Translating Predictive Models for Alzheimer’s Disease to Clinical Practice: User Research, Adoption Opportunities, and Conceptual Design of a Decision Support Tool

Maura Bellio

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Medical Physics and Biomedical Engineering
Centre for Medical Image Computing
University College London

August 16, 2021
I, Maura Bellio, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.
Nothing in life is to be feared, it is only to be understood. 
Now is the time to understand more, so that we may fear less. 

Marie Skłodowska Curie
Abstract

Alzheimer’s Disease (AD) is a common form of Dementia with terrible impact on patients, families, and the healthcare sector. Recent computational advances, such as predictive models, have improved AD data collection and analysis, disclosing the progression pattern of the disease. Whilst clinicians currently rely on a qualitative, experience-led approach to make decisions on patients’ care, the Event-Based Model (EBM) has shown promising results for familial and sporadic AD, making it well positioned to inform clinical decision-making. What proves to be challenging is the translation of computational implementations to clinical applications, due to lack of human factors considerations. The aim of this Ph.D. thesis is to (1) explore barriers and opportunities to the adoption of predictive models for AD in clinical practice; and (2) develop and test the design concept of a tool to enable EBM exploitation by AD clinicians. Following a user-centred design approach, I explored current clinical needs and practices, by means of field observations, interviews, and surveys. I framed the technical-clinical gap, identifying the technical features that were better suited for clinical use, and research-oriented clinicians as the best placed to initially adopt the technology. I designed and tested with clinicians a prototype, icompass, and reviewed it with the technical teams through a series of workshops. This approach fostered a thorough understanding of clinical users’ context and perceptions of the tool’s potential. Furthermore, it provided recommendations to computer scientists pushing forward the models and tool’s development, to enhance user relevance in the future. This thesis is one of the few works addressing a lack of consensus on successful adoption and integration of such innovations to the healthcare environment, from a human factors’ perspective. Future developments should im-
Abstract

prove prototype fidelity, with interleaved clinical testing, refining design, algorithm, and strategies to facilitate the tool’s integration within clinical practice.
Impact Statement

This Ph.D. thesis contributes new knowledge for a high-impact process on the clinical translation of computational tools, particularly on the delivery of predictive models for Alzheimer’s Disease (AD) to clinical use. This work studies the opportunities and design of a Computerised Decision-Support System (CDSS) built on a specific model, the Event-Based Model, as a valuable aid in AD clinical decision-making. It first provides formative understanding of current clinical needs and practices in clinical AD, identifying if, where and in what form a decision-support tool could fit (Chapter 3). Results show how barriers, such as data availability, type and stage of patients, or trust in the algorithm could hamper clinicians’ adoption of such a tool. This can bring more advantages to specialised settings in AD practice, and to support specific steps of the decision-making process. Understanding these conceptual requirements early in the development process provided key insights on the role of CDSS for the AD context and a set of design and adoption recommendations to inform the prototype design of a CDSS. Furthermore, this work evaluates the conceptual design of a CDSS which is built on previously identified requirements (Chapter 4). Testing the prototype with early adopters reinforced existing understanding of the context and purpose of such a tool, its limitations, whilst providing richer feedback on intended use, and on more detailed design aspects. Finally, the clinical testing of the tool helped define the expected impact, and strategies for its integration in the clinical setting (Chapter 5). Whilst this is a project focused on a specific algorithm and one single tool example, it acknowledges a wider challenge of getting innovative computational models adopted in clinical practice. This is exemplified in the framework proposed in Chapter 6, with the intent to provide a
structured approach to researchers and practitioners facing similar challenges. This project was developed as part of EuroPOND [1], an EU Horizon 2020 Project contributing to knowledge on progression of neurodegenerative diseases by developing computational methods to represent and predict their evolution. The work was supported by the industry partner icometrix (https://icometrix.com). They develop methods and tools that promote quantification and better understanding of imaging data for clinical use, with the goal to improve patients’ care. The collaboration with them was key to the clinical application vision and mission, and their expertise in developing clinical tools invaluable for informing the process. Icometrix have also benefited from this project’s outcomes, particularly from unique insights on potential stakeholders and their needs, an initial design concept, and a comprehensive set of requirements that represents the building blocks for their next product development phases. This work was presented in high-impact Conferences and Journals at each step of its development, particularly in the fields of health, behavioural science, and human-computer interaction for digital health. This ensured invaluable feedback and input from the scientific community. On some occasions, the work was jointly presented with the icometrix team at conferences’ exhibitions, maximising the impact of this work on both Academic and Industry representatives.

**Journal papers**


nitive datasets: considerations and challenges’ In *Alzheimer’s & Dementia: Diagnosis, Assessment & Monitoring* (2020) [3]

**J.4** Bellio, M., Oxtoby, N. P, Ribbens, A., Alexander, D. C., Blandford, A. ‘Design and concept testing of icompass: a tool supporting clinicians to better understand the evolution of Alzheimer’s Disease in individuals’ *Submitted to: International Journal of Human-Computer Studies, Special Issue on Advances in Human-Centred Dementia Technology* (2021)

**Conference Abstracts**


**C.4** Bellio, M., Oxtoby, N. P, Ribbens, A., Alexander, D. C., Blandford, A. ‘A framework to translate computational approaches for chronic conditions into a clinical tool: the icompass case’ *CompAge* 2020 – poster

**Talks and Workshops**

**T.1** Bellio, M., Furniss D., Oxtoby, N. P., Garbarino S., Firth, N. C., Ribbens, A., Alexander, D. C., Blandford, A. ‘Delivering disease progression models to
clinical practice: a user study on potential adoption and technology transfer’

*UCL Institute of Healthcare Engineering – Autumn Symposium* 2018 – Early career researchers talk


**Dissertations**

Acknowledgements

When you embark in a long-term journey, such as a Doctorate, you don’t quite know what to expect ahead, and it’s like the discovery steps that make up an exploratory research. Many are the people that have made this journey more exciting, inspiring, bearable, fascinating, worth it.

First of all, I want to thank my multidisciplinary team of Supervisors, who understood earlier and better than me I was the right person for this project. They launched me in this challenge, and followed my steps until the end, with incredible patience and professionalism. Thank you Prof. Daniel Alexander and Dr. Neil Oxtoby for training me on the technical aspects of this project and for reminding me to see the bigger picture of things; thank you Dr. Annemie Ribbens for the time at icometrix and for bringing in the practical application perspective; thank you Prof. Ann Blandford for your guidance on the human factors side of things, for sharing your approach to work and life, and for helping me discover my place in this field.

I had the honour to receive advice from truly inspiring clinicians and researchers, such as Dr. Zuzana Walker and Dr. Keir Yong. Thanks for your enthusiasm in all of our clinical conversations and for your words of encouragement. A big portion of this experience has been being part of the EuroPOND project, a diverse and solid ecosystem of researchers, with whom I have shared discoveries, achievements, brainstorming sessions, and trips. You are too many to name, but all special! Thanks also to Dr. Sara Garbarino, Dr. Dominic Furniss, and Dr. Nicholas Firth, who shared only part of my route. The real fuel of this project, however, is represented by all the clinicians and technical experts that participated in my
studies: thank you for your time and curiosity. You truly helped me shaping the contributions to this work.

My time at UCL has been invaluable, thanks to the friendly people I shared joys and struggles with. I have been lucky to be part of three different teams, that gave me rich feedback and support. Thanks to my CDT colleagues, in particular to Elena, Alessia, and Laura; thanks to my CMIC friends: Rica, Kiko, Pere, Beppe, Ioana, Marzia; thanks to my UCLIC people: Laura, Tao, Dilisha, Roxana, Julia, Niels, Tu, Jeremy, Nikki, Olga. And thanks to my new team outside UCL, Ctrl Group, a bunch of brilliant minds that made my transition into my first job seamless and exciting, in such unprecedented times.

I also want to thank my old and new friends, who cheered me up and distracted me from work, when it was much needed. Special thanks to Elena, Inna, Luca, Anthony, Giulia, Francy, Sara, Serena, Irene; the One HealthTech community; the London Heathside Athletic Club.

And finally, thanks to the Family that I have and the family that I chose. Thank you Tilde, Maurizio, and Alessio for being my lighthouse from afar, my best and first supporters in life. You’ve shaped me into the determined woman I am today. Thank you Filippo, for your love, for being my life companion and my strength. I would not be here without your tireless support.
Contents

List of Abbreviations 25

1 Introduction 27
   1.1 General aims ........................................... 28
   1.2 Thesis outline and contribution ...................... 29

2 Background 31
   2.1 Alzheimer’s Disease ................................. 31
      2.1.1 Description and Impact .......................... 31
      2.1.2 The Care Pathway ................................ 32
      2.1.3 Importance of Early Detection ................... 36
      2.1.4 AD markers ...................................... 38
   2.2 Predictive Models .................................... 42
      2.2.1 Artificial Intelligence Systems for Healthcare .... 42
      2.2.2 Introduction to Predictive Models ............... 42
      2.2.3 Predictive Models for AD ........................ 44
      2.2.4 Shared Resources and Projects .................. 47
   2.3 Clinical Decision-Support Systems .................... 48
      2.3.1 Clinical data visualisation ....................... 50
      2.3.2 Data visualisation in AD ........................ 55
      2.3.3 Interfaces and systems .......................... 55
   2.4 Design and Adoption of CDSS ......................... 58
      2.4.1 Technology adoption and diffusion of innovation .. 59
Contents

F.2 Activity 1: cards ................................................. 293
F.3 Activity 2: cards ................................................. 295
F.4 Information and consent ....................................... 297
F.5 Background survey ............................................ 302

G Requirements ..................................................... 307

H Future work proposal ........................................... 317
  H.1 Suggested Phases ............................................ 317
  H.2 Suggested Future Studies ................................... 320
List of Figures
2.1

Neurofibrillary tangles and senile plaques in a healthy neuron (left)
and a diseased neuron (right). Illustration by Bob Morreale [4]. . . . 32

2.2

NICE Pathways: Dementia assessment and diagnosis [5]. . . . . . . 34

2.3

The seven stages of Alzheimer’s Disease, by Dr. Barry Reisberg of
New York University (source [6]). . . . . . . . . . . . . . . . . . . 37

2.4

Comparison between stages as defined by NIA-AA 2011 framework, IWG 2014, and NIA-AA 2018 relative to cognitive staging.
Adapted from [7]. . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

2.5

Representation of the progression cascade of biomarkers characterising Alzheimer’s evolution in the hypothetical model from Jack et
al. [8]. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44

2.6

Validation of EBM staging based on patient’s diagnostic categories.
The bars represent the proportion of patients classified at that stage.
Colours refer to the diagnostic category: cognitively normal (light
blue), MCI (black), and AD (orange). From Young et al. [9] . . . . 45

