TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

PATIENT RECRUITMENT, FEASIBILITY EVALUATIONS AND USE OF ELECTRONIC HEALTH RECORDS IN CLINICAL TRIALS

A Nordic Approach

Niina Laaksonen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1595 | MEDICA – ODONTOLOGICA | TURKU 2021

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

PATIENT RECRUITMENT, FEASIBILITY EVALUATIONS AND USE OF ELECTRONIC HEALTH RECORDS IN CLINICAL TRIALS

A Nordic Approach

Niina Laaksonen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1595 | MEDICA – ODONTOLOGICA | TURKU 2021

University of Turku

Faculty of Medicine
Institute of Biomedicine
Pharmacology, Drug Development and Therapeutics
Drug Research Doctoral Programme (DRDP)
Turku, Finland

Supervised by

Professor Emeritus Risto Huupponen, MD, PhD
Institute of Biomedicine
University of Turku and
Turku University Hospital
Turku, Finland

Professor Emeritus Mika Scheinin, MD, PhD
Institute of Biomedicine
University of Turku and
Turku University Hospital
Turku, Finland

Adjunct Professor Juuso Blomster, MD, PhD
Department of Cardiology
University of Turku and
Turku University Hospital
Turku, Finland

Reviewed by

Adjunct Professor Tarja Laitinen, MD, PhD
Director of Research
Tampere University Hospital
Tampere, Finland

Professor Miia Turpeinen, MD, PhD
Research Unit of Biomedicine
University of Oulu and
Oulu University Hospital
Oulu, Finland

Opponent

Adjunct Professor Juhana J. Idänpään-Heikkilä, MD, PhD
University of Helsinki
Helsinki, Finland
Medical Director, GSK Finland

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-951-29-8696-5 (PRINT)
ISBN 978-951-29-8697-2 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama, Turku, Finland 2021

To my Family

UNIVERSITY OF TURKU
Faculty of Medicine
Institute of Biomedicine
Pharmacology, Drug Development and Therapeutics
NIINA LAAKSONEN: Patient Recruitment, Feasibility Evaluations and Use of Electronic Health Records in Clinical Trials – A Nordic Approach
Doctoral Dissertation, 135 pp.
Drug Research Doctoral Programme (DRDP)
November 2021

ABSTRACT

Clinical trials constitute an important cornerstone for the development of new drugs. Patient recruitment is one of the main challenges in clinical trials. Pharmaceutical companies apply feasibility evaluations to identify potential countries, investigators and study sites for their trials and to evaluate their potential for successful patient recruitment. Electronic health records (EHR) maintained by health care providers are regarded as one potential tool for improving patient identification and recruitment for clinical trials.

This study investigated patient recruitment and trial feasibility evaluations in the Nordic countries and the role and usability of EHR data in those processes. The pharmaceutical industry's view was investigated by conducting semi-structured qualitative interviews of 21 respondents from Finland, Sweden, Norway and Denmark. Additionally, the usability of one commercial EHR research platform for identifying patients from Turku University Hospital's EHR system was tested in comparison with a manual search.

The success or failure of patient recruitment was influenced by many sponsor-related, investigator/site-related, patient-related, collaboration-related and start-up-related factors. Most trials had recruited their patients by reviewing the hospitals' EHR data, but its use was much less frequent already during the feasibility evaluation phase. Feasibility evaluation was found to be a complex and time-consuming process for estimating the number of potential trial patients. The sponsors did not use EHR tools for such evaluations, mainly because of legislative barriers. Although the EHR data search tools have limitations in accuracy, they were seen to have great potential for identifying trial participants from the hospital EHR, for example by reducing the manual work.

The comprehensive data in the EHR systems in the Nordic countries offer a possibility for more accurate identification of trial participants in the feasibility evaluations and may thus contribute to the success of recruitment. The data protection legislation and its interpretation should be harmonized for the use of EHR data. Continuous improvements in the EHR systems' technical accuracy and data quality will be needed to enhance the successful use of EHR data in future clinical trials.

KEYWORDS: patient recruitment, feasibility evaluation, clinical trials, pharmaceutical industry, electronic health records, EHR.

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Biolääketieteen laitos
Farmakologia, lääkekehitys ja lääkehoito
NIINA LAAKSONEN: Potilasrekrytointi, toteutettavuuden arviointi ja elektronisten potilastietojärjestelmien hyödyntäminen kliinisissä tutkimuksissa – Pohjoismainen näkökulma
Väitöskirja, 135 s.
Lääketutkimuksen tohtoriohjelma
marraskuu 2021

TIIVISTELMÄ

Kliiniset lääketutkimukset ovat uusien lääkkeiden kehityksen kulmakivi. Tutkimuspotilaiden rekrytointi on merkittävä haaste näissä tutkimuksissa. Lääkeyritykset tekevät toteutettavuusarviointeja tunnistaakseen potentiaalisia tutkimukseen osallistuvia maita, tutkijoita ja tutkimuskeskuksia ja arvioidakseen niiden mahdollisuuksia onnistua potilaiden rekrytoinnissa. Terveydenhuolto-organisaatioiden ylläpitämät elektroniset potilastietojärjestelmät (EHR) ovat tässä eräs mahdollinen työkalu.

Tässä tutkimuksessa tutkittiin potilaiden rekrytointia ja tutkimusten toteutettavuusarviointeja Pohjoismaissa ja EHR:n roolia ja käytettävyyttä näissä prosesseissa. Näitä tekijöitä tekijöitä tutkittiin lääketeollisuuden näkökulmasta laadullisilla teemahaastatteluilla (21 haastateltavaa Suomesta, Ruotsista, Norjasta ja Tanskasta). Yhden kaupallisesti saatavilla olevan EHR-hakutyökalun tarkkuutta halutun potilasjoukon löytämisessä verrattiin perinteiseen, manuaaliseen hakuun Turun yliopistollisen sairaalan potilastietojärjestelmästä.

Potilaiden rekrytoinnin onnistumiseen tai epäonnistumiseen vaikutti moni toimeksiantajaan, tutkijaan/tutkimuskeskukseen, potilaaseen ja tutkimuksen aloitustoimenpiteisiin liittyvä tekijä sekä näiden tahojen yhteistyö. Valtaosassa tutkimuksista tutkittavat rekrytoitiin keskuksen omista potilaista EHR:a hyödyntäen, mutta EHR:n käyttö potilasmäärän arvioinnissa ennen tutkimuksen alkua oli vähäistä. Toteutettavuusarvioinneissa tehdyt potilasmäärien arviot nähtiin monimutkaisina ja aikaa vievinä prosesseina. Toimeksiantajat eivät käyttäneet EHR-työkaluja lainkaan, pääasiassa tietosuojalainsäädäntöön liittyvistä syistä. Vaikka EHR-hakutyökalujen tarkkuudella on rajoitteensa, niitä voidaan hyödyntää esimerkiksi vähentämään manuaalista työtä potilaiden identifioinnissa.

Terveydenhuollon kattavat EHR-järjestelmät tarjoavat Pohjoismaissa hyvän mahdollisuuden tutkimuspotilaiden tarkempaan identifiointiin, joka omalta osaltaan vaikuttaa rekrytoinnin onnistumismahdollisuuksiin. Tietosuojalainsäädäntöä ja sen tulkintoja on harmonisoitava EHR:n käytön hyödyntämiseksi. EHR-hakujen teknistä tarkkuutta ja tiedon laatua on edelleen parannettava sen menestyksekkään käytön lisäämiseksi tulevaisuuden kliinisissä tutkimuksissa.

AVAINSANAT: potilasrekrytointi, toteutettavuusarviointi, kliiniset tutkimukset, lääketeollisuus, elektroniset potilastietojärjestelmät, EHR.

Table of Contents

Abbreviations

AF	Atrial fibrillation
AI	Artificial intelligence
CDW	Clinical data warehouse
CRA	Clinical research associate
CRO	Contract research organization
CTA	Clinical Trial Application
DKMA	Danish Medicines Agency (Denmark)
eCRF	Electronic Case Report Form
EHR	Electronic health record
EHR4CR	Electronic health records for clinical research initiative
EMA	European Medicines Agency
FDA	U.S. Food and Drug Agency
FIMEA	Finnish Medicines Agency (Finland)
FPI	First patient in
GCP	Good clinical practice
ICD9	The International Classification of Diseases, 9th revision
ICD10	The International Classification of Diseases, 10th revision
i-HD	The European Institute for Innovation through Health Data
IMA	Icelandic Medicines Agency (Iceland)
IMI	Innovative Medicines Initiative
LPI	Last patient in
MPA	Medical Products Agency (Sweden)
NLP	Natural language processing
NoMA	Norwegian Medicines Agency (Norway)
RCT	Randomized clinical trial
WHO	World Health Organization

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Laaksonen N., Bengtström M., Axelin A., Blomster J., Scheinin M., Huupponen R. Success and failure factors of patient recruitment for clinical trials and the role of the electronic health records – a qualitative interview study in the Nordic countries. *Manuscript*.

II Laaksonen N., Bengtström M., Axelin A., Blomster J., Scheinin M., Huupponen R. Clinical trials site identification practices and the use of electronic health records in feasibility evaluations. *Clinical Trials*, 2021; doi:10.1177/17407745211038512. Advance online publication.

III Laaksonen N., Varjonen J-M, Blomster M., Palomäki A., Vasankari T., Airaksinen J., Huupponen R., Scheinin M., Blomster J. Assessing an Electronic Health Record research platform for identification of clinical trial participants. *Contemporary Clinical Trials Communications*, 2021; 21: 2451–8654.

1 Introduction

Randomized clinical trials constitute a cornerstone for the development of new drugs. Drug development is a very long process in which clinical trials are often the costliest and most time-consuming phase. Possible challenges in the conduct of clinical trials are directly reflected in the duration of the entire drug development process, and thus have a major influence on how quickly new treatments can be delivered to patients.

Recruitment of participants – volunteer patients - into clinical trials has been widely recognized as a critical challenge for decades (Hunninghake, Darby, and Probstfield 1987; Prescott et al. 1999; Spaar et al. 2009; Bentley et al. 2019). Success in patient recruitment implies that the trial recruits its patients according to a plan (both in terms of time and the number of patients) set out before the trial's commencement. Barriers to recruitment have been investigated extensively, but the solutions that have been identified still remain incomplete (Treweek et al. 2018) as there are numerous factors that can influence the success of recruitment. Because of the multitude of recruitment success and failure factors, it has proved to be very difficult to predict the success of recruitment (Bruhn et al. 2019; White and Hind 2015), and only approximately one in two or three clinical trials manages to recruit its participants as planned (McDonald et al. 2006; Sully, Julious, and Nicholl 2013). In addition to the recruitment failure factors, also the success factors were investigated in the present study, because so far, fewer reports have been focused on these issues.

Only limited information is available on the recruitment success and failure factors in the Nordic countries. In a Swedish study, clinical investigators were surveyed on these issues (Isaksson et al. 2019), but information covering all Nordic countries is lacking. Previous qualitative research on the topic has mainly concentrated on identifying investigators' or patients' perceptions, but the current study aims to shed light on the views of the pharmaceutical industry, i.e the organizations that are actually conducting almost all large-scale clinical trials on pharmaceuticals (hereafter called "trial sponsors"). It is important to clarify the views of the industry, as approximately 65% of all clinical drug trials conducted in

the Nordic countries are industry-sponsored (DKMA 2020; FIMEA 2020; IMA 2020; MPA 2020; NoMa 2020).

Many factors that have an impact on patient recruitment are already present in a clinical trial before its start. According to Briel and colleagues, almost 90% of the reasons for poor recruitment could have been anticipated already in the planning phase of a randomized clinical trial (Briel et al. 2016). Therefore, in order to obtain a more comprehensive picture of the challenges of patient recruitment, feasibility evaluations were included in the topics covered by the current study.

Trial feasibility evaluations aim to generate an understanding of and predictions for the execution scenarios of the planned study and to estimate the availability of eligible patients in different countries, to assess the overall timelines and cost of the trial, to identify and select potential investigators and trial sites, and to evaluate their potential for patient recruitment. Despite their obvious importance, less research has been devoted to feasibility evaluations than to patient recruitment. Especially at a Nordic level, there are very few published studies on feasibility evaluations (Dombernowsky, Haedersdal, et al. 2017).

Technological advances have recently introduced some novel features into patient recruitment and feasibility evaluations. Trial participants are increasingly recruited by employing digital technologies, such as social media and internet pages, while traditional methods such as newspaper advertisements are less frequently relied upon. Secondary use of health information contained in electronic health records (EHR) has increased in clinical trials during the recent years, with the emphasis being on the recruitment process (Mc Cord et al. 2019).

The use of EHR has been noted to enhance recruitment of trial participants, even if some critical views have also been expressed. Overall, the use of all types of data in the decision-making has increased and keeps increasing, including the use of EHR. The Nordic countries are in the front line concerning the secondary use of EHR data (Nordforsk 2019). The successful use of EHR data for better predicting patient recruitment or to relieve problems in recruitment could have a major positive impact on the Nordic countries' competitiveness in clinical trials. Therefore, research on the use of EHR in clinical trials is important and justified in the Nordic context.

This study investigated the nature and magnitude of possible problems in patient recruitment, the success and failure factors impacting on recruitment, and the patient recruitment methods employed in clinical trials carried out in the Nordic countries. The use of EHR in patient recruitment and in feasibility evaluations was evaluated, and the functionality and accuracy of an EHR research platform for finding eligible clinical trial participants from the EHR system of Turku University Hospital were tested.

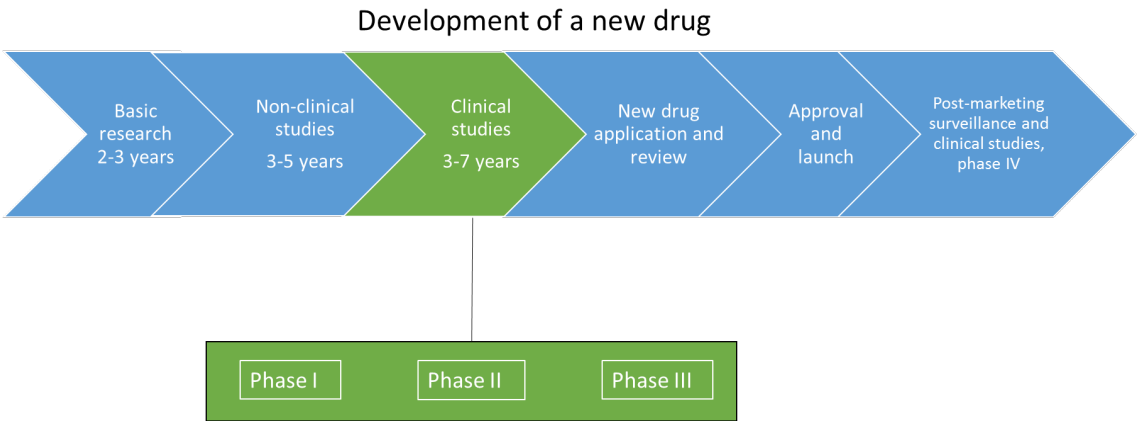
2 Review of the Literature

2.1 Clinical trials in drug development

Randomized clinical trials represent the foundation for evidence-based medicine and the development of new drugs. The development of a new drug from an idea to a product on the market usually takes approximately ten years and is stringently regulated (Taylor 2015). Traditionally, three distinct phases of clinical trials must be passed before a company can apply for marketing authorization (Figure 1): In phase I, the safety, tolerability and pharmacokinetic properties of the drug candidate are initially evaluated in healthy volunteers or sometimes in patients; furthermore, the pharmacodynamic effects of the drug candidate in humans are also examined, if possible. Phase II trials are usually performed in patients with the target disease, in order to identify the therapeutic range and dosing schedule, to tentatively explore the efficacy of the study drug and to obtain information on its safety and tolerability in the target population. Phase III trials are the most time-consuming phase, and include pivotal confirmatory trials of the efficacy and safety of the treatment in the target population (European Medicines Agency 1998). Trials belonging to the different phases are in reality being performed partially in parallel, and the phases are increasingly combined (phase I/II or II/III trials) in order to reduce the administrative burden caused by separate trials and to speed up the process. Because of recent advances in clinical medicine and research methods, the strict division of trials into phases has softened (Tenhunen, Turpeinen, and Kurki 2017). Examples of such advances are the application of personalized medicine, adaptive trial designs and new types of trials targeting on a certain genetic mutation, and not on a distinct clinical disease entity, i.e. basket trials.

The number of trial participants increases when the drug candidate moves forward in the clinical drug development pipeline: only some dozens of trial participants may be included in the first phase I trials, whereas thousands of patients may be enrolled in the pivotal phase III trials (U.S. Food and Drug Administration 2018). Clinical trials may also be performed with drugs that already have marketing approval; such trials may be called phase IV trials or post-marketing surveillance. The main objectives of phase IV trials are to check the drug's performance in a real-life setting, to evaluate its long-term risks and benefits and to discover any rare side

effects. In a phase IV trial, any rare or delayed effects of the drug can be observed in a much larger population of patients and over a much longer period of time compared to formal phase III trials.



The total cost for developing a new drug from an idea to a marketed product currently amounts to approximately 2.6 billion US dollars, on the average, but this estimate also includes the costs of failed projects (DiMasi, Grabowski, and Hansen 2016). The costs have increased – almost tripled in two decades (PhRMA 2019). Clinical development is the most expensive part of the drug development process. Depending on the therapeutic area, the average cost of a single clinical trial may vary from 1.4 million US$ (phase I) to 53 million US$ (phase III) (Sertkaya et al. 2016). A recent review estimated that the costs of randomized controlled trials ranged from 0.2 to 611.5 million US$ (Speich et al. 2018).

Globally, the volume of ongoing clinical trials of new pharmaceuticals is huge. There were 36 743 new clinical drug trials registered in the U.S. National Library of Medicine's Clinical Trials Registry, www.ClinicalTrials.gov, during the year 2020. Oncology, infectious diseases and central nervous system disorders are the most investigated therapy areas, with cancer trials being the ultimate leader with almost 5700 drugs in development, from the preclinical stage to those seeking marketing approval (Pharmaintelligence Informa 2019).

Most clinical trials are conducted in North America and in Europe, with regard to both the number of trial sites and the enrollment numbers (Song, Chee, and Kim 2019), but there seems to be an ongoing shift, especially of phase III clinical trials, to other geographical regions (Drain et al. 2018), especially to some Asian countries (Ali et al. 2019). This is mainly because of high trial costs and barriers to patient

recruitment in North America and in Europe (Scorr Marketing 2017; Song, Chee, and Kim 2019). This challenges the European countries to implement new strategies to maintain their role in the field of clinical trials.

Europe is traditionally regarded as an important region to conduct clinical trials because of the quality of its health care systems and its investigators' experience in conducting trials. Germany, France, Italy, the UK and Spain are regarded Europe's top five healthcare markets (Gehring et al. 2013). According to the statistics (2008 - 2017) of the U.S. National Library of Medicine, European countries were involved in 62% of all phase II and in 66% of all phase III trials, while their involvement in all phase I trials was 42% (Song, Chee, and Kim 2019). However, concerns have been raised on the declining trend of clinical trials being conducted in Europe. For example, the European Commission stated that the number of clinical trial applications (CTAs) had decreased by 25% in the EU as a whole from 2007 to 2011 (European Commission 2012). In a more recent study, a 12% decline in 2007–2011 was identified, but from 2014 to 2015, the number of CTAs increased markedly (by 10%) after a period of stagnation from 2012 to 2013 (Dombernowsky, Hædersdal, et al. 2017). According to that study, there seemed to be a positive development in the clinical trial landscape in Western Europe. However, it remained unclear whether this constituted a transient fluctuation or a new trend.

The number of clinical trials carried out in the Nordic coutries has decreased over the past years. In 2007, the total number of CTAs in the Nordic countries was 1147, while in 2011 it had fallen to 865 (Bengtström and Nybond, Scanbalt web page, date accessed 25Feb2021). The number of CTAs has continued to decline, as in 2019 the total number of CTAs in the Nordic countries was 817. The Nordic CTAs of 2019 represent 2.5% of all clinical trials registered in the U.S. National Library of Medicine Clinical Trials Registry in the same year (Table 1).

According to the statistics provided by the national authorities in the Nordic countries, almost three quarters of all new clinical trials in the Nordic countries in 2019 were phase II or III trials (Table 1). Two thirds (65%) of all clinical trials were industry-sponsored in the Nordic countries (Table 1), whereas globally, the share of industry-sponsored trials was only one third of all trials conducted (Atal et al. 2015; Drain et al. 2018).

When the number of clinical trials is related to the number of inhabitants in each country, Finland, Sweden, Norway and Denmark are all in the global list of top 20 countries (Drain et al. 2018). This density of clinical trial sites per million people was noted as being the 4th highest in Denmark (361 trial site-years/10^6 population), while in Sweden, Norway and Finland the respective densities were 253, 229 and 216. As a comparison, the USA had the highest site density (540), but for example, while Germany had a score of 325, two other European countries important for conducting clinical trials, the UK and Italy, were not on the top 20 list.

Table 1. Number of industry-sponsored and investigator-initiated clinical drug trials in the Nordic countries in the year 2019.

	Denmark	Sweden	Finland	Norway	Iceland	Total
Sponsorship						
Industry-sponsored	185 (58%)	161 (69%)	106 (77%)	73 (61%)	5 (100%)	530 (65%)
Investigator-initiated	134 (42%)	74 (31%)	31 (23%)	48 (39%)	0	287 (35%)
Phase [a]						
Phase I	38	34	13	15	0	100 (12%)
Phase II	123	66	39	46	3	277 (34%)
Phase III	121	85	61	44	2	313 (38%)
Phase IV	57	28	24	15	0	124 (15%)
Total	319	235	137	120	5	817

[a] In Denmark, trials in multiple phases are recorded in multiple categories. In Sweden, there were 22 trials for which no information was available on phase, or trials covering multiple phases.
Data from the Annual Reports 2019 of authorities in Finland (FIMEA), Sweden (MPA), Norway (NoMA), Denmark (DKMA) and Iceland (IMA).

2.2 Challenges of clinical trials

As the average age of the world's population has increased, also drug development has shifted to focus on chronic and complex diseases (DiMasi, Hansen, and Grabowski 2003). It is challenging to demonstrate sufficient efficacy and safety in an environment with multiple potentially confounding factors associated with concomitant medical conditions. This, together with increasingly complex regulatory requirements, longer-lasting trials and expected cost increases have increased the pressure to improve and streamline the clinical trial processes.

Four main challenges have been identified for conducting clinical trials: patients, information systems, site staff training and funding (Sung et al. 2003). More recently, a global report reviewing 156 published articles on the barriers to clinical trial conduct identified also restrictive interpretations of privacy laws and lack of transparency as important barriers. The lack of transparency refers to the information made publicly available on the authorisation, conduct, and results of a clinical trial. Privacy laws have been drafted, in part, due to the public's lack of trust in the healthcare system regarding the handling of sensitive personal information (Djurisic et al. 2017). Other barriers identified by Djurisic et al. were inadequate knowledge of clinical research and trial methodologies by the site staff, lack of funding, excessive monitoring, complex regulatory requirements and inadequate clinical

research infrastructures. Many initiatives have been started to overcome these barriers. For example, the burden caused by excessive monitoring has been reduced by endorsing regulations with a risk-based approach to monitoring. With this approach, the extent of monitoring activities can be adjusted according to needs.

As stated, many factors contribute to the success or failure of a clinical trial. Failure factors include the lack of capacity to demonstrate the efficacy and safety of the investigated drug (because of an inadequately conducted trial; performance failure of the drug in an adequately conducted trial does not mean that the trial has failed), lack of funding, incorrect eligibility criteria, failed patient recruitment, additional costs caused by difficulties of recruitment, the patients' concerns and excessive trial burden for the patients, drop-outs because of poor patient retention, and underpowered trials (Fogel 2018).

In a survey conducted in the UK, the main sources of inefficiency in trial conduct were judged by the clinical trial units. The top inefficiency factor before the start of recruitment of participants was obtaining approvals and completion of contracts. The top inefficiency factor after the start of recruitment was failure to meet recruitment targets, reported by 44% of the respondents. This was seen to reflect the over-optimistic or inaccurate estimates made by the sites. Other inefficiency factors were data management, including case report form design and delays in resolving data queries with the sites (26% of the respondents), and preparation and submission for publication (21% of the respondents) (Duley et al. 2018).

2.3 Patient recruitment in clinical trials

Patient recruitment is seen as one of the major challenges in conducting clinical trials (Bentley et al. 2019). It has been reported as the main factor for causing trial discontinuations (Briel et al. 2016; Kasenda et al. 2014) and it is also the main reason for premature trial termination (van den Bogert et al. 2017; Kasenda et al. 2014). Several researchers have tried to reveal the proportion of recruitment failures in trials conducted. For example, McDonald et al. examined 114 clinical trials conducted in the UK during 1994–2002 and concluded that fewer than one in three (31%) of the trials reached their planned recruitment target (McDonald et al. 2006). In more recent studies, the proportion of unsuccessfully recruiting trials has been about one half of the trials investigated (Sully, Julious, and Nicholl 2013; Walters et al. 2017).

In cancer trials (419 oncology trials sponsored by the National Cancer Institute (NCI) Cancer Therapy Evaluation Program in the US), recruitment was successful in 62% of the trials throughout phases I–III (Cheng, Dietrich, and Dilts 2010). In that study, only 29% of the phase III cancer trials succeeded in their recruitment as planned. However, in another study investigating 7776 phase II and III cancer trials in adult patients, 80% success in recruitment was reported (Stensland et al. 2014). It

is known that the success of recruitment is influenced by both the trial phase and the therapeutic area (Carlisle et al. 2015; Cheng, Dietrich, and Dilts 2010; Lamberti et al. 2012). For example, oncology trials have been noted to have the shortest enrollment timelines relative to plan, while central nervous system and metabolic/endocrine trials have the longest actual enrollment timelines (Lamberti et al. 2012).

