An Agent-Based Approach to Real-Time Patient Identification for Clinical Trials

Gareth Tyson[1], Adel Taweel[1], Simon Miles[1], Michael Luck[1], Tjeerd Van Staa[2], and Brendan Delaney[1]

[1] King's College London, UK
[2] General Practice Research Database, UK

Abstract. Patient recruitment for clinical trials is expensive and has been a significant challenge, with many trials not achieving their recruitment goals. One method that shows promise for improving recruitment is the use of interactive prompts that inform practitioners of patient eligibility for clinical trials during consultation. This paper presents the ePCRN-IDEA recruitment system, which utilises an agent-based infrastructure to enable real-time recruitment of patients. In essence, whenever patients enter a clinic, the system compares their details against eligibility criteria, which define the requirements of active clinical trials. If a patient is found to be eligible, a prompt is raised to notify the user. In this way, it becomes possible for recruitment to take place quickly in a cost effective manner, whilst maintaining patient trust through the involvement of their own health care practitioner.

1 Introduction

Clinical trials are the gold standard by which medical research is evaluated. They are used to study various aspects of medical science, as well as being a vital stage in the deployment of new drug treatments. Currently, however, such trials are frequently unsuccessful at recruiting sufficient patients. A review of the UK Medical Research Council found that only 31% of trials actually recruited to their planned target, with 30–40% of costs arising during the recruitment phase alone [1]. This is because discovering and contacting eligible potential recruits is both logistically and legally challenging. Consequently, many research projects take far longer to complete than is desirable, resulting in an unnecessary burden for those who could potentially benefit from the results.

The main challenge for patient recruitment lies in locating and contacting patients in a sufficiently timely manner to allow them to participate. However, the ease with which this can be done varies dramatically with the type of trial; for instance, recruitment can be challenging for trials that have high recruitment targets or complex eligibility criteria. Currently, recruitment is performed in a highly laborious manner, which is ill-suited to the above situations. It often involves a human recruitment agent visiting clinics in an attempt to locate suitable patients (e.g. asking practitioners or searching local medical records). This creates significant overhead as it is both slow and costly, as well as non-scalable

for most trials. For example, a trial investigating rare ailments might need an extensive number of visits to locate sufficient patients.

In consequence, it is of paramount importance to address the recruitment challenges within trials to ensure the future efficacy of medical research. It is therefore necessary to find a scalable way by which eligible patients can be discovered. To address this, we propose replacing human agents with software agents that permanently reside within clinics, with the aim of searching for patients who might satisfy trial eligibility criteria. Through this, the agent could inspect patient information in real-time to ascertain eligibility before presenting notifications to local practitioners. Importantly, by using a software agent this could be done rapidly within a consultation before a patient has left, thereby shortening the recruitment lifecycle (as well as enabling trials based on incidental cases). The paper's contributions are therefore as follows:

– A critique of existing clinical trial recruitment approaches, highlighting that current techniques are both slow and expensive.
– An agent-based distributed architecture called ePCRN-IDEA that enables real-time recruitment of patients, whilst avoiding the key limitations of existing approaches.
– A procedure by which software agents can guide the recruitment of patients to their most appropriate trials.

The rest of the paper is structured as follows; first the background to the research is discussed in Section 2 before detailing the ePCRN-IDEA recruitment system in Section 3. Following this, a discussion is presented in Section 4, alongside future work and conclusions.

2 Background and Related Work

2.1 Recruitment in Clinical Trials

Clinical trials are a challenging stage in the research of clinicians due to the complexity of recruiting patients for participation. Many types of trials can suffer from such difficulties; for instance, trials that have potential recruits who are widely distributed over many clinics (e.g. primary care) are extremely difficult to recruit for due to the intense resource requirements. Similarly, trials with certain types of patient eligibility criteria can be difficult to recruit for; for example, a trial dealing with incidental/acute conditions would need a practitioner to immediately identify patients in real-time. This can be highly challenging, especially when dealing with complex eligibility criteria or trials that require immediate actions (e.g. a change of drug treatments). This has a significant impact on medical research and stunts potential life-saving advances. Recruitment is currently performed in a number of ways; examples include:

– *Recruitment Visits*: Using recruiters to visit or contact practices so that they can search local medical records and/or ask local practitioners.