2.7

Prediction of conversion. (A) From MCI to AD; (B) from normal
to MCI. The different lines represent groups of patients at stage 0
(blue), stages 1-3 (green), stages 4-5 (orange), stages 6-10 (cyan),
stages 11-14 (magenta). From Young et al. [9] . . . . . . . . . . . . 46

2.8

Outflow example representation of events and temporal progression
pathways [10] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

2.9

Blob chart (A), Bar (B), and Radial (C) visualisation of predicted
diagnoses. From [11]. . . . . . . . . . . . . . . . . . . . . . . . . . 54


2.10 Symptoms related to different types of dementia and their intercon-
nections, represented as pathways. From Rita Maldonado Branco
[12]. ................................................................. 56

2.11 AD progression visualisation, including MMSE scores, symptoms,
and diagnostic labels, on a timeline. From [13]. ............................ 57

2.12 MRI-based visual rating for the assessment of atrophy progression
in AD. Adapted from Harper et al. [14] ................................. 57

2.13 PredictND. The areas of the dashboard include a bar reporting
uploaded input (A), a section reporting the hierarchical decision-
making outcome (B), the expected accuracy of DSI (C), distribution
of an individual biomarker (D), and the influence each measure has
in the classification (E). From [15]. ...................................... 58

2.14 The User-Centred Design Cycle. From Usability.gov [16]. ....... 62

2.15 Methods used in the design-thinking process [17]. .............. 63

2.16 Thesis roadmap. ...................................................... 71

3.1 Chapter 3 content detail from the thesis roadmap. ................. 74

3.2 The input visual stimulus used in the semi-structured interviews
(credits: icometrix early implementation). ................................. 78

3.3 Two examples of the six output visual stimuli used in the semi-
structured interviews (credits: icometrix early implementation). ... 79

3.4 Data collection scheme. .................................................. 80

3.5 Combination of workflow’s representations: individual and MDT.
Barriers are represented by numbers 1 - 4. (1) Availability and
sufficient quality of data; (2) Diagnostic uncertainty; (3) Lack of
outcome from MDT; (4) Uncertainty in prognosis. ........................ 83

3.6 Survey’s design process. ............................................... 93

3.7 Characterisation of sample. (A) Types of different clinical profes-
sionals (Neurologists, Psychiatrists, others), by context (R+C, C-
O); (B) Respondents’ country of work. .................................... 94
List of Figures

3.8 Sum of responses to the question: “What sources of data are more relevant to you in assessing the stage of disease? (Select all that apply)”. Results based on 21 respondents (R+C $n=16$; C-O $n=5$). 95

3.9 Data available to participants, related to 4 groups of markers: (A) cognitive tests, (B) MRI, (C) CSF, (D) Blood or Genetic test. Sum of responses is reported for R+C (dark blue) and C-O (light blue). 98

3.10 Agenda presented at the workshop. 101

3.11 Conceptual map representing the process of discussion during the workshop and exemplifying the relationship between areas and themes. 103

3.12 Affinity diagram generated from Activity 1. Eight items are reported in the boxes (EBM = Event-Based Model; DEM = Differential Equation Model), whilst the clinical need they might address is reported in red. 104

3.13 An approximate representation of the dimensions describing the technical-clinical gap, and the area that the future tool should cover to meet translational opportunities. 111

4.1 Chapter 4: content detail from the thesis roadmap. 116

4.2 Hand sketches for the a) input and b) output screens. 133

4.3 First design iteration of the pathway output tile. 135

4.4 Digital wireframe for input and output screens in its second iteration. 136

4.5 Example of (a) icompass input screen and (b) details call-out for one of the cognitive tests. 138

4.6 Example of icompass output screen. 139

4.7 Details of icompass output screen. 140

4.8 Sankey diagrams representing changes in clinical decision along the three sub-tasks, a) collectively, and for b) neurologists and c) psychiatrists separately. 144

4.9 Scores are related to judgements on a 5-point Likert scale for questions linked to the task. 145
List of Figures

4.10 CASSM analysis of entities and attributes. U = user; I = interface; S = system; s/c = set/change; c/d = create/delete.

4.11 Clinical workflow, representing current practice and possible icompass integration (marked with an orange circle).

5.1 Chapter 5: content detail from the thesis roadmap.

5.2 Example of card used for the misalignment activity.

5.3 Summary of the post-it notes from all groups contributing to Activity 1 on misalignments between clinical mental model and icompass functioning and information provided. Colour codes: Orange = QR; Green = CA; Light blue = MD; Pink = FD.

5.4 Summary of the post-it notes from all groups contributing to Activity 2 on sorting current and future clinical needs based on the two parameters feasibility and relevance. The figure has illustrative purposes only, to visually communicate the contribution of different groups by looking at the colours’ distribution.

6.1 Framework to translate computational, data-driven approaches to clinical tools.

A.1 Flowchart representing example (A) research questions and steps regarding (B) test selection, (C) quality control and standardisation, and (D) computational/statistical methods. Example research questions (A) correspond closely to test selection (B), while subsequent processes outlined in steps C and D are broadly relevant across questions and tests. Example measures are reported in italics and section headings underlined. Dashed arrows indicate revisiting steps, for example, revisiting test selection owing to missing data [3].
List of Tables

2.1 Design requirements for CDSS, across four publications: Horsky et al., 2012 [18]; Kilsdonk et al., 2017 [19]; Miller et al., 2018 [20]; and Sutton et al., 2020 [21]. .................................................. 66

3.1 Gagnon’s factors organised by Heeks’ dimensions. ...................... 81

3.2 Comparison of content for three user profiles. ............................. 85

3.3 Assessment of Heeks’ dimensions for the two contexts. [*critical score when equal or higher than 5.] ................................. 90

3.4 Table summarising barriers and facilitators related to the exploratory phase of the project. .................................................. 112

4.1 Structure of requirements’ tables. .............................................. 118

4.2 Technical factors: system quality requirements. ......................... 119

4.3 Technical factors: information quality requirements. .................. 120

4.4 Technical factors: information delivery requirements. ................ 122

4.5 User factors: system’s intended use requirements. ..................... 123

4.6 User factors: system understanding requirements. ..................... 124

4.7 Organisation factors: structure requirements. ............................ 125

4.8 Organisation factors: Environment requirements. ....................... 126

4.9 Organisation factors: Interoperability requirements. ................... 126

4.10 List and description of the four case studies derived from the technical experts’ workshop. ................................. 128
4.11 Comparative-competitive analysis of three tools that have developed an interface to display results from predictive models in relation to AD. ........................................... 131
4.12 Clinical agreement with icompass output for the three sub-tasks. . . 143
4.13 Overview of intended use of icompass from clinicians, grouped into areas of advantage. ......................................................... 148
4.14 Overview of surface misfits (U = user; S = system). .................. 153
5.1 Sample and background knowledge description .......................... 163
5.2 Overarching themes across all activities and relative findings. ...... 172
6.1 Table summarising requirements for adoption of CDSS for AD. .... 184
List of Abbreviations

**AD**  Alzheimer’s Disease. 19, 27, 31–33, 35–37, 39, 40, 45, 47, 48, 55, 56, 69, 73, 75, 83, 100, 127, 137, 140, 161, 181, 182, 184–186, 188, 193, 196, 201–203

**AI**  Artificial Intelligence. 42, 43, 49, 65, 67, 68, 118, 155, 193

**CASSM**  Concept-based Analysis of Surface and Structural Misfits. 60, 143, 150, 151


**CSF**  Cerebrospinal Fluid. 41, 95, 97, 137

**DPM**  Disease Progression Model. 43, 100, 104, 161, 182, 186, 198, 202

**DRGm**  Design-Reality Gap model. 76, 79, 86, 89

**EBM**  Event-Based Model. 44, 46, 52, 73–75, 79, 80, 84, 87–89, 112, 113, 119, 121, 122, 124, 133, 141, 150, 151, 166, 183, 184, 188, 201, 202

**EHR**  Electronic-Health Records. 50, 52, 53, 196

**GDPR**  General Data Protection Regulation. 126, 175, 177

**GP**  General Practitioners. 33–35

**HCI**  Human-Computer Interaction. 67, 68, 196, 202
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IT</strong></td>
<td>Information Technology</td>
<td>59, 60, 65, 112, 154, 185</td>
</tr>
<tr>
<td><strong>IWG</strong></td>
<td>International Working Group</td>
<td>19, 37, 38</td>
</tr>
<tr>
<td><strong>MCI</strong></td>
<td>Mild Cognitive Impairment</td>
<td>19, 36, 38, 41, 45, 83, 98, 111, 137, 146</td>
</tr>
<tr>
<td><strong>MDT</strong></td>
<td>Multi-Disciplinary Team</td>
<td>35, 76, 80, 82, 87, 113, 124, 147, 170, 192</td>
</tr>
<tr>
<td><strong>ML</strong></td>
<td>Machine Learning</td>
<td>43, 49, 67, 118</td>
</tr>
<tr>
<td><strong>MRI</strong></td>
<td>Magnetic Resonance Imaging</td>
<td>35, 40, 47, 55, 88, 95–97, 106, 133, 146, 149, 175</td>
</tr>
<tr>
<td><strong>NIA-AA</strong></td>
<td>National Institute of Aging and Alzheimer’s Association</td>
<td>19, 37, 38, 94</td>
</tr>
<tr>
<td><strong>NICE</strong></td>
<td>National Institute of Health and Care Excellence</td>
<td>33, 35, 40, 205</td>
</tr>
<tr>
<td><strong>UCD</strong></td>
<td>User-Centred Design</td>
<td>61, 64</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Dementia is a global problem with huge significance for individuals, families and healthcare systems, affecting over 46 million people [22]. In Alzheimer’s Disease (AD), the most common form of dementia, earlier stages are widely believed to be the most effective for therapeutic interventions [23]. Various clinical indicators, called markers, are considered reliable detectors for these early stages, and they come from different sources: medical imaging, cognitive tests, analysis of biological fluids, and many others [24]. At the moment, clinical professionals consider this snowstorm of information qualitatively, interpreting their relevance to the disease based on clinical expertise. Quantitative methods offer the potential to combine information more consistently and accurately to deliver better understanding of disease status and development. To this end, computational researchers have developed statistical (or predictive) models, integrating information from large cross-sectional data sets of biomarkers into quantifiable longitudinal pictures of disease development.

Given that the proposed innovation has originated from a computational perspective of understanding the evolution of AD, and the potential clinical utility has emerged from those studies, the key challenge is to bring this research into practice. The ambition is to identify translational opportunities and directions to exploit disease progression models so that it benefits clinicians treating dementia on the front line, as there is no such tool in use at the moment. Not only is it important to frame an approach and technique to facilitate this transition, but also to establish whether
(and under what circumstances) this might be valuable to the end-users.