One out of every two or three trials had to prolong their recruitment period because of poor recruitment (Bower, Wilson, and Mathers 2007; Campbell et al. 2007; Lamberti et al. 2012; McDonald et al. 2006). Nonetheless, sometimes even prolonging the recruitment period is not a sufficient measure to reach the recruitment target of the trial. Carlisle et al. investigated all phase II and III trials registered as closed in the National Library of Medicine Clinical Trial Registry until the year 2011. They reported that in 19% of the closed trials, the reason for premature termination was failed patient recruitment or recruitment completed with less than 85% expected enrolment, seriously compromising the statistical power of these studies (Carlisle et al. 2015).

Scientific breakthroughs in genomic research and biomarkers have paved the way for personalized treatments. This has changed, and will increasingly change also the landscape of clinical research in the future. The recruitment of patients into future trials is not expected to become easier because of the need to find even more targeted patient groups, having a certain genetic mutation or representing a subtype of a certain disease (Medidata 2020).

2.3.1 Most common recruitment methods

Only a limited number of different recruitment methods are being applied in clinical trials. On the average, four different recruitment methods are utilized within one trial (Lamberti et al. 2012). This was concluded in a study evaluating 151 trials in different parts of the world, mostly phase II and III trials in the cardiovascular, central nervous system, metabolic/endocrine, oncology, and respiratory therapy areas.

The selection of the recruitment methods in a clinical trial depends on multiple factors and should always be tailored to trial-specific purposes. Therefore, there is no single justification available for the superiority of certain methods or even for their popularity. Some frequently employed recruitment methods are outlined in Table 2.

A survey with 102 respondents from mid-size to large pharmaceutical, biotechnology, and medical device companies identified that 72% of the patients that had enrolled in their clinical trials were drawn from the sites' own patients. For example, such trial participants were recruited during their normal care visits, by

using in-practice recruiting tools such as posters, or by having members of the trial staff review the patients' charts and contact the possibly eligible patients to inquire about their interest to join the trial (ISR Report 2014).

The most popular recruitment methods were investigated in a study with 87 global phase I to phase III trials (excluding healthy volunteer trials). The trials were conducted mostly in physicians' practices, academic study sites or hospitals. The trials represented six therapeutic areas of which the top three were oncology (34 trials), central nervous system or neuroscience (16 trials) and cardiovascular/metabolic (10 trials). Patient brochures, patient letters, physician referrals and chart prescreening were the most commonly used recruitment methods, used in 47%, 46%, 29% and 21% of the trials, respectively. Advertisements (television, radio, print) were used in 20% of the trials, social/digital media in 18%, mailings in 16%, websites in 16%, patient communities/patient support advocacy groups in 11%, EHR in 11%, enhanced trial matching services in 3% and other methods in 7% of the trials (Lamberti et al. 2020).

Traditional recruitment methods are widely used, and include referrals from other physicians (Dew et al. 2013; Fletcher et al. 2012; Johnson, Niles, and Mori 2015), advertisements published in newspapers, on bulletin boards, on the radio or TV (Lamberti et al. 2012), and flyers posted or distributed in high-traffic areas within the health care system (Johnson, Niles, and Mori 2015; Ping et al. 2008). The traditional methods have become less effective for the purposes of recruitment of trial participants (Treweek et al. 2018). The use of traditional recruitment methods is expected to decrease in the future, while the importance of digitalized, non-traditional recruitment methods is increasing (ISR Report 2014; Treweek et al. 2018). There are various digital recruitment methods which can be applied; examples are database tools (searching patients from the databases with commercial or in-house devised tools or from disease registers), social media (Facebook, Twitter, YouTube) or trial-specific web pages that guide potential trial patients to contact the recruitment sites (Blatch-Jones et al. 2020).

The use of EHR in recruitment has been reported to be an efficient and quick method (Johnson, Niles, and Mori 2015). This will be discussed in more detail in section 2.5.4. Additionally, solutions based on artificial intelligence (AI) are being increasingly used to assist investigators and to make patients aware when appropriate trials are available. Natural language processing (NLP) and machine learning to profile patients based on their EHR data can be used for targeting trial information to those patients who are most likely to enroll in and complete a given trial (Fogel 2018).

Patient portals have also been created; these are databases into which prospective volunteer patients can register to be notified about possibly suitable clinical trials. Examples of such portals are Trialx (www.trialx.com, USA) and ResearchMatch

(www.researchmatch.org, USA). However, these apply only for residents of the US. In the Nordic countries, patients can review ongoing cancer trials in the Nordic region in www.nordicnect.org, or they can find out about ongoing trials at www.nta.nordforsk.org, which helps to identify the locations of trials listed at www.clinicaltrials.gov, or alternatively they can volunteer to be contacted if there are suitable trials recognized for them by www.centerwatch.com/clinical-trials. Many clinical trial units and organizations have also developed patient portals for their ongoing recruiting trials where potential participants can establish a contact through the sites' web pages (Dwyer-White et al. 2011).

Effect and efficiency of different recruitment methods

The efficiency of patient identification and/or enrollment or the time and effort used for recruitment have been evaluated in multiple ways (Lai and Afseth 2019). In spite of all this research, there is still limited information on the impact of different recruitment methods on the success of the recruitment (Adamson, Hewitt, and Torgerson 2015; Blatch-Jones et al. 2020; McAnulty 2009). Some gains from the use of non-traditional recruitment methods (for example, social media, digital recruitment, databases, and registers) in terms of enrollment rates have been noted (Lamberti et al. 2020; McAnulty 2009). Furthermore, the use of EHR (see chapter 2.5.4.) and registers have been reported to increase the effectiveness and efficiency of recruitment (Tan, Thomas, and MacEachern 2015). For large trials, aiding recruitment with specific types of patient health information obtained from wearables has been recognized to be valuable, for example through smartphone applications (Perez et al. 2019). On the other hand, solutions leveraging patient health data are becoming an important tool for recruitment in trials on rare diseases and small target populations (Bremond-Gignac, Lewandowski, and Copin 2015; Thacker, Wegele, and Pirio Richardson 2016).

Some studies have, however, failed to show any improved efficiency when using new digital recruitment methods (Ilori et al. 2020; Weng, Bigger, et al. 2010). For example, Facebook advertising has been reported not to dramatically increase recruitment when keeping the costs the same as with traditional methods (Juraschek et al. 2018). Treweek et al. also obtained mixed results in terms of recruitment efficiency in trials with non-traditional (i.e. digital) methods (Treweek et al. 2013).

As a conclusion, based on a literature review of the research conducted on different recruitment methods for trial participants, the challenges of patient recruitment have been widely recognized, and extensive research has been performed to overcome these challenges (Frampton et al. 2020).

Table 2. Studies on various recruitment methods and their effects on the recruitment of clinical trial participants.

	Trial investigated	**Recruitment methods**	**Effect**
Treweek et al. 2018	68 Randomized clinical trials (RCT) from 12 countries	-Telephone reminders to people who do not respond to a postal invitation. -Open design trials (where participants know which treatment they are receiving in the trial)	Improved recruitment
Preston et al. 2016	Systematic review on 11 RCTs investigating recruitment strategies	-Methods with a dedicated resource (e.g. a clinical recruiter or automated alert system) for identifying suitable participants	Improved recruitment
Johnson, Niles, and Mori 2015	RCT (type-2 diabetes)	-Targeted mailing for patients identified in the electronic health records (EHR) -Flyers and clinician referrals	Improved recruitment Not sufficient recruitment
Treweek et al. 2013	45 RCTs	-Telephone reminders to non-respondents -Use of opt-out rather than opt-in procedures for contacting potential participants -Open design trials	Improved recruitment
Caldwell et al. 2010	37 randomized or quasi-randomized trials	-Strategies that increased people's awareness of the health problem: an interactive computer program, an education session, health questionnaire, or a video -Monetary incentives -Increasing patients' understanding of the trial process -Recruiter differences -Various methods of randomisation and consent design	Improved recruitment No difference in recruitment
Weng et al. 2010	Large diabetes RCT	-Diabetes Registry and Clinical Data Warehouse to recruit participants for a diabetes clinical trial	Clinical Data Warehouse more efficient than Diab Registry
Avenell et al. 2004	Randomized controlled comparison nested within a placebo-controlled trial	-Open design trials	Improved recruitment

2.3.2 Consequences of poor patient recruitment

Scientific failure, lack of statistical power and ethical implications

Failures in patient recruitment have multiple consequences for clinical trials. The scientific relevance of the trial is compromised if patient enrollment cannot support the intended testing of a scientific hypothesis and the power calculations performed. Trials that cannot provide the expected scientific benefit are wasteful; they deplete the resources of clinical research and expose the participants to the trial procedures in vain (Kitterman et al. 2011). Clinically important effects may not be recognized if they are reported as statistically non-significant due to lack of statistical power. Thus, failures in patient enrollment may lead to situations where a potentially effective intervention is abandoned before its true value is recognized (Fletcher et al. 2012). Poor recruitment (Kasenda et al. 2014) and over-estimated recruitment targets (Briel et al. 2016; Carlisle et al. 2015) were reported to be the main reason for premature trial discontinuation.

Failed recruitment also raises ethical concerns: patients invest their time to participate in a trial that cannot be completed due to insufficient recruitment (Watson and Torgerson 2006) and investigators expose their patients to interventions, and still the trial fails to achieve its goals and uncertainty remains as to whether the intervention was clinically effective and sufficiently safe (Treweek et al. 2013). Poor patient recruitment may also increase the risk for fraud and other types of misconduct in clinical trials (Herson 2016).

Trial delays

Poor recruitment often results in delays in trial conduct. Any hold-ups in the drug development process will inevitably also delay patients' access to new treatments. Delayed trials are usually more costly, with fewer trials being carried out with the limited funds available (Watson and Torgerson 2006). Countermeasures to combat the delays, such as prolongation of the recruitment period, inclusion of additional trial sites or implementation of additional recruitment methods will also increase the overall cost of the trial (Bower, Wilson, and Mathers 2007).

2.3.3 Factors influencing patient recruitment

Because successful patient recruitment is such a critical success factor for all clinical trials, the reasons for recruitment success and especially those for failure have been extensively investigated. An overview of the findings published in the relevant literature is presented in Table 3.

Table 3. Success and failure factors of patient recruitment identified in previous literature.

Success factors	Failure factors
-Investigators' personal interest in research (van Staa et al. 2014)	-Investigators' lack of strong personal interest (van Staa et al. 2014)
-Therapeutic trials (compared to non-therapeutic trials) -Trials including pre-trial accrual assessment (Schroen et al. 2010)	-Lack of patients -Internal site issues - More patients refusing to participate than expected (McDonald et al. 2006)
-A clearly defined "system" of recruitment -Engagement of other staff in the hospital -Short time from ethics approval to first recruit -Provision of a dedicated trial coordinator (Levett et al. 2014)	-Previous poor site performance -Slow approval process -Strong treatment preferences -Unrealistic recruitment target -Protocol implementation at site -Lack of site engagement and experience -Busy staff (Bruhn et al. 2019)
-Increasing the number of recruiting centres -Good centre support -Using processes embedded in clinical practice -Good communication between the stakeholders (Dickson et al. 2013)	-Fewer eligible participants than anticipated -Patients' preference to receive active treatment rather than allocation to the control group -Lack of support staff -High staff turnover (Dickson et al. 2013)
-Scientific purpose of the trial -Simple study protocol -Correct ethical practices -Quality of communication with the study organization -Degree of participation in investigators' meetings (Hjorth et al. 1996)	-Negative publicity by media -Lack of patient education about clinical trials -Complex study designs (R. Kadam et al. 2016)
-Patients willing to participate -Revision of the eligibility criteria -Routine procedures of research activities (regular check of patient logs and pre-screening) -Sufficient staff resources (Team, Bugeja, and Weller 2018)	-Strict exclusion criteria -Burdensome study for the patients -Barriers in technical issues (need for additional training) -Patients' unwillingness to participate -Difficulty in combining roles of researcher and clinician -Resource shortages at site (Team, Bugeja, and Weller 2018)

Two thirds of all recruitment problems were identified promptly after the start of the trial: once the trial activities at the sites were initiated, early recruitment (within the first quarter of the scheduled recruiting time) was reported to be slower than anticipated in 63% of the investigated trials, usually because of fewer eligible patients than expected, internal problems at the site such as staff inavailability and fewer patients agreeing to participate than was originally expected (McDonald et al. 2006). Some signals of potential recruitment failure have been identified: previous poor site performance, slow approval process, strong staff/patient treatment

preferences, unrealistic site recruitment target, slow trial protocol implementation at the site, lack of staff engagement and/or research experience, and busy site personnel (Bruhn et al. 2019). Some typical factors affecting success in patient recruitment are described below.

Long development time from site selection to the start of recruitment

A long development time from site selection (more precisely from the letter of intent agreement) to the start of actual patient recruitment predicts poor recruitment success. For example, a US study with 419 phase I-III cancer trials initiated during 2000–2004 concluded that trials with a development time less than 12 months recruited more often as planned than trials with a development time longer than 12 months (Cheng, Dietrich, and Dilts 2010). It should be noted that the average time spent between the pre-study visits to the start of patient recruitment are longer in Western Europe (13.3 months) than for example in North America (7.4 months) (Lamberti et al. 2013).

Narrow eligibility criteria and complex trial protocols

Randomized clinical trials with very narrow eligibility criteria more often fail in their recruitment compared to trials with less restrictive eligibility criteria (Briel et al. 2019). This feature may have an increasing influence on the success of recruitment in the future, when patients with certain genetic alterations or biomarkers are sought for participation.

Clinical trial protocols have tended to become more complex (Getz 2014), which may also increase the burden for both patients and investigators. Higher burden for patients and trial staff have in turn been associated with poor recruitment (Briel et al. 2019). For example, burdensome (painful, uncomfortable, lengthy, or discomforting) trial procedures were noted to hinder the recruitment of patients (Roberts, Waddy, and Kaufmann 2012). Complex trial protocols may impose a burden for the investigators, requiring more time and resources (Briel et al. 2019). Complex study protocols do not encourage investigators to perform efficient recruitment. Therefore, it should be carefully considered which tests, procedures and clinical outcomes are mandatory in order to meet the trial's objectives.

Difficulties in finding patients

Difficulty in finding eligible patients is one of the major factors hindering successful recruitment (Dickson et al. 2013; Johnson, Niles, and Mori 2015). Patients are not found at the time of recruitment, or they are revealed to be ineligible at screening. In

a study performed in palliative care, the researchers interviewed the trial staff and noted that there were five major barriers for recruitment. One of those barriers was a difficulty of locating eligible patients, because the data needed were not recorded in the site's database or were only recorded on paper. The other four factors were the severity of the illness, protectiveness of family members and care providers, seeking patients from multiple settings, and lack of resources for recruitment tasks (Hanson et al. 2014).

An interview of trial staff identified diagnostic and care pathways as recruitment barriers in a trial of a drug to treat Alzheimer's disease. Up-to-date patient records were missing and data access problems hampered the screening efforts (Clement et al. 2019).

Patients' preferences and attitudes

The patients' state of health, their personal preferences concerning different treatment options, their financial concerns e.g. about the insurance coverage and their safety concerns may hinder their willingness to participate in trials (Hanson et al. 2014). Mistrust in clinical research, and doubts about data privacy and security issues have also been reported as barriers to patient recruitment (Kalkman et al. 2019). Education programs and generally increasing the public's knowledge on the conduct of clinical trials have been suggested to raise awareness, reduce fears, and dispel myths about trial participation (Jones et al. 2007).

Investigators' resources, commitment, experience and confidence

There are reports on logistical problems faced by investigators such as lack of adequate time and resources to devote to the research (Spaar et al. 2009), as well as their possible lack of equipoise and/or interest in the scientific question at hand (Roberts, Waddy, and Kaufmann 2012). Lack of resources for conducting clinical trials has also been noted in the Nordic countries: the trial staff surveyed in Sweden found it difficult to allocate sufficient time for the conduct of clinical trials, which in turn hindered the recruitment of patients (Isaksson et al. 2019).

Lack of clinical trial experience, insufficient knowledge of clinical research, the complexities of patient-clinician relationships (the possible pressure experienced by the patient to participate in the trial if suggested by his treating physician), clinicians' perceptions on the study drug and the trial's effect on normal clinical work (interruptions and lack of credit) have also been reported as impediments to successful patient recruitment. In addition, many investigators find it difficult to combine the roles of a recruiter and a research clinician (Fletcher et al. 2012).

In a trial conducted with patients with a mental illness, five main challenges concerning patient recruitment were identified when interviewing the trial investigators: misconceptions about trials, lack of equipoise, misunderstanding of the trial arms, variable interpretations of the eligibility criteria and paternalism (Howard et al. 2009).

Organizational and country-level recruitment barriers

Organizational and country-level barriers for successful patient recruitment have also been identified. Organizational or national norms, structures and processes were in some cases seen to seriously hinder patient recruitment (Adams, Caffrey, and McKevitt 2015). Examples of such factors are competition for the same type of patients between different trials in the same geographical region or insufficient communication between the clinicians and investigators of the region, reducing the number of referrals.

2.4 Feasibility evaluations in clinical trials

The feasibility evaluation of a clinical trial is a process of assessing the practicability of a clinical trial in a defined setting before the execution of the trial. On the global and regional levels, trial feasibility evaluations are often organized by the global affiliates of the sponsoring pharmaceutical company or they may be performed by their local subsidiaries or CROs (Dombernowsky et al. 2017).

A feasibility evaluation is an important part of the trial, carried out in order to understand and predict the trial execution scenarios and availability of eligible patients in different countries, to estimate the overall timeline and costs of the trial, to identify and select potential investigators and trial sites, and to evaluate their potential for successful patient recruitment. A typical feasibility evaluation process is presented in Figure 2. Feasibility evaluations should be separated from feasibility studies, i.e. short and small-scale clinical trials sometimes performed by trial sponsors before the launch of the main trial. Feasibility studies are performed to estimate important trial parameters and characteristics that will be needed to design and streamline an upcoming clinical trial (Arain et al. 2010). This thesis only concentrates on feasibility evaluations, and thus feasibility studies are not discussed further here.

The aim of a feasibility evaluation is to create the best possible starting point for the set-up and execution of a clinical trial and to identify all avoidable obstacles for successful conduct of the upcoming trial. A systematic review of trials that were discontinued due to poor recruitment identified that 89% of the reasons for poor recruitment could have been anticipated in the planning phase of the randomized

clinical trial (Briel et al. 2016). A well performed feasibility evaluation improves a trial's opportunities to proceed as planned. Therefore, an upstream approach in patient recruitment, i.e. being better prepared for patient recruitment already before the trial, has been increasingly recognized (Dombeck et al. 2020; Kadam et al. 2016) and recommended (Huang et al. 2018; White and Hind 2015). The report from Huang and colleagues described a framework for such planning, presenting three aspects essential to strategic recruitment planning efforts: (1) trial design and protocol development, (2) trial feasibility evaluation and site selection, and (3) efficient communication between the stakeholders during the recruitment. Their survey findings suggested focusing on preparing a comprehensive recruitment plan rather than concentrating on specific recruitment activities and tools. The conduct of any feasibility evaluations should be streamlined in order for the sites to better plan their work and to allocate their resources adequately.

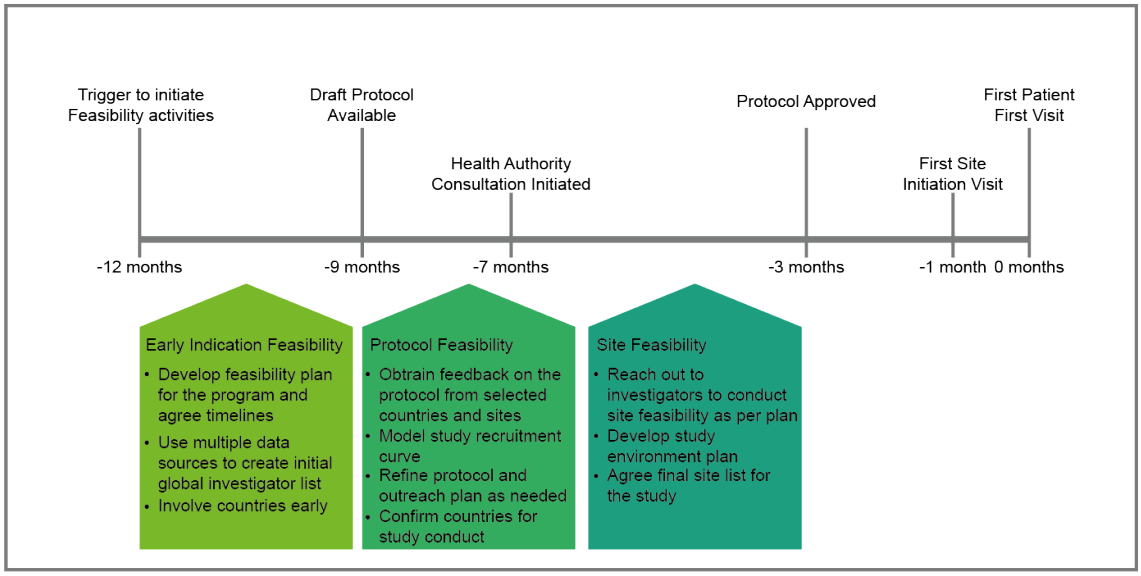


Figure 2. Typical process of a trial feasibility evaluation. Modified from Pharmaceutical Outsourcing 2017. https://www.pharmoutsourcing.com/Featured-Articles/333830-Clinical-Trials-A-Data-Driven-Feasibility-Approach/, visited on 20Feb2021.

2.4.1 Site identification

When allocating trials to different countries, the headquarters and regional departments of pharmaceutical companies are the key decision makers for country selection, whereas local subsidiaries mainly decide on the selection of the study sites (Dombernowsky et al. 2017). Upon deciding on the trial allocation, the headquarters primarily value timely patient recruitment and the quality of trial data to be produced.

In the site identification process, also the recruitment-related qualities are prominent: study populations' availability, timely patient recruitment, resources at the sites as well as the site personnel's interest and commitment (Dombernowsky et al. 2017).

In the site identification, trial sponsors favor those sites with whom they have already had previous collaboration; in fact less than one third of the sites will be new to the sponsor (Lamberti et al. 2018). Trial sponsors are already aware of the site's way of working and there will be data available on its previous performance. Old sites are also ready to start the trial approximately 2.5 months earlier than sites that are new to the sponsor (Harper et al. 2017). In a survey with 591 respondents from biopharmaceutical and CRO companies, Harper and colleagues reported that the identification of an "old" site takes 3.5 weeks, whereas an average of 6.5 weeks is needed to identify a "new" site. In addition, the site selection process is shorter with the old sites compared to new sites (5.2 and 7.9 weeks, respectively) (Harper et al. 2017), as indicated in Figure 3.

Other methods for identifying trial sites are recommendations from other investigators or study teams, and the use of personal networks and proprietary databases (Lamberti et al. 2018). Several technical tools have been increasingly applied for this purpose, such as commercial investigator databases (Clinical Research News Online 2019). However, there is no single data source which can be accessed to identify suitable trial sites. Site identification still very heavily relies on low-tech methods, i.e. there are no truly evidence-based approaches (Harper et al. 2017).

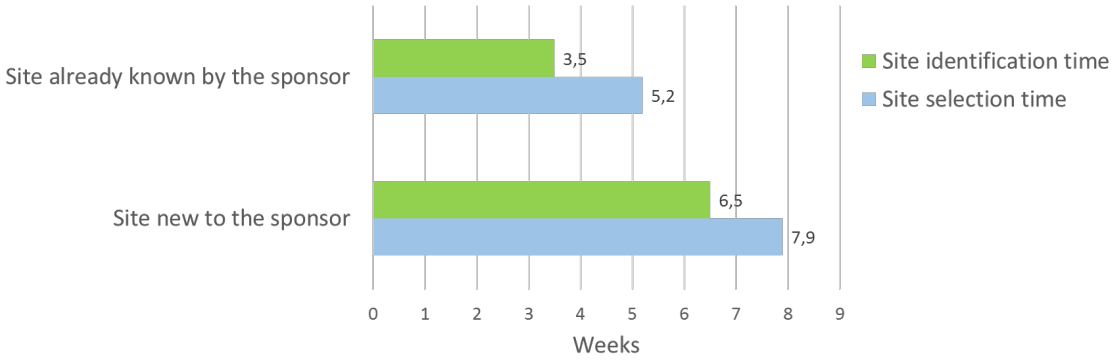


Figure 3. Average site identification, site selection and trial start-up times (in weeks) with new sites and sites with previous collaboration with the sponsor. Figure based on information from Harper et al. 2017.

2.4.2 Evaluation of access to patients

Evaluation of the sites' access to eligible patients is an important part of the trial planning (Dombeck et al. 2020). Site feasibility questionnaires are commonly sent to potential investigators to inquire their interest and capabilities to conduct the trial. Their capabilities to recruit eligible patients are also evaluated: how many patients could they potentially recruit and what would be their recruitment methods.

Trial sponsors often visit the potential investigators or contact them by phone to discuss the patient count estimates and other trial-related matters. According to previous literature, the estimates on potential recruitable patients are often over-optimistic (White and Hind 2015), as they are not based on factual data (Roberts, Waddy, and Kaufmann 2012). Investigators most often form their patient count estimates based on their own or their colleagues' experience from similar trials conducted earlier (White and Hind 2015).

How recruitment targets are decided varies between trials and sites. Setting up the recruitment target for a site is not just a number but has much broader implications: the set target will subsequently define whether or not the site succeeded in its recruitment, and that number has significance in that it will affect the motivation to recruit, especially if it is set unrealistically high at the outset (Bruhn et al. 2019). Bruhn and colleagues have highlighted that the setting up of recruitment targets does not receive much attention in the early stages (before the trial), they are often set artificially and for some trials appear to have been set too low across sites (Bruhn et al. 2019).

Most investigators are aware of the consequenses of poor recruitment for their future possibilities to conduct trials (White and Hind 2015). An interview study conducted with US investigators who had successfully conducted clinical trials highlighted the value of good previous performance in clinical trials as being a plus for their reputation (Dombeck et al. 2020). In order to maintain a good reputation, the preparations for an upcoming trial were perceived as one of the most important factors for success, by trying to avoid the greatest obstacles for patient recruitment. Investigators' preparations included thinking through their ability to recruit the promised number of participants prior to joining a trial, and also choosing or declining trials accordingly, in order to ensure that they maintain their reputation for conducting high-quality research and thus continue to be offered trial opportunities.

