

# VU Research Portal

## Supporting trial recruitment and design by automatically interpreting eligibility criteria

Milian, K.

2014

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Milian, K. (2014). *Supporting trial recruitment and design by automatically interpreting eligibility criteria*.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# SAMENVATTING

Deze dissertatie is een kleine bijdrage aan een groter doel van het verbinden van heterogene medische data bronnen. Omdat klinisch onderzoek vaak door lage participatie vertraagd wordt, was de belangrijkste focus van het gepresenteerde werk op het ondersteunen van patiënten werving en het ontwerpen van toelatingscriteria voor nieuwe onderzoeken. Elk klinisch onderzoek richt zich op het werven van een specifieke soort patiënten, gedefinieerd door een verzameling van criteria. Dit werk is gericht op het automatisch interpreteren van deze criteria voor het ondersteunen van verificatie van patiëntgeschiktheid en het creëren van een database met gerelateerde criteria. We hebben ons gefocust op klinisch onderzoek dat zich richt op borstkanker, maar veel van onze methoden kunnen in verschillende domeinen hergebruikt worden.

Deze dissertatie onderzoekt vier hoofdvragen. De eerste vraag is "In hoeverre kunnen we de inhoud van medische teksten zoals toelatingscriteria van klinisch onderzoek vast leggen?". De tweede vraag gaat een stap verder: "Kunnen we queries genereren uit de vrije tekst van toelatingscriteria om te kunnen beoordelen of een patiënt in aanmerking komt?". Dit soort evaluatie vereist de aanwezigheid van bijbehorende data in patiëntendossiers. Omdat medische applicaties vaak met incomplete data moet werken hebben we onderzoek gedaan naar de impact van dit probleem op onze specifieke taak. De volgende vraag is: "Biedt de evaluatie van incomplete data hulp bij het bepalen van de geschiktheid van een patiënt?". Als laatste hebben we onderzoek gedaan naar de mogelijkheid om een bibliotheek te maken van correlerende criteria en zijn we ingegaan op de laatste vraag: "In hoeverre kunnen we de inhoud van het corpus van toelatingscriteria vastleggen ter ondersteuning voor het bevragen van relevante criteria?" We

benaderden deze vragen door het analyseren van een grote set criteria. We hebben gebruik gemaakt van de geobserveerde specificiteit van hun taal, medische terminologie, normen binnen de gezondheidszorg en semantische redenering om de mogelijkheid van het automatisch evalueren van patiënt geschiktheid voor klinisch onderzoek te onderzoeken. Dit onderzoek leidde tot de volgende bevindingen. Ten eerste demonstreerden we een pijplijn van bewerkingsstappen voor het transformeren van tekstuele toelatingscriteria naar een gestructureerde en uiteindelijk berekenbare vertegenwoordiging waarmee de patiënt geschiktheid automatisch geverifieerd kan worden. De gestructureerde representatie bestaat uit een combinatie van taalpatronen, concepten uit standaard medische terminologieën en genormaliseerde metingen. De evaluatie van de expressiviteit van deze structurerende aanpak, handmatig uitgevoerd met een domein expert, demonstreerden een bijna complete dekking op de beschouwde reeks onderzoeken. Het is een automatisch proces dat van nature uitdagender en interessanter is vanuit een informatica perspectief, omdat het mogelijk zou zijn de oplossing te schalen. Het programma wat patronen detecteert behaalt voldoende prestatie. Alhoewel het ontdekken van de correlatie tussen criteria een grotere uitdaging is en dit aspect verder onderzoek vereist. Vervolgens bestudeerden we de volgende stap in de pijplijn. We demonstreerden hoe de gestructureerde criteria kunnen worden omgezet naar queries. Om de herbruikbaarheid van onze aanpak te verbeteren zijn we ervan uitgegaan dat de data is weergegeven volgens een van de standard data representatie normen (openEHR). Verder hebben we de experimenten beschreven die we hebben gedaan met de patiënten data van de Maastricht kliniek in Nederland. Evaluatie van de prestatie van de query generatie voor een aantal specifieke soorten van tekstuele criteria toonde positieve resultaten, maar ook hier kwam een behoefte naar voren voor verder onderzoek. Een interessante uitkomst van dit experiment was de observatie dat het evalueren van zelfs een incomplete verzameling van criteria de inspanning van het handmatig nazien van de overige criteria aanzienlijk kan verminderen.

Tot slot hebben we ons gericht op het automatisch opbouwen van een bibliotheek van gestructureerde criteria wat semantisch zoeken mogelijk maakt naar relevante criteria en onderzoeken. Om dit te bereiken hebben we het vorige algoritme toegepast. Dit algoritme verwerkt tekstuele criteria

op een corpus van borstkanker onderzoeken. De fractie van automatisch gestructureerde criteria geven aan dat het algoritme voor patroonherkenning een uitbreiding nodig heeft om het toe te kunnen passen op criteria die bestaan uit meerdere patronen. Desalniettemin, de verkregen resultaten lieten ons toe om criteria verder te verwerken. We hebben een methode ontworpen die criteria vergelijkt op hun restrictiviteit en ze op basis hiervan sorteert. Dit algoritme kan gebruikt worden om criteria versoepeling te ondersteunen tijdens de ontwerpfasen van een onderzoek. De daaropvolgende evaluatie van de aanpak vanuit het medisch perspectief moedigde verder onderzoek aan. De resulterende bibliotheek van criteria kan het hergebruik van gestructureerde representaties en het zoeken naar gerelateerde maar bredere criteria en relevante onderzoeken met nauwkeurige queries faciliteren.

Deze dissertatie presenteert een interessante use case die demonstreert hoe de specificiteit van een domein en de gintegreerde resultaten uit veel gebieden gebruikt kunnen worden om relevante methodes aan te leveren voor specifieke gebruikers. Vanuit een biomedisch klinisch onderzoek perspectief kan dit werk potentieel leiden tot efficiëntere werving van patiënten voor klinisch onderzoek, en snellere afronding van onderzoeken en levering van wetenschappelijk bewijs.