– *Advertisements*: Using posters, web sites, mailing lists or newspaper adverts describing what is required for the trial.
– *Practitioners*: Using practitioners to suggest patients who might be eligible. Alternatively, practitioners may be asked to look for patients (during consultations) who match the criteria in real-time.

Clearly, the above approaches are slow and often quite expensive. The predominant approach of *employing recruiters* to contact and visit clinics is laborious and often infeasible, especially when dealing with the above types of trials. For instance, different clinics often use different database systems, thereby requiring a high degree of training for recruiters; something that is exacerbated by the limited in-house IT training in many clinics. Further, attempts to standardise this database search procedure (e.g. MIQUEST) are often poorly understood by researchers, whilst such standards can often be undermined by variations in coding and medical dictionaries (e.g. if multiple databases use different semantics). Consequently, there is little automated support for finding eligible patients, often leaving recruitment very much as an ad-hoc process that falls outside of recognised standards; Moreover, there is little infrastructure available that can be exploited by clinicians to assist in the stage.

As an alternative to the recognised approaches, Embi et. al. [2] propose the use of a Clinical Trial Alert (CTA) system, through which practitioners could be notified while they are in consultation with a patient who is eligible for a trial. Their work shows that significant increases in recruitment could be achieved. However, their pilot study was only deployed in a single clinic with a single trial, in an attempt to study the improvements in recruitment. Other similar techniques have also seen only limited large-scale testing, e.g. [3]. Consequently, a number of issues (e.g. scalability) have not been investigated, leaving the sole focus on the recruitment outcomes. To address these concerns, this paper therefore focusses on more infrastructural aspects for enabling a large-scale trial alert system, specifically through the use of intelligent agents.

2.2 Agents in Healthcare

Agents have emerged as a prominent technology for handling a range of real-world problems [4]. An agent can be defined as "a computer system that is situated in some environment, and that is capable of autonomous action in their environment in order to meet its delegated objectives" [5]. Agents in healthcare have seen widespread investigation; in general, their use attempts to address endemic issues such as (*i*) distributed information and expertise, (*ii*) unpredictable dynamics, and (*iii*) uncertainty in reasoning and data.

For example, MAID [6] is an agent-based system for integrating heterogeneous data sources within a hospital environment. The hospital studied had 24 departments, each using their own information systems. To address this, agents were constructed to interoperate with each system to monitor changes and retrieve data for insertion into a central repository. In a subsequent work, HealthAgents [7] went beyond MAID to also enable decision support, specifically

for diagnosing brain tumours. A range of agent-based systems have also been proposed for handling distributed expertise. These includes using agents to enable better communication between healthcare workers based on ambient information, e.g. their role, location etc. [8], as well as using agents to remotely monitor patients [9][10]. These systems also often involved data analysis; S(MA)^{2}D [10], for instance, uses statistical analysis to cluster patients into similar groups. This ability to scalably perform data analysis in real-time, clearly, also shows potential for enabling the type of eligible patient identification discussed previously. Despite this, so far little work has been performed into using agents to improve clinical trial recruitment. Consequently, the rest of this paper explores exploiting the properties of agents to enable real-time recruitment of patients to trials.

3 ePCRN-IDEA Recruitment System

This section presents the ePCRN-IDEA recruitment system, which is designed to enable real-time identification of eligible patients. We first present the overall architecture before describing the individual components, focussing on the behaviour of the clinic-based agents.