It has also been noted that because of the increased use of site-performance data and data-driven decisions based on that performance, some sites have become very cautious and predict their patient targets far too low even if they would be able to recruit many more patients (Bruhn et al. 2019).

There are some examples of successful trial sites who invest considerable time on assessing their recruitment capabilities. For example, the Clinical Trial Unit at the University of Louisville, Louisville, Kentucky, USA, has a defined process for

responding to incoming feasibility questionnaires: before sending their responses to the sponsor, the site submits its own supplementary questions to the trial sponsor in order to be able to respond to the questionnaire as reliably as possible. They also routinely use their patient database for patient count estimations (Lale Akca, oral communication at the SCOPE Europe meeting, 2020).

Patient data may not always be readily available for the investigators for estimating the potential number of trial participants. Because of data privacy regulations and concerns (Fernández-Alemán et al. 2013), especially the data contained in a hospital's EHR systems may not be accessible as freely as would be needed for rapid and accurate feasibility evaluation purposes. Some query tools have been created to overcome this barrier. For example, the EHR4CR (Electronic Health Records for Clinical Research) query tool was developed as an European Union Innovative Medicines Initiative (IMI) project for querying patient data with certain criteria and for obtaining patient counts as a result (Doods et al. 2014; De Moor et al. 2015) . No identifiable patient data needs to be shared, only the number of potentially eligible patients in the hospital is queried. Hospital site personnel in Germany, UK, Switzerland and France were interviewed on how they experienced the use of the EHR4CR query tool and the benefits it provided (McCowan et al. 2015). According to McCowan and colleagues, the respondents (n=37) experienced the tool as welcome and beneficial when estimating the potential number of recruitable patients.

As part of the feasibility evaluations, the trial sponsors also try to evaluate the potential number of patients that each site will be able to recruit. Setting precise recruitment targets for the sites is difficult because of the many parameters that influence the sites' access to patients. The use of data from earlier trials is one method which is applied to predict the sites' potential for recruiting patients into a new trial (Clinical Research News Online 2019; Lamberti et al. 2018). Innovative approaches have also been established, for example, by visualising the site's activity through the number of laboratory samples that the site had taken in earlier clinical trials (Yang et al. 2018). The researchers made assumptions on the site's recruitment potential in the new trial based on the number of samples taken from similar patients by that site.

Predicting a trial site's future performance is often difficult. Currently, there is no reliable tool to estimate in advance which sites will be able to succeed in their recruitment (van den Bor et al. 2017). The large number of zero-recruiting sites in actual trials makes all too visible the difficulty of such predictions. No sponsor will intentionally select zero-recruiting sites for their next trial. Nonetheless, during the 2010s, the proportion of zero-recruiting clinical trial sites has increased globally from 11% to 14.3% (Lamberti et al. 2012, 2020). In Western Europe, the percentage was even more disturbing: the average share of sites enrolling no patients was 19.1%.

Alarmingly, the number of zero-recruiting sites has increased in Western Europe over the past several years, whereas their number has decreased elsewhere, for example in many countries with emerging economies (Lamberti et al. 2012, 2020).

If prediction of recruitment success is difficult for one single trial site, it is even more difficult for a whole trial. Multiple statistical models have been developed for predicting patient accrual, especially in multicenter trials, but none of them has been recognized as fully functional for this purpose (Barnard, Dent, and Cook 2010; Zhang and Long 2012). However, during recent years, the leading pharmaceutical companies have developed some promising solutions (Clinical Research News Online 2019). New prediction tools use the sites' performance data from previous trials (such as enrollment success, number of enrolled patients, and speed of enrollment) with algorithms developed with AI techniques.

2.4.3 Site selection

A Nordic survey with 83 respondents evaluating the trial site selection process concluded that 84% (confidence interval ± 8%) of the pharmaceutical companies and CROs evaluated recruitment-related factors as the most important site-related qualities when selecting study sites (Dombernowsky et al. 2019). In a survey of site selection properties, performed on the European level, the most important factors influencing site selection were the investigators' previous trial performance, experience and motivation to conduct the trial, access to potential trial participants in the area and the ease of gaining regulatory approval (Gehring et al. 2013). In contrast, in the emerging markets outside of Western Europe and North America, the most important factor for selecting sites and investigators was to determine whether good clinical practice (GCP) standards had been established at those sites (Scorr Marketing 2017).

2.5 EHR in clinical trials

2.5.1 EHR definition

An electronic health record, EHR, is "an electronical format, longitudinal health information on individual patients recorded by health care personnel and to be used in the patient care" (Häyrinen, Saranto, and Nykänen 2008). A typical individual EHR may include the patient's medical history, diagnoses, treatment plans, immunization dates, medication records, and laboratory and other test results, including those derived from imaging investigations (FDA 2018).

According to a World Health Organization (WHO) report on eHealth in Europe, the technological transformation of patient records from paper to electronic format

has increased substantially over the past two decades. Of all European countries, 59% have a national EHR system and 69% of those have legislation governing its use (World Health Organization 2016a). All Nordic countries have national EHR systems covering almost 100% of their generated patient data (Bonomi 2016; Nordforsk 2019). On a global level, the existence of national EHR systems is less common: only 47% of all countries (n=57) included in the analysis reported having a national EHR system (World Health Organization 2016b).

The data contained in an EHR are both in structured and unstructured format. Structured data types include coded data, values from pre-populated lists, or data entered into fields requiring specific alpha-numeric formats. The amount and quality and the coding used for the structured data vary significantly between different EHR systems and between different hospitals. The most common items in EHR systems that are found in structured format are information on demographics, diagnoses, procedures and laboratory findings, whereas the most underrepresented items are specific scores and classifications and medical history (Doods et al. 2014).

The WHO has encouraged countries to develop their national EHR systems also for research purposes (World Health Organization 2016a). The use of EHR data for research purposes is seen to enhance the transparency of research and to generate improved efficiency of the use of the publicly funded EHR systems. The results of successful secondary use of EHR can also confer valuable information for the data providers, i.e. for the health care organizations, thus adding value to their health information (Vikström et al. 2019).

In addition, regulatory bodies overseeing the development and use of pharmaceuticals have expressed a positive stance on the use of electronic patient data for enhancing clinical trial conduct and have provided written guidance on their expectations regarding clinical source data existing in electronic format (EMA 2010; FDA 2013, 2018; MHRA 2015; PMDA 2015). These factors have encouraged the use of EHR systems for different purposes other than those for which they were originally designed to serve.

2.5.2 Regulations and guidelines for EHR secondary use

The secondary use of EHR, their levels of adoption, data quality, and ease of use and time for obtaining the information for research purposes varies widely between different countries (Van Velthoven et al. 2016). Van Velthoven and colleagues investigated 16 countries around the world; for example, they found that obtaining approval for extracting data from an EHR system for research was moderately easy in most of the countries assessed. Exceptions were India and South Africa, where it was difficult to obtain approval, and Austria, where it was not possible to obtain approval as data lodged in an EHR system were not allowed to be used for secondary

purposes. The times required for obtaining approvals for the secondary use of EHR data were about 3 months (China), 3 to 6 months (Czech Republic, India, Indonesia, the Netherlands, Poland, South Korea, and United Arab Emirates), 6 to 12 months (Australia, Brazil, Italy, Mexico, and Saudi Arabia), and more than one year in South Africa. Typically, the process was found to be lengthy and the time needed was dependent on the number of different sites to be included in a study (Van Velthoven et al. 2016).

The use of patient data for secondary purposes is stringently regulated. Restrictive privacy laws were seen to impede the flow of patient information that could help investigators to identify patients who might be offered the opportunity to enrol into clinical trials (Djurisic et al. 2017). In 2008, Duley and colleagues had already reported the same issue: an over-restrictive interpretation of privacy laws without evidence of patient benefit was found to be one of the five major barriers for the conduct of randomized clinical trials (Duley et al. 2008).

A robust infrastructure of policies, standards, and best practices for the secondary use of EHR data (Safran et al. 2007) and requirements for clarifying the legislation regulating its secondary use have been presented (Bahr and Schlünder 2015). For example, Bahr and Schlünder suggested a code of practice for EHR secondary use in Europe. The purpose of this code of practice would be to resolve issues in a way that would balance the need to make research possible and the need to protect the patients' privacy at the same time. The European Institute for Innovation through Health Data (i-HD) has worked towards harmonizing the secondary use of EHR data in Europe (Kalra et al. 2017). One of the priorities in 2021-2025 of the European Union Commission is the creation of a European Health Data Space (https://ec.europa.eu/health/ehealth/dataspace_en), which is planned to promote better exchange and access to different types of health data, including the secondary use of EHR data.

The Nordic countries (Finland, Sweden, Norway, Denmark, and Iceland) have long traditions and trusted reputations in conducting clinical trials and in maintaining their hospitals' patient data in electronic format (Bonomi 2016). The countries are increasingly executing new regulations on the secondary use of health data. Most regulations do not explicitly define how EHR data can be used for patient recruitment and feasibility evaluations; instead, this must be interpreted by the stakeholders. In Sweden, new legislation is under preparation for the use of data particularly for feasibility evaluations. A summary of the main items of the legislation pertaining to the secondary use of EHR data in the different Nordic countries is presented in Table 4.

Table 4. Summary of the legislation concerning the secondary use of EHR data in clinical trials in the Nordic countries.

	Legislation involved in using EHR for clinical trial patient recruitment and feasibility evaluations
All	EU General Data Protection Regulation 2016/679
Finland	The Act on the Secondary Use of Health and Social Data 552/2019 Biobank Act 688/2012 (new revision under preparation)
Sweden	The Act on Patient data 2008/355 (new revision under preparation to clarify the patient data use in feasibility evaluations)
Norway	Health Research Act: Act 2008-06-20 no. 44 Health Registers Act: Act 2014-06-20 no. 43 Health Records Act: Act 2014-06-20 no. 44.
Denmark	Danish Health Act (Sundhedsloven) Executive Order for the Health Act, 1067 (given 29/06/2020)

The similarities between the Nordic countries in terms of health care infrastructures offer advantages also for the similar secondary use of EHR data in clinical trials. The relatively high quality of the patient data is made possible in part by the comprehensive health care systems, unique social security numbering systems and the high level of technical expertise in these countries (Nordforsk 2019). It may well be possible to combine patient registries located in different Nordic countries, because of their similar infrastructure, health care systems and data quality (Maret-Ouda et al. 2017). In spite of the many similarities, there are also differences between the Nordic countries in collecting, processing and sharing EHR data and its use for secondary purposes. Unfortunately, there is very limited research on this topic (Vikström et al. 2019).

2.5.3 Various ways to use EHR data in clinical trials

In spite of numerous reports on promising possibilities offered by the use of EHR data in clinical trials, the secondary use of EHR data is still not commonly applied, at least not as much as anticipated (Cowie et al. 2017; Lai and Afseth 2019). In 2008, 14 case categories were presented on how EHR data were or could be used in clinical trials, the three areas most highlighted being drug safety and surveillance, clinical trial recruitment and support in regulatory approval. Many of the EHR software systems available already in the beginning of the 2000s had possibilities to capture the data necessary for clinical trial recruitment and drug safety surveillance (Kim, Labkoff, and Holliday 2008). The EHR system capabilities for assisting trial recruitment were also presented, for example in a study testing five different EHR systems in German hospitals (Schreiweis et al. 2014).

There are many benefits to the use of EHR data in clinical trials including the streamlining of clinical research processes in hospitals, improvement of data quality by reducing the number of transcription errors, evaluating the feasibility of trial protocols, estimating the availability of potential trial participants and enhancing drug safety and early identification of safety events (Nordo et al. 2017). Real world evidence reference groups can also be formed from the patient data in the EHR system, for use in intervention trials where no placebo treatment is applicable (Franzén et al. 2016). In addition, instead of manually transcribing trial-related data from the EHR system into electronic case report forms (eCRF), the process can be expedited by capturing the EHR data directly into the eCRF (Nordo et al. 2017). It has been claimed that about 13–35% of the data requested by the eCRFs were available by default in the EHR system (El Fadly et al. 2011; Köpcke et al. 2013). There are reports on such direct data capture from EHRs to eCRFs published after that, but no information is available on whether the coverage has increased from that date. According to some service providers (such as www.protocolfirst.com, Salt Lake City, UT, USA), approximately 30–60% of the data can be directly transferred from EHRs to eCRFs depending on the trial, but there was no further evidence available to confirm that estimate (Nordo et al. 2019).

In current clinical trial settings, EHRs are most commonly employed for the recruitment of trial participants and for long-term outcome evaluations (Mc Cord et al. 2019).

2.5.4 Use of EHR data in patient recruitment

As the quality, structure and contents of EHR systems (section 2.5.1) and the legislation for their secondary use (section 2.5.2) vary between countries and continents, also the use of EHRs for patient recruitment varies significantly. The Tufts Center for the Study of Drug Development investigated 151 trials performed globally and reported that only 8% of the trials applied an EHR review for recruiting patients (Lamberti et al. 2012), while seven years later, 11% of the trials used EHRs in recruitment (Lamberti et al. 2020). Correspondingly, the patient charts were reviewed in 21% of the trials.

When searching through the EHR system, different recruiters may understand the eligibility criteria in different ways and identify different individuals as being eligible than their colleagues (Howard et al. 2009). Hilton et al. investigated this with a vignette method. They created 20 imaginary patients with health data and asked the recruiters to identify the eligible patients among them. By looking at the same data, the recruiters were unanimous on only seven patients. The authors concluded that how the recruiters manually identify patients in the hospital's EHR system is a very subjective process (Hilton et al. 2016). However, there are also studies claiming

that the recruitment process is not heavily dependent on the recruiters (Caldwell et al. 2010).

Lai and Afseth described the identification of trial participants with automated EHR screening. Some trials used alert systems, mailings or calls. Alerts were sent to physicians or to study staff, investigators or study nurses. This kind of use of alerts was seen as advantageous. When EHR systems were only used for filtering potential trial participants, and no alert system was used, lower recruitment yields were encountered (Lai and Afseth 2019). The recruitment was also found to be lower if the patients were contacted by letter instead of a face-to-face visit after identifying them from the hospital's EHR system (Edwards et al. 2019).

Patient preferences have also been investigated. Patients were generally positive for being contacted through an EHR search by the investigator who then inquired about their interest in participating in a clinical trial (Beskow, Brelsford, and Hammack 2019). With a traditional method, they would be contacted by their own treating physician at the request of the investigator.

Benefits of EHR use

The use of EHR data has been reported to enhance recruitment in clinical trials. The main benefits have been reported to be accuracy of finding eligible trial participants, speed of the identification process, time saved in work spent for recruitment, holistic coverage of the potential trial participants in the hospital, and possibly lower recruitment costs per patient.

Accuracy and efficiency

The use of EHR systems has been reported to be an accurate and efficient way to improve the identification of trial participants (Denburg et al. 2019; Penberthy et al. 2010; Rollman et al. 2008). This has been noted especially when implementing EHR searches with an alert systems, i.e. with automated notification to the investigator when a patient with the suitable eligibility criteria is identified by the EHR system (Cardozo et al. 2010; Herasevich et al. 2011; Rollman et al. 2008).

In one investigation, traditional recruitment approaches (clinician referrals and posting of flyers) were compared to EHR recruitment (targeted mailings to patients with uncontrolled diabetes identified from the EHR system) (Johnson, Niles, and Mori 2015). In that study, 77% of the participants were recruited with the EHR method and it appeared to produce a more representative and appropriate sample than other recruitment methods under study.

The applicability of EHR systems for identifying patients into trials varies by the disease and the intervention: some diseases are recorded in EHR systems more accurately than others, and this influences the accuracy of identifying the patients

(Dugas et al. 2010). For example, the accuracy of identification of type 2 diabetes by ICD9 coding (International Classification of Diseases, Ninth Revision) was 96.6% in the EHR system of the University of New Mexico Health Sciences Center, (Albuquerque, New Mexico, USA), whereas for Parkinson's disease it was only 55% (Thacker, Wegele, and Pirio Richardson 2016). It should be noted that also negative views on the efficiency of EHR in patient recruitment have been presented (Effoe et al. 2016; Treweek et al. 2013).

Speed and time saving

Faster processing times, fewer man-hours and fewer working days because of the use of EHR in recruitment have been cited (Penberthy et al. 2010; Rollman et al. 2008; Weng, Tu, et al. 2010). The value of EHR for reducing the burden of the trial staff has also been highlighted, compared to the time needed for a manual review of the records of possibly eligible patients (Jonnalagadda et al. 2017; Thadani et al. 2009).

In some recent research, artificial intelligence (AI) has been adopted as a tool for patient identification in EHR systems, with promising results (Beck et al. 2020; Calaprice-Whitty et al. 2020). AI techniques enable the review and use of the content present both in structured and in unstructured format in the EHR system, after training the computer with algorithms (Shivade et al. 2014). For example, Calaprice-Whitty et al. explored AI (the mendel.ai solution) in trying to identify patients for two oncology trials that had earlier failed in their recruitment. Their study resulted in a faster and more comprehensive process for identifying eligible patients. With AI, they identified 24–50% more patients than with a manual search, and furthermore, the system was also able to identify all those patients that had been found manually but with much less time spent (Calaprice-Whitty et al. 2020). They also highlighted an important patient perspective on the speed in identifying potential patients: had the AI approach been applied in the first place in those cancer trials at the time of their failed recruitment, not only would more patients have been screened for the opportunity to participate but they could also have been offered this opportunity sooner, which could have altered the course of their disease.

Diversity of participants

The enormous amount of data being collected in the patients' medical records has been found to confer additional value when integrated and stored in data warehouses allowing all data from an organization with numerous inpatient and outpatient facilities to be integrated and analyzed. (R. Scott Evans, Lloyd, and Pierce 2012). It has been concluded that automation of the comprehensive EHR system offers an opportunity to assure that all patients have an opportunity to be evaluated for participation in clinical trials (Leather et al. 2020; Penberthy et al. 2010).

Minority participation in clinical trials is considered crucial for trial validity (Beresniak et al. 2015). In a study evaluating trials in patients with an anxiety disorder, an EHR-based recruitment method detected non-white trial participants five times more often than waiting-room recruitment (Rollman et al. 2008). Success in the recruitment of minority participants was also noted for example in an intervention trial aimed at promoting physical activity in patients with type II diabetes (Johnson, Niles, and Mori 2015).

Recruitment costs

The total cost per patient was found to range from 44 to 2000 US$ with patients identified through an EHR system (Mc Cord et al. 2019), whereas the range was 41–6990 US$ with traditional recruitment methods (Speich et al. 2018). This has subsequently been confirmed; EHR-based methods (60 US$) were more cost-effective per patient than non-EHR based methods (107 US$) in a randomized clinical trial of 294 adult patients with gout (Miller et al. 2020).

Schroy et al. demonstrated the cost-effectiveness of EHR-based recruitment: the average per-patient cost when using an electronic opt-in referral method with manual screening from the EHR system was US$ 129, the cost of a referral letter method with manual screening from the EHR was US$ 1967, and the cost of an investigator call with manual screening from the EHR was US$ 156. When combined with automated patient identification from the EHR system, the call method was even less expensive, at US$ 99 per patient (Schroy et al. 2009).

Challenges in using EHR data

Despite multiple benefits, many challenges in using EHR data for patient recruitment have been identified (Weiskopf et al. 2013). The most commonly reported challenges include the lack of EHR standardization, the quality of EHR data, ethical, privacy and data-security considerations, considerable infrastructure costs and challenges in interoperability, i.e. possibilities to use EHR data reciprocally between different hospitals

Lack of rules and regulations for EHR secondary use

The lack of standards and regulations for secondary use of EHR data have been claimed to hinder the effective use of EHR systems for trial recruitment (Weng et al. 2012), and the need to harmonize all regulations to allow the secondary use of EHR data has been identified many times (Harrer et al. 2019; Safran et al. 2007; Weng et al. 2012). Even within the Nordic countries, the legislation is not compatible; some of the countries prohibit the secondary use of EHR data for identifying trial participants when investigators aim to cover the whole hospital or the whole region.

There may also be features in the local interpretation of the regulations that hinder the process (Duley et al. 2008).

Currently, investigators cannot directly approach the patients identified by the algorithm in the hospital's EHR system unless the patients have previously given consent to being contacted regarding trial opportunities. Patients must also be offered the opportunity to opt out of such use of their data (Raman et al. 2018). EHR alerts can be used to partly circumvent this ban; an alert first prompts the treating physician to ask for such consent from the potentially eligible patient, and subsequently the patient can be contacted. However, alert fatigue may result if this method is used excessively (Lai and Afseth 2019).

Ethical aspects

Despite the varying laws and regulations of different countries, the ethical views and considerations are generally similar across different cultures and healthcare systems (Häyrinen, Saranto, and Nykänen 2008). For example, some unresolved ethical considerations include patients' consent for the secondary use of their EHR data (Coorevits et al. 2013), assurance of the secure and controlled access to hospital data (Coorevits et al. 2013; Lehnbom, Mclachlan, and Brien 2013) and how and who should actually contact possibly eligible patients after identifying them from the EHR system (Callard et al. 2014).

Even if patients' awareness of clinical research and their acceptance of using their EHR data for contacting them for trial opportunities have increased, many individuals still have concerns regarding the secondary use of their health data (Kalkman et al. 2019). This is an important aspect to keep in mind. Public acceptance of the secondary use of health data needs to be obtained, and all aspects of patient privacy and human subject protection policies should be clarified, at the local, national and global levels, in order to optimally exploit the full potential of EHR systems in clinical research (Kalkman et al. 2019; Weng et al. 2012).

Technical challenges and data availability and quality

Many reports have cited the technical difficulties, for example in transcribing patient eligibility criteria to search items, in order to perform a query within an EHR system. Various standards and guidelines have been published to harmonize the formal presentation of the eligibility criteria (Chondrogiannis et al. 2017; Doods et al. 2013; Weng, Bigger, et al. 2010; Y. Zhang, Zhang, and Shang 2017).

The kinds of methods that are only able to search items in the structured data of the EHR system present a common technical challenge: the search may not be comprehensive enough. According to a study conducted in five German university hospital EHR systems with 15 different trial protocols, only 35% of the patient characteristics were available in the structured data elements (Köpcke et al. 2013). When testing the Austrian national EHR system, 61% of the criteria could be mapped

to structured data (Augustinov and Duftschmid 2019). The most prevalent elements in the structured data were the diagnosis codes (37% of all eligibility criteria out of 1120), procedure codes (10%), and medication codes (8%) (Gulden et al. 2019).

Some reports have demonstrated the possibilities for identifying suitable patients only by processing structured data (Miotto and Weng 2015), but many studies have reported insufficient patient identification because the automated search has excluded all information that is present in the free-text fields of the EHR (Lai and Afseth 2019; Majeed, Car, and Sheikh 2008). With the aid of new methods, such as AI and natural language processing (NLP), this form of unstructured data could be included in the searches (Jonnalagadda et al. 2017; Meystre et al. 2019). It has been predicted that machine learning, followed by NLP will gain popularity in identifying patients in EHR data (Shivade et al. 2014).

The quality of the data contained in EHR systems varies between hospitals, health care systems and countries (R. S. Evans 2016). Missing data or inclusion of incorrect data impair the trustworthiness of the EHR search results. Thus, because of variations in data quality, the exact numbers of potentially eligible patients may be difficult to compare reliably between different hospitals. Interoperability (i.e. how compatible the data are between hospitals) is an important factor for the efficient use of EHR systems for clinical trials. Without interoperability, the data of one hospital have only limited value for use in a clinical trial (Weiner and Embi 2009).

Secondary use of data

Clinical patient data are recorded for the needs of the patient's medical care, and the data needs for secondary purposes such as clinical trials are not necessarily considered (Weiner and Embi 2009). This may cause significant gaps in the structure and contents between the data documented during patient care and the data required for patient eligibility assessments for trials (Butler et al. 2018). Clinicians routinely record the data into the EHR system when treating their patients, and they may not consider any benefit for their purposes in recording additional data that would serve better some other uses of the EHR data (Morrison et al. 2014).

2.5.5 Use of EHR data in feasibility evaluations

There has been less research conducted on the use of EHR data for trial feasibility evaluations than is available for EHR use in the recruitment of trial participants (Kearney et al. 2018). Nonetheless, EHR systems are used rather similarly for both purposes: to identify potential trial participants. Therefore, also the benefits and challenges of using EHR systems in feasibility evaluations are largely the same as for using them for recruitment. One example of how EHR searches could be used in

feasibility evaluations would be in elements of trial planning, i.e. how many eligible patients each site would be able to recruit for an upcoming trial (Huang et al. 2018).

One important difference relates to time constraints: in a feasibility evaluation, also historical data may be very useful, and the records do not necessarily have to be contemporaneous. It is also easier to ensure compliance with privacy protection when the data subjects do not have to be contacted, and the data can be truly anonymized.

The use of EHR systems in feasibility evaluations was reported for example by Beck et al. who explored the performance of an AI clinical trial matching tool for eligibility screening for four different clinical breast cancer trials in Arkansas, USA. The technical capabilities showed promise in identifying eligible patients in the hospital's EHR system, with good sensitivity and with time savings compared to a manual review (Beck et al. 2020).

3 Aims

The general aim of the research presented in this thesis was to elucidate the challenges faced in the identification and recruitment of patients participating in clinical trials of pharmaceuticals in the Nordic countries, and to evaluate whether patient recruitment might be improved by the exploitation of the electronic health record (EHR) systems employed by health care providers. The research was carried out by investigating clinical drug trials conducted by the pharmaceutical industry in the Nordic countries, and by assessing the functionality of one commercially available EHR research platform linked to the clinical data warehouse of Turku University Hospital.

Detailed research questions were:

1. What is the pharmaceutical industry's view on the success of recruitment into clinical drug trials in the Nordic countries, and what is the role of EHR in the recruitment of trial patients?
2. What factors can be identified as determinants of successful or failed patient recruitment for clinical trials?
3. How does the pharmaceutical industry perform feasibility evaluations when searching for potential trial sites and evaluating the sites' access to eligible patients?
4. Do EHR data have a role in identifying and selecting trial sites and study patients - currently and in the near future?
5. How accurately can a currently employed EHR research platform identify the same patients discovered with a manual search from a hospital's database, and what are the reasons for possible discrepancies?