1.1 General aims

This Ph.D. project aims to address the research need of translating computational disease progression models into clinical practice, through the design of graphical visualisations of these models for specialised AD clinicians, to support their decision-making. A user-centred design approach is adopted, suitable to develop technologies that are based on users’ needs and practice, and consequently more likely to succeed and get adopted. There are many variables that need to be taken into account in a complex socio-technical system when developing and implementing a new tool, such as the context, availability of resources, needs, specialisations of users, to mention some. Understanding the roles and influence of these factors, contributes to frame the gap between technical capability and clinical application, defining opportunities and barriers for the novel technology to be adopted and disseminated. Moreover, defining user-informed system requirements and strategies to visualise disease progression data will promote clinical adoption without disrupting the working routine, although providing the additional benefits disease progression models are designed for. The general objectives identified are:

- To frame the gap between technical capability and clinical application. This implies acquiring an understanding of both the clinical and computational context, in terms of needs, practices, concepts, mental models, barriers, and opportunities;

- To develop an understanding of the design concept of a tool that gives clinicians access to predictive models for Alzheimer’s Disease;

- To create appropriate low and high fidelity design prototypes, enabling users to provide feedback;

- To provide recommendations to model developers on enhancing the utility and relevance of models.
1.2 Thesis outline and contribution

This thesis is structured as follows, and contributions for each chapter are highlighted in italic:

- In Chapter 2, I present the background literature. Various areas of research are touched on, and particularly: AD clinical content, computational research on disease progression models, and user-centred design for decision-support tools. Parts of this chapter have been published in the following: [J.3].

- In Chapter 3, I report results from exploratory research on clinical and technical characteristics. This chapter contributes to the understanding of current clinical needs and practice, definition of different types of context and users, and models’ potential in supporting clinical decision-making. Parts of this Chapter have been published and presented in the following: [J.1, J.2, C.1, C.2, T.1, D.1].

- In Chapter 4, a set of requirements is defined based on results from Chapter 3, that informed the initial design concept for the tool. Based on this, I report on the design thinking process, including various low-fidelity iterations of the prototype, and describing the prototype version representing the first testable tool. I then present a study where the prototype is tested with relevant specialists, which provides further understanding on the value of this concept, but also highlights new opportunities and barriers to consider in the next stage of the tool’s development. Parts of this Chapter have been published and presented in the following: [J.4, C.3].

- In Chapter 5, I report how results from Chapter 4 are addressed by different technical experts in a series of workshops. Results from this study provide actionable guidelines on how to improve the prototype from various perspectives (algorithm, interface design, clinical application, and regulation and quality), that inform a strategy for the optimal implementation of a working system, but also of the tool’s integration in the actual setting. Parts of this Chapter have been presented in [C.4].
• In Chapter 6, I present the overall discussions from this work. Here, I also propose a framework that generalises this work to the wider problem of how to best translate computational models for chronic conditions into tools that are used and adopted by clinicians. This framework aims to represent a guide to facilitate and exemplify the translational process from technical innovation to clinical use, which includes both perspectives at every stage of the development, thus promoting a successful adoption and evaluation of possible limitations. Parts of this Chapter have been published and presented in the following: [J.2, J.3, C.4, W.1].

• In Chapter 7, I summarise the conclusions to this work.
Chapter 2

Background

This chapter provides the background knowledge to the project and to support the experimental chapters. Due to the multidisciplinary nature of the project, the presented topics are quite diverse, but interconnected. First, clinical information on AD and current practice is reported. The second section introduces the reader to computational approaches and resources to advance knowledge on AD and related biomarkers. The third section introduces the reader to the concept of Computerised Decision-Support System (CDSS) and ways in which data visualisation strategies have been integrated in these systems. The fourth section describes the user-centred design approach, generally and in the context of CDSS, with fundamental considerations on the design and adoption of these systems in clinical practice. Finally, the last section summarises the scope and motivation of this thesis, illustrating the manuscript’s roadmap in a flowchart that will act as a signposting throughout the subsequent chapters. Parts of this chapter are included in the perspective paper [J.3] [3].

2.1 Alzheimer’s Disease

2.1.1 Description and Impact

The functional and symptomatic consequences of AD are represented by a plethora of cognitive deficits, among which episodic memory is the most characteristic, although not always the main symptom [25]. Other frequent deficits that might emerge contextually or consequentially are aphasia (language disruption), agnosia
(impaired visual recognition not due to sensory problems), apraxia (disturbance of motor acts not due to deficits of the motor system), attention or executive functions alterations [26]. With the disease progressing in time and severity, other behavioural, cognitive, and neuropsychiatric impairments appear, up to the point when patient’s social and daily life activities are completely disrupted.

Figure 2.1: Neurofibrillary tangles and senile plaques in a healthy neuron (left) and a diseased neuron (right). Illustration by Bob Morreale [4].

AD is at the third position worldwide for medical costs and it currently affects more than 46 million people, a number that is forecast to increment up to 130 million people by 2050 [27]. Whilst causes and treatments are still unknown, focusing on care pathways could uncover gaps in the disease management and how this process could be enhanced towards better care and decision-making.

2.1.2 The Care Pathway

The uncertainty characterising AD has unavoidable effects on care pathways. Some examples are a heterogeneous, at times controversial, approach from clinical professionals and different centres, unpredictability along the disease course, and various professional figures intervening on the clinical route [28]. Given this project’s aim to enhance AD clinical practice, it is worth looking at what is currently done based on literature and regulations. A “care pathway” is defined by the European Pathway Association as “a complex intervention for the mutual decision making and
organisation of care processes for a well-defined group of patients during a well-defined period” [29]. Each country adopts different strategies to tackle this issue, such as a stronger focus on primary care in Australia [30], compared to more emphasis on research dedicated to early detection of the disease in UK [28]. One of the most urgent strategies is a timely identification of AD, to allow best practice for interventions, but also to support patients and their families [31, 32]. Another one is a person-centred approach, stressed in various global discussions, encouraging assistance systems to consider each time the spectrum of co-existing clinical conditions and patients’ preferences in care choices [33, 22, 34]. However, this might cause an even more segmented experience. An additional strategy looked into the development of an Integrated Care Pathway, meaning a coordinated interaction and integration of various services, according to the patients’ needs and situation [28]. AD specifically is a candidate for a phase-oriented care pathway, as opposed to a fixed-time care. This means that the care pathway is structured in a way that allows a continuous, flexible, and responsive service, throughout the patient’s condition [22]. To facilitate an understanding of the disease steps to the general audience, the British National Institute of Health and Care Excellence (NICE) regularly develops flowcharts, capturing the key-points of dementia care pathways [5] at various stages and considering all the figures involved (see Figure 2.2). Key points along the disease journey represent an analytic support in clarifying the condition’s issues and requirements. Four main steps from Figure 2.2 should be pointed out: (1) Early identification of symptoms and first contact with non-specialised services (see box 1); (2) Assessment procedure (box 3); (3) Diagnosis (box 4); (4) Follow up, support and interventions (boxes 5, 6, 7).

**Early symptoms and first contacts.** The first element starting the process is represented by the patient or caregiver seeking help due to alarming issues. Memory is the most common sign of probable AD [32]. However, it often gets misdiagnosed as normal ageing, and recognised only through other triggering events, such as hospitalisation, life threatening events, or emergencies [35]. The professional figure contacted at this stage is the General Practitioners (GP) [28]. This initial step faces
a number of disruptive barriers, such as inadequate awareness of the disease, low referral rate to higher specialisation, and reluctance in seeking help.

**Assessment procedure.** At this stage, various investigations are conducted to draw a picture of the clinical condition. The types and relevance of these indicators will be thoroughly presented at a later point. It is important to gather information from different sources and to integrate this data in a consistent way. This phase also includes investigations for possible comorbidities, meaning clinical conditions that are likely to be present contiguously with the principal one, and differential diagnoses, that involves excluding conditions that cannot co-exist. The professional figure leading this stage is the GP, with the role of referring patients to the correct facility. The principal ones are memory clinics, specialised centres for the investigation and treatment of memory disorders and related medical conditions [36]. These centres are led by a multidisciplinary team, generally including (but not limited to)
neurologists, psychiatrists, nurses, geriatricians, psychologists, and healthcare specialists. In England, the “Map of Medicine - dementia pathway” is a tool for GPs and other primary care professionals to facilitate them in planning patients’ journey. However, most of the times such guidelines are not taken into account [37]. Another barrier at this stage is a poor understanding of patients’ experience: this causes anxiety and confusion, that inevitably affects the clinical status and the relationship with caregivers and professionals [28].

**Diagnosis.** The diagnosis includes both the medical work of integrating and interpreting collected evidence from the patient, and the communication between patients and professionals. Regarding the first, a careful examination of clinical history, functional impairment, and cognitive assessment contributes to an informed diagnosis. Cases are often discussed by a team of experts in meetings called Multi-Disciplinary Team (MDT) meetings. Of interesting importance is the one around Magnetic Resonance Imaging (MRI), given the impact that imaging and the associated biomarkers are gaining. It is usually attended by 10 to 15 different specialists (e.g. neuroradiologists, neurologists, nurses, psychologists, students), and chaired by a neuroradiologist consultant [38]. For each patient discussed, the neurologist introduces the case, then the neuroradiologist comments on the MRI scan. This stimulates the debate amongst all participants, until a consensus on the patient diagnosis and/or next medical steps is reached. MDT meetings have been proven to benefit clinical outcomes, multidisciplinary collective thinking, and evidence-based decisions [38], so they are included into the clinical practice guidelines stated by NICE [39]. However, the discussion is characterised by wide use of qualitative and approximate observations [38].

AD is an inclusive diagnosis and follows standardised criteria that are regularly revised [40, 22]. However, data reports that amongst all AD patients, only half of them are correctly diagnosed and treated [41], probably due to the gate-keeper role of GPs, less prone to make a diagnosis compared to neurologists, or due to denial from families [42]. At this point, the patient might be referred to specialised clinics and assessments. Communicating such a long-term condition to the patient is itself
a critical step and studies have reported on the impact on families and caregivers [43, 28].

Follow up, support, interventions. The following stages are closely dependent on the specific diagnosis and the eventual co-existing medical conditions. If patients are referred to higher specialised centres, they undergo further examinations, to decide upon treatment options and follow-up plans. Many are the cases of uncertain diagnosis, that gets classified as MCI and where no further support is provided. What is generally expected at this stage is prescribing medications to alleviate some of the symptoms, although psychosocial and community support proved to be fundamental integration to a pharmacological intervention [43].