3.1 Overview

The core goal of the ePCRN-IDEA recruitment system is to improve patient recruitment. In order to do so, researchers must formally define the eligibility criteria of participants, then distribute it to software agents that reside locally on practitioners' computers. These software agents listen to interactions between the practitioner's local Electronic Healthcare Record (EHR) database and the user in an attempt to locate patients who are eligible for trials. Importantly, this occurs in real-time during the consultation, thereby allowing a pop-up to be generated, notifying the practitioner of the patient's eligibility. In this way, the patient can be instantly consulted regarding the trial and, if interested, recruited via a web interface. The key architectural entities in the system are as follows:

– *LEPIS:* An agent that resides at primary care practices and investigates the eligibility of any present patients, termed the Local Eligible Patient Identification Service.
– *CCS:* A point of storage and distribution that allows clinical researchers to inject new trials into the system, termed the Central Control Service.
– *CTMS:* A website that handles the actual recruitment process once an eligible patients has been discovered, termed the Clinical Trial Management System.

Prototype implementations of all these components have already been developed. Fig. 1 provides an overview of these, as detailed in the rest of this section.
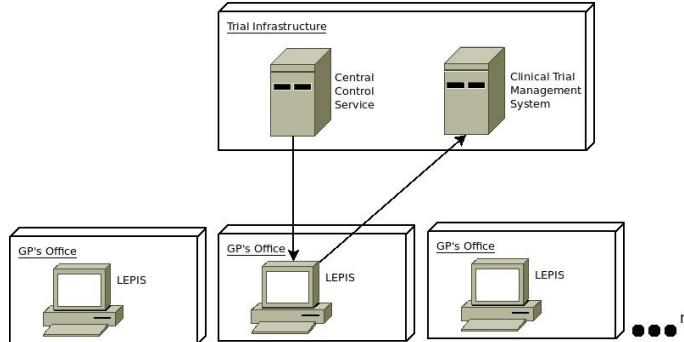
Fig. 1. Overview of ePCRN System

3.2 Central Control Service (CCS)

A *trial store* maintains a repository of active trials within the system, as entered by any clinical researchers wishing to recruit patients. Trials are stored using a standard model (the PCROM standard [11]), which defines the data format used to represent the various aspects of the trials. This includes a description of the trial as well as the eligibility criteria, which can be based on a variety of aspects ranging from a pre-computed list of eligible patient identifiers to complex diagnosis information. The trial store is realised within the ePCRN-IDEA system through the Central Control Service (CCS), which is a service used to manage all the trials. When a new trial is created, it is injected into the system via the CCS and stored in a MySQL database back-end before being made accessible to the appropriate (and authorised) parties — namely, the recruitment agents. These are accessed securely using an encrypted SQL connection initiated by each agent to the CCS.

3.3 Local Eligible Patient Identification Service (LEPIS)

The Local Eligible Patient Identification Service (LEPIS) is a Java-built software agent that resides on every practitioner's PC (who is authorised to recruit patients). The purpose of this agent is to actively discover any eligible patients who might be seen by the practitioner. It therefore operates as a replacement for the existing human agents who visit clinics in an attempt to locate suitable patients. Every LEPIS agent is required to obtain two distinct sets of information. First, it must acquire information about patients in real-time as they enter the clinic. Second, LEPIS must also acquire the necessary trial information (from the CCS) so that it can compare patients against each trial's eligibility criteria; ideally, eligibility should also be computed locally to avoid unnecessary delays or bottlenecks.

Accessing Patient Information To discover the eligibility of a patient for a trial, it is first necessary for LEPIS to gain access to any pertinent information

related to the patient. Whenever a patient enters a clinic, the practitioner opens his or her medical record using the Electronic Healthcare Record (EHR) database on their desktop computer. This medical record contains a range of demographic and medical information about the patient, allowing practitioners to retrieve information about the patient during the consultation, as well as enter new information. This offers an existing platform through which a recruitment agent can access information about patients. We have modified a popular EHR system, *Vision* [12], to interact with LEPIS. Whenever a patient record is opened or modified, the information is passed to LEPIS (through a standard file using a shared XML schema) so that it can attempt to locate trials for which the patient may be eligible. Information is coded using standard Read Codes and Multilex Drug Codes to allow LEPIS and the EHR to understand each other. LEPIS is therefore given real-time access to information about any patients who are currently in consultation. Although, evidently, eligibility criteria is limited to those attributes provided by the EHR, which can vary based on both policy and EHR implementation (many EHR vendors exist).