4 Methods

4.1 Data sources

4.1.1 Patient recruitment and feasibility evaluations

The information used to investigate patient recruitment factors (I) and trial feasibility evaluations (II) in the Nordic countries was derived from the interviews of 21 respondents representing private enterprises operating within the field of the pharmaceutical industry and clinical trial CROs in the Nordic countries.

Before the interviews, the participants were requested to select the two most important pre-market clinical drug trials with which they had been involved in 2015-2018 (i.e. started < 4 years before the interviews in 2019), and where at least one Nordic country was included. No requirements were set to select the trials based on the outcome (success or failure) of recruitment.

The planned number and the actual number of the recruited patients were collected for each Nordic country. Thereafter, all participants were asked to judge whether recruitment in the Nordic countries had succeeded or failed (for ongoing recruitments; on schedule or delayed). If the number of patients recruited in the Nordic countries reached the planned number (a 95% value of the target was allowed), the recruitment was regarded successful. It was possible to compensate for reduced numbers in one Nordic country by exceeding the initial target in another country. If the recruitment target had not been reached or had been reached only after a prolongation of the recruitment period, the recruitment was classified as failed.

The respondents were asked to identify key factors for the recruitment success or failure: the recruitment success factors of successfully recruited trials and the failure factors of trials regarded as failed in recruitment.

For the trials that failed in recruitment, the participants were also asked to evaluate the recruitment's contribution to a trial delay, compared to other possible delay factors (on a scale from 1 = 'patient recruitment had no effect on trial delays' to 4 = 'patient recruitment was the major trial-delaying factor').

In addition to discussing the success or failure of the recruitment, the respondents were to describe how the patients were identified, how the sites were identified and whether EHR systems had had a role and what kind of role in their trials.

An interview guide (Appendix 1) was created and sent to the respondents in advance of the interview. It contained five categorical questions and eight open questions/themes.

The categorical questions were as follows:

- Trial profile (clinical phase, therapeutic area, participating Nordic countries, and number of patients planned/recruited in the Nordic countries) (I)
- Trial overall status (trial completed (or ongoing) according to planned schedule, Yes or No) (I)
- Delays caused by patient recruitment compared to delays caused by other trial-delaying factors (ranging from 'patient recruitment was the major trial-delaying factor' to 'patient recruitment had no effect on trial delays') (I)
- Patient recruitment outcome (recruitment succeeded / failed, or for ongoing recruitments: recruitment on schedule / delayed) (I, II). If the recruitment target had been reached only after prolongation of the recruitment period, the recruitment was classified as failed.
- Benefit (very much / much / little / no benefit) of using EHR in patient recruitment (I) and in feasibility evaluations (II).

The open questions/discussion themes were as follows:

- Trial feasibility evaluation activities preceding the start of patient recruitment (II)
- Trial site identification methods (II)
- Use of EHR in trial site identification (II)
- Patient identification methods (I, II)
- Use of EHR in patient identification (I, II)
- Key recruitment success/failure factors (I)
- Factors causing trial delays (I)
- Respondent's expectations for the future use of EHR in the recruitment of trial participants (I) and in trial feasibility evaluations (II)

The interview guide was tested with one pilot interview, which was included in the analysis, as no need for major modifications was noted in this pilot round.

Seven trials selected by the participants did not meet the trial selection criteria (for example, the trial had not yet started) and one participant had only chosen one trial. Therefore, 34 trials were discussed in the interviews (I, II).

4.1.2 EHR Research platform assessment

A commercially available EHR research platform, Insite, (III) was tested for its capability to find recruitable patients. For this functionality testing, the patient data in the EHR system of Turku University Hospital, a tertiary care hospital serving a population of 870 000 in South-West Finland, was used. The patients to be identified were to be diagnosed with atrial fibrillation (AF) before the year 2013 and with at least one incidence of stroke, transient ischemic attack (TIA) or intracranial hemorrhage during the years 2003-2012. Their detailed inclusion and exclusion criteria are presented in Table 5. The diagnoses were classified according to ICD10 (The International Coding of Diseases, U.S. National Center for Health Statistics, https://www.cdc.gov/nchs/icd/) categorization, which has been applied in Finland since 1996.

The EHR Research platform's capability for identifying patients was compared with a manual search (see section Reference data), in which the patients had been identified manually from the hospital's EHR system by researchers reviewing electronic patients' charts, one at a time.

Table 5. Inclusion and exclusion criteria for identifying patients with a manual search of the hospital's patient records

Criteria for Initial screening
1. Disturbances in cerebral blood flow at any point in the years 2003–2012: I60.0–I60.9, I61.0–I61.9, I62.0–I62.9, I63.0–I63.9, I64.0–I64.9, I65.0–I65.9, I66.0–I66.9, I69.0–I69.9 or G45.0–G45.9, G46.0–G46.9 or S06.0–S06.9
2. Diagnosis of atrial fibrillation or atrial flutter (AF), I48
Inclusion criteria
1. Stroke, TIA, intracranial bleeding during 2003-2012
2. AF
Exclusion criteria
1. Intracranial bleeding (S06) diagnosed before AF
2. Post-operative AF only related to cardiac surgery procedure [a]
3. Suspected TIA (G45) but not confirmed by a neurologist [a]
4. Diagnosis of transient global amnesia (G45.4) without evidence of cerebrovascular event
5. Patients with data not available electronically
6. Patients living in the catchment area for less than a year after the Index event [a]

[a] The criterion was not applicable when queried with the InSite EHR Research Platform and with Turku University Hospital's Clinical Data Warehouse query tool.

EHR Research platform

An automated EHR Research tool created in Europe is InSite, a research platform maintained by Custodix N.V., Belgium, Daughter Company of TriNetX, USA. Through the platform, a hospital's EHR data can be utilized for research purposes, for example, for validating clinical trial protocols and for identifying patients who are potentially eligible for clinical trials (Doods et al. 2014). Thus, users can form queries based upon the trial protocol's inclusion and exclusion criteria and obtain counts of patients who match these criteria. Data processing is performed under the hospital's control and the users of the EHR Research Platform can only see aggregated results, i.e. patient counts.

Reference data

Traditionally, detection and preselection of eligible patients for clinical trials has been done by identification with manual searches from electronic patient records. This 'manual search' is able to provide reliable and controlled results and is considered a standard method for patient identification. The reference dataset used in this study was collected in the Fibstroke study (Palomäki 2017), which was a retrospective cardiovascular register study evaluating the associations of AF with stroke, TIA and intracranial hemorrhage. Strokes, TIA events and intracranial bleeding events were collectively referred to as 'Index events' in this assessment.

The reference dataset was formed from the EHR data of patients with AF diagnosed before the year 2013 and with at least one incidence of an Index event during the years 2003-2012. More detailed eligibility criteria are presented in Table 5.

The reference data had been identified manually from the hospital's EHR system by researchers reviewing electronic patients' charts, one at a time, according to a data listing derived from the EHR system of Turku University Hospital.

Hospitals' Clinical Data Warehouse

As the EHR Research Platform only provides search results expressed as patient counts, the Clinical Data Warehouse (CDW) service of Turku University Hospital was asked to provide a comparison of the patient counts provided by the platforms on a patient identifier level. The CDW was used to verify whether the EHR research platform could accurately identify, at the individual level, the patients in the hospital's CDW, and to compare the EHR query results with the results of the manual search.

The CDW of Turku University Hospital is a structured repository processed for data extraction and therefore optimal for research purposes. The CDW process has been described elsewhere (Auria Clinical Informatics 2018).

Both the manual search and the query performed with the EHR Research platform, as well as the CDW query, used the same EHR data of Turku University Hospital as the source data for the searches.

4.2 Study populations

4.2.1 Interview respondents

The respondents interviewed for the investigation of patient recruitment and trial feasibility evaluation processes were employees of pharmaceutical companies and clinical CROs operating in Finland (7 participants), Sweden (5 participants), Denmark (5 participants) and Norway (4 participants). The interviewees represented senior-level employees in 17 different mid-size and large pharmaceutical companies and 4 CROs working in the Nordic countries. In four cases, there were two respondents from the same company, but from a different country. The respondents' experience in clinical trials and years of being employed by the current company are presented in Table 6.

Table 6. Characteristics of the participants in the interviews, n=21.

Country	
Finland	7
Sweden	5
Denmark	5
Norway	4
Position in the company	
Clinical Study Management	6
Clinical Operations	6
Clinical Site Management	4
Feasibility and Recruitment positions	5
Gender	
Female	17
Male	4
Company type	
Pharmaceutical company	17
CRO	4
Served current employer	
0–5 years	6
6–10 years	7
11–20 years	6
>20 years	2
History of working with clinical trials	
<10 years	2
10–20 years	9
>20 years	10

4.2.2 Patients

The reference data contained information on 2166 individual patients identified manually from the EHR system of Turku University Hospital. All patients fulfilled the eligibility criteria as confirmed by the researchers of the Fibstroke study (Palomäki 2017).

4.3 Study design

4.3.1 Patient recruitment and feasibility evaluations

In the interviews, a qualitative semi-structured design was applied (I, II) (DiCicco-Bloom and Crabtree 2006). The data analysis contained also quantitative elements with descriptive statistics. The interviews were conducted between March and July of 2019.

The participants were recruited by e-mail invitations through suggestions of the Nordic Pharma Industry associations, Pharma Industry Finland, Läkemedelsindustriföreningen (Sweden) and Legemiddelindustrien (Norway). Participants in Denmark were recruited through personal industry contacts by the research team. Respondents were considered eligible if they were working for a pharmaceutical company or a clinical CRO representing the industry and were involved as sponsors in conducting phase I-III clinical drug trials with patients. The participants were to have an impact on the site identification and patient recruitment process in their company, which was confirmed before the interviews. Participants only involved in phase I trials with healthy volunteers were excluded. As background information, the participants' titles, experience in clinical trials, time of employment in their current company and contribution to site identification and evaluation processes were collected.

Twenty-eight interviewee candidates were contacted; one refused to participate and three candidates did not respond to the e-mail request, and another three did not fulfil the inclusion criteria. Purposive sampling was applied to ensure that professionals across all of four countries were included. The inclusion of participants was continued consecutively until saturation was achieved, i.e. until no new meanings to the categories were captured from the interviews (Hennink, Kaiser, and Marconi 2017).

4.3.2 EHR Research platform assessment

A commercially available EHR research platform, InSite, was assessed in order to evaluate how accurately it could identify the same patients that had been identified

with a conventional manual search (III). In addition, the capability to find the dates when AF and the corresponding index events had been diagnosed and to identify the temporal relationship of the two diagnoses (i.e. which one occurred first, AF or the index event) were examined and compared with the reference dataset (III). The temporal relation between the two diagnoses was assessed because of its essential informative value for searching of study subjects and because of the complexity of formalizing an adequate time constraint query in a structured, digital format.

The Insite platform has its origins in the European Union's IMI EHR4CR project (De Moor et al. 2015), where a European public-private partnership aimed to develop a computerized platform enabling the secondary use of EHR data collected from EHRs over its network (Girardeau et al. 2017). The platform allows researchers to interact with an anonymized copy of the hospital's EHR system for validating and optimizing clinical trial protocols and for accelerating the recruitment of trial participants. Data processing is performed under the hospital's control and the users of the EHR Research Platform can only see aggregated results, i.e. patient counts.

The platform was integrated with Turku University Hospital's clinical data warehouse (CDW). Doods and colleagues have previously described the general practices for the installation of the platform, configuration to the networks and consideration of the local data protection regulations (Doods et al. 2014).

The platform's functionalities were examined in the spring and summer of 2018 with the same patient eligibility criteria (Table 5) as those employed in the reference study, after formalizing the criteria into structured items. Examples of the transcription of the eligibility criteria to searchable items are presented in Appendices 2 and 3.

4.4 Data analysis

4.4.1 Patient recruitment and feasibility evaluations

Participant profiles, trial profiles and responses to categorical questions were collected using REDCap, version 9.1.12 data management software (Harris et al. 2009) and analyzed descriptively. All other interview contents were transcribed verbatim and managed with NVivo software, version 12 plus (QSR International Inc., USA).

Inductive qualitative content analysis (Elo and Kyngäs 2008) was applied for analyzing the success and failure factors of participant recruitment (I) and feasibility evaluations (II). The transcripts were read multiple times to obtain an overall impression of their contents. Only the manifest content (the items actually uttered by the participants) of the interviews was analyzed.

The coding process was initiated by identifying factors impacting (positively or negatively) on the recruitment (I). Thereafter, the sub-categories started to form as

similarities and differences in the codes were noted. The codes were collated to sub-categories that were further grouped and abstracted into categories by classifying them into items having similarities or conjunctive causes. The categorization, as derived from the data, was made together with another researcher, in order to develop a mutual understanding of the meanings of the codes. The role and potential use of EHR in patient recruitment were coded similarly.

All coding and formation of the sub-categories and main categories of the recruitment success and failure factors (I) were completed and the analysis was finalized before the initiation of the analysis of the feasibility evaluation data (II). For the feasibility evaluation analysis (II), all transcripts were first re-read in order to gain an overall understanding of the items related to the feasibility evaluations. The data were coded based on the research questions and the codes sharing the same content area were abstracted into sub-categories and further grouped to categories and main categories.

4.4.2 EHR research platform assessment

Three mutually exclusive categories were formed to explore the differences between the results provided by the two search methods (III): ‘Patients identified with CDW, but not with the manual search’, ‘patients not identified with CDW, but identified with the manual search’ and ‘patients identified with both methods’ (Figure 4).

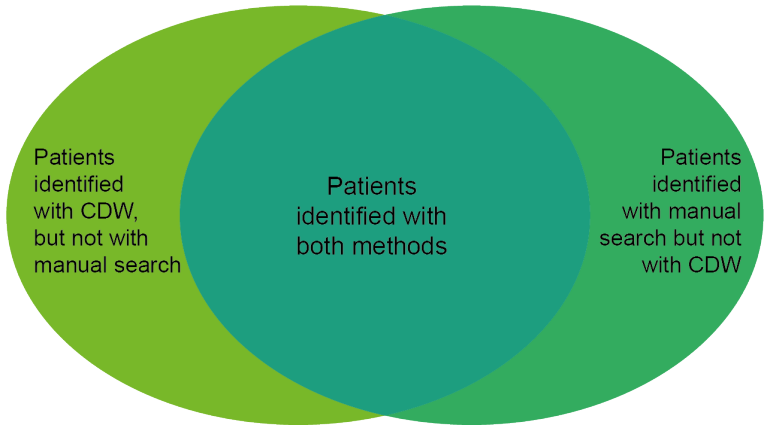


Figure 4. Testing the differences of the searches: three categories were formed. Schematic presentation, not in scale.

For testing the temporal relations of the two diagnoses (III), AF and the index event, three groups of patients were formed: ‘patients with their first incidence of AF before the Index event’, ‘patients with their first incidence of AF after the Index event’ and

'patients with their first incidence of AF and Index event at the same time' (Figure 5). The time window was one day, i.e. in order not to be concurrent, the first AF had to occur at least one day before or one day after the first diagnosis of the Index event.

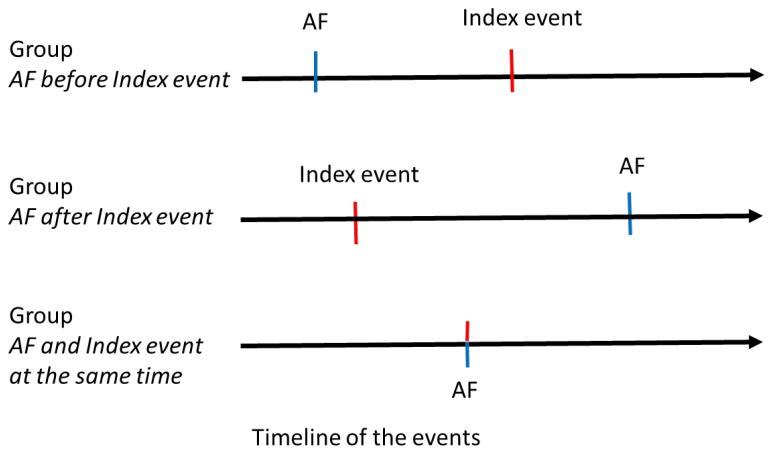


Figure 5. Testing the temporal relation of the patient's two diagnoses. Three patient groups were formed depending on which event they had first, AF or the index event, or whether those two events had occurred at the same time. The time window was one day (24 hours).

4.4.3 Statistical analyses

Categorical data were summarized as event counts (I, II), patient counts (III) and percentages (I, II, III). No further statistical analyses were performed because of the descriptive nature of the study.

4.5 Ethical considerations

Participation in the interview (I, II) was voluntary and the participants were informed verbally and with a written information sheet about the aim of the study and its practical implementation. The participants gave their verbal consent for the interview. Permission to record the discussions was requested from the participants, and recording was accepted by them in the beginning of the interviews. The participants were informed that the interview outcomes would be published without revealing the identity of the respondents or their companies.

For the assessment of EHR search tool functionality (III), patients' medical records contained in Turku University Hospital's CDW were reviewed. A research permission for a register-based study was obtained from Turku University Hospital.

5 Results

5.1 Trial demographics

Table 7 lists the trials selected by the participants for the interviews examining the recruitment factors (I) and feasibility evaluations (II) in the Nordic countries.

Each Nordic country appeared in at least half of the trials covered by the interviews: out of 34 trials, Finland, as a country, was involved in 21 trials (62%), Sweden in 20 trials (59%), Denmark in 19 trials (56%) and Norway in 17 trials (50%).

Most of the trials were phase III trials (65%) and trials with adult patients (91%). The most prominent therapeutic areas were oncology (24%), neurology (18%) and endocrinology (18%). Detailed trial characteristics are listed in Table 7. The numbers of planned and recruited patients per therapeutic area are presented in Appendix 4.

Table 7. Trial characteristics in the interviews. n=34

Phase	
Phase I	3 (9%)
Phase II	9 (26%)
Phase III	22 (65%)
Patient type	
Adults	31 (91%)
Pediatric patients	3 (9%)
Therapeutic area	
Oncology	8 (24%)
Neurology	6 (18%)
Endocrinology	6 (18%)
Cardiology	4 (12%)
Gastroenterology	3 (9%)
Pulmonology	3 (9%)
Psychiatry	1 (3%)
Nephrology	1 (3%)
Dermatology	1 (3%)
Immunology	1 (3%)
Recruitment status	
Recruitment ongoing	12 (35%)
Recruitment ended	22 (65%)
Trial status	
Trial ongoing	15 (44%)
Trial ended	19 (56%)
Trial schedule	
Trial on schedule	21 (62%)
Trial delayed	13 (38%)
Number of patients	
Number of patients planned vs enrolled in all trials	6931/6249 (90%)
Number of patients planned vs enrolled in successfully recruited trials	5573/5445 (98%)
Number of patients planned vs enrolled in failed recruitments	1358/804 (59%)

5.2 Patient recruitment (I)

5.2.1 Success of patient recruitment

Of all trials evaluated, 50% (17 trials) were reported as having succeeded in their recruitment of participants, whereas the other half were considered to have failed to meet their initial recruitment timelines. When dividing the trials into clinical development phases I, II and III, 83% of phase I and II trials were reported to have succeeded in their recruitment. The reported success rate was much lower in phase III trials (32%) (Table 8). The most successful therapeutic areas in terms of recruitment of participants were oncology and neurology (Table 8). Cardiology trials

were seen to have the lowest success rate in their recruitment; none of them recruited as planned. Nonetheless, cardiology trials were reported to have recruited 68% of the planned numbers of patients, whereas failed recruitments within all therapeutic areas reached, on the average, 59% of the targeteted numbers of participants (Appendix 4). Comparisons between phases or therapeutic areas should remain tentative and cautious, as the numbers of trials in each of them were small (Table 8).

Table 8. Recruitment success and failure in the trials covered by the interviews.

	Success	Failure
Phase		
Phase I and II	10 (83%)	2 (17%)
Phase III	7 (32%)	15 (68%)
Therapeutic area		
Oncology	6 (75%)	2 (25%)
Neurology	5 (83%)	1 (17%)
Endocrinology	2 (33%)	4 (67%)
Cardiology	0 (0%)	4 (100%)
Gastroenterology	2 (67%)	1 (33%)
Pulmonology	2 (67%)	1 (33%)
Other	1 (25%)	3 (75%)

5.2.2 Patient recruitment as a trial delay factor

Of the 34 trials discussed, 13 trials encountered delays from their original schedule (Table 7). The interview participants evaluated that patient recruitment was the '*major reason for the delays*' in 9 trials (69%), whereas in the four remaining delayed trials, slow recruitment had *'no impact on the trial delays'* (31%). The reasons reported for trial delays in those four trials were slow ethics committee procedures (one trial), slow start by the sites (two trials) and delayed availability of the investigational medicinal product (one trial).

5.2.3 Success and failure factors in patient recruitment

Four main categories (sponsor related, site/investigator related, patient related, sponsor-site-patient collaboration related factors) were derived from the data representing both success and failure factors, whereas a fifth category (factors related to start-up activities) represented only failure factors. The main categories and their principal sub-categories are listed in Table 9.

Table 9. Success and failure factors of patient recruitment in clinical drug trials.

Sponsor related	Site/ Investigator related	Patient related	Collaboration related	Start-up related [a]
Trial protocol	Access to patients	Patients´ medical need for new treatments	Sponsor-site-patient collaboration	Ethics Committee evaluation
Trial preparation and feasibility evaluations	Investigators' motivation for trials and commitment to recruitment Site resources, set-up and experience	Patients´ role in their care and attitudes to clinical trials		Site contracts

5.2.4 Role of EHR in patient recruitment

When the respondents, i.e. representatives of pharmaceutical companies and clinical CROs, were asked how the trial sites found their patients for trials, site EHRs emerged as the most important source, used in 85% of the trials (29 out of 34 trials, Table 10).

Nordic trial sites were perceived to use EHRs and patient registers in their recruitment more often than those located in other European countries, although large variation was reported in how EHRs were used by the sites: Some sites "knew" the patients that they have in their EHR system, and recruited them without any further searching from the EHR, whereas some sites searched their entire hospital's EHR data by filtering with certain eligibility criteria with the help of the hospital's IT department.

Table 10. Reported sources of potential trial subjects, n=34 trials. Most trials used more than one recruitment source

From where did the investigators identify trial participants?	
Electronic Health Records	29 (85%)
Referrals within/from outside of the hospital	9 (26%)
Patient register or biobank	6 (18%)
Social media (Facebook etc.) and web-based recruitment tools	5 (15%)
Advertisements in newspapers and magazines	5 (15%)
Patient organizations (advertisements, public lectures)	5 (15%)

Other tools, such as referrals within/from outside of the hospital, patient registers, patient organizations, social media, web-based recruitment tools and/or traditional advertisements, were much less frequently used than EHR. The other tools were most commonly used along with the EHR data, and often only after realizing that the recruitment target would not be reached solely with the site´s own patients. For example, only six trials used patient registers, such as national cancer registers or registers for diabetes, cardiac diseases or biopsies, or a dedicated register for a

chronic, rare, progressive disease. However, most of them were regarded as useful for finding potential trial subjects in those few trials.

The respondents stated that the use of EHRs in patient recruitment was not equal in all Nordic countries because of national differences in data protection legislation and its interpretation. They also highlighted the importance of the quality of the EHR data; the data should always be up-to-date, correct and comprehensive in order to find potential trial participants.

5.3 Feasibility evaluations (II)

5.3.1 Site identification practices and the role of EHR systems in site identification and selection

The main categories formed for feasibility evaluation practices and examples on quotations are presented in Appendices 5 and 6, respectively.

Changes in the feasibility evaluation process

As viewed by the participants, changes in the landscape of clinical trials, for example the increased need to find certain types of patients with specific mutations, laboratory values or rare diseases, has challenged the feasibility evaluation process and the data needed in evaluations. Also the increased use of various types of electronically available data has changed the evaluation process. It was presumed that the feasibility evaluation process, especially the contents and application of site feasibility questionnaires would change within the near future because of the large amounts of data that have become and are becoming available; the feasibility evaluation process was perceived as becoming faster and more precise. Some participants found it worrying that investigators did not have sufficient time or interest to respond to site feasibility contact requests; many sites were viewed as not having time to conduct clinical trials. Even if they were identified by the sponsors, no collaboration emerged.

Site identification on a global level

Overall, it seemed that site identification is based on information on previous performance of the site, not on defining where the suitable patients are. Site identification procedures were recognized on two different levels, global and local. At the global level, various databases were seen to have a major role in identifying potential countries and sites for the conduct of clinical trials. Companies used their own or joint databases on sites´ previous performance (e.g. DQS DrugDev, IQVIA

Company, USA), trial and investigator databases (e.g. Citeline, Informa Plc., London, UK or Global Data, Global Data Plc., London, UK), or data available in public repositories (e.g. National Library of Medicine's Clinical Trials Registry, www.clinicaltrials.gov). Companies were also reported to share non-confidential data on potential trial sites with each other for feasibility evaluation purposes. Only a few participants mentioned that, on the global level, prospective countries were also identified by employing commercial EHR technology platforms based on EHR data from health care providers (such as TriNetx, TriNetX LLC, USA).

Some disadvantages were also highlighted with respect to data-based decision-making; countries and sites not existing in the databases will not end up being considered for participation even if they might have good possibilities to conduct certain trials. The respondents perceived that the Nordic countries lack visibility on the global level because of their small populations and low trial volumes and were not marketing their capabilities sufficiently well to the global decision-makers of pharmaceutical companies. The local Nordic subsidiaries, who were aware of their own country's advanced technological systems for using health care data for clinical trial purposes, were seen to play a key role in the marketing of their countries to the global decision-makers.

Site identification on a local level

According to many interview participants, on the local level, investigator databases were seldom used in the Nordic countries for site identification. Instead of using investigator databases for site identification, some participants used them for evaluating the validity of the patient number estimates provided by the investigators (see chapter 5.3.2). EHR tools were not applied at all at the local level. EHR query tools were only used for testing purposes by the sponsors in the Nordic countries.

