform other studies to triangulate our findings.

Conclusion

We developed a model that describes variations of how stakeholders meet requirements for patient recruitment to clinical trials. This model was based on a qualitative development process. The model describes necessary information flow between investigators, patients and physicians, and specifically predicts how the flow of information is affected by the projected use of personal health records.

References

1. Sung, N.S., et al., *Central Challenges Facing the National Clinical Research Enterprise.* JAMA, 2003. 289(10): p. 1278-1287.

2. Mowry, M. and D. Constantinou, *Electronic Health Records: A Magic Pill?* Applied Clinical Trials 2007.

3. Canavan, C., et al., *Integrating Recruitment into eHealth Patient Records.* Applied Clinical Trials, 2007.

4. Rubin, R., et al., *The Diabetes Prevention Program: recruitment methods and results.* Controlled Clinical Trials, 2002. 23: p. 157-71.

5. Sinackevich, N. and J.-P. Tassignon, *Speeding the Critical Path.* Applied Clinical Trials, 2004.

6. Sullivan, J., *Subject Recruitment and Retention: Barriers to Success.* Applied Clinical Trials, 2004.

7. Lara, P.N.J., et al., *Prospective evaluation of cancer clinical trial accrual patterns: identifying potential barriers to enrollment.* J Clin Oncol, 2001. 19(6): p. 1728-33.

8. Schain, W., *Barriers to clinical trials, part 2: knowledge and attitudes of potential participants.* Cancer, 1994. 74: p. 2666-71.

9. Mansour, E., *Barriers to clinical trials, part 3: knowledge and attitudes of health care providers.* Cancer, 1994. 74: p. 2672-5.

10. Fink, E., et al., *Selection of patients for clinical trials: an interactive web-based system.* Artificial Intelligence in Medicine, 2004. 31(3): p. 241-254.

11. Carlson, R., et al., *Computer-based screening of patients with HIV/AIDS for clinical-trial eligibility.* Online J Curr Clin Trials, 1995.

12. Mandl, K. and I. Kohane, *Tectonic shifts in the health information economy.* N Engl J Med 2008. 358(16): p. 1732-7.

13. Butte, A.J., D.A. Weinstein, and I.S. Kohane, *Enrolling patients into clinical trials faster using RealTime Recuiting.* Proc AMIA Symp, 2000: p. 111-5.

14. Google. *Google Privacy Policy*. 2008 [cited 2008; Available from: http://www.google.com/privacypolicy.html.

15. Smolij, K. and K. Dun, *Patient health information management: searching for the right model.* Perspect Health Inf Manag, 2006. 3: p. 10.