Accessing Trial Information To allow a LEPIS agent to compute a patient's eligibility, it must first gain access to trial information. Ideally, this should be stored locally to enable real-time eligibility checks within a consultation. However, evidently, this is largely infeasible with the huge number of active trials running; e.g. clinicaltrials.gov currently lists well over 100,000 trials. Consequently, it is necessary for each agent to independently select the most appropriate trials for its clinic and practitioner. Each agent thus maintains a set of trials T of size n, as limited by the host's local resources. An agent therefore selects n based on the capabilities of its host, by performing eligibility checks on a random set of trials repeatedly for one second; n is then set as the number of iterations. It then keeps a persistent record of all patient identifiers, Read Codes and Multilex Codes provided by the EHR to build up a profile of the clinic. Using this information, n trials are retrieved from the CCS through the following process:

1. LEPIS attempts to retrieve a set of n trials from the CCS containing:
   a) p trials that includes a known patient registered within the clinic ($p =< n$);
   b) if $p < n$, c trials that includes coded information previously encountered within the clinic ($c =< n - p$); and
   c) if $p + c < n$, r randomly selected trials ($r =< n - p - c$).
2. Remove any trials that are fully recruited.

These two steps are repeated throughout an agent's lifetime with a configurable interval, which is set to 24 hours by default.

Computing Eligibility and Generating a Popup When LEPIS acquires patient information from the EHR, it must compare it against the eligibility criteria of any known trials. This is a simple process that currently involves iteratively computing eligibility for each known trial and then selecting a random one if multiple are found. A popup is then generated to notify the user. Fig. 2 shows a screenshot of the user interface.
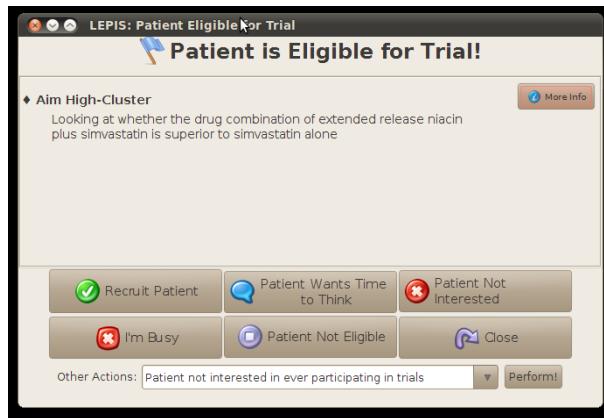
Fig. 2. Screenshot of LEPIS Popup

3.4 Clinical Trial Management System (CTMS)

If a patient is interested in being recruited for a particular trial, it is then necessary to actually perform the recruitment procedure. This is not handled by the local agent; instead, an external website is used, called the Clinical Trial Management System (CTMS). The CTMS is securely accessed by the practitioner and then used to register the patient's interest in being recruited. Any necessary steps can then be taken, e.g. contacting the patient, recording information etc.

4 Discussion, Future Work and Conclusion

The ePCRN-IDEA recruitment systems differs substantially from previous work in that it is agent-based. Our early evaluation shows that the agent-based approach is promising and has several potential advantages over traditional client-server approaches (e.g [2]). Through the use of agents, intelligence is decentralised within the system so that both computation and decision making is independently performed by each agent.