The integrated approach provided by care pathways presents both benefits and risks. Benefits are represented by improved care consistency, quality, responsiveness, and efficiency. This is due to well-defined guidelines, quality controls, resource planning, and team coordination [22]. On the other hand, though, a too mechanical and generalised approach might shadow the interest for the single person, their feelings and expectations. What has been proven as representing an advantage in all the steps of the care pathways is the early detection of the disease [44]. The next section will present current definitions of AD stages and why early detection is believed to be so crucial at the moment, in research and clinical practice.

2.1.3 Importance of Early Detection

Alzheimer’s Disease is a progressive, long-term condition characterised by insidious onset. This implies that it is difficult to mark the start even from visible symptoms only. Clinically relevant alerts are represented by patient’s complaints or family feedback. The first signs affect language, word finding, or naming. With time, daily tasks are increasingly affected by visuospatial impairments, preventing them from moving easily around a familiar environment. Along with functional disabilities, personality and behaviour are affected over the years, leading to irritability, delusions, hallucinations, loss of social skills, until the final stages, with limb and verbal unresponsiveness, incontinence, and compromised immune system
2.1. Alzheimer’s Disease

The duration of the whole course varies individually, but might last between 10 to 20 years [45].

Figure 2.3: The seven stages of Alzheimer’s Disease, by Dr. Barry Reisberg of New York University (source [6]).

To provide a frame of reference, different staging criteria have been proposed. The most conventional in both clinical practice and trials are: the seven stages of Reisberg [46], the neuropathological stages of Braak and Braak [47], and the frameworks from the International Working Group (IWG) and National Institute of Aging and Alzheimer’s Association (NIA-AA), a less fine-grained but more commonly used staging system. Reisberg’s stages (Figure 2.3) were defined in 1999 [46], to depict the AD continuum. Each stage is described in terms of functional, self-care, and daily activities changes, from normal ageing to the most severe AD. In Braak & Braak’s staging system [47], the clinical course of the disease is divided into six histopathological stages. The extracellular distribution of neurofibrillary tangles and neuropil threads enabled the identification of a progressive pattern of neurodegeneration. However, this study does not match biology with clinical evidence, thus limiting its applicability in research. NIA-AA is a less fine-grained pathological classification, but more familiar to clinical professionals and the general population. In this system (Figure 2.4), the AD continuum is divided into four clinical stages,
the last three being levels of severity for the declared pathology (mild, moderate, severe). At the left end of the disease course, a stage defined as Mild Cognitive Impairment (MCI) covers a category of patients that do not satisfy AD criteria, but present with memory complaints [48]. These patients are prone to develop the disease with a probability of 50% to 70% in the following 5 years [42]. Sperling et al. [32] describe AD course starting with a decade-long asymptomatic phase, involving biological and histological changes, but no clinical evidence. From this stage, the disease can evolve into a clinical phase, showing initial signs of subjective cognitive impairments or MCI, until AD criteria are met.

![Figure 2.4: Comparison between stages as defined by NIA-AA 2011 framework, IWG 2014, and NIA-AA 2018 relative to cognitive staging. Adapted from [7].](image)

The current primary goal is an early detection of Alzheimer’s and related risk factors [25, 32]. Preclinical stages seem to be more sensitive to treatments that can slow down or, in the future, possibly reverse the course of the disease [32]. Moreover, early detection can support patients to plan and educate themselves and families, while still able to understand the disease’s implications [42].

### 2.1.4 AD markers

To better understand how AD can be identified early, when other symptoms are not visible, various disease indicators, called markers, are of key importance in clinical practice and research. Markers can be classified in clinical, imaging, and biochemical [25, 49]. Their effective potential is given by combinations of these indicators, and longitudinal monitoring of the disease course [49]. In fact, they
can now be combined to make an in-vivo diagnosis of typical AD, whilst only an ex-vivo brain analysis could once confirm the pathology [25]. The most relevant markers to this project will now be described and contextualised.

**Clinical markers.** Clinical markers from cognitive tests are used near-ubiquitously to understand the impact of neurodegenerative disease on patients. Standardised cognitive tests aim to measure impairment objectively adjusting for demographic factors, while being relatively cheap, widely available, quick to administer, minimally invasive, and with quantifiable reliability for their use in clinical work [50]. A complete assessment is typically composed of several tasks, each intended to examine a specific function or domain, such as memory, attention, executive function, language, and visuospatial processing. The most widely adopted is the Mini-Mental State Examination (MMSE) [51], covering basic temporal orientation, recall of words and concepts, and basic apraxic abilities. Other batteries, like Alzheimer’s Disease Assessment Scale (ADAS-cog) [52] and Rey Auditory Verbal Learning Test (RAVLT) [53], inspect more in depth critical domains in AD. The most relevant cognitive tests, described by subscales, respective task, scoring system, and target domain measured by that specific subscale is summarised in Appendix A.1 [3].

It is not possible to completely isolate measurements for individual domains, as test performance is multifactorial (e.g. poor memory might be attributable to impaired attention or impaired visual processing, rather than a primary memory deficit). If the purpose is to characterise cognitive domains, tests are better supported in combination with other tests. On the other hand, there is a surge in composite scores derived from batteries of tests in AD research, which might mitigate individual test idiosyncrasies [54, 3]. In their recent review, Schneider & Goldberg [55] identified 12 composite scales that have been used in clinical trials to assess cognitive function. However, cognitive composites are still prone to issues [55], as their performance over individual measures of cognitive domains is not guaranteed, or psychometric properties are often overlooked, since they are normally not designed for longitudinal assessment. Lim et al. [56] mention the importance of
evaluating the sensitivity of each scale contributing to the composite, as it can affect the overall sensitivity of the composite. Moreover, domains relevant to early clinical symptoms of AD are often underrepresented [55]. For example, while episodic memory deficits are one of the most common indicators of preclinical AD, non-memory domains may be susceptible to pathological changes during the preclinical phase [3]. Overall, a cognitive-composite approach might be appropriate in clinical trials and disease progression monitoring [57], but current measures are still substantially limited in their validation and psychometric assumptions [55].

**Imaging markers.** Imaging markers, in particular from Magnetic Resonance Imaging (MRI) are related to identification of brain atrophy and they currently play a substantial role in the early detection of AD. MRI is a non-invasive imaging technique, mainly developed for medical investigations to acquire images and information on the internal structure of the human body. An accurate definition can be derived from words in the acronym: magnetic refers to the source producing the signal, which is a magnet and magnetisation; resonance indicates the method used to excite and detect the signal, which in this case is radio-frequency coil; imaging explains the process of transforming the signal into images [58]. In brief, this technique produces anatomical images from the signal generated by particles in the brain, releasing electromagnetic energy when exposed to a radio-frequency pulse. Images of different sequences (determined by the amount of radio-frequency pulses) are visually characterised by grey-scale colours defining different types of tissues or compounds. The most common sequences for AD are T1-weighted, T2-weighted, and Fluid Attenuated Inversion Recovery (FLAIR).

The reason why MRI is being extensively used in neurodegenerative diseases is due to the non-invasive and in-vivo procedure. Compared to other imaging techniques, MRI shows high spatial sensitivity, especially in detecting inflammations, trauma, infections or vascular accidents. The European and American guidelines, including UK NICE, elected MRI as one of the gold-standard in assessing patients with suspected dementia [59, 60, 61]. Brain atrophy is the main indicator in AD, confirmed in post-mortem studies [59]. It is visible as a shrinkage of the grey mat-
ter tissue. Hippocampus, a small brain structure in the medial-temporal lobe, is an efficient marker, as its boundaries permit a smoother volume segmentation from the scan, as well as ventricles’ volume. The course of degeneration can vary, but in AD it usually generates in the medial temporal lobe, involving the entorhinal and cingulate cortex, hippocampus, and amygdala. It then spreads progressively through the limbic lobe and temporal neocortex, until it generalises to the majority of neocortical associations. The progression of the disease affects subsequently the whole temporal lobe, together with parietal and frontal lobes. Atrophy patterns are closely related with clinical symptoms, as they reflect the patient’s impaired behaviour due to affected brain areas. MRI presents also some disadvantages: long acquisition times, motion artefacts, exclusion of patients with metallic or implanted device, and high economical costs. Limitations in MRI use for AD relate to impossibility in detecting the molecular information, and differential diagnosis, as other factors can be the cause of brain atrophy [62].

**Biochemical markers.** A biomarker is defined by Henley and colleagues [24] as “an indicator of the presence or extent of a biological process that is directly linked to the clinical manifestations and outcome of a particular disease”. AD biomarkers derive from two main sources: Cerebrospinal Fluid (CSF) and blood analysis. CSF biomarkers are: Aβ, Total tau (T-tau), and Phosphorylated tau (P-tau). Aβ<sub>1–42</sub> reflects the burden caused by amyloid deposits in the extracellular space. Low levels of it in CSF are a warning for insufficient clearance of Aβ in the brain, resulting in plaque formation. T-tau accounts for the general degeneration, with levels usually increasing with normal ageing. In AD patients this value is well above the healthy population standards. Moreover, T-tau is a valid predictor of conversion from MCI to AD [40]. P-tau indicates disruption in the axonal transportation and is the elective marker for differentiating AD from non-AD dementia [63]. Like T-tau, higher levels of P-tau compared to healthy population represent abnormality. A downside in the use of these biomarkers is that the cut-off scores vary depending on the analysis kit, and on the specific population [64, 65]. Blood analysis is preferred to CSF, being less invasive, but it still requires further validation [49]. Valid blood biomarkers are
and tau-related plasma levels, followed by blood indicators of inflammation or cell ageing and death.

2.2 Predictive Models

2.2.1 Artificial Intelligence Systems for Healthcare

Data mining applications such as Artificial Intelligence (AI) are starting to play a fundamental role in healthcare. Here, large volumes of markers are being generated every day, thanks to improved computing technologies (e.g. imaging techniques, storage, tracking sensors). However, this data is not always “knowledge rich” [66]. Additionally, clinicians can easily be overwhelmed by the volume and heterogeneity of information they have to handle, which inhibits decision making. In line with the assumptions of this specific project, AI can uncover patterns in noisy and high dimensional datasets, to produce actionable knowledge. These techniques have been used for years in sectors like finance or business, but the complexities of healthcare made it a relative latecomer to this scene. AI has to adapt to a completely new set of challenges, such as data heterogeneity, high-dimensionality, specialised domain knowledge, and more stringent privacy requirements [67]. In return, the expectation is to deliver unprecedented value: foster precision medicine, generate tailored diagnostic and treatment plans, predict readmissions at the hospital, or monitor the progression for disease subtype.