The Nordic countries have limited numbers of investigators, and most of them are already known to the local subsidiaries. Instead, local intelligence, e.g. understanding local practices, treatment paths and healthcare systems by the pharmaceutical companies and CROs were applied. Local knowledge was perceived as valuable in site identification, and it was considered impossible to capture that from any database, but was instead perceived as being based on the experience of the sponsor's local country representatives. They "knew" their countries.

The effect of previous collaboration

As reported by the respondents, in two thirds of the trials covered by this study (23 out of 34 trials, Table 11), the method for identifying sites was based on previous collaboration, possibly supported by other identification methods. Previous collaboration between the trial sponsor and the trial sites did not as such guarantee

recruitment success: almost half of the trials covered here (15/34) solely used sites with previous experience, but one third (5/15) of them still failed in their recruitment. Based on the data, it seems that the site selection methods are not explicitly related with recruitment success or failure. However, successfully recruiting trials more often used multiple recruitment methods than trials with failed recruitment, and trials relying on previous collaboration more often succeeded in recruitment than failed.

Table 11. Site identification methods (n=34 trials). A single trial could use multiple methods.

	Number of trials	Recruitment succeeded	Recruitment failed
Existing contacts with sites and investigators + possibly other supporting identification methods	23	14	9
Only existing contacts with sites and investigators	15	10	5
Sites suggested by KOL, NCI, PI[a] or other investigators	7	5	2
Recommendation from within the sponsor company	4	1	3
Sites known to treat certain types of patients but no previous collaboration	3	0	3
Public database (such as www.clinicaltrials.gov)	3	2	1
Internet search (such as Google)	2	1	1
Investigator network	1	0	1
Publication database review	1	1	0
Commercial investigator database	1	1	0
Not known	3	0	3

[a] KOL, Key opinion leader; NCI, National Coordinating investigator; PI, principal investigator

5.3.2 Evaluation of the sites' access to patients

Assessing the access to patients, i.e. the sites' capabilities in finding trial subjects, was seen as a complex process during feasibility evaluations, both to the sponsors and to the investigators. Most investigators were perceived as not having enough time, interest or information for undertaking proper feasibility evaluations, and because of that, their estimates on potential patient counts were quite often seen to be significantly unrealistic. Most investigators were anticipated to have based their patient count estimates mostly on their previous experience, not employing EHR data or statistics from previous trials to support their assumptions.

In contrast, some sites were perceived to be able to estimate their recruitment capabilities correctly. There were also some examples of investigators using EHR in assessing their recruitment capabilities (presented in section 5.3.3.)

It became very clear during the interviews that trial sponsors did not usually accept the investigators' estimates as such, but tried to evaluate their validity in many

ways. Investing time in the feasibility discussions with the investigators was considered important. Many respondents highlighted that the feasibility evaluation should always be bi-directional: providing proper information to the investigators and listening to their feedback and justifications about trial feasibility.

Many respondents had only carried out a dialog with the investigators to obtain the rationale behind the investigator-estimated patient counts. Some interview participants admitted that they had not performed sufficiently thorough evaluations on how the investigators had ended up with certain numbers of predicted trial subjects.

By using previous site performance data in their own or in commercial investigator databases, the sponsors were seeking confidence in the investigator-estimated patient counts. This was perceived as a quick and objective way to validate the investigators' estimates. However, it was considered possible that such historical data were not comparable with the requirements of a new trial, or did not contain enrolment numbers of the site under evaluation, which complicated the validation.

5.3.3 EHR in evaluating access to patients

Investigators using EHR data

The sites using EHR data were considered attractive by the sponsors: they could promptly justify their estimates of potential trial subjects, which together with their earlier recruitment performance offered them a clear advantage. In fact, those sites seemed to be regarded as more reliable in their patient count estimates even if information whether their estimates actually were more accurate than the estimates of those not using EHR data was absent. The interview participants presented some examples of sites who used their EHR data in patient count estimations. For a feasibility evaluation, these sites pre-screened their potential trial subjects in the hospital's EHR system. If they did perform this already in the feasibility evaluation phase, it was perceived as beneficial for both the site and the trial sponsor: the sponsors received more reliable information on the sites' recruitment capabilities, and the sites saved time at the launch of the trial as they already had the patients pre-screened, which expedited the start of recruitment.

The participants highlighted that the contribution of EHR data in feasibility evaluations is indication-dependent. In trials on chronic diseases, EHR may give information on actual patients potentially identified as suitable for the trial, whereas in acute diseases, e.g. stroke, EHR data could be used to reveal how many such patients have been seen by the site in the recent past and hence to estimate the number of potential trial subjects in the near future.

As became evident by the interviewed participants, there were legislative differences between the Nordic countries on how investigators can gain access to EHR systems for trial feasibility evaluation purposes. Furthermore, local hospital practices were seen to mean that sometimes investigators were not allowed to promptly obtain estimates based on actual EHR data for their feasibility evaluations.

Verification of trial patients' availability by sponsors

Some sponsors had asked the sites to justify their patient count estimates and to show that they had made a search in the EHR system. Several reasons why sites did not use EHR data emerged. Patient count estimates could require monetary compensation paid by the investigator to the hospital's IT department. Some hospitals required internal approval before the data search, which would have resulted in unacceptable delays as the time frame to reply to feasibility questionnaires is rather short. Overall, to get patient count estimates beyond the investigator's own patients required additional time and effort from the investigator, without any monetary compensation for this work at a time when there was no guarantee that the investigator would be selected to participate in the trial.

The use of EHR data for reviewing the availability of potential trial subjects was not only a choice to make or not to make by the investigators. As viewed by the participants, there are legislative restrictions for example in the access to and in the secondary use of patient data in the Nordic countries which regulate how investigators can utilize EHR systems for this purpose.

It became very clear that trial sponsors themselves did not commonly use EHR query tools for evaluating potential patient counts in the Nordic countries. The main reason was seen in the legislation restricting access to patient data for such use. For aggregate EHR data (only patient counts), some participants mentioned that the hospital management's interpretation of the legislation and prevailing attitudes were the biggest obstacles to their use by the sponsor. Only a few participants were aware of platforms that enable sponsors to view patient counts in the EHR systems of different hospitals/countries, and had piloted for example the InSite EHR research platform (formerly Custodix, Belgium, currently part of TriNetX, USA), but did not continue this use because of its lack of data regarding the Nordic sites.

Most participants stated that the use of EHR query tools by trial sponsors had not increased as expected or desired.

Advantages of Nordic trial sites and future development

The participants evaluated that Nordic sites could and should distinguish themselves from other countries by employing advanced technical solutions and processes for

the efficient secondary use of their health care data for clinical trials. The Nordic sites were seen to be the most competitive in complex trials with a need for database searches for suitable trial participants, either from disease registries or from EHR data.

Most participants underlined that the need for patient EHR data in the identification of trial participants will become emphasized in the future, especially in trials on rare diseases and in trials on targeted medicines. The participants perceived that the use of EHR data would expedite obtaining the patient count estimates and thus improve the accuracy of the estimates.

When requested to describe the ideal future set-up for querying the EHR data, the participants wished to have access to larger entities than single hospitals; this could involve Nordic-wide EHR data lakes with an emphasis on highly secured data protection and a sufficient coverage of the whole Nordic population, with 27 million inhabitants.

5.4 Identification of trial participants with an EHR platform (III)

In order to perform a valid comparison, the EHR Research Platform and the CDW query were first confirmed to search similarly (Table 12).

The EHR Research Platform identified 5859 eligible patients, whereas the patient count obtained with the manual search was 2166 (Table 12). The CDW search identified 5840 patients.

Table 12. Number of patients found in Turku University Hospital's EHR system as queried with a manual search, with the EHR Research Platform and with the hospital's Clinical Data Warehouse (CDW) query tool. The sum of proportions of the index events within the cohort exceeds 100% as patients could have multiple index events.

	Manual search (Fibstroke)	EHR Research Platform (InSite)	CDW query
Patients with diagnosis of atrial fibrillation (AF) any time before 2013 and Index event in 2003–2012	2166	5859	5840
Composition of intracranial vascular incidents within the cohort:			
Patients with AF and stroke	1755 (81%)	4807 (82%)	4806 (82%)
Patients with AF and TIA	428 (20%)	1389 (24%)	1396 (24%)
Patients with AF and intracranial bleeding	313 (14%)	625 (11%)	596 (10%)

TIA, Transient ischemic attack

Patients identified with CDW but not manually

There were 2.7 times more patients identified with CDW as identified with the manual search, 5840 patients and 2166 patients, respectively (Table 13). Three reasons responsible for this discrepancy were found:

1) There were exclusion criteria that could not be transformed into searchable items.
2) There was an inability of the CDW to distinguish between suspected and confirmed diagnoses in the structured data. For example, the diagnosis of AF or the Index event could be preliminary, unconfirmed or only suspected, as described in the free text field of a patient's EHR. In the manual search, only those patients with confirmed diagnoses were collected.
3) Some diagnoses were only noted in the free-text fields and were therefore only identified in the manual search.

Table 13. Patient counts obtained with the CDW query compared to the manual search.

		Patients identified with CDW query (n=5840)	
		Yes	No
Patients identified with manual search (n=2166)	Yes	2033	133
	No	3674	-

Patients identified manually but not with CDW

There were 133 patients not identified with the CDW query, but found with the manual search. After investigating a sample of 50 patients, an explanation was found: most of the patients in the sample (41 patients) were diagnosed with the index event before the CDW was applied in the hospital, therefore being impossible for the CDW to locate. The remaining nine patients in the sample had been collected with the manual search even though the patients had an exclusion criterion ('Intracranial bleeding diagnosed before AF') recorded in the EHR, but for unknown reasons they had not been excluded from the manual search. Based on the results, it was concluded that the CDW did not miss identifying any eligible patients that were in the EHR.

Patients found with both methods

There were 2033 patients identified with both the CDW query and the manual search. The identity of those patients was confirmed to be the same in both searches.

Time constraint analysis

The CDW query found the three temporal categories (AF diagnosed before, after or at the same time with the Index event) in different proportions compared to the manual search (Table 14). This was due to the different search logics and missing data in the structured data: the manual search collected the actual diagnosis dates for both AF and the Index events, using unstructured data. For the structured data in the EHR, as used by the CDW query, the diagnosis codes for each hospital stay were fixed to the date of the patient's discharge from the hospital, even if many of the diagnoses were established earlier during that hospital stay. This clearly impaired the precision of the diagnosis dates. Several cases were also identified where a diagnosis of AF was only added into the structured EHR data at the time of hospitalization because of the Index event, although a review of the unstructured data revealed that AF had actually been diagnosed earlier.

Table 14. Patient population cohort with AF and index event divided into three categories based on the temporal relations of these two diagnoses. The search results were different due to the different search logic and an inappropriately narrow time window (one day) for these diagnoses.

	Patients with AF[a] *before* Index event	**Patients with AF *at the same time* with Index event**	**Patients with AF *after* Index event**
Manual search (n=1002)	533 (53%)	270 (27%)	199 (20%)
CDW[b] query (n=1002)	412 (41%)	283 (28%)	307 (31%)

[a] AF= atrial fibrillation

[b] CDW=Clinical Data Warehouse

6 Discussion

6.1 General discussion

The aim of this study was to investigate the challenges encountered in patient recruitment for clinical trials in the Nordic countries, and whether the recruitment could be improved by using EHR data. It was important to conduct this study as there was limited information available on the recruitment success and failure factors and on the use EHR data in recruitment and feasibility evaluations in the Nordic countries.

It was considered beneficial to explore the views of the pharmaceutical industry, as two thirds of all clinical drug trials conducted in the Nordic countries are currently industry-sponsored (Table 1). The Nordic countries are in the front line in the secondary use of EHR data (Bonomi 2016; Nordforsk 2019). It was seen as important to assess the current and future roles of EHR data in the recruitment of trial participants and in site feasibility evaluations, and to evaluate how accurately an EHR research platform could identify potential trial participants. Optimization of the use of EHR queries for patient recruitment and feasibility evaluations might have major positive effects on the Nordic countries' competitiveness in attracting clinical trials.

The study findings indicated that in addition to factors related to the trial investigators/sites, patients, and the collaboration between these stakeholders, the sponsors were very well recognizing their own contribution to the recruitment success or failure (I). This study also revealed that EHR data are commonly used in patient recruitment for clinical drug trials (I), but they are rarely used for trial feasibility evaluations before trial commencement (II). Lastly, EHR systems and data have limitations in accuracy, but it was concluded that they have great potential for identifying trial participants, while reducing the manual workload (III).

The current study was thus able to answer the predefined research questions and to generate new information on this topic in a Nordic setting.

6.2 Methodological considerations

In order to investigate the current research questions, Nordic representatives of pharmaceutical companies and clinical CROs were interviewed, and one commercially available EHR research platform was tested for its functionality.

A qualitative approach, in this case interviews, was chosen to obtain in-depth understanding of the pharmaceutical industry's views on the factors that influence patient recruitment success and failure (I) and trial site feasibility evaluations (II). The use of qualitative methods has been encouraged as a means to obtain a better understanding of and overcoming the challenges of patient recruitment for clinical trials (Elliott et al. 2017; Hennessy et al. 2018). Identification of recruitment success and failure factors and elucidation of the conduct of feasibility evaluations would not have been possible with other methods than qualitative interviews.

Finland, Sweden, Denmark and Norway share many similarities in their cultural features, health care systems with national EHR, economies and living standards, with similar influences of these features on disease prevalence and outcomes, quality of care and the inhabitants' possibilities to receive care. With their total of 27 million inhabitants, they constitute a substantial geographical region for conducting clinical trials. Therefore, for the transferability of the results, it was considered advantageous to conduct the study across these national borders. The interview participants represented pharmaceutical companies and contract research organizations over a wide scale and presented trials in various therapeutic areas and in all pre-market phases of clinical drug development, expressing and presenting heterogeneous views on the items discussed.

In the EHR query tool evaluation, the categorical data were summarized as patient counts and percentages. Therefore, the analysis method provided a guide to the accuracy of the EHR research platform in being able to identify potential trial participants. The use of the hospital's CDW enabled a comparison of the EHR research platform and a manual search of the patient records.

The challenges raised in the interviews regarding the use of EHR data and EHR query tools (I, II) were supplemented by the findings recognized in the assessment of the EHR research platform (III).

6.3 Patient recruitment in clinical trials (I)

6.3.1 Proportion of trials recruiting successfully and recruitment's impact on trial delays

Failed patient recruitment was seen as the greatest single factor causing trial delays. Of 13 trials that were considered as being delayed, the interview respondents stated that slow patient recruitment was the main reason for their trial delays in nine trials. Even if the number of delayed trials evaluated was rather small, these results indicate that recruitment challenges are the most prominent reason for trial delays in the Nordic countries. This result conforms with earlier reports concluding that poor patient recruitment is one of the main reasons for trial delays (CenterWatch 2009).

Half of the trials in this study (17 trials out of 34) reached their original recruitment target in the Nordic countries, but the other half failed. The success of recruitment was emphasized in phase I and II trials (83% success rate), whereas the recruitment failures occurred mostly in phase III trials (32%). As the small number of trials and the qualitative design of the present study do not allow generalizations, further investigations will be needed to confirm the overall recruitment success rate, and also whether recruitment into phase I and II trials in the Nordic countries is more favorable than generally. In previous reports, the recruitment success rates have ranged from approximately 30% in all kinds of trials up to 62% in cancer trials (Bower, Wilson, and Mathers 2007; Cheng, Dietrich, and Dilts 2010; McDonald et al. 2006). Variable recruitment success in different clinical phases has been reported for example by Carlisle et al. They reported that phase II trials were less successful in their recruitment than phase III trials (Carlisle et al. 2015).

6.3.2 Recruitment success and failure factors

Success of patient recruitment in clinical trials was seen as being dependent on many factors. Sponsor-related, investigator/site-related and patient-related factors, complemented by collaboration-related factors, were identified in this study. Recruitment success and failure factors were mostly complementary to each other, except for the start up-related factors: the Ethics committee application practices and site contracting were only identified as failure factors. None of the sponsor representatives stated that those factors were significant contributors to their trials' recruitment success.

Similar findings have been presented previously; for example, teamwork (Peckham et al. 2018; Strong et al. 2016), investigators' research experience (Getz 2011), implementation of a clearly defined "system" of recruitment, engagement of other staff in the hospital, time from ethical approval to first recruitment, and the provision of a dedicated trial coordinator (Levett et al. 2014) have been reported to be significant recruitment success factors. Williams et al. reported that the conduct of investigator training at the very beginning of the trial complemented by subsequent follow-up contacts and face-to-face meetings increased the success of recruitment (Williams et al. 2014). This aspect was not identified in our study. In an interview study with experienced U.S.-based investigators, good planning of the recruitment methods, the use of multiple recruitment methods, collaboration and commitment were perceived as tools to overcome possible recruitment challenges (Dombeck et al. 2020).

Hanson and colleagues listed five effective recruitment strategies; i.e. systematic screening of patient lists, thoughtful messaging to make research relevant for the patients, flexible protocols to accommodate patients' needs, support from clinical

champions, and the additional resources of a trials' cooperative group (Hanson et al. 2014). In a Swedish study, the study personnel perceived that a trial protocol that is easy to merge into the daily routines, a smooth consent process and committed personnel were likely to enhance the success of recruitment (Isaksson et al. 2019).

Many of the recruitment success and failure factors identified in this study are items that are being purposefully formed (i.e. the trial protocol) or systematically evaluated (i.e. sites' access to patients), and which can be influenced before the trial commencement (Figure 6). Therefore, in order to minimize the impact of recruitment failure factors and to maximize the potential of the success factors, investments should be made into conducting adequate feasibility evaluations before the launch of the trial.

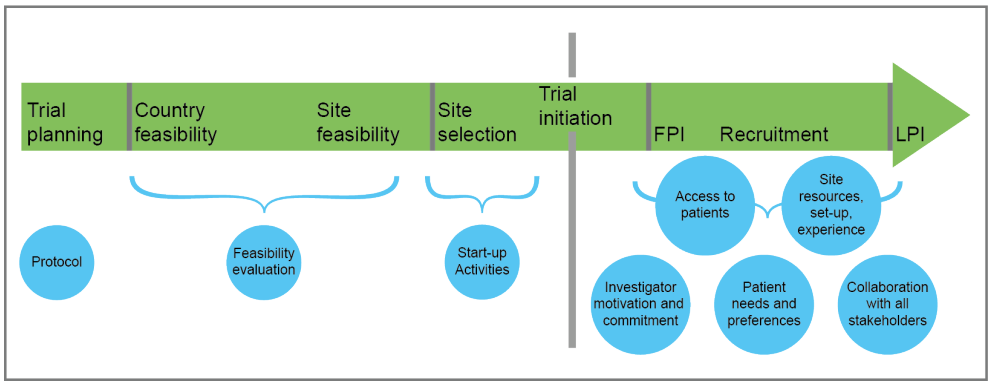


Figure 6. Recruitment-related factors inserted in the timeline of a clinical trial, based on their occurrence. FPI=First patient in, LPI= Last patient in.

6.4 Feasibility evaluations in clinical trials (II)

As stated in section 2.3.3, long development times from site selection to the start of recruitment can impair the success of recruitment. One explanation for this is the changing conditions at the sites during that time. A study conducted in an academic hospital in Western Cape, South Africa, concluded that justifications on patient count estimates made a long time previously had changed before the trial commencement, thus possibly complicating the recruitment (Burgess and Sulzer 2011). The researchers reviewed all feasibility evaluation questionnaires received by the hospital in the years 2004–2009. The site had to wait an average of 12.9 months (from 3 up to 33 months) after returning the feasibility evaluation questionnaire to the sponsor before the site was allowed to start recruiting the first patient. However, in a broader study with 400 clinical trial protocols reviewed worldwide, the time from site identification to trial start was calculated to be shorter, on an average 5–6 months (Lamberti et al. 2018). There is therefore an obvious need to develop the feasibility evaluation process so that it is both faster and more accurate.

6.4.1 Country and region selection

Success in recruitment of trial participants can be defined as the trials' capability to recruit certain numbers of eligible patients in the period set out before the trial commencement.

Success in recruitment in ongoing trials is key for the Nordic countries to be selected also for new clinical trials, as sites that have previously recruited successfully are favored. Countries and regions are increasingly selected based on their earlier performance data, contained in different databases. This was also noted in a large survey with 485 industry, CRO and Clinical Trial Unit respondents in 34 countries worldwide (Gehring et al. 2013). In the report, 83% of the respondents would have been 'much more likely' to include a site that had all relevant investigator-related and hospital-related information readily available.

A challenge for the Nordic countries is their relatively limited visibility in such databases because of the limited numbers of completed trials. If the numbers of clinical trials keep decreasing in the Nordic countries, also their visibility in the databases will decrease from the current level. In the future, the need to identify certain types of patients more precisely will be important. The visibility of low-volume trial countries, such as the Nordic countries, could be enhanced with advanced technology and by efficient secondary use of patient data, such as EHR data, to meet the needs in the changing landscape of clinical trials. This would also require changes in the legislation, i.e. allowing access to the aggregated patient data for investigators and third parties.

According to our respondents, site selection appears to be based on trust and relationships built in previous collaborations, but electronic data provide important support in the selection process. On a local level, the sponsors use electronic data (data on previous site performance) for validating the patient count estimates rather than for identifying the sites. The site selection methods do not seem to be explicitly associated with the success or failure of patient recruitment. This investigation should have been performed on the site level (and not on the trial level) in order to make more distinct conclusions about the significance of site identification methods for the success of recruitment.

In the Nordic countries, 65% of all clinical trials are currently industry-sponsored (Table 1). The Nordic representatives of the pharmaceutical companies seem to have a key role in marketing their countries to the global management teams that decide on the country selection. The local presence of the subsidiaries of the pharmaceutical companies and their marketing efforts to the global teams were seen to be important and even necessary in order to increase the number of clinical trials being conducted in the Nordic countries. Marketing efforts should also be increased by the Nordic countries as societies, even if some efforts have

already been made (see e.g. the Ministry of Foreign Affairs in Denmark, https://investindk.com/publications/clinical-trials-in-denmark).

It is seen as important to maintain and even increase the numbers of clinical trials conducted in the Nordic countries. Gaining access to new medicines already during their clinical development is considered as the greatest benefit gained by the communities that host global clinical trials. The early implementation of new treatments many times, but of course not always, offers opportunities to deliver the best clinical care to the populations of the Nordic countries. There are also other benefits; access to external expertise (Rosas et al. 2011), assistance in the development of the health care infrastructure, and the benefits for the local economy (Scorr Marketing 2017). For example, the use of investigational study medication instead of standard medication has been calculated to confer large savings for the communities (Kaló et al. 2014).

6.4.2 Site identification and selection

At the Nordic level, no query tools utilizing EHR data had been used for site identification. In addition, there was very limited use of investigator databases. In 15 trials out of the 34 trials covered by the interviews, the trial sites were selected purely based on an earlier collaboration (Table 11). The fluency of the collaboration was also noted as one central recruitment success factor (I) (Table 9). However, of the trials where sites were only selected because of earlier collaboration, 10 out of 15 succeeded in their recruitment but five of them failed (Table 11). Thus, collaborating with sites known from earlier trials did not guarantee successful recruitment, as recruitment success was found to be dependent on many additional factors (I). An accurate prediction of the recruitment success is a demanding challenge (Bruhn et al. 2019) in spite of previous collaboration, because there are always many other factors than access to patients that influence the overall success of recruitment. Examples of these are investigators' time resources and motivation, and the patients' willingness to join a study. Therefore, electronic data can be used to support estimating and validating the patient counts, but it can only partly solve the challenges of patient recruitment.

6.4.3 Complex feasibility evaluation process

The most frequently mentioned recruitment factor related to investigators and sites in this study was their access to patients (I). The existence of possibly eligible patients and the sites' capabilities to identify those patients need to be evaluated before trial commencement; it is an important criterion for site selection (Dombernowsky et al. 2019; Gehring et al. 2013).

Since lack of suitable patients was one of the most commonly mentioned reasons for failed recruitment (I), it is obvious that there have been difficulties, inefficiencies and uncertainties in the process of evaluating the sites' access to patients. This was confirmed when analyzing the views of the respondents about the feasibility evaluation process (II). Improvements need to be made in the accuracy of the patient count estimates of the feasibility evaluations.

The importance of communication has not been replaced by the use of data, but various types of data sources have been deployed along with the communication when evaluating the sites' recruitment projections. This study indicates that accurate patient count estimates are crucially important for two reasons:

1) A thoroughly evaluated and data-driven recruitment target of each trial site that has been set to reflect the reality will improve the chances for successful recruitment. Even if the site would have a good coverage of potentially eligible patients but the recruitment target had been set over-optimistically, the site would be considered as having failed in its recruitment of trial participants.
2) The success in recruitment can be viewed later in the site performance databases used by trial sponsors, increasing the site's possibilities to participate in new upcoming clinical trials. This is a key factor for the Nordic countries to be selected for participation in new clinical trials also in the future.

The planned numbers of patients should be set realistically, not too high, but not too low, based on the investigators' experience and existing evidence. It seems that investigators often estimate their potential patient counts predominantly based on their subjective experience, and only in some cases are the estimates backed up by evidence from the EHR system. Trial sponsors may try to validate the investigators' estimates against the previous site performance data, but this may not always be sufficient, because the earlier trials and their patient cohorts may not be directly comparable with the upcoming trial and its needs.

Those sites routinely using EHR data in making their patient count estimates and promptly providing the estimates were regarded as reliable and the most attractive from the viewpoint of the trial sponsors. All clinical trial sites in the Nordic countries should be allowed to review their hospital EHR data as a routine practice to evaluate their patient count estimates for an upcoming trial. The EHR review process should be both fluent and prompt. This would save time for the trial staff and would create a more factual basis for the dialogue between the sponsor and the investigators when setting up the recruitment targets for the sites. There will always be trials that cannot recruit patients from the EHR system, such as trials on acute stroke, but also in such

trials, the patient count estimates can be improved to reflect the number of similar patients who were treated in a certain time period in the past.