Primarily, we have used the approach to enable superior scalability. This is critical, especially for patient recruitment in primary care. Even for common diseases, eligible patients are thinly spread across many practices, which potentially number several thousand. In the UK, alone, there are more than 10,000 practices. A client-server approach could be realised in two ways: either all patient information could be transmitted to a server, where it centrally computes eligibility; or all trial information could be transmitted to all clinics for local computation. The former is non-scalable as well as dangerous in terms of privacy and security. The latter, however, is also highly non-scalable as the number of trials (e.g. greater than 100,000) alongside the size of each trial description (e.g. 0.5 MB) makes it impossible for all clinics to know of all trials. Consequently, to address this, we embed intelligence within the agents to learn how to best select trials

for their host clinic, exploiting the local knowledge (and computational abilities) of each agent, rather than burdening a central point.

From our initial phase-1 prototype we have identified a number of future lines of work. First, we aim to complete a full system deployment within the UK primary healthcare system, thereby enabling a detailed quantitative evaluation. Beyond this, we also intend to extend the agent capabilities. Key research lines include, (*i*) *inter-agent collaboration*: allowing agents to build societies to better enable information and resource sharing (e.g. based on disease areas, localities); (*ii*) *interface adaptation*: allowing agents to learn (and share) the behaviour of users to adapt interaction; and (*iii*) *trial negotiation*: allowing agents to negotiate with each other to best distribute trials based on runtime conditions.

References

1. A. McDonald, R. Knight, M. Campbell, V. Entwistle, A. Grant, J. Cook, D. Elbourne, D. Francis, J. Garcia, I. Roberts, and C. Snowdon, "What influences recruitment to randomised controlled trials? a review of trials funded by two uk funding agencies," *Trials*, vol. 7, no. 1, p. 9, 2006.
2. P. J. Embi, A. Jain, J. Clark, S. Bizjack, R. Hornung, and C. M. Harris, "Effect of a clinical trial alert system on physician participation in trial recruitment," *Arch Intern Med.*, vol. 195, no. 19, 2005.
3. B. L. Rollman, G. S. Fischer, F. Zhu, and B. H. Belnap, "Comparison of electronic physician prompts versus waitroom case-finding on clinical trial enrollment," *Journal of General Internal Medicine*, vol. 23, no. 4, 2008.
4. M. Luck, P. McBurney, and C. Preist, *Agent Technology: Enabling Next Generation Computing (A Roadmap for Agent Based Computing)*. AgentLink, 2003.
5. M. Wooldridge, *Introduction to Multiagent Systems*. John Wiley & Sons, Inc., 2009.
6. R. Cruz-Correia, P. Vieira-Marques, P. Costa, A. Ferreira, E. Oliveira-Palhares, F. Araújo, and A. Costa-Pereira, "Integration of hospital data using agent technologies - a case study," *AI Commun.*, vol. 18, August 2005.
7. H. González-Vélez, M. Mier, M. Julià-Sapé, T. Arvanitis, J. García-Gómez, M. Robles, P. Lewis, S. Dasmahapatra, D. Dupplaw, A. Peet, C. Arús, B. Celda, S. Van Huffel, and M. Lluch-Ariet, "Healthagents: distributed multi-agent brain tumor diagnosis and prognosis," *Applied Intelligence*, vol. 30, 2009.
8. M. D. Rodríguez, J. Favela, A. Preciado, and A. Vizcaíno, "Agent-based ambient intelligence for healthcare," *AI Commun.*, vol. 18, pp. 201–216, August 2005.
9. V. Koutkias, I. Chouvarda, and N. Maglaveras, "A multiagent system enhancing home-care health services for chronic disease management," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 9, pp. 528–537, dec. 2005.
10. A. Rammal, S. Trouilhet, N. Singer, and J.-M. Pécatte, "An adaptive system for home monitoring using a multiagent classification of patterns," *Int. J. Telemedicine Appl.*, vol. 2008, pp. 3:1–3:8, January 2008.
11. S. M. Speedie, A. Taweel, I. Sim, T. N. Arvanitis, B. Delaney, and K. A. Peterson, "The primary care research object model (pcrom): A computable information model for practice-based primary care research," *Journal of the American Medical Informatics Association*, vol. 15, no. 5, 2008.
12. "Vision." http://www.inps4.co.uk/vision/.