2.2.2 Introduction to Predictive Models

One application of AI in medicine is the development of predictive models, which aim to integrate information from large cross-sectional data sets of markers into quantifiable longitudinal pictures of disease development [68]. Not only can these models describe the probable evolution of a pathology, but they can also outline a sequence of well-defined stages along this evolution timeline [69, 70]. Both staging and prediction give information on the classification of a particular individual with respect to the reference population, supporting pre-clinical, early detection of a condition. However, these two concepts differ in terms of technical and clinical implications [69, 71]. Staging identifies where a patient is at one point in time,
2.2. Predictive Models

based on markers values compared to the reference population. It is technically relevant to classify patients for modelling purposes, and has clinical importance for diagnosis and for assigning different treatments to patients at a specific stage. The temporal relationship between stages is key in prediction. This second concept, in fact, requires a higher amount of data, collected at different points in time, to build that disease trajectory that describes how patients’ stage will evolve as the disease progresses. This information is clinically relevant when monitoring interventions, when describing cluster of patients that follow different evolution paths, and in planning for future follow-ups. As such, they can contribute to the development of disease-modifying treatments, and design of clinical trials [72]. Chronic and progressive diseases are the best suited ones to this scope, and in fact computational researchers have experimented in a variety of conditions (e.g. type 2 diabetes and Parkinson’s [73, 74], or Chronic Obstructive Pulmonary Disease [75]).

Methods used to develop predictive models vary. Basic statistics include regression [76], where clinical markers at one time are used to predict values at a later point. Recent advances in AI and Machine Learning (ML) have moved the discipline towards increasingly automated processes, especially in assessing associations, classifications, and relationships amongst data [77], based on supervised and unsupervised ML techniques. Supervised learning techniques learn mappings from input data to outcome from a large set of labelled data; unsupervised learning aims to identify patterns from unlabelled data. Unsupervised learning techniques are used to develop Disease Progression Models (DPMs), which aim to learn patterns of disease progression, i.e. temporal series of changes, from data sets with no a-priori labelling of patient stage. Some DPMs use discrete staging [78]; others aim to capture continuous temporal evolution, such as continuous models [79, 69], or spatio-temporal models [71]. I have illustrated key examples of models used, depending on research questions and processes regarding test selection, quality control and standardisation, and computational/statistical methods in Bellio et al. [3] and reported in Appendix A.2.
2.2.3 Predictive Models for AD

In AD, different models have been proposed to identify staging and progression of this condition. Initial ones worked only on given assumptions, for example providing a-priori event position of biomarkers [59]. Some others needed pre-set definitions of abnormality thresholds for each biomarker [61]. The model from Jack et al. [8] described a qualitative and hypothetical “disease signature” of the events’ cascade characterising AD in the earlier stages, based on a set of scalar biomarkers, such as CSF, imaging, and cognitive data (Figure 2.5). In some other cases, the sequence of biomarkers was pre-defined on existent clinical stages [80], resulting in a general qualitative definition of the disease. Recent models can overcome some of the limitations here presented, thanks to additional resources, initiatives and world-scale research efforts that will be presented in the following section. The ones described here, and that are most relevant to the purposes of this project, are Event-Based Model (EBM) [78], Subtype and Stage Inference (SuStaIn) [81], and latent time joint mixed effect models (LTJMM) [82].

![Figure 2.5: Representation of the progression cascade of biomarkers characterising Alzheimer’s evolution in the hypothetical model from Jack et al. [8].](image)

**Event-Based Model.** The EBM [78, 9] is a probabilistic, data-driven approach to describe staging and prediction for familial and sporadic AD. It is data-driven since
2.2. Predictive Models

it does not require a-priori knowledge about patient’s staging; it is probabilistic because it evaluates the probability of a patient to be classified at a particular stage. This procedure was initially tested with familial AD [78], then adapted for sporadic AD [9]. In EBM, every biomarker is described as normal, when it does not report alterations caused by the disease, or abnormal. An event is described as the switch of a biomarker from normal to abnormal status. Thus, the EBM principle conceives disease progression as a discrete and linear sequence of events. After determining the sequence, each subject can be assessed, according to the highest probability of belonging to that stage. To validate the model, Young et al. [9] evaluated how well the EBM staging mapped the labels assigned to patients in the dataset (respectively, cognitively normal, MCI, and Alzheimer’s disease). As Figure 2.6 reports, EBM could clearly separate cognitively normal individuals (light blue) from AD patients (orange), with an accuracy higher than 99%. Longitudinal validation matched between prediction of progression and follow-up data at the time points of 12 and 24 months.

![Figure 2.6: Validation of EBM staging based on patient’s diagnostic categories. The bars represent the proportion of patients classified at that stage. Colours refer to the diagnostic category: cognitively normal (light blue), MCI (black), and AD (orange). From Young et al. [9]](image)

One of the important features of this model is the possibility to predict patients’ conversion from cognitively normal status to MCI, or from MCI to AD. This is
Chapter 2. Background

represented in Figure 2.7 as the amount of changes in diagnosis amongst different diagnostic groups, along time. Higher EBM stage value is a significant predictor for conversion in both the cases. The most informative predictors of conversion within 2 years were rate of atrophy, particularly for hippocampus and entorhinal cortex, and cognitive scores.

Figure 2.7: Prediction of conversion. (A) From MCI to AD; (B) from normal to MCI. The different lines represent groups of patients at stage 0 (blue), stages 1-3 (green), stages 4-5 (orange), stages 6-10 (cyan), stages 11-14 (magenta). From Young et al. [9]

Limitations of the EBM are caused by uncertainty due to heterogeneity of the population, in terms of different disease course or clinical characteristics (e.g., genetic predisposition to AD), sampling density (when events occur closely to each other, they will be based on less data points compared to events that are further apart), and outliers (individuals who do not have typical AD). The contribution offered by this model to clinical application is three-folded: early diagnosis and prognosis, stratification of patients, and conversion information. The first one is aimed at promoting personalised medicine and monitoring of drug trials. A fine-grained disease staging has the potential to support clinicians’ decision making, as well as patient’s management. Finally, conversion supports information on disease progression. The robustness and relative simplicity of this method has promoted translation in various other disease areas, such as Huntington Disease [83], and Multiple Sclerosis [84].
Other disease progression models. Two other models contributing to a further understanding of disease mechanisms are SuStaIn [81], and the Bayesian LTJMM [82]. SuStaIn is a data-driven algorithm which disentangles heterogeneous disease populations into phenotypical clusters characterised by similar disease progression. It has been applied to genetic Fronto-Temporal Dementia (FTD), and AD. Particularly for AD, it has shown the potential to stratify the population based on structural MRI, into three clusters (typical, cortical, and subcortical). Even patients at early stages presented strong alignment to a particular subtype, and this has been reproduced with independent datasets. The output that SuStaIn can provide is a data-driven taxonomy of subgroups and stages, but also a prediction of conversion risk. This would bring advantages in precision medicine, patients’ stratification for clinical trials, and decision-support in clinical practice [81]. LTJMM is a mixed effects model, which aims to reconstruct a long-term picture of the disease, based on short-term data availability. This approach is advantageous to understand the disease course, when normally it requires recruiting multiple patients at different stages of the disease, with obvious biases [82]. This model contributes with patient-specific predictions of long-term biomarkers trajectories. Its efficacy would be beneficial for treatments’ assessment and clinical trials; however, clinical practice advantages could be envisioned, in supporting timely decisions and improved clinical pathway.

2.2.4 Shared Resources and Projects

In order to train computational models, extensive and longitudinal datasets are needed. Various initiatives collect multicentre, multimodal, longitudinal data on AD and other types of dementia. Examples are: Alzheimer’s Disease Neuroimaging Initiative (ADNI) [85], Layton Aging & Alzheimer’s Disease Centre (LAADC) [86], and National Alzheimer’s Coordinating Centre (NACC) [87]. Other projects focus on specific populations or cohorts, such as VITA [88], and large biobank studies (e.g. UK Biobank [89]). One of the most prominent datasets in AD research, and the main one used to train the EBM, is ADNI. ADNI was launched in 2004, as a longitudinal multicentre American study funded by 20 companies (including pharma and non-profit organisations), and other foundations. The project aims to
identify clinical, biochemical, genetic, and imaging markers to guide early detection of the disease and support treatment development, prediction of disease progression and trajectory.

As a consequence to this increased and organised data availability, many multidisciplinary research projects originated, to advance clinical AD understanding through innovative clinical and computational applications. The one this Ph.D. thesis contributes to is EuroPOND (grant EP/J020990/01), a Horizon 2020 research and innovation programme sponsored project. Its aim is to develop computational models to describe and support the understanding of a range of neurodegenerative conditions. Thanks to the availability of large datasets, related not only to AD, but also to Multiple Sclerosis, Prion Disease, normal ageing, and preterm brain development, the models shed a new light on disease-related patterns of progression. The motivation underlying this project and the ultimate clinical impact refers to new possibilities in describing specific disease signatures, improving diagnosis, particularly early detection of the disease, guiding the design of clinical trials, and supporting treatment decision and patients’ management. A multidisciplinary group of researchers is involved, including computational scientists, neurologists, and neuroscientists.

Notwithstanding the huge potential of these innovations, their ultimate success relies on how easily clinical specialists can access them, their seamless inclusion in the workflow, and a consistent adoption of the technology. The next section presents current knowledge on how these aspects have been considered in the translation of AI innovations to clinical practice.

2.3 Clinical Decision-Support Systems

One common way to deliver predictive models to the clinical context is by means of digital tools called Computerised Decision-Support System (CDSS). CDSS are pieces of software which, based on a given clinical input, produce output that clinicians can use to support their decision-making [90, 21, 91]. CDSS are normally integrated into various devices, such as desktop, mobile devices (tablet, phone), bio-
Clinical Decision-Support Systems

CDSS is an umbrella term that includes very different tools, with different underlying rules, output, and purpose. Horsky et al. [18] described six main types of decision-support intervention: alerts/reminders and ordering support for real-time interaction with the system, and guidelines, forms, clinical context, and clinical pathways for tasks which are more related to management and reasoning. Berlin et al. [90] proposed a taxonomy of five categories to classify different CDSS: Context, Knowledge and Data Source, Decision Support, Information Delivery, and Workflow. Context indicates setting and all related aspects of using the tool, but also the objectives and the end user. One interesting aspect of this category is timing: 41% of CDSS were used at the point of care, meaning during the clinician-patient session, while the remaining provided decision-support outside the point of care or in between visits. Section 2.4 will outline how defining this aspect is fundamental in designing such tools and integrating them in the clinical workflow. Regarding knowledge and data source, CDSS can be knowledge based, when they are built on rules deriving from clinical evidence, and non-knowledge based, when making use of data mining techniques such as AI, ML, or pattern recognition [21]. This last one often presents issues around data source and interpretability which will be discussed at a later point. Decision support refers to the nature of information provided (e.g. immediate action, medication dosing, prevention), which is closely linked to information delivery and how the information is provided (e.g. pushed notifications, pulled information with action required from the user). The last one, workflow, although hard to characterise, concerns all issues related with the integration of the tool in the setting, such as additional actions or staffing needed to operate it, and if there is flexibility for when to use it or a strict time window [90].