Reviewing of EHR data, both in the recruitment phase but especially already in the feasibility evaluation phase, should be enabled by harmonized legislation in all Nordic countries, simultaneously meeting all relevant standards for personal data protection. The legislation and data protection authorities in all Nordic countries should also describe the policies on how and by whom the sites can contact potential trial participants identified in the EHR system. Harmonized Nordic procedures, allowing the use of Nordic-level data lakes, could enable an ideal use of existing data reserves, clinical trial experience and competence for new clinical trials.

6.5 EHR in recruitment and feasibility evaluations (I–III)

6.5.1 EHR as a source for recruiting patients

In this study, 85% of all trials covered by the interviews recruited their participants from the sites' own patients, i.e. by viewing potential trial participants from the site's EHRs. Based on these results, the role of EHR data is very important for the recruitment of trial participants. It should be highlighted that in this study, the recruitment from the site's own patients mostly meant viewing potential trial participants from the site's EHR, which may not be the case in other regions, due to the utilization rate of EHR systems (Jha et al. 2008; World Health Organization 2016b).

The use of EHR data was encountered far more frequently than other recruitment methods in the trials of this study. The other methods were referrals from other physicians, disease registers, social media, advertisements and patient organizations.

Use of EHR data is regarded as an efficient tool in the recruitment of trial participants (Patrão et al. 2015; Zimmerman et al. 2018). It was reported that the incidence of actual enrolment timelines being less than or equal to planned enrolment timelines has increased from 47% to 77% between the years 2012 and 2020 (Lamberti et al. 2020). There is also evidence of a slight improvement of the recruitment success in trials: the number of trials reaching or exceeding their enrolment goals has slightly increased during recent years (Lamberti et al. 2020). It is possible that this apparent improvement is due to the implementation of novel recruitment methods, such as EHR queries and other digital solutions. The development of new recruitment methods should clearly be continued. It needs to be emphasized that EHR queries as a recruitment method may not be suitable for all trials, and the selection of recruitment methods should be made carefully and adjusted to the disease being investigated.

Clinical quality registers have a long history in the Nordic countries (Emilsson et al. 2015), but interestingly, only six trials out of the 34 covered by the interviews had used patient registers, such as national cancer registers or registers for diabetes, cardiac diseases, and biopsies to identify potential trial subjects for the recruitment of participants into clinical trials. As also identified in the current study, Tan et al. reported similar findings on the benefits of using registries for successful recruitment of trial participants (Tan, Thomas, and MacEachern 2015). Even if it was not within the scope of this study, the findings encourage also the recruitment of trial participants by accessing these existing comprehensive registries in the Nordic countries.

6.5.2 EHR's role in evaluating sites' capabilities to recruit patients

According to the interviews, EHRs were rarely reviewed in the course of the feasibility evaluations carried out before trial commencement; the sites were perceived to have mainly estimated their potential patient counts based on their earlier experience, complemented by a study-specific dialogue with the sponsor. Some exceptions were noted: some sponsors requested the sites to show their EHR evidence to support their estimates and some sites reviewed their hospital's EHR data automatically as part of their feasibility evaluation process. Sites that based their patient count estimates on EHR information were regarded as attractive by the sponsors, because the estimates were perceived to be more trustworthy. As a future research topic, it would be interesting to investigate sites who systematically use EHR data in their feasibility evaluations and sites who do not, and to compare their recruitment success.

It does seem that it would be beneficial for sites to be able to demonstrate, also for themselves and for their host organizations, how many patients they can realistically recruit for a trial (Nasser, Grady, and Balke 2011). The use of EHR data also in the feasibility evaluation phase, not simply in the beginning of the recruitment phase, and the analysis of the available patient population based on EHR evidence should be implemented as a routine practice in all clinical trial sites throughout the Nordic countries.

6.5.3 Legislative barriers in the use of EHR data for recruitment and feasibility evaluations

Legislative differences between the Nordic countries and local hospital practices on how investigators can gain access to EHR systems seem to be the greatest challenge in using EHR data in the hospitals by the investigators (Bonomi 2016). This concerns

both EHR searches conducted for recruitment purposes and EHR data reviews for counts of potential trial particpants already in the feasibility evaluations. Predominantly, investigators are only allowed to review the EHR data of their own patients, but in some hospitals, at least during the interviews of this study, the investigators have been able to perform hospital-wide data queries on the EHR system with the help of the hospital's IT department.

Reform of the relevant legislation and/or harmonization of the interpretations of the existing legislation are crucial steps to create clear and unified rules on the secondary use of EHR data for identifying potential trial participants, whether it occurs already in the feasibility evaluation phase or only after the start of recruitment. For feasibility evaluations, one relevant solution for efficient use of patient data for site and/or country identification would be the use of dedicated computer software such as Insite. Such software, capable of providing only aggregated data, would protect the individual patient data but at the same time would allow the researchers to make data-driven decisions for allocation of the trials. In recruitment of trial paricipants, it is more challenging to simultaneously protect the privacy of patient data and to use it for contacting prospective study patients. One solution would be the use of an independent outside actor, for example a biobank as in Finland, or another independent and trustworthy body, to contact the potential trial participants.

Similar legislative barriers have been reported for example in the UK and Canada (Callard et al. 2014; LeBlanc et al. 2013). In some cases, these barriers have been overcome with a consent-to-contact model: the investigators were allowed to filter the pseudonymized patient data in the data lake from the EHR and then received the contact information of those patients in the list who had given their consent to be contacted for research purposes. However, this solution was applied only at one site at a time, and the efficient and publicly accepted use of EHR data for trial participant identification and recruitment would require explicit, enabling legislation (Kalkman et al. 2019) on a national or even EU level. It should also be noted that overcoming legislative barriers would not solve all of the challenges in the secondary use of EHR data: the EHR systems and procedures for securely obtaining EHR data for research purposes should be further developed to allow the adoption of this kind of use (Luis Fernández-Alemán et al. 2013; Powell et al. 2017).

6.5.4 EHR query tools as used by the trial sponsors

Commercially available data-protected EHR query tools enable also trial sponsors to view EHR data as aggregated patient counts obtained from the hospitals' EHR systems (Coorevits et al. 2013). The sponsor representatives participating in the interviews of this study regarded such query tools as very welcome and much needed

to support patient identification, but their use was not seen to have increased as rapidly as expected. Current legislation or the interpretation of the legislation was seen to hinder the use of such tools. In the Nordic countries, it seems that trial sponsors do not routinely use EHR query platforms, as there are not yet enough sites applying these technologies. This opportunity should be facilitated, together with proper data protection and transparency.

6.5.5 EHR query tool functionality

Additional challenges in the secondary use of EHR data for clinical trials are caused by insufficiencies in the functional properties of current query tools. In this study, we assessed one commercially available EHR query tool for identifying potential trial participants in a hospital EHR system. By using the same criteria as in the preceding traditional manual search, the EHR research platform identified 2.7 times more patients than the manual search (5840 vs 2166 patients).

A similar comparison between an EHR search and a manual search was earlier performed by Davila and colleagues: they searched for HIV patients with certain eligibility criteria in the hospital's EHR system in Houston, Texas, USA. The vast majority, 84% (n=1514), of the patients in the cohort of HIV patients were found to meet the inclusion criteria based on the EHR search. After in-person screening, only half of these (52%, n=781) were considered as being eligible according to the criteria specified for the trial in question (Davila et al. 2017). Even if the EHR review identified also patients that were later found to be ineligible, it limited the number of patient records to be reviewed manually, thus reducing the workload (Ateya, Delaney, and Speedie 2016; Davila et al. 2017; Schmickl et al. 2011) and increasing the speed of patient screening.

A comparison between a manual search and EHR screening with a free-text search engine was conducted in 102 patients with chronic obstructive pulmonary disease in Rochester, Minnesota, USA (Schmickl et al. 2011). Compared to the manual search, which identified 18 possibly eligible patients, the EHR search identified 25 patients, with 100% sensitivity, 92% specificity, 100% negative predictive value, and 72% positive predictive value. However, this study sample was rather small, so the results should be interpreted with caution and cannot be generalized to other diseases, trials and settings.

Transforming elibility criteria into search queries

In the current comparison study, 5 out of 8 of the eligibility criteria could be transformed into formal search queries. In a European study with 23 trial protocols evaluated in 24 hospitals around Europe, a median of 55% (38–89%) of the

eligibility criteria could be formalized into a structured query format compatible with EHR research tools (Claerhout et al. 2019). In a study conducted with 228 primary care clinical trial protocols in the UK, 74% of the eligibility criteria presented as elemental statements were considered to be associated with structured data (Ateya, Delaney, and Speedie 2016). Köpcke et al. were able to transcribe 55% of the eligibility criteria into search queries but assessed also the availability of the patient data in the EHR system, which reduced the percentage of the total EHR data usability. In their study with 15 clinical trial protocols including a total of 351 eligibility criteria, they concluded that the total completeness of EHR data for recruitment purposes was 35% (Köpcke et al. 2013).

The transformation of the eligibility criteria into computable queries has been extensively investigated in recent years, and several guides and prototypes have been created to assist researchers to write computable eligibility criteria and to execute them against the EHR data (Claerhout et al. 2019; Zhang et al. 2018).

The search with the EHR research platform in this study contained diagnoses of AF, stroke, TIA and intracranial bleeding as classified with the ICD10 categorization. Even if the criteria were successfully technically transformed into search queries in the EHR research platform, as a result, the search algorithm also included preliminary, unconfirmed or only suspected diagnoses in addition to the confirmed cases. When interpreting such EHR query results in the future, it should be kept in mind that EHR data reflect the data collected during clinical patient care and may not necessarily correspond to all of the needs for their secondary use (Köpcke et al. 2013).

EHR data in structured format

In addition, the information contained in the non-structured EHR data introduced inconsistencies in the search (Weng, Bigger, et al. 2010): the EHR research platform searched only the structured data, thus missing some important information available in the free-text fields. Girardeau et al. had tested the same Insite EHR research platform in several university hospitals in France and Germany, by exploring three clinical trials with a total of 67 eligibility criteria. According to these investigators, 20% of the criteria corresponded to non-structured data and 81% of all criteria could be mapped to terminologies used in the Insite EHR research platform, and 75% could be mapped to local terminologies (Girardeau et al. 2017). EHR-based eligibility screening has been cited to be feasible, if it is able to evaluate more than three out of every four of the patient selection criteria in the EHR (Ateya, Delaney, and Speedie 2016).

AI and NLP techniques can assist in searching for information also in the non-structured data, even though these approaches are not yet widely applied (Lamberti

et al. 2019). When EHR research platforms are developed further, the implementation of these techniques should be considered. Lamberti et al. surveyed the use of AI in clinical trials and reported that out of 402 respondents (pharmaceutical and biotechnology companies) only 34 had used AI for patient identification and recruitment. However, this was still the largest single use of these technologies in clinical trials (Lamberti et al. 2019).

When the EHR search is complemented with AI techniques, this can decrease the subsequent manual workload (Ni, Wright, et al. 2015) and increase even further the efficiency of patient screening (Ni, Kennebeck, et al. 2015). Ni and Kennebeck reported a 90% decrease in the time needed for identification of potential trial participants as well as a 450% increase in efficiency. Also NLP has been reported by some authors to be an accurate tool to extract eligibility criteria from the clinical notes contained in EHR and to automatically pinpoint patients possibly eligible for a clinical trial (Meystre et al. 2019).

Already in its current form, the evaluated EHR research platform was able to identify all patients that were identified manually. Thus, by recognizing the current limitations in their use, current EHR query tools can be used to facilitate and expedite the manual work needed in patient identification from EHR data.

Implications for the future use of EHR in patient recruitment

The secondary use of EHR data in patient recruitment for clinical trials in the Nordic countries should be increased in at least two ways (Figure 7):

1) Investigators should be permitted to have good access to the EHR data of their hospital to make realistic estimates of the numbers of potential trial participants.
2) Trial sponsors should be permitted to review EHR data through EHR query platforms, ideally to obtain an overall picture of the potential trial subjects at the site level, on the national level or even on the Nordic level.

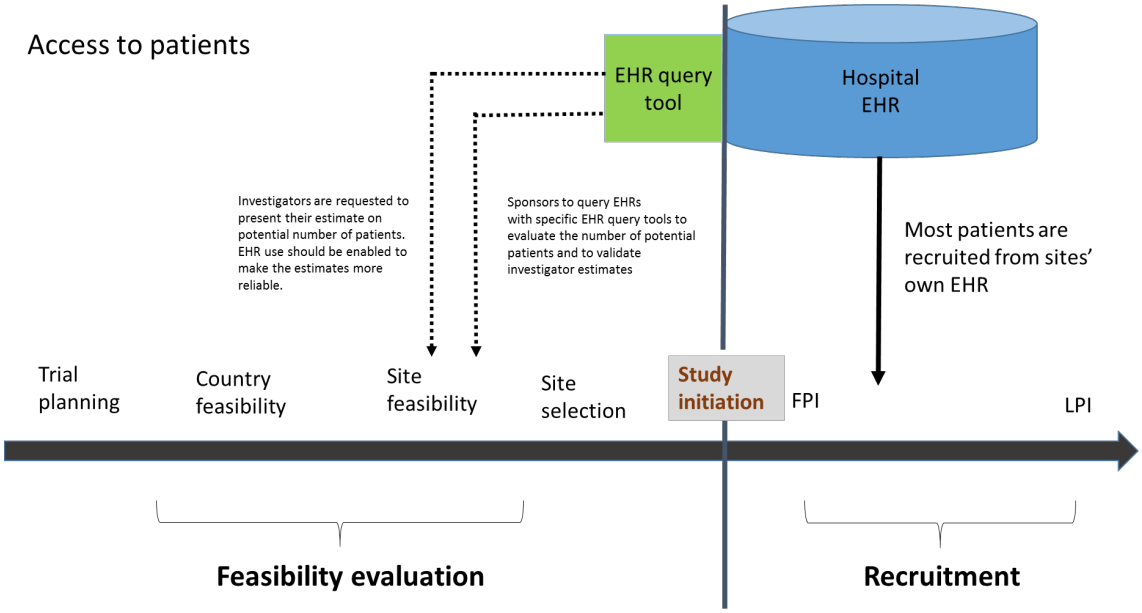


Figure 7. Assessing a key recruitment success factor, "access to patients". Most patients were recruited from the sites' own patients by reviewing the hospitals' EHR systems. Still, they were rarely used in feasibility evaluations. EHR use should be enabled also in the feasibility evaluations for preparing more accurate estimates on potential trial participants. FPI=First patient in, LPI=Last patient in.

Both options require enabling legislative conditions for effective, transparent, and data protected secondary use of EHR data. Legislative barriers were highlighted many times by the participants of this study. As the Nordic countries share many similarities in their cultural features and health care systems, and their economies and living standards are rather similar, the possibilities for conducting clinical trials should also be harmonized, achieving synergy benefits to all Nordic countries.

The ultimate goal of clinical research should always be to benefit the patients. Existing patient data should be used more efficiently for site identification and patient count estimations, to foster the efficient development of new treatments and medicines for the patients and to improve the possibilities of Nordic patients to participate in clinical trials.

Finally, were the EHR data used in trial feasibility evaluations incorrect or misleading, this could lead to wrong decisions. Thus, no matter what data are used, their use is only beneficial as long as they reflect the real life situation and have high quality (Lee 2017). As well as the technical developments required from the system providers, the health care personnel that in the first place record the patient data into the EHR systems have an important role in the future use of EHR data. Their motivation should be promoted and they should be encouraged to adopt straight-

forward and fluent data recording practices. Maintaining and developing good coverage in all EHR data, with high quality, is one of the most critical requirements for EHR secondary use, and it can be viewed as a positive factor for the Nordic countries when they are competing for inclusion in future clinical trials.

6.6 Strenghts and limitations

Patient recruitment and feasibility evaluations

The investigated sample contained data on 34 clinical trials, which can be regarded a small number for justifying any generalization of the quantitative results. Therefore, the numerical estimates of the recruitment success rates and the level of EHR use should remain tentative and cautious. Still, the amount of data obtained in the interviews was extensive and provided an in-depth analysis of the current situation as viewed by the participating industry representatives. After 11 interviews, the study findings started to develop and after 21 interviews, it seemed that the topics had been covered broadly enough to be able to derive reliable conclusions. The selected qualitative approach was the only possible method for the identification of recruitment success and failure factors, and to investigate the conduct of feasibility evaluations and the role of EHR.

Even if the trial selection criteria were defined *a priori*, it was up to the respondents to select which of their trials would be discussed in the interviews. This may be viewed as a possible limitation of the study. To minimize this limitation, all respondents were instructed in the same way by one researcher.

Inductive content analysis (I, II) was applied (Elo and Kyngäs 2008), i.e. the data analysis was not performed according to any pre-determined categories derived from the literature, as there was no established theoretical framework available for presenting the trial sponsors' views. Using the research questions (recruitment success/failure) as an initial coding frame may have slightly limited the inductive approach of the analysis.

Reflexivity (Tong, Sainsbury, and Craig 2007) was considered when evaluating the trustworthiness of the findings; the researchers' background and own perceptions may impact on the results and cause bias. This was acknowledged and the interviews were conducted by following a pre-formulated interview guide in a strictly similar manner with all respondents. In addition, the analysis results were examined in a critical manner by the other authors.

The recruitment success factors (I) were requested to be identified from trials perceived as successful by the respondents. Similarly, the respondents were asked to describe failure factors from trials regarded as unsuccessful in their recruitment of participants. The question on the recruitment failure factors was not evident in the

interview guide, which should perhaps have been revised about this at the outset. However, this had probably no or only minor effects on the results as the interviews were conducted in a systematic manner, in the same way for all interviewees, requesting them to list the failure factors for each trial that was perceived by him/her as having recruited unsuccessfully.

The factors influencing to recruitment (I) was investigated from both directions: recruitment success factors in successfully recruiting trials and failure factors in trials that failed in their recruitment. Nonetheless, when viewing the situation from both directions, the same four main categories were recognized, which strengthens the view that the key recruitment factors were indeed identified in this qualitative investigation.

EHR Research platform assessment

The analysis method for assessing the EHR research platform (III) was descriptive in nature. The reference study included patients with suspected medical emergencies requiring emergency care. In a different reference study conducted with patients not in need of urgent medical treatment or without unconfirmed diagnoses, the number of suspected diagnoses would probably have had less influence on the patient counts.

The transcription of the patient selection criteria into searchable items in the EHR Research Platform was undertaken by one researcher and into CDW queries by another researcher. In spite of that, similar patient counts were reached, which strengthened the validity of the criteria formalization process.

6.6.1 Suggestions for further research

The findings of this study provide the following suggestions for further research:

- To quantitatively identify how great a problem the clinical trial patient recruitment is in the Nordic countries, and whether phase I and II trials recruit patients better than reported elsewhere.
- To investigate whether there are differences in the recruitment success between sites who use EHR in their feasibility evaluations compared to sites who do not.
- To assess whether the EHR Research platform can identify patients with chronic diseases or other stable conditions more accurately than those with medical emergencies.

7 Summary/Conclusions

The results of the current interview-based studies and EHR search experiment indicate that:

- Most clinical trials recruited their participants from the sites' own patients by reviewing the hospital's EHR data when the site started to recruit. Reviews of the EHR data were much less frequently conducted during the feasibility evaluation phase. The utilization of EHR data was limited by legislative and other barriers.
- The patient recruitment success and failure factors in the Nordic countries are similar to those reported earlier in other regions. Examples of these are a complex protocol, investigator motivation, access to patiens and the patients' willingness to join trials. Previous collaboration between the trial sponsor and the trial sites does not as such guarantee recruitment success, because recruitment success or failure is multifactorial.
- The recruitment success and failure factors are interconnected. The industry representatives well recognized their own role in contributing to the success or failure of recruitment.
- The sites using EHR data were considered attractive and more reliable by the sponsors: they could promptly justify their estimates of potential trial subjects, which together with their earlier recruitment performance offered them a clear advantage over other sites.
- The sponsors did not commonly use EHR query tools for evaluating potential patient counts in the Nordic countries because of restrictive local legislation and lack of Nordic sites' data in the tested EHR search tools.
- Secondary use of EHR data in clinical trials, both in the recruitment phase but especially already in the feasibility evaluation phase, is limited by concerns on data protection and regulated by national legislation on the secondary use of patient data. These should be priority issues in all Nordic countries.

- The Nordic countries are facing challenges in being selected into clinical trials because of their limited visibility (due to low volumes of clinical trials) in site performance databases. The increasing need to identify certain types of patients more precisely will offer new opportunities for visibility, which the Nordic countries could supply with their comprehensive EHR data.
- The use of EHR data for clinical research seems to yield more patients in comparison to manual searches of patient records. An understanding of how the structured digital data represent and reflect the eligibility criteria presented in the trial protocols is a driver for improvement. Despite their current limitations, the utilization of EHR databases appears to be a feasible means to identify potentially eligible patients for clinical trials. Continuous improvements in the EHR systems' technical accuracy and their data quality will be needed to enhance the successful use of EHR data in future clinical trials.

Acknowledgements

This study was carried out while the author was a Ph.D. student at the Institute of Biomedicine of the University of Turku during 2015–2021.

I express my sincere gratitude to Professors Emeriti Risto Huupponen and Mika Scheinin who found my suggestion for a research topic interesting and important, and devoted their time to guide me in my studies. I had previous collaboration with both of them already in 2001, in my first position at Turku Bioscience Center, at the University of Turku, and I am very grateful that our paths crossed again.

I wish to express my deepest gratitude to my principal supervisor Risto Huupponen for all support and advice he has given me and for being available whenever I had questions and challenges in my work. I am also very grateful to my supervisor Mika Scheinin for sharing his enormous knowledge of clinical trials and for his valuable comments meant to develop my skills in conducting research and in academic writing. I feel privileged to learn from their experience. I am also indebted to my third supervisor, Adjunct Professor Juuso Blomster, who introduced me to working with electronic health record (EHR) search tools. He encouraged and advised me whenever needed. I am grateful for his belief in my ability to contribute to the ambitious goal of testing of the EHR search tool. My supervisors' support and valuable comments have made this work possible.

My warmest thanks also go to Professor Päivi Rautava and Adjunct Professor Mia Bengtström for their guidance as members of my Ph.D. steering group.

My best thanks go to Professor Miia Turpeinen and Adjunct Professor Tarja Laitinen for their professional and smooth pre-examination of the thesis. Their constructive comments helped me to further improve this work.

I wish to express my sincere thanks to all co-authors of the original publications: to Associate Professor Anna Axelin from the Department of Nursing Science, University of Turku, for her valuable support in qualitative methodologies and for challenging me in the qualitative analyses; to Mia Bengtström from Pharma Industry Finland for ongoing support and inspiration and professional contacts with the Nordic clinical research experts; Mr. Juha-Matti Varjonen, Database System Specialist at Turku University Hospital, Mrs. Minna Blomster, BioBank Data Director at Auria Clinical Informatics, and Adjunct Professor Antti Palomäki,

Research Manager Tuija Vasankari and Professor Juhani Airaksinen at the Heart Centre of Turku University Hospital for collaborating in the research.

I also wish to thank Mr. Tom Southerington (University of Turku) for his advice regarding legislation on the secondary use of EHR data, Ms. Eliisa Löyttyniemi (University of Turku) for her advice on statistical aspects and Dr. Ewen MacDonald for checking the English language of this thesis.

I express my special thanks to those pharmaceutical industry representatives in Finland, Sweden, Norway and Denmark who volunteered to be interviewed for the study.

This work was financially supported by the Drug Research Doctoral Programme of the University of Turku, and I wish to thank the Director and the Coordinator of the Programme, Professor Ullamari Pesonen and Dr. Eeva Valve, for their support and encouragement. Grants from the Finnish Pharmaceutical Industry Research Foundation and Turku University Hospital were important to be able to cover the publication costs of the original articles in open-access journals.

I wish to thank my first managers, Ms. Ulla Varhiala (Turku Bioscience Center, University of Turku) and Dr. Tuija Keinonen (Oy MedFiles Ltd.) for introducing me into the field of clinical research. After 20 years in this business, I still have strong passion for working with clinical drug trials. I am indebted to my current employer TFS Health Science and my managers Mesfin James, Simone Völkle and Viviana Castagna for their encouragement and interest and for providing me the opportunity to conduct Ph.D. studies as part-time and full-time student.

I also wish to thank all my friends, Anna, Teija, Minna, Maaret, Henna, Sanna, Riia and Terhi, and all my relatives, especially my cousin Heli and her family, for their sincere interest and support during my studies. I owe my warm thanks to my mother Liisa and my brother Niko and his family. They have always supported me and I have always been able to count on them. My memories go to my deceased father Markku, who would have been very proud of me had he seen the completion of my thesis. I also thank my mother-in-law Leena for taking care of the children and for letting me concentrate on writing. Finally, my warmest thanks go to my family, my husband Jari and our children Maija and Mikko. Your patience, understanding and encouragement made this work possible.

Merimasku, October 2021
Niina Laaksonen

References

Adams, Mary, Louise Caffrey, and Christopher McKevitt. 2015. Barriers and Opportunities for Enhancing Patient Recruitment and Retention in Clinical Research: Findings from an Interview Study in an NHS Academic Health Science Centre. *Health Research Policy and Systems* 13(1):8. doi: 10.1186/1478-4505-13-8.

Adamson, Joy, Catherine E Hewitt, and David J Torgerson. 2015. Producing Better Evidence on How to Improve Randomised Controlled Trials. *BMJ (Clinical research ed.)* 351: h4923. http://www.ncbi.nlm.nih.gov/pubmed/26408289 (August 30, 2019).

Ali, Sheraz, Oluwaseun Egunsola, ZaheerUd Din Babar, and SyedShahzad Hasan. 2019. Clinical Trials in Asia: A World Health Organization Database Study. *Perspectives in Clinical Research* 10(3): 121. doi: 10.4103/picr.PICR_109_18.

Atal, Ignacio, Ludovic Trinquart, Raphaël Porcher, and Philippe Ravaud. 2015. Differential Globalization of Industry-and Non-Industry-Sponsored Clinical Trials. *Plos One* 10(12). https://doi.org/10.1371/journal.pone.0145122.