Advantages of CDSS are therefore not only related to the knowledge brought for example from disease progression models into the clinical routine. Sutton et al. [21] in their review outlined a number of benefits that go beyond the diagnostic and decision support, touching on aspects such as workflow and documentation improvement, patient safety, reduction of human error and cognitive workload, cost
containment, and administrative automation. However, these do not come without downsides. For example, Garg et al. [92] reported how the effect on patients’ outcomes is not clear, although practitioners’ performance is improved.

2.3.1 Clinical data visualisation

One of the main features of CDSS is how data is presented and visualised. Data visualisation can amplify the cognitive perceptions of users, reducing the load caused by multivariate sets of data and maximising users’ performance [93]. There are many reasons why efficient data visualisations are so relevant to the medical field. They can inform decision making, through intuitive and easy recognition of relevant information [94], reducing the amount of possible errors [95]. They can also improve understanding, navigation, and management of multivariate data, specifically on diagnosis, therapy and prognosis [96]. In this last case, data visualisation is a great support in evaluating patients’ conditions, temporal changes and effectiveness of therapies [97]. A well-designed visualisation have proved to contribute towards these aims [98].

The first attempt of data visualisation for clinical purposes dates back to 1855, when the physician John Snow adopted it as a way to understand the causes of cholera in London [99]. In the late 90s, one of the first electronic visual summary of multiple patients’ data was developed by Powsner and Tufte [100]. From that point, various Electronic Health Records (EHR) emerged to support clinicians’ work [93]. Later on, neuroimaging graphical workspaces introduced 3D images inspection, contextualisation, and interpretation supported by databases [101]. The concept of combining multimodal information was included in the work of Joshi and Szolovits [102], where a 100-dimensional information space was represented through a radial starbust. Multidimensional graphic solutions were differently tested in Cube, a platform where data from a sample of patients is organised in clusters of parallel diagrams, in order to magnify and recognise pathological patterns within the population [103]. Another example is reported by Faiola and colleagues [104, 105]. Their innovative interface for intensive care units used different graphical and system techniques to support clinicians in reducing information overload and decision-
2.3. Clinical Decision-Support Systems

making time.

Literature on data visualisation can inform how to best represent data optimised for CDSS. The features that are frequent in disease progression models include the following: multidimensional and multivariate data; time-oriented information; events and classification; predictions, probability, and uncertainty.

**Multidimensional and multivariate data.** Various techniques suggest how to better represent multiple variables from big data sets, ranging from simple box-plots, correlational matrices, or radial tree [106], to parallel coordinates, with each axis representing one variable as in ClustVis [107]. Probably the most efficient method is the use of interactive and filterable tables, allowing users to personalise and focus their search [108]. There are clear advantages in supporting multivariate indicators through intuitive visualisation techniques, although a multitude of variables might overload the user. Different strategies can reduce the dimensionality of data sets, as demonstrated in the work from Cavallo and Demiralp [109], suggesting visualisation alternatives to facilitate the use and exploration of output from a dimensionality-reduction system.

**Time-oriented information.** Visualising the time-course of data as a graphical representation facilitates the monitoring of disease progression, patient management, and evaluation of treatments’ efficacy [95]. Temporal representation in medicine [110] has explored a plethora of well-known techniques, like radar maps, conceptual trees, fish-eye lenses, dynamic queries, scatterplots, and charts. The common goal of these techniques is the abstraction of raw information to higher levels of interpretation. Shahar and colleagues [111], introduced the concept of temporal-abstraction task. In their framework, KNAVE, longitudinal clinical data are brought to a higher level of abstraction, thanks to a system based on previous knowledge databases. The novelty of this work is to provide high-elaborated temporal data, in an easy and interactive platform, editable and domain-specific. A similar project is VISITOR [112], a knowledge-based architecture for longitudinal temporal data representation, generating complex elaborations from raw data. System assessments
have shown reduction in clinical evaluation times and higher accuracy. Temporal information is key in those settings requiring fast decisions and constant monitoring of life-related indicators, such as Intensive Care Unit (ICU). Various works have addressed these needs, focusing on visualisations to provide contextual knowledge and allow immediate interpretation of multivariate data, as in the case of MIVA [104].

**Events and classification.** A model such as the EBM provides event-related data along time. Visualisations of events embedded in the temporal representation allow the extraction of patterns of event recurrence within and between subjects, and the graphical exemplification of the interaction between events and time. In PatternFinder, Fails and colleagues [113] formulate an architecture that works on a database of over 26,000 medical events. Each event is characterised by type and value parameters. A series of events, interleaved by time-lapses, is a temporal pattern. The user can customise both events and time spans, representing temporal patterns on a timeline. The concept of temporal pattern is applied also in Outflow [10], a graphical representation of a series of events and their pathways (Figure 2.8). The aim of this work is to visualise multiple event-based data simultaneously and analyse event properties together with the tendency of progression for each event. The graphical representation is a state transition graph (Outflow graph), a visual metaphor quite common in computer science to exemplify system status and status changes. Being able to visualise different temporal patterns, gives the opportunity for classifications. This is proposed in DICON [114], a system that can cluster groups of similar patients, based on algorithms working on EHR data.

**Predictions, probability, and uncertainty.** Recent advances in predictive algorithms have triggered research on how to visualise those outcomes to facilitate their interpretation, particularly for drug discoveries and disease management. Osuala et al. [11] guide the reader through alternative visualisations for predictive models, informed by consultations with relevant healthcare professionals and iterations. They made use of blob charts, with each blob representing a diagnosis and the size of the
2.3. Clinical Decision-Support Systems

Figure 2.8: Outflow example representation of events and temporal progression pathways [10]

blob informing on its probability, of bar and radial charts (Figure 2.9). Their proposed visualisations conveyed the same information, but differently encoded. Lazic and colleagues [115] focused on researching and testing graphical methods to represent disease progression over time in Huntington disease, supported by vector and path plots. The Multiple Sclerosis project BioScreen collected clinical data from patients and compared them with a reference population, to allow comparisons and account for possible prognosis and risk scores [116]. An advantage in exploring risk prediction and disease trajectories is represented by the implementation of interactive interfaces, such as in the work of Li & Arandjelović [96], or in Krause et al. [117], where their system Prospector allowed the user to interact with a set of features and observed how the overall prediction was influenced. One particularly interesting feature in disease progression models is the ability to represent a prediction pathway. The system COWpath [118] is an example aimed at facilitating evidence-based care through extracting common clinical pathways from EHR data. Pathway visualisations have shown to support predictions for key events, as represented in the work of Yamashita et al. [119]. CareCruiser is another example of clinical pathways’ visualisation and prediction, to optimise clinicians’ treatment
choices based on medical evidence [120]. Despite the impressive accuracy reached by predictive algorithms, it is undoubtedly true that there is some levels of uncertainty, and this information should be clear to the user when adopting such systems. The visual representation of uncertainty was studied in the work from Greis et al. [121], particularly when the prediction was based on multisourced data types.

Graphical representations of expected-risk models have shown a positive impact on clinical outcomes [97], providing guidance for valid tools development and opportunities for improving healthcare by combining the quantitative resources brought by technical advances and their translation into intuitive visualisations [96].

Figure 2.9: Blob chart (A), Bar (B), and Radial (C) visualisation of predicted diagnoses. From [11].
2.3.2 Data visualisation in AD

AD specialists currently interpret disease outcomes qualitatively, relying on clinical expertise. However, some recurrent features can be detected, the two most evident being: classification and progression. An example of visualisation strategies targeting disease classification is given by Weintraub et al. [50], highlighting individual early cognitive profile, as it should facilitate recognition of more prominent cognitive traits at early stages of the disease. Another interesting classification depicts the sequence of emerging symptoms in different types of dementia along the disease course, and how these are interconnected (see Figure 2.10) [12]. The majority of AD clinical visualisations are crafted as plain lists of symptoms relative to each stage, being that the 7 stages of Reisberg [46] or a 3-level stages (mild, moderate, severe). Some other representations combine the disease pathway (indicated as text) on a bidimensional plan, including diagnostic labels, a timeline, and sometimes referring that to the specific marker’s trajectory (Figure 2.11). The method from Ferreira et al. [122] and later on by Harper et al. [14] proposed a more quantitative classification representation, thanks to increased availability of high resolution data in medical imaging. They defined a diagnostic tool of visual rating scales based on representative samples of atrophy from consecutive MRI scans. Whilst the former targeted more specifically the classification through MRI visual rating, the latter used this method to support diagnosis and progression assessment (Figure 2.12).

2.3.3 Interfaces and systems

Various systems, such as CDSS, dashboards or web-based applications, have been developed for healthcare, translating predictive models through visualisations to enable their clinical accessibility. Some relevant examples come from the Intensive Care Unit (ICU) environment, characterised by time-pressure on decision-making and the need to understand possible disease course. EHDViz [98] is a clinical dashboard which gives the user access to high-risk and real-time monitoring of clinical data, trend predictions, and other interactive features. As a limitation, it requires the user to be familiar with programming languages. Other examples come from applications in the heart disease field [123, 124].
Figure 2.10: Symptoms related to different types of dementia and their interconnections, represented as pathways. From Rita Maldonado Branco [12].

Alzheimer’s context does not have the same requirements as an emergency department, where fundamental variables are time-pressure and point-of-care delivery. AD management would rather benefit from visualisations that allow a better exploration of disease patterns and progression, to inform long and short term care plans [96]. Therefore, more suitable options are dashboards including information on time-oriented data, probability and risk predictions [11], and event classification [125]. A few examples of systems disclosing AD prediction analyses started to appear in recent years. Amyloid-risk model [126, 127] was implemented in a basic interface, providing the user with a probability of amyloid-positive index, based on
2.3. Clinical Decision-Support Systems

Figure 2.11: AD progression visualisation, including MMSE scores, symptoms, and diagnostic labels, on a timeline. From [13].

Figure 2.12: MRI-based visual rating for the assessment of atrophy progression in AD. Adapted from Harper et al. [14]

a set of more accessible biomarkers, such as clinical, and genetic. PredictAD [128] and its successor PredictND [129] were developed to support decision making and early detection of the disease. They run on an algorithm developed to generate a Disease State Index (DSI) and its graphical counterpart, the Disease State Fingerprint (DSF). When a new patient’s biomarkers are given as input, the system compares the individual against the heterogeneous training population and provides DSI and DSF. PredictAD (Figure 2.13) is the graphical interface supporting clinical use of these models, and the one more close to this project’s goals. PredictAD made use of multivariate data to feed an advanced model of disease progression, including a graphical metaphor and accounting for interactivity. Differently from these two examples, ADappt [130] was developed including clinical feedback in the process,
and considering ways to communicate AD output to patients and caregivers. This web-based tool was intended for clinicians in memory clinics treating early stages AD patients, and provides estimations of risk to progress to more severe stages. It included a shared decision-making module, to facilitate the communication of test results to patients and support a communal care plan, informed by data and clinical expertise.

![Figure 2.13: PredictND. The areas of the dashboard include a bar reporting uploaded input (A), a section reporting the hierarchical decision-making outcome (B), the expected accuracy of DSI (C), distribution of an individual biomarker (D), and the influence each measure has in the classification (E). From [15].](image)

### 2.4 Design and Adoption of CDSS

In recent decades, almost all CDSS have failed when moved to clinical practice [131, 132, 18, 20]. Looking at possible causes, a common pitfall is that these systems often lack Human Factors considerations [133]: poor fit within the workflow, context, and attention to clinical needs [91] open a gap between design and reality. According to Heeks [132], the severity of this gap can be assessed along seven dimensions that are representative of the socio-technical context and can predict the likelihood of the system failing or succeeding. The bigger the gap, the greater the change needed to successfully adopt the innovation and, consequently, the chances
of failure. This framework reinforces the importance of realigning “work as imagined” and “work as done” [134], a necessary effort that innovators should consider in the “design thinking” phase of delivering a system that is truly fit for purpose [134].

2.4.1 Technology adoption and diffusion of innovation

There is a body of work addressing issues on adoption of Information Technology (IT) systems in healthcare. Technology adoption was defined by Ammenwerth [131] in relation to two conditions: “For voluntary used system, IT adoption is reflected in the usage of the IT system; for mandatory used systems, IT adoption is reflected in the overall user acceptance”. However, this term is often interchanged with acceptability or acceptance [135]. Renaud [136] suggested that whilst acceptability is the “attitude towards a technology, and it is influenced by various factors”, adoption is a multi-phased process, with differences along the pre and post-use continuum.

Innovation always brings elements that trigger a change in the standard practice. Therefore, poor uptake of technologies could be improved by the study of IT adoption and fit with the socio-technical context. Frameworks on the analysis of technology adoption, such as the Technology Acceptance Model (TAM) [137], the Theory of Planned Behaviour (TPB, [138]), or Innovation Diffusion Theory (IDT, [139]), were developed for already implemented technologies, not always considering important social and contextual variables. Therefore they do not provide insights into the formative understanding of how the work “could be done based on the relationships between information environment elements in a work system” [20]. Nadal et al. [135] proposed the Technology Acceptance Lifecycle (TAL) to encourage practitioners and researchers to consider acceptance as a process that evolves along the technology development journey. According to Lindley et al. [140], an early evaluation of implications for adoption is desirable to prevent failure. This concept refers to research-based speculations on the future potential adoption of a novel technology that would support design thinking and identify contextual fit with the socio-technical environment. The HOT-fit (Human, Organisation, Technology
fit) framework proposed by Yusof et al. [141] evaluated the fit amongst various
dimensions related to human, organisational, and technology factors. Similarly,
Gagnon et al. [142] proposed a set of factors (barriers and facilitators) to the imple-
mentation of a new system in the medical setting. These factors are derived from
case studies on IT adoption by clinical professionals and are overall divided into
four main areas: factors related to IT, individual and professional factors, human
environment, and organisational environment.

The presented works demonstrated that the most important facilitators in IT
adoption are the context of implementation and users’ perceptions of technology
potential. Regarding the first, an important requirement is the impact of a novel
technology on the workflow [143]. It is necessary to fit the novel system into an ex-
sting workflow, without disrupting users’ beliefs, but promoting safety, data quality,
and work efficiency. Regarding the latter, another reason to invest in early engage-
ment with users is to consider how they build concepts and mental models around
the technology, to guide its design and implementation. Blandford et al. [144] pro-
posed a systematic approach, the Concept-based Analysis of Surface and Structural
Misfits (CASSM), to qualitatively define conceptual fit between mental processes of
the users and how the technology is implemented. Areas of misfit should inform de-
sign implications and improvements. Furthermore, clarifying these misalignments
and mental models can support the characterisation of a complex reality such as
clinician-patient-IT, as described by Smith and Koppel [145]. They presented dif-
f erent scenarios in which this interplay was not balanced, to guide IT uptake. It is
not only important to evaluate the expected use of a system, but also the unexpected
uses of it, as they could result in interaction issues that were not originally consid-
ered [146]. This is particularly true for safety-critical applications. Campos et al.
[146] proposed a 4-phase model where the interaction between the user and the de-
vice is modelled to include only relevant information resources, to limit the risk of
unexpected paths. This approach, called resource-based analysis, has the advantage
to study the usability of a tool that is not yet a full developed system, while avoiding
confirmation biases and assumptions from the evaluators.
Technology adoption leads to the importance of diffusion of innovation. Rogers [139] defined it as “the process in which an innovation is communicated through channels over time among the members of a social system”. Diffusion of innovation is influenced by multiple factors: context [147], leaders and champions [148], or network [148]. First and foremost, however, the innovation process is guided and influenced by people, where a particular key role is covered by “early adopters”. Lettl et al. [149] described how this group of users can unlock and disseminate the take up of radical innovations. Early adopters are often skilled professionals [150], embedded in interdisciplinary contexts, with facilitated access to specialised resources, and whose needs anticipate those of their peers, making them motivated and open to innovation [149]. Identifying this user group and their context in relation to the system’s purposes can facilitate the introduction of the target innovation in the workflow, collaborate in defining elements to better adapt the technology to a wider context, and diffuse the innovation to their peers.

2.4.2 User-centred design

This project adopts a user-centred design approach, and the reasons are here explained. Involving intended users at all stages of the design process promotes efficiency, effectiveness and safety [149, 151], according to User-Centred Design (UCD) principles (ISO 9241-210:2010 [152]). UCD is a cyclical and sequential process (Figure 2.14), which starts from understanding the needs and context of use, to the requirements that will inform the initial prototype. The prototype will be then evaluated with the end users, until a satisfactory solution is reached.

For each stage of the cycle, different methods are applied. This is well reported by Pontis [17] (Figure 2.15). The main methods for an initial understanding of user’s needs and practice, and those applied in this thesis, are: field observations, semi-structured interviews, and focus groups or workshops. Field observation is an ethnographic approach in which the researcher observes the phenomenon happening [153], usually in the context where the technology is or will be used. Different parameters are involved: the setting is strictly determined by the research question; participants can be aware or not that they are being observed (respectively called
overt or covert observation - although covert method is not much used in UCD and involves ethical issues), and the observer can play three different roles (passive observer, actively involved as a participant, or minimally invasive). It is fundamental to plan in advance what to observe and how to record it (e.g. photo, audio or videos, and written notes). Observations are frequently accompanied by interviews or questionnaires, to further clarify the observed material. *Semi-structured interviews* are a particular type of interviews where the main structure is defined, but the conversation can extend to broader discussion [153]. They are optimal in gathering individual’s insights on current work or on new technologies, although subjective responses might be biased or incorrect. It is important to plan them carefully in advance, to ensure a logical ordering. *Focus groups and workshops* are data collections involving multiple participants. They are guided discussion around a theme of interest, led by a moderator, who can be internal or external to the group. Focus groups are useful to discuss sensible topics, elicit different perspectives, and getting a wide range of insights [154]. Similarly, workshops are run with multiple participants, but the aim is to create knowledge, or reshaping existing one. Workshops involve use of material, maps, activities, and creation of ideas, and are very useful at the beginning of the design process, to stimulate idea and contents generation [155, 156].

**Figure 2.14:** The User-Centred Design Cycle. From Usability.gov [16].
Methods for later stages in the cycle are aimed to test the design concept for a system or prototype. Rapid (or low-fidelity) prototyping is quite effective in early stages of the design process, since it can produce prompts to bring into research for validating design ideas with the end users, saving time and resources [157]. These can be video, paper-based, storyboards, and mock-ups. Common testing methods for this context are user studies or usability inspections. The former involves the end-user and is meant to collect both quantitative measures (e.g. time on task, accuracy, ratings), and qualitative (interviews) [158], whereas usability inspections are performed with experts and can include cognitive walk-through (an expert talks-through during the interaction with a system), or heuristic evaluations (assessment of a system based on a predefined set of heuristics) [159]. The most suitable method should be decided case-by-case, according to the research question and the specific study’s characteristics.
2.4.3 Designing for CDSS

A user-centred design approach can bring clear advantages in developing IT for healthcare. Designing clinical tools including specialists’ feedback at all stages of the development not only ensures that needs and requirements of such a complex system are carefully considered, but also promotes a much easier uptake and use of proposed technologies [160]. Craft and colleagues [97] conducted an evaluation study on how a clinical diagnostic system used in intensive care units would affect the workers’ practice. This study highlighted struggles to include technology in the clinical routine, despite meeting evident clinical needs. Conversely, Faiola and Newlon [104], in their initial evaluation of a similar interface for intensive care units, reported good responses from the workers, in terms of efficacy and usability.

The importance of human-factor considerations is well represented in the work of Yang et al. [91]. They conducted a field study to describe the current decision process on heart-pump implant, to define opportunities for the integration of a CDSS to support this process. A qualitative UCD approach is also adopted in ADappt [130], where clinical specialists and patients’ feedback was iteratively included in the design and implementation of the AD diagnostic tool.

Whilst every application differs, some general guidance on how to best design such tools have been provided from reviews of various works and other application areas [18, 19, 20, 21]. The ones that are common across these works are: system’s use and quality, usability issues, and structure (Table 2.1). System’s use and quality includes guidelines around how the system should deliver information, the knowledge and training required to use it, and quality of the information provided. Usability guidelines cover general requirements such as reduced workload or efficiency, design conventions, and interaction specifications, such as prescriptive approach. One fundamental topic is trust, which will be discussed more extensively in the next subsection (2.4.4). Structure guidelines refer to all those elements that facilitate the integration of the tool in the workflow, collaboration, interoperability, and scalability. Clear standards in structure facilitate the use and adoption of the tool from small to wider scale.
Despite guidance and best practice, designing for CDSS represent a complex problem, that will inevitably require trade-offs in design decisions and implementations. As explained by Tatar [161] in her design tensions framework, when many dimensions need to be taken into account, then design is not intended as a problem-solving exercise, but a compromise to balance conflicts in the design that cannot be solved. These tensions need to be acknowledged and explained in the design process.