Ateya, Mohammad B, Brendan C Delaney, and Stuart M Speedie. 2016. The Value of Structured Data Elements from Electronic Health Records for Identifying Subjects for Primary Care Clinical Trials. *BMC Medical Informatics & Decision Making* 16:1. doi: 10.1186/s12911-016-0239-x.

Auria Clinical Informatics. 2018. Data Refinery Book. https://github.com/auriadih/data-refinery-book/blob/master/ACI_book-2018-09-24.pdf (April 15, 2020).

Avenell, Alison et al. 2004. The Effects of an Open Design on Trial Participant Recruitment, Compliance and Retention – a Randomized Controlled Trial Comparison with a Blinded, Placebo-Controlled Design. *Clinical Trials* 1(6): 490–98.

Barnard, Katharine D, Louise Dent, and Andrew Cook. 2010. A Systematic Review of Models to Predict Recruitment to Multicentre Clinical Trials. *BMC Medical Research Methodology* 10(1): 63. https://doi.org/10.1186/1471-2288-10-63.

Beck, J. Thaddeus et al. 2020. Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center. *JCO Clinical Cancer Informatics* (4): 50–59. doi: 10.1200/CCI.19.00079.

Bengtström, Mia, and Leena Nybond. Clinical Trials in the Nordic and Baltic Countries - Scanbalt. Scanbalt web page. https://scanbalt.org/scanbalt-news/clinical-trials-nordic-baltic-countries/ (February 25, 2021).

Bentley, Colene et al. 2019. Conducting Clinical Trials—Costs, Impacts, and the Value of Clinical Trials Networks: A Scoping Review. *Clinical Trials* 16(2): 183–93.

Beresniak, Ariel et al. 2015. Improving Performance of Clinical Research: Development and Interest of Electronic Health Records. *BioMed Research International.* vol. 2015, Article ID 864549. https://doi.org/10.1155/2015/864549.

Beskow, Laura M., Kathleen M. Brelsford, and Catherine M. Hammack. 2019. Patient Perspectives on Use of Electronic Health Records for Research Recruitment. *BMC Medical Research Methodology* 19(1):42. doi: 10.1186/s12874-019-0686-z.

Blatch-Jones, Amanda et al. 2020. Using Digital Tools in the Recruitment and Retention in Randomised Controlled Trials: Survey of UK Clinical Trial Units and a Qualitative Study. *Trials* 21(1): 304.

van den Bogert, Cornelis A. et al. 2017. Recruitment Failure and Futility Were the Most Common Reasons for Discontinuation of Clinical Drug Trials. Results of a Nationwide Inception Cohort Study in the Netherlands. *Journal of Clinical Epidemiology* 88: 140–47.

Bonomi, Sabrina. 2016. The Electronic Health Record: A Comparison of Some European Countries. Lecture Notes in Information Systems and Organisation, Springer Heidelberg, 33–50. https://link.springer.com/chapter/10.1007/978-3-319-28907-6_3 (April 4, 2021).

Bower, Peter, Sue Wilson, and Nigel Mathers. 2007. Short Report: How Often Do UK Primary Care Trials Face Recruitment Delays? *Family Practice* 24(6): 601–3.

Bremond-Gignac, Dominique, Elisabeth Lewandowski, and Henri Copin. 2015. Contribution of Electronic Medical Records to the Management of Rare Diseases. *BioMed Research International* 2015. 954283. doi: 10.1155/2015/954283.

Briel, Matthias et al. 2016. A Systematic Review of Discontinued Trials Suggested That Most Reasons for Recruitment Failure Were Preventable. *Journal of Clinical Epidemiology* 80: 8–15.

Briel, Matthias, Benjamin Speich, Erik Von Elm, and Viktoria Gloy. 2019. Comparison of Randomized Controlled Trials Discontinued or Revised for Poor Recruitment and Completed Trials with the Same Research Question: A Matched Qualitative Study. *Trials* 20(1): 800.

Bruhn, Hanne et al. 2019. Estimating Site Performance (ESP): Can Trial Managers Predict Recruitment Success at Trial Sites? An Exploratory Study. *Trials* 20(1). https://doi.org/10.1186/s13063-019-3287-6 (May 22, 2019).

Butler, Alex et al. 2018. The Data Gap in the EHR for Clinical Research Eligibility Screening. *AMIA Joint Summits on Translational Science* 2017: 320–29.

Calaprice-Whitty, Denise et al. 2020. Improving Clinical Trial Participant Prescreening With Artificial Intelligence (AI): A Comparison of the Results of AI-Assisted vs Standard Methods in 3 Oncology Trials. *Therapeutic Innovation and Regulatory Science* 54(1): 69–74.

Caldwell, Patrina H.Y., Sana Hamilton, Alvin Tan, and Jonathan C Craig. 2010. Strategies for Increasing Recruitment to Randomised Controlled Trials: Systematic Review. *PLoS Medicine* 7(11). www.plosmedicine.org (May 22, 2019).

Callard, Felicity et al. 2014. Developing a New Model for Patient Recruitment in Mental Health Services: A Cohort Study Using Electronic Health Records. *BMJ Open* 4(12).

Campbell, M. K. et al. 2007. Recruitment to Randomised Trials: Strategies for Trial Enrolment and Participation Study. The STEPS Study. *Health Technology Assessment* 11(48): iii.

Cardozo, Eden, William J. Meurer, Barbara L. Smith, and Jolie C. Holschen. 2010. Utility of an Automated Notification System for Recruitment of Research Subjects. *Emergency Medicine Journal* 27(10): 786–87. https://emj.bmj.com/content/27/10/786 (January 19, 2021).

Carlisle, Benjamin, Jonathan Kimmelman, Tim Ramsay, and Nathalie MacKinnon. 2015. Unsuccessful Trial Accrual and Human Subjects Protections: An Empirical Analysis of Recently Closed Trials. *Clinical Trials* 12(1): 77–83.

CenterWatch. 2009. *State of the Clinical Trials Industry 2009: A Sourcebook of Charts and Statistics*. https://www.amazon.com/State-Clinical-Trials-Industry-CenterWatch/dp/1930624549 (April 3, 2021).

Cheng, Steven K., Mary S. Dietrich, and David M. Dilts. 2010. A Sense of Urgency: Evaluating the Link between Clinical Trial Development Time and the Accrual Performance of Cancer Therapy Evaluation Program (NCI-CTEP) Sponsored Studies. *Clinical Cancer Research* 16(22): 5557–63. https://clincancerres.aacrjournals.org/content/16/22/5557 (September 15, 2020).

Chondrogiannis, Efthymios et al. 2017. A Novel Semantic Representation for Eligibility Criteria in Clinical Trials. *Journal of Biomedical Informatics* 69: 10–23.

Claerhout, Brecht et al. 2019. Federated Electronic Health Records Research Technology to Support Clinical Trial Protocol Optimization: Evidence from EHR4CR and the InSite Platform. *Journal of Biomedical Informatics* 90. https://doi.org/10.1016/j.jbi.2018.12.004 (April 29, 2019).

Clement, Clare et al. 2019. Challenges to and Facilitators of Recruitment to an Alzheimer's Disease Clinical Trial: A Qualitative Interview Study. *Journal of Alzheimer's Disease* 69(4): 1067–75.

Clinical Research News Online. 2019. Right-Sizing Site Selection: Predicting Enrollment. https://www.clinicalresearchnewsonline.com/news/2019/06/25.

Coorevits, Pascal. et al. 2013. Electronic Health Records: New Opportunities for Clinical Research. *Journal of Internal Medicine* 274(6): 547–60. http://doi.wiley.com/10.1111/joim.12119 (January 3, 2020).

Cowie, Martin R. et al. 2017. Electronic Health Records to Facilitate Clinical Research. *Clinical Research in Cardiology* 106(1):1-9. doi: 10.1007/s00392-016-1025-6

Davila, Jessica A. et al. 2017. Feasibility of Identifying out of Care HIV-Positive Patients in a Hospital Setting and Enrolling Them in a Retention Intervention. *HIV Clinical Trials* 18(2): 75–82.

Denburg, Michelle R. et al. 2019. Using Electronic Health Record Data to Rapidly Identify Children with Glomerular Disease for Clinical Research. *Journal of the American Society of Nephrology* 30(12): 2427–35. https://jasn.asnjournals.org/content/30/12/2427 (June 30, 2020).

Dew, Alexander et al. 2013. Recruitment Strategy Cost and Impact on Minority Accrual to a Breast Cancer Prevention Trial. *Clinical Trials* 10(2): 292–99.

DiCicco-Bloom, Barbara, and Benjamin F Crabtree. 2006. The Qualitative Research Interview. *Medical Education* 40(4): 314–21.

Dickson, Sylvia et al. 2013. Reflecting on the Methodological Challenges of Recruiting to a United Kingdom-Wide, Multi-Centre, Randomised Controlled Trial in Gynaecology Outpatient Settings. *Trials* 14(1):389. doi: 10.1186/1745-6215-14-389.

DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen. 2016. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *Journal of Health Economics* 47: 20–33.

DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski. 2003. The Price of Innovation: New Estimates of Drug Development Costs. *Journal of Health Economics* 22(2): 151–85.

Djurisic, Snezana et al. 2017. Barriers to the Conduct of Randomised Clinical Trials within All Disease Areas. *Trials* 18(1):360. doi: 10.1186/s13063-017-2099-9.

Dombeck, Carrie B. et al. 2020. Continued Investigator Engagement: Reasons Principal Investigators Conduct Multiple FDA-Regulated Drug Trials. *Contemporary Clinical Trials Communications* 17: 100502.

Dombernowsky, Tilde, Merete Haedersdal, Ulrik Lassen, and Simon F. Thomsen. 2017. Clinical Trial Allocation in Multinational Pharmaceutical Companies – a Qualitative Study on Influential Factors. *Pharmacology Research and Perspectives* 5(3): 1–14.

Dombernowsky, Tilde, Merete Haedersdal, Ulrik Lassen, and Simon Francis Thomsen. 2019. Criteria for Site Selection in Industry-Sponsored Clinical Trials: A Survey among Decision-Makers in Biopharmaceutical Companies and Clinical Research Organizations. *Trials* 20(1): 708. https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3790-9 (August 19, 2020).

Dombernowsky, Tilde, Merete Hædersdal, Ulrik Lassen, and Simon Francis Thomsen. 2017. Development in the Number of Clinical Trial Applications in Western Europe from 2007 to 2015: Retrospective Study of Data from National Competent Authorities. *BMJ Open* 7(7).

Doods, Justin et al. 2014. A European Inventory of Common Electronic Health Record Data Elements for Clinical Trial Feasibility. *Trials* 15: 18. https://www-ncbi-nlm-nih-gov.ezproxy.utu.fi/.

Doods, Justin, Kirstin Holzapfel, Martin Dugas, and Fleur Fritz. 2013. Development of Best Practice Principles for Simplifying Eligibility Criteria. *Studies in Health Technology and Informatics.* 192:1153.

Drain, Paul K., Robert A. Parker, Marion Robine, and King K. Holmes. 2018. Global Migration of Clinical Research during the Era of Trial Registration. *PLoS ONE* 13(2) doi: 10.1371/journal.pone.0192413. (February 24, 2021).

Dugas, Martin et al. 2010. Routine Data from Hospital Information Systems Can Support Patient Recruitment for Clinical Studies. *Clinical Trials* 7(2): 183–89.

Duley, Lelia et al. 2008. Specific Barriers to the Conduct of Randomized Trials. *Clinical Trials* 5(1): 40–48. http://ctj.sagepub.com (September 4, 2019). doi: 10.1177/1740774507087704

Edwards, Zoe, Michael I. Bennett, Duncan Petty, and Alison Blenkinsopp. 2019. Evaluating Recruitment Methods of Patients with Advanced Cancer: A Pragmatic Opportunistic Comparison. *International Journal of Pharmacy Practice* 27(6): 536–44.

Effoe, Valery S. et al. 2016. The Use of Electronic Medical Records for Recruitment in Clinical Trials: Findings from the Lifestyle Intervention for Treatment of Diabetes Trial. *Trials* 17(1): 496. https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-016-1631-7 (January 19, 2021).

Elo, Satu, and Helvi Kyngäs. 2008. The Qualitative Content Analysis Process. *Journal of Advanced Nursing* 62(1): 107–15.

EMA. 2010. Reflection Paper on Expectations for Electronic Source Data and Data Transcribed to Electronic Data Collection Tools in Clinical Trials. www.ema.europa.eu (January 18, 2021).

Emilsson, L. et al. 2015. Review of 103 Swedish Healthcare Quality Registries. *Journal of Internal Medicine* 277(1): 94–136.

European Commission. 2012. Proposal for a Regulation of the European Parliament and of the Council on Clinical Trials on Medicinal Products for Human Use, and Repealing Directive 2001/20/EC. https://eur-lex.europa.eu (February 24, 2021).

European Medicines Agency. 1998. ICH E8 General Considerations for Clinical Studies, European Medicines Agency. https://www.ema.europa.eu/en/ich-e8-general-considerations-clinical-studies (February 1, 2021).

Evans, R. S. 2016. Electronic Health Records: Then, Now, and in the Future. *Yearbook of medical informatics* (Suppl 1): S48–61.

Evans, R. Scott, James F. Lloyd, and Lee A. Pierce. 2012. Clinical Use of an Enterprise Data Warehouse. AMIA Annual Symposium proceedings 189–98. /pmc/articles/PMC3540441/ (April 1, 2021).

El Fadly, Abden Naji et al. 2011. Integrating Clinical Research with the Healthcare Enterprise: From the RE-USE Project to the EHR4CR Platform. *Journal of Biomedical Informatics* 44(SUPPL. 1): S94–102. http://dx.doi.org/10.1016/j.jbi.2011.07.007.

FDA. 2013. Guidance for Industry Electronic Source Data in Clinical Investigations. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm (January 18, 2021).

FDA. 2018. Use of Electronic Health Record Data in Clinical Investigations Guidance for Industry. https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htmand/orhttps://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htmand/or (April 29, 2019).

Fletcher, Ben et al. 2012. Improving the Recruitment Activity of Clinicians in Randomised Controlled Trials: A Systematic Review. *BMJ Open* 2(1): 496.

Fogel, David B. 2018. Factors Associated with Clinical Trials That Fail and Opportunities for Improving the Likelihood of Success: A Review. *Contemporary Clinical Trials Communications* 11: 156–64.

Frampton, Geoff K. et al. 2020. Digital Tools for the Recruitment and Retention of Participants in Randomised Controlled Trials: A Systematic Map. *Trials* 21(1):478.

Franzén, Stefan et al. 2016. Evaluation of the Use of Swedish Integrated Electronic Health Records and Register Health Care Data as Support Clinical Trials in Severe Asthma: The PACEHR Study. *Respiratory Research* 17(1):152. doi: 10.1186/s12931-016-0461-1.

Gehring, Marta et al. 2013. Factors Influencing Clinical Trial Site Selection in Europe: The Survey of Attitudes towards Trial Sites in Europe (the SAT-EU Study). BMJ *Open* 3(11): 2957. http://www.sbg-marcom.ch/sat-eu/ (November 16, 2018).

Getz, Kenneth. 2011. Predicting Successful Site Performance. Applied Clinical Trials Online. https://www.appliedclinicaltrialsonline.com/view/predicting-successful-site-performance (January 15, 2021).

Getz, Kenneth. 2014. Improving Protocol Design Feasibility to Drive Drug Development Economics and Performance. *International Journal of Environmental Research and Public Health* 11(5): 5069–80.

Gulden, Christian et al. 2019.Extraction and Prevalence of Structured Data Elements in Free-Text Clinical Trial Eligibility Criteria. *Studies in Health Technology and Informatics*, IOS Press, 258:226–30.

Hanson, Laura C et al. 2014. Strategies to Support Recruitment of Patients with Life-Limiting Illness for Research: The Palliative Care Research Cooperative Group. *Journal of pain and symptom management* 48(6): 1021–30.

Harper, Beth et al. 2017. Tufts CSDD-GoBalto Site Selection Market Research (START II). https://www.oracle.com/a/ocom/docs/dc/gobalto-start-mini-reports-lpd100790097.pdf (August 19, 2020).

Harrer, Stefan, Pratik Shah, Bhavna Antony, and Jianying Hu. 2019. Artificial Intelligence for Clinical Trial Design. *Trends in Pharmacological Sciences* 40(8): 577–91.

Harris, Paul A. et al. 2009. Research Electronic Data Capture (REDCap)-A Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support. *Journal of Biomedical Informatics* 42(2): 377–81.

Häyrinen, Kristiina, Kaija Saranto, and Pirkko Nykänen. 2008. Definition, Structure, Content, Use and Impacts of Electronic Health Records: A Review of the Research Literature. *International Journal of Medical Informatics* 77(5): 291–304.

Herasevich, Vitaly, Matthew S. Pieper, Juan Pulido, and Ognjen Gajic. 2011. Enrollment into a Time Sensitive Clinical Study in the Critical Care Setting: Results from Computerized Septic Shock Sniffer Implementation. *Journal of the American Medical Informatics Association* 18(5): 639–44. https://pubmed-ncbi-nlm-nih-gov.ezproxy.utu.fi/21508415/ (December 9, 2020).

Hjorth, Martin et al. 1996. Physicians' Attitudes Toward Clinical Trials and Their Relationship to Patient Accrual in a Nordic Multicenter Study on Myeloma. *Controlled Clinical Trials* 17(5):372-86. doi: 10.1016/s0197-2456(96)00022-0.

Huang, Grant D. et al. 2018. Clinical Trials Recruitment Planning: A Proposed Framework from the Clinical Trials Transformation Initiative. *Contemporary Clinical Trials* 66: 74–79.

Hunninghake, Donald B., Charles A. Darby, and Jeffrey L. Probstfield. 1987. Recruitment Experience in Clinical Trials: Literature Summary and Annotated Bibliography. *Controlled Clinical Trials* 8(4 SUPPL. 1): 6–30.

Isaksson, Eva et al. 2019. Identifying Important Barriers to Recruitment of Patients in Randomised Clinical Studies Using a Questionnaire for Study Personnel. *Trials* 20(1): 618.

Jha, Ashish K et al. 2008. The Use of Health Information Technology in Seven Nations. *International journal of medical informatics* 77(12): 848–54.

Johnson, Elizabeth J., Barbara L. Niles, and Deanna L. Mori. 2015. Targeted Recruitment of Adults with Type 2 Diabetes for a Physical Activity Intervention. *Diabetes Spectrum* 28(2): 99–105. https://spectrum.diabetesjournals.org/content/28/2/99 (January 19, 2021).

Jones, Jennifer M. et al. 2007. Identifying Motivations and Barriers to Patient Participation in Clinical Trials. *Journal of Cancer Education* 21(4): 237–42.

Jonnalagadda, Siddhartha R. et al. 2017. Text Mining of the Electronic Health Record: An Information Extraction Approach for Automated Identification and Subphenotyping of HFpEF Patients for Clinical Trials. *Journal of Cardiovascular Translational Research* 10(3): 313–21.

Kadam, Rashmi Ashish et al. 2016. Challenges in Recruitment and Retention of Clinical Trial Subjects. *Perspectives in Clinical Research* 7(3): 137–43. doi: 10.4103/2229-3485.184820.

Kadam, Rashmi Ashish et al. 2016. Challenges in Recruitment and Retention of Clinical Trial Subjects. *Perspectives in Clinical Research* 7(3): 137. doi: 10.4103/2229-3485.184820.

Kalkman, Shona et al. 2019. Patients' and Public Views and Attitudes towards the Sharing of Health Data for Research: A Narrative Review of the Empirical Evidence Clinical Ethics. *Journal of Medical Ethics* 0: 1–11. http://dx.doi.org/10.1136/medethics-2019-105651 (May 28, 2020).

Kaló, Zoltán et al. 2014. Contribution of Clinical Trials to Gross Domestic Product in Hungary. *Croatian Medical Journal* 55(5): 446–51.

Kalra, Dipak. et al. 2017. The European Institute for Innovation through Health Data. *Learning Health Systems* 1(1): e10008. http://doi.wiley.com/10.1002/lrh2.10008 (February 27, 2020).

Kasenda, Benjamin et al. 2014. Prevalence, Characteristics, and Publication of Discontinued Randomized Trials. *Journal of the American Medical Association* 311(10): 1045–51.

Kearney, Anna et al. 2018. Development of an Online Resource for Recruitment Research in Clinical Trials to Organise and Map Current Literature. *Clinical Trials* 15(6): 533–42.

Kim, Daijin, Steven Labkoff, and Samuel H. Holliday. 2008. Opportunities for Electronic Health Record Data to Support Business Functions in the Pharmaceutical Industry-A Case Study from Pfizer, Inc. *Journal of the American Medical Informatics Association* 15(5): 581–84.

Kitterman, Darlene R., Steven K. Cheng, David M. Dilts, and Eric S. Orwoll. 2011. The Prevalence and Economic Impact of Low-Enrolling Clinical Studies at an Academic Medical Center. *Academic Medicine* 86(11): 1360–66.

Köpcke, Felix et al. 2013. Evaluation of Data Completeness in the Electronic Health Record for the Purpose of Patient Recruitment into Clinical Trials: A Retrospective Analysis of Element Presence. *BMC medical informatics and decision making* 13: 37.

Lai, Yan See, and Janyne Dawn Afseth. 2019. A Review of the Impact of Utilising Electronic Medical Records for Clinical Research Recruitment. *Clinical Trials* 16(2):194-203.

Lamberti, Mary Jo et al. 2012. Evaluating the Impact of Patient Recruitment and Retention Practices. *Drug Information Journal* 46(5): 573–80.

Lamberti, Mary Jo et al. 2018. Assessing Study Start-up Practices, Performance, and Perceptions Among Sponsors and Contract Research Organizations. *Therapeutic Innovation and Regulatory Science* 52(5): 572–78.

Lamberti, Mary Jo et al. 2019. A Study on the Application and Use of Artificial Intelligence to Support Drug Development. *Clinical Therapeutics* 41(8): 1414–26.

Lamberti, Mary Jo et al. 2020. Benchmarking Patient Recruitment and Retention Practices. *Therapeutic Innovation and Regulatory Science* 1: 3.

Lamberti, Mary Jo, Cindy Brothers, Dan Manak, and Ken Getz. 2013. Benchmarking the Study Initiation Process. *Therapeutic Innovation and Regulatory Science* 47(1): 101–9.

Leather, David A. et al. 2020. Real-World Data and Randomised Controlled Trials: The Salford Lung Study. *Advances in Therapy* 37(3): 977–97.

LeBlanc, Jodi et al. 2013. Impact of a Permission to Contact (PTC) Platform on Biobank Enrollment and Efficiency. *Biopreservation and Biobanking* 11(3): 144–48.

Lee, Lisa M. 2017. Ethics and Subsequent Use of Electronic Health Record Data. *Journal of Biomedical Informatics* 71: 143–46. http://dx.doi.org/10.1016/j.jbi.2017.05.022 (May 18, 2019).

Lehnbom, Elin C., Andrew J. Mclachlan, and Jo Anne E. Brien. 2013. A Qualitative Study of Swedes' Opinions about Shared Electronic Health Records. *Studies in Health Technology and Informatics*, IOS Press, 192:3–7.

Levett, Kate M., Christine L. Roberts, Judy M. Simpson, and Jonathan M. Morris. 2014. Site-Specific Predictors of Successful Recruitment to a Perinatal Clinical Trial. *Clinical Trials* 11(5): 584–89.

Luis Fernández-Alemán, José, Inmaculada Carrión Señor, Pedro Ángel Oliver Lozoya, and Ambrosio Toval. 2013. Methodological Review Security and Privacy in Electronic Health Records: A Systematic Literature Review. *Journal of Biomedical Informatics* 46(3):541-62.

Majeed, Azeem, Josip Car, and Aziz Sheikh. 2008. Accuracy and Completeness of Electronic Patient Records in Primary Care. *Family Practice* 25(4): 213–14.

Maret-Ouda, John, Wenjing Tao, Karl Wahlin, and Jesper Lagergren. 2017. Nordic Registry-Based Cohort Studies: Possibilities and Pitfalls When Combining Nordic Registry Data. *Scandinavian Journal of Public Health* 45(17_suppl): 14–19.

Mc Cord, Kimberly A. et al. 2019. Current Use and Costs of Electronic Health Records for Clinical Trial Research: A Descriptive Study. *CMAJ Open* 7(1): E23–32.

McAnulty, John D. 2009. Bringing Patient Recruitment into Our Digital World. *Drug Information Journal* 43(4): 501–8. http://link.springer.com/10.1177/009286150904300416 (November 24, 2020).

McCowan, Colin et al. 2015. Using Electronic Health Records to Support Clinical Trials: A Report on Stakeholder Engagement for EHR4CR. *BioMed Research International* 2015: 707891.

McDonald, Alison M et al. 2006. What Influences Recruitment to Randomised Controlled Trials? A Review of Trials Funded by Two UK Funding Agencies. *Trials* 7(1): 9.

Medidata. 2020. Rare Disease Clinical Development: Novel Approaches to Overcoming Operational Challenges. https://www.medidata.com/en/rare-disease-clinical-development-white-paper-reg

Meystre, Stéphane M. et al. 2019. Automatic Trial Eligibility Surveillance Based on Unstructured Clinical Data. *International Journal of Medical Informatics* 129: 13–19.

MHRA. 2015. MHRA Position Statement and Guidance Electronic Health Records. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470228/Electronic_Health_Records_MHRA_Position_Statement.pdf (January 18, 2021).

Miller, Hailey N et al. 2020. Use of Electronic Recruitment Methods in a Clinical Trial of Adults with Gout. *Clinical Trials* 18(1):92-103.

Miotto, Riccardo, and Chunhua Weng. 2015. Case-Based Reasoning Using Electronic Health Records Efficiently Identifies Eligible Patients for Clinical Trials. *Journal of the American Medical Informatics Association* 22(e1): e141–50.

De Moor, Georges et al. 2015. Using Electronic Health Records for Clinical Research: The Case of the EHR4CR Project. *Journal of Biomedical Informatics* 53: 162–73.

Morrison, Zoe et al. 2014. National Evaluation of the Benefits and Risks of Greater Structuring and Coding of the Electronic Health Record: Exploratory Qualitative Investigation. *Journal of the American Medical Informatics Association* 21(3): 492–500.

Nasser, Nariman, Deborah Grady, and C. William Balke. 2011. Commentary: Improving Participant Recruitment in Clinical and Translational Research. *Academic Medicine* 86(11): 1334–35.

Ni, Yizhao, Stephanie Kennebeck, et al. 2015. Automated Clinical Trial Eligibility Prescreening: Increasing the Efficiency of Patient Identification for Clinical Trials in the Emergency Department. *Journal of the American Medical Informatics Association* 22(1): 166–78.

Ni, Yizhao, Jordan Wright, et al. 2015. Increasing the Efficiency of Trial-Patient Matching: Automated Clinical Trial Eligibility Pre-Screening for Pediatric Oncology Patients. *BMC Medical Informatics and Decision Making* 15:28. doi: 10.1186/s12911-015-0149-3.

Nordforsk. 2019. A Vision of a Nordic Secure Digital Infrastructure for Health Data: The Nordic Commons. www.nordforsk.org (November 26, 2020).

Nordo, Amy Harris et al. 2017. A Comparative Effectiveness Study of ESource Used for Data Capture for a Clinical Research Registry. *International journal of medical informatics* 103: 89–94.

Nordo, Amy Harris et al.. 2019. Use of EHRs Data for Clinical Research: Historical Progress and Current Applications. *Learning Health Systems* 3(1): e10076.

Palomäki, Antti. 2017. *Real-Life Challenges of Stroke Prevention in Patients with Atrial Fibrillation – The FibStroke Study*. Annales Universitatis Turkuensis D 1227. http://urn.fi/URN:ISBN:978-951-29-6775-9.

Patrão, Diogo F.C., Michel Oleynik, Felipe Massicano, and Ariane Morassi Sasso. 2015. Recruit-An Ontology Based Information Retrieval System for Clinical Trials Recruitment. *Studies in Health Technology and Informatics*, IOS Press, 216:534–38.

Peckham, Emily et al. 2018. Successful Recruitment to Trials: Findings from the SCIMITAR+ Trial. *Trials* 19(1):53. https://doi.org/10.1186/s13063-018-2460-7

Penberthy, Lynne, Richard Brown, Federico Puma, and Bassam Dahman. 2010. Automated Matching Software for Clinical Trials Eligibility: Measuring Efficiency and Flexibility. *Contemporary Clinical Trials* 31(3): 207–17.

Pharmaintelligence Informa. 2019. Pharma R&D Annual Review 2019. https://pharmaintelligence.informa.com (February 24, 2021).

PhRMA. 2019. Annual Membership Survey. https://www.phrma.org (November 17, 2020).

Ping, Siu et al. 2008. Effectiveness of Recruitment in Clinical Trials: An Analysis of Methods Used in a Trial for Irritable Bowel Syndrome Patients. *Contemporary Clinical Trials* 29(2): 241–51.

PMDA. 2015. Conformance Guide on Electronic Study Data Submissions. http://www.pmda.go.jp/ (January 18, 2021).

Powell, G. A. et al. 2017. Using Routinely Recorded Data in the UK to Assess Outcomes in a Randomised Controlled Trial: The Trials of Access. *Trials* 18(1):389.

Prescott, R. J. et al. 1999. Factors That Limit the Quality, Number and Progress of Randomised Controlled Trials. *Health Technology Assessment* 3(20):1-143.

Preston, Nancy J et al. 2016. Strategies Designed to Help Healthcare Professionals to Recruit Participants to Research Studies. *Cochrane Database of Systematic Reviews* (2) Art. No.: MR000036. doi: 10.1002/14651858.

Raman, Sudha R et al. 2018. Leveraging Electronic Health Records for Clinical Research. *American Heart Journal* 202: 13–19.

Roberts, J, S Waddy, and P Kaufmann. 2012. Recruitment and Retention Monitoring: Facilitating the Mission of the National Institute of Neurological Disorders and Stroke (NINDS). *Journal of vascular and interventional neurology* 5(supp): 14–19.

Rollman, Bruce L., Gary S. Fischer, Fang Zhu, and Bea Herbeck Belnap. 2008. Comparison of Electronic Physician Prompts versus Waitroom Case-Finding on Clinical Trial Enrollment. *Journal of General Internal Medicine* 23(4): 447–50.

Rosas, Scott R. et al. 2011. Evaluating Research and Impact: A Bibliometric Analysis of Research by the NIH/NIAID HIV/AIDS Clinical Trials Networks. *PLoS ONE* 6(3).

Safran, Charles et al. 2007. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association* 14(1): 1–9.

Schmickl, Christopher N. et al. 2011. The Accuracy and Efficiency of Electronic Screening for Recruitment into a Clinical Trial on COPD. *Respiratory Medicine* 105(10): 1501–6.

Schroen, Anneke T. et al. 2010. Preliminary Evaluation of Factors Associated with Premature Trial Closure and Feasibility of Accrual Benchmarks in Phase III Oncology Trials. *Clinical Trials* 7(4): 312–21.

Schroy, Paul C. et al. 2009. A Cost-Effectiveness Analysis of Subject Recruitment Strategies in the HIPAA Era: Results from a Colorectal Cancer Screening Adherence Trial. *Clinical Trials* 6(6): 597–609. http://journals.sagepub.com/doi/10.1177/1740774509346703 (January 19, 2021).

Scorr Marketing. 2017. *Global Trends Survey Report*. https://cdn.sanity.io/files.

Sertkaya, Aylin, Hui-Hsing Wong, Amber Jessup, and Trinidad Beleche. 2016. Key Cost Drivers of Pharmaceutical Clinical Trials in the United States. *Clinical Trials* 13(2): 117–26.

Shivade, Chaitanya et al. 2014. A Review of Approaches to Identifying Patient Phenotype Cohorts Using Electronic Health Records. *Journal of the American Medical Informatics Association*. 20(2): 221-230.

Song, Seung Yeon, Deborah Chee, and Eunyoung Kim. 2019. Strategic Inclusion of Regions in Multiregional Clinical Trials. *Clinical Trials* 16(1): 98–105.

Spaar, Anne et al. 2009. Recruitment Barriers in a Randomized Controlled Trial from the Physicians' Perspective-A Postal Survey. *BMC Medical Research Methodology.* 9(14). https://doi.org/10.1186/1471-2288-9-14 (April 3, 2021).

Speich, Benjamin et al. 2018. Systematic Review on Costs and Resource Use of Randomized Clinical Trials Shows a Lack of Transparent and Comprehensive Data. *Journal of Clinical Epidemiology* 96: 1–11. http://www.jclinepi.com/article/S089543561730851X/fulltext (January 18, 2021).

van Staa, Tjeerd Pieter et al. 2014. The Opportunities and Challenges of Pragmatic Point-of-Care Randomised Trials Using Routinely Collected Electronic Records: Evaluations of Two Exemplar Trials. *Health Technology Assessment* 18(43): 1–146.

Stensland, Kristian D. et al. 2014. Adult Cancer Clinical Trials That Fail to Complete: An Epidemic? *JNCI: Journal of the National Cancer Institute* 106(9): 229. https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/dju229 (February 8, 2021).

Strong, Sean et al. 2016. 'The Trial Is Owned by the Team, Not by an Individual': A Qualitative Study Exploring the Role of Teamwork in Recruitment to Randomised Controlled Trials in Surgical Oncology. *Trials* 17(1): 1–9. https://link.springer.com/articles/10.1186/s13063-016-1341-1 (June 25, 2020).

Sully, Ben G.O., Steven A. Julious, and Jon Nicholl. 2013. A Reinvestigation of Recruitment to Randomised, Controlled, Multicenter Trials: A Review of Trials Funded by Two UK Funding Agencies. *Trials* 14(1): 166.

Sung, Nancy S. et al. 2003. Central Challenges Facing the National Clinical Research Enterprise. *JAMA* 289(10): 1278.

Tan, Meng H., Matthew Thomas, and Mark P. MacEachern. 2015. Using Registries to Recruit Subjects for Clinical Trials. *Contemporary Clinical Trials* 41: 31–38.

Taylor, David. 2015. The Pharmaceutical Industry and the Future of Drug Development. *Pharmaceuticals in the Environment*. Royal Society of Chemistry. doi: 10.1039/9781782622345-00001 (February 25, 2021).

Team, Victoria, Lyndal Bugeja, and Carolina D Weller. 2018. Barriers and Facilitators to Participant Recruitment to Randomised Controlled Trials: A Qualitative Perspective. *International Wound Journal* 15(6): 929–42. http://doi.wiley.com/10.1111/iwj.12950 (April 3, 2021).

Tenhunen, Olli, Miia Turpeinen, and Pekka Kurki. 2017. Kliinisen Lääketutkimuksen Uudet Tutkimusasetelmat. *Duodecim* 133(6): 599–605. https://www.duodecimlehti.fi/duo13612 (February 1, 2021).

Thacker, Tapan, Ashley R. Wegele, and Sarah Pirio Richardson. 2016. Utility of Electronic Medical Record for Recruitment in Clinical Research: From Rare to Common Disease. *Movement Disorders Clinical Practice* 3(5): 507–9. http://doi.wiley.com/10.1002/mdc3.12318 (July 1, 2020).

Thadani, Samir R. et al. 2009. Electronic Screening Improves Efficiency in Clinical Trial Recruitment. *Journal of the American Medical Informatics Association* 16(6): 869–73.

Treweek, Shaun et al. 2013. Methods to Improve Recruitment to Randomised Controlled Trials: Cochrane Systematic Review and Meta-Analysis. *BMJ Open* 3(2): e002360. doi: 10.1136/bmjopen-2012-002360

Treweek, Shaun et al. 2018. Strategies to Improve Recruitment to Randomised Trials. *Cochrane Database of Systematic Reviews*. 2(2):MR000013. doi: 10.1002/14651858.MR000013.pub6.

U.S. Food and Drug Administration. 2018. Step 3: Clinical Research. https://www.fda.gov (March 31, 2021).

Van Velthoven, Michelle Helena et al. 2016. Feasibility of Extracting Data from Electronic Medical Records for Research: An International Comparative Study. *BMC Medical Informatics and Decision Making* 16(1): 90. doi: 10.1186/s12911-016-0332-1.

Vikström, Antti et al. 2019. Secondary Use of Electronic Health Records: Availability Aspects in Two Nordic Countries. *Health Informatics Management Journal* 48(3): 144–51.

Walters, Stephen J et al. 2017. Recruitment and Retention of Participants in Randomised Controlled Trials: A Review of Trials Funded and Published by the United Kingdom Health Technology Assessment Programme. *BMJ Open* 7: 15276. http://www.isrctn.com/.

Weiner, Mark G., and Peter J. Embi. 2009. Toward Reuse of Clinical Data for Research and Quality Improvement: The End of the Beginning? *Annals of Internal Medicine* 151(5): 359–60.

Weiskopf, Nicole G, George Hripcsak, Sushmita Swaminathan, and Chunhua Weng. 2013. Defining and Measuring Completeness of Electronic Health Records for Secondary Use. *Journal of Biomedical Informatics*. 46(5): 830-836.

Weng, Chunhua, J. Thomas Bigger, et al. 2010. Comparing the Effectiveness of a Clinical Registry and a Clinical Data Warehouse for Supporting Clinical Trial Recruitment: A Case Study. AMIA Symposium 2010: 867–71. www.researchmatch.org/ (November 24, 2020).

Weng, Chunhua et al. 2012. Using EHRs to Integrate Research with Patient Care: Promises and Challenges. *Journal of the American Medical Informatics Association* 19(5): 684–87.

Weng, Chunhua, Samson W. Tu, Ida Sim, and Rachel Richesson. 2010. Formal Representation of Eligibility Criteria: A Literature Review. *Journal of Biomedical Informatics* 43(3): 451–67. http://dx.doi.org/10.1016/j.jbi.2009.12.004.

White, David, and Daniel Hind. 2015. Projection of Participant Recruitment to Primary Care Research: A Qualitative Study. *Trials* 16(473). https://doi.org/10.1186/s13063-015-1002-9.

World Health Organization. 2016a. From Innovation to Implementation. EHealth in the WHO European Region. http://www.euro.who.int/en/ehealth (June 30, 2020).

World Health Organization. 2016b. Global Diffusion of EHealth: Making Universal Health Coverage Achievable: Report of the Third Global Survey on EHealth. https://www.who.int (February 3, 2021).

Zhang, Hansi et al. 2018. Computable Eligibility Criteria through Ontology-Driven Data Access: A Case Study of Hepatitis C Virus Trials. AMIA Symposium 2018: 1601–10.

Zhang, Xiaoxi, and Qi Long. 2012. Modeling and Prediction of Subject Accrual and Event Times in Clinical Trials: A Systematic Review. *Clinical Trials*. 9(6):681-8.

Zhang, Yinsheng, Guoming Zhang, and Qian Shang. 2017. Computer-Aided Clinical Trial Recruitment Based on Domain-Specific Language Translation: A Case Study of Retinopathy of Prematurity. *Journal of Healthcare Engineering*, vol. 2017, Article ID 7862672.

Zimmerman, Lindsay P. et al. 2018. A Novel Patient Recruitment Strategy: Patient Selection Directly from the Community through Linkage to Clinical Data. *Applied Clinical Informatics* 9(1): 114–21.

Appendices

Appendix 1. Interview guide.

Success of patient recruitment in clinical drug trials in Nordic countries - use of electronic data for site identification and patient identification

Thank you for your interest in participating in the interview. It will be done via phone or skype and will contain the following themes. For preparation, please write down your notes in advance. The interview will take max 1 hour.

Please think about 2 of your most important clinical drug trials during 2015-2018 in which you were/are involved and which included one or more Nordic country.

	Trial 1	**Trial 2**
1. Which phase (I-IV) and disease was investigated?		
2. Which countries were involved?		
3. How many patients per Nordic country were planned to be included vs how many were actually included?		
4. Is the patient recruitment completed (or ongoing) according to planned schedule? • If Yes, define key success factors for recruiting patients as planned • If No, how much was the recruitment period prolonged from planned? (in months)		
5. Is the trial completed (or ongoing) according to planned schedule? • If Yes, define key success factors for keeping timelines • If No, how much was the trial completion delayed from planned? (in months)		

6. Which were the major factors causing delays in the trial?		
7. Please compare delays caused by patient recruitment to delays caused by other trial-delaying factors? Answer options: • Patient recruitment was the major trial-delaying factor • Patient recruitment delays had some effect on trial delay • Patient recruitment was a minor trial-delaying factor • Patient recruitment had no effect on trial delays		
8. Please consider trial activities preceding the recruitment period: Was the recruitment started on schedule or were there delays in preceding start-up activities? Why? If start of recruitment period was delayed, was it compensated by efficient recruitment? Yes/No		
9. Have you done protocol feasibility to ensure that the inclusion/exclusion criteria are optimal?		
10. How was the site identification and patient identification done? I.e. From where and how did the study team find out about potential sites and potential patients? Include also comment if Patient organizations were used and whether it was considered useful.		

11. What electronic data (eg. electronic databases, registers, electronic health records or other electronic tools) were used for trial site identification and for patient identification? Are there country-specific differences in using electronic data for these purposes?		
12. Was there benefit from using them? Yes/No • If yes, how much (little/ much/very much) and what kind of benefit? • If No, why?		
These final questions are not restricted to the "2 most important trials" referred to in questions 1.-12. but in general: • What kind of electronic solutions would you regard valuable in future clinical trials when identifying suitable trial sites or trial patients? • In your experience, which types of trials have been most successful in finding suitable patients? Why?		

Operational environment and the role of the respondent, also these will be asked during the interview:

- *What is your job description and history in the company and impact on the site identification and patient recruitment process?*
- *What is your history (in years) in conducting clinical trials?*

All answers will be treated with highest confidentiality. No names of interviewees or their companies will be identifiable from the published results.

Thank you very much for your collaboration!

Appendix 2. Example of the eligibility criteria translated to the computed copy for the EHR Research platform.

1. First diagnosis of i48
2. The following events
   'First diagnosis of i60 or i61 or i62 or i63 or i65 or i66 or i69 or G45 or G46 or S06 happened
   10 years before now' [a, b, c, d, e]
3. NOT: the following events [e]
   'Any diagnosis of G45.4'
4. NOT: the following events [f]
   'First diagnosis of S06 happened at least 1 days before these events: First diagnosis of i48'

[a] *Criteria were separated by "OR", so the query included all patients having Stroke, TIA or Intracranial bleeding.*

[b] *I64 was excluded from the script as it was not searchable as such in InSite. However, I64 is mapped to the closest ancestor, ICD10CM:I60-I6964, so I64 was included in the patient counts.*

[c] *'Before now', referes to the Reference date 31 Dec 2012. All patients having at least one Index event within 10 years preceding 31 Dec 2012 were searched.*

[d] *'First', means the first occurrence date of AF and the Index events to identify patients who have at least one event of both diagnoses.*

[e] *Exclusion criterion no. 4*

[f] *Exclusion criterion no. 1*

Appendix 3. Example of the CDW search: Patients diagnosed with the Index event for the first time (2003-2012) and ever diagnosed with AF (before the end of 2012): patient n = 5840.

```
# ====================================================
# Set variable dgn_list for all diagnosis codes:  Ischemic stroke, TIA and Bleeding
dgn_list   <- c("^I60.[0-9]","^I61.[0-9]","^I62.[0-9]","^I63.[0-9]","^I64.[0-9]","^I65.[0-9]","^I66.[0-9]","^I69.[0-9]","^G45.[0-9]","^G46.[0-9]","^S06.[0-9]")
# ====================================================
# Load diagnoses for above codes Ischemic stroke ja Tia, bleeding
dgn      <- load_diagnoosit_pg(con_pg     = con_pg
                               ,diagnoosi  = dgn_list
                               ,date_end   = as.Date("2012-12-31")
                               ,sql = T
                               )
# Above load_diagnoosit_pg R function generates following SQL script:
SELECT *  from func.get_content(_this_table                  := NULL::stage_uraods.mv_diagnoosi,
                               _criteria_list                := ARRAY[['diagnoosi', ^I60.[0-9]|^I61.[0-9]|^I62.[0-9]|^I63.[0-9]|^I64.[0-9]|^I65.[0-9]|^I66.[0-9]|^I69.[0-9]|^G45.[0-9]|^G46.[0-9]|^S06.[0-9]']],
                                _criteria_list_condition     : = 'and',
                                _hetu_list                    := '',
                                _kohortti_id                  := '',
                                _date_start                   := '',
                                _date_end                     := '2012-12-31',
                                _limit                        := NULL,
                                _sql                          := '1',
                                _pseudonym                    := '0',
                                _encryption                   := 'md5',
                                _salt                         := '',
                                _research_name                := '',
                                _keep_default_research_name   := '0',
                                _source_db_name               := '',
                                _print_result                 := '1',
                                _research_column_list         := '*',
                                    )
# ====================================================
# First diagnosis day of Ischemic stroke, TIA or Bleeding in 2003 - 2012, diagnoses given before 2003 excluded
dgn_min_2003_2012 <- dgn[,.SD[which.min(DGN_PVM)],PERSON_ID][year(DGN_PVM)%in% 2003:2012]
# ====================================================
# Patients diagnosed with stroke, TIA or Bleeding for the first time in 2003-2012
> dgn_min_2003_2012[,list(otsikko = "Aivotapahtuma",pot_n = length(unique(PERSON_ID)))]
       otsikko pot_n
1: Aivotapahtuma 31088# ====================================================
# Set variable for all diagnosis codes: AF
dgn_list_i48   <- c("I48")
# ====================================================
# Load diagnoses for above code AF
dgn_i48      <- load_diagnoosit_pg(con_pg     = con_pg
                                  ,diagnoosi  = dgn_list_i48
                                  ,date_end   = as.Date("2012-12-31")
                                  ,sql = T)
# ====================================================
# Patients diagnosed with stroke, TIA or Bleeding for the first time in 2003-2012 and ever diagnosed with AF (until the end of 2012):
dgn_min_2003_2012_i48 <- dgn_min_2003_2012[PERSON_ID%in% dgn_i48[,.N,PERSON_ID]$PERSON_ID]
dgn_min_2003_2012_i48[,list(otsikko = "Aivotapahtuma + AF",pot_n = length(unique(PERSON_ID)))]
           otsikko pot_n
1: Aivotapahtuma + AF  5840# ============================================
```

Appendix 4. Planned and recruited patients per therapeutic area. Table includes only trials with completed recruitment (n=22).

	Patients planned	**Patients recruited**
Therapeutic area		
Oncology	90	99 (110%)
Neurology	85	104 (122%)
Endocrinology	5590	5346 (96%)
Cardiology	228	156 (68%)
Pulmonology	54	10 (19%)
Other [a]	96	23 (24%)
Total	**6143**	**5738 (93%)**

[a] ”Other” includes trials in psychiatry, nephrology and gastroenterology.

Appendix 5. Main categories and sub-categories formed in feasibility evaluation study (II).

Feasibility/ Site identification			Estimation of number of potential trial participants at sites		
General	**Global level**	**Nordic level**	**Investigators' estimates**	**Sponsors' estimates**	Nordics as viewed by the sponsors
Use of data in feasibility has increased during past history	Site visibility in data is crucial for countries to be selected	Investigator databases not used in Nordics	Most investigators base their patient count estimates on previous experience, with no support from EHR data	Dialog with investigators crucial in feasibility evaluation	Investigators cannot participate in conducting trials due to lack of time
Due to technology and data available, site feasibility process will change in future	Risk of incorrect conclusions due to erroneous data	Local intelligence crucial for site identification	Many investigators lack time for proper feasibility evaluation	Sponsors try to validate investigator evaluations with dialog and various type of data	Nordics are slow starters with good data and advanced technology
	Need of strong marketing by local Nordic pharma hubs based on data	Current site identification through old contacts suggestions of KOLs/investigators	Time invested to EHR review pays back at trial initiation	Sponsors value sites providing EHR data as evidence on potential patient counts	Nordics can niche themselves by their electronic data enabling also complex and rare indication trials
		Informing investigators properly is crucial	Sites presenting EHR evidence are the interesting ones	EHR data platforms for sponsors not used in Nordics	National, data protected EHR query systems preferred
			EHR can be used for pre-screening but also historical data to evaluate potential for recruitment	Sponsors increased need and interest in EHR data to focus the patient estimates	
			Legislative limitations for sites EHR use		

Appendix 6. Examples on quotations, codes and categories from the Feasibility evaluation interviews (II).

Examples on quotations	Code	Sub-category	Category	Main category
"Because it is just not talking about volumes, we are talking about targeted medicine. We are talking about patients with specific mutations. We need to be able to pinpoint where those exact patients are. That is something that is here and will also be more and more in future, I am sure." ID5	The increased need to find specific patients	Increased need for data	Changed needs of data in the feasibility evaluations	**Changing landscape of feasibility evaluations**
"That can be good or bad. Data driven. You can´t argue with data. I mean if you have the correct data, that can be a good tool (for finding sites), but if you don't, you can make some great mistakes." ID14	Risk of incorrect conclusions due to erroneous data	Conclusions from erroneous data	Increased amount of data causing considerations on trust on available data	
"I would say to my colleagues that one should be aggressive (in marketing) and not just accept the thing that we just don't get the trials. It is our obligation as directors in Finland to constantly market (our country), and not just wait whether we get (the trials) or not." ID15	Need for active marketing	Role of Nordic subsidiaries of the pharmaceutical companies	Site identification on global level	**Site identification in two layers**
"No, we don't use any databases because Denmark is a small country. We have had so many studies so we already know them (the sites)." ID20	Databases not used for site identification	Site identification	Site identification in local level	
"And very important is the communication to the potential investigators. Then they know what they are saying "yes" to. That is very important for the feasibility evaluation." ID17	Communication with the investigators	Communication in site identification process		
"It is always when they (trials) are initiated, that is usually the first time when they (investigators) really start to think about the study and where to find the patients.. ..I don't think they take time until studies are initiated." ID1	Investigators estimating patient counts too late	Approximate estimation of patient counts	Investigators evaluating number of potential trial subjects	**Evaluation of sites' access to patients**
"They (the site) were very good in prediction of the study. So we can set up good expectations for Sweden. We did know we have this many patients (at the site)… it is not just a guess, or not even a validated guess… it is a fact that we know, that we have this number of patients." ID9	Trust on investigator estimates because of data evidence	Data evidence provided by the investigators		

"Q: How do you evaluate the investigator's promises to recruit this many patients? A: We take that with caution. They are always more optimistic than the reality is. The physicians are generally very busy; they don't have time to go through all the details in the protocol. That is why we really have to be aware of potential challenges. That is why we perform the feasibility evaluation, anyway. We need to ask very carefully and particularly about those challenges." ID17	Discussions about the challenges	Dialogs with the investigator	Sponsors evaluating the number of potential trial subjects	
"They are requested to do a database search. To make sure they have a database and have looked in the database on how many patients they would recruit." ID14	Request for EHR search	Requesting site's EHR evidence		
"Because you can basically see sites (from the database) that are delivering and not delivering. You are not dependent on what someone else is telling you. You are looking at the data that are in the systems." ID5	Estimating site projections against previous performance data	Investigator databases and previous performance data		
"That said, we also have lot of challenges to ensure data privacy and then also to make the data owners (feel) comfortable how the data are being used. So there are lots of challenges, but also lots of opportunities." ID22	Challenges and opportunities in using EHR data	Use of EHR query tools		
"I don't think some of the hospitals in our region fully understand the competitive environment. I was in South Europe last week, meeting with a few hospitals, and they have done very very well, but they are also very aggressively selling the hospital. We don't do that in the Nordics. Even though we are so much better" ID9.	Competence in clinical trials not marketed	Competence	Competitive factors	**Characteristics of the Nordic countries in feasibility evaluations**
"Well, it is mostly hospital patient record type of data that we need in our trials. It should be structured in a way that enables data searches. And most preferably so that it would be nation-wide (data lake), not one hospital at a time." ID7	Searchable patient data in big data lakes needed	Need for EHR data	Future of the EHR query systems	

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2021

ISBN 978-951-29-8696-5 (PRINT)
ISBN 978-951-29-8697-2 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)