2.4.4 Challenges to the use of AI in healthcare

CDSS systems built on AI algorithms, as for the present project’s main goal, face additional challenges in the design, uptake, and use, compared to those listed in the previous section. There is no doubt that AI systems for healthcare need transparency and explainability more than other sectors, because of personal data handling, ethics, and legal requirements [162, 163]. Hence, the process of translation might be slower and require extensive evidence, compared to other AI applications, such as finance.

Two topics are of main concern around the use of AI for clinical decision-making: data quality, and explainability. Data “is the fuel of AI” [164], therefore its characteristics and quality are fundamental to determine the accuracy of outputs [165]. Harwich & Laycock describe five factors that determine data quality: integrity, timeliness, coverage, completeness, and validity. Validity and integrity of data can be disrupted in the data-collection process, and particularly in the classification and labelling used to train AI algorithms. Data entry should also be accurately timed, accounting for the real timestamp, and not delayed, as this might alter patterns or progression outcomes. Although some algorithms can handle missing data, coverage and completeness are paramount to account for accuracy and fairness of outcomes [166]. The quality of algorithms is also key [160]. They should account for alternative and complex scenarios, be scalable to various tasks, and take into account contextual factors that might be relevant to decision-making or recommendations [160]. Strategies to improve data and algorithms quality point to the design of IT systems, that can not only facilitate the collection and development
<table>
<thead>
<tr>
<th>System’s attributes and quality (Technology)</th>
<th>Guidelines</th>
<th>Examples</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>System understanding; required training</td>
<td>[18, 19, 21]</td>
<td></td>
</tr>
<tr>
<td>Information delivery</td>
<td>Perceptual grouping and data relationship; consistent terminology and wording; notification timing</td>
<td>[18, 19]</td>
<td></td>
</tr>
<tr>
<td>Information quality</td>
<td>Completeness; relevance; format of information</td>
<td>[19, 21]</td>
<td></td>
</tr>
<tr>
<td>System status</td>
<td>System logs; performance</td>
<td>[18, 21]</td>
<td></td>
</tr>
<tr>
<td>Usability</td>
<td>General usability</td>
<td>Ease of use; efficiency; minimal disruption</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>Information adaptation</td>
<td>Acceptable density of information on screen; information ‘hubs’</td>
<td>[18, 20]</td>
<td></td>
</tr>
<tr>
<td>User control</td>
<td>Advice rather than commands; alternatives rather than stops</td>
<td>[18, 21]</td>
<td></td>
</tr>
<tr>
<td>Design conventions</td>
<td>Text entries; visual cues; responding procedure</td>
<td>[18]</td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>Motivation to use; expectations and beliefs</td>
<td>[19, 21]</td>
<td></td>
</tr>
<tr>
<td>Structure (context)</td>
<td>Workflow integration</td>
<td>Usability evaluation; workflow modelling; time constraints</td>
<td>[18, 19, 21]</td>
</tr>
<tr>
<td>Third party access</td>
<td>Adoption of standards; scalability</td>
<td>[18, 21]</td>
<td></td>
</tr>
<tr>
<td>Interoperability</td>
<td>Mitigate sub-specialty blindness; provide clarity at different steps; facilitate team awareness of patient’s status</td>
<td>[20]</td>
<td></td>
</tr>
<tr>
<td>Collaborative support</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.1:** Design requirements for CDSS, across four publications: Horsky et al., 2012 [18]; Kilsdonk et al., 2017 [19]; Miller et al., 2018 [20]; and Sutton et al., 2020 [21].
2.4. Design and Adoption of CDSS

process, but also monitor the quality post-hoc [165, 163]. Another approach from Mitchell et al. [167] suggests “model cards”, a systematic and standardised documentation of ML algorithms, disclosing relevant information on models’ details, training dataset, factors and metrics, but also intended use and ethical implications. Such structured disclosure provides additional advantages to all those involved in the decision-making on when and how to use such systems (e.g. policy makers, organisation), to facilitate communication between them and the technical experts.

A detailed description of ML algorithms and their functioning is not always sufficient or intuitive to those who need to use these systems, but belong to a different area of expertise, such as clinicians. In this case explainability of those systems and their outcomes is paramount. Explanations are defined by Lombrozo as “the currency in which we exchange beliefs” [168]. From a ML perspective, that currency is represented by translating AI procedures into “understandable terms to a human” [169, 163]. We seek explanation to facilitate our learning, find causes, and generalise insights to similar problems [170]. There are two classes of explanations: ante-hoc, when the explanations are incorporated in the model (e.g. decision trees or linear models), and post-hoc, when a secondary model is implemented to explain the main one (e.g. LIME [171]) [172]. Healthcare is a particularly challenging space, as the demand for explanation is much higher compared to other applications [173], and where explanations depend more profoundly on domain knowledge and the needs of the specific context. In fact, it is not only about translating the outcome from an algorithm into a tool, but translating a concept between two very different fields. To translate a concept, we need to take into account each side’s beliefs, mental models, and terminology. More importantly, understanding how an AI system generates suggestions can incentivise clinicians’ trust in that system, a fundamental requirement to their adoption and use [162]. Once again, HCI contribution has proven critical in investigating explainability. For instance, Cai et al. [174] demonstrated that instead of improving the performance of imperfect algorithms, researchers could work on improving users’ interaction with the system. This will stimulate users to build appropriate mental models through the interaction with the
algorithm, and therefore gather those explanations in a progressive enquiry process. Yang et al. [133] reported insights from design testing and field observations on how to seamlessly enhance the decision-making processes in a heart-surgery department thanks to the use of AI. Researchers have built on this consolidated knowledge on AI and explainability to produce guidelines for the design of explainable AI systems. Wang et al. [175] proposed a theoretical framework where different types of explanation techniques (Bayesian probability, similarity, queries, visualisations) are matched to different human reasoning processes. This schema is mitigated by possible caveats in human decision-making and how explainable AI techniques can leverage. Regarding the user-interface design, Amershi et al. [176] proposed a set of 18 practical design guidelines based on HCI research, and mapped along the different stages of interaction of human and the machine (i.e. initially, during interaction, when wrong, and over time). They tested these guidelines on various AI scenarios through an heuristic evaluation, a usability method where experts examine whether a set of guidelines are met in a given interface or system. Although more work is needed in defining optimal ways of promoting explainability and trust for AI systems, these initial guidelines provide a common ground to design tools that are trusted and adopted by the end users.

Building on top of the guidelines classification presented in 2.1, these additional considerations for the implementation of AI in healthcare provide guidance on the approach that can be adopted when designing these novel systems from the ground up. Given the central role of data and explainability, it becomes necessary to understand how clinicians are currently using data they have available in the first place, why they use a certain type of data, and what are their quality standards according to their clinical workflow and needs. Then, the focus shifts to the other side of the gap: the data provided by the novel system. Here it is necessary to look at how it complies with the standards expected by healthcare and, most importantly, what additional novel and clinically relevant information it contributes with. Principles just presented around the explainability (involving users early in the process [19], testing design guidelines [176], focussing on contextual design [20]) and the
2.5 Summary and thesis motivation

Alzheimer’s Disease is still poorly understood in terms of treatment and causes. Nevertheless, its increasing burden on the healthcare sector and society is currently motivating researchers from many different fields to tackle the problem from very different perspectives. Disease progression models represent an innovative approach in promoting and advancing AD understanding, specifically by enhancing trends and patterns from big population data that would not be accessible to clinicians otherwise. Whilst these outputs can directly support research and clinical trials recruiting, the goal of this project is to understand how, and in which form, this multidisciplinary approach can be translated to the individual, personalised care environment. The uptake of technology innovation in AD clinical practice is widely unexplored, given the failure of previously proposed technologies. New computational approaches for AD are characterised by inherent technical barriers such as data quality and use, and explainability of their outcomes (as reported in Section 2.4.4). However, one main cause of failure can be attributed to the lack of engagement with clinical needs and their current workflow, particularly early on in the development and design process. Human-Computer interaction and the user-centred design framework represent a methodological base to accompany a more solid design and development plan for a CDSS to be adopted and used in the clinical environment. In particular, the focus is first on understanding the user, to guide the development process, always keeping end-users’ feedback in the loop. This project follows the same rationale, with necessary adaptations that are motivated by the ground-breaking nature of the computational models considered in this work, and the multidisciplinary characteristics.

The next chapters will follow a roadmap (Figure 2.16) that draws from established frameworks such as User-Centred Design and Design Thinking [177] and
Pontis et al. [17]. This roadmap is also inspired by the guidelines and statements from literature in the design of CDSS and translation of AI into clinical practice (Sections 2.4.3 and 2.4.4). This visual will guide the reader through the various stages of the project, outlining the methods used at each stage. The roadmap particularly illustrates how stages are linked in the process, and how knowledge gathered at one point informs subsequent ones. The exploratory phase includes the presented background literature (Chapter 2), and the initial formative research on user needs and technical potential (Chapter 3). Results from this phase naturally lead into the design thinking and testing phase (Chapter 4), where requirements are discussed and integrated in the design process. Here, the core contribution is the testing of the design concept with the clinical users, to validate those initial needs, and produce further understanding on how to guide the next steps of the development process. This is covered in the design and review phase (Chapter 5), where results from the clinical testing are discussed with teams of technical experts, to define subsequent strategies and inform the future implementation of the tool. This roadmap will be re-contextualised at the beginning of each research chapter. Chapter 6 and 7 will respectively outline the overall discussion and conclusions.
2.5. Summary and thesis motivation

**Figure 2.16:** Thesis roadmap.
Chapter 3

Exploratory Phase

This chapter includes:

1. An initial exploratory study on AD current clinical needs and practices, with input on the potential for use of an EBM in the care setting (Study 1);

2. A more focused survey study on the use and sources of clinical data that are needed to assess the status and progress of the condition (Study 2);

3. A technical workshop to clarify the opportunities that various predictive models and configurations of the EBM can offer, and the expected clinical use from a technical perspective (Study 3).

Preliminary results for 1 were presented in [C.1, C.2, T.1]. Full method and results for Study 1 (section 3.1) were submitted to [J.1]. Outcomes from 1 informed part of the discussion published in [J.2]. Part of the study design and the data collection in [D.1] was used in 1, although different analysis have been conducted for the present work.

Previous research reports how introducing a technical innovation in the healthcare setting is often problematic and many applications have not shown their full potential in this space [132, 20, 91]. The inherent complexity of clinical environments makes the development of CDSS considerably challenging. The biggest hurdle is, however, that the majority of CDSS so far have focused on the technical aspects, neglecting the importance of developing a system that is truly fit for purpose [134], integrated in the clinical ‘systems of work’ [20], and built on the specific
clinical needs for that context. A different approach, based on Human Factors research, prior to the development of a system, can help defining those aspects, which will inform opportunities for adoption and development of a future clinical system [20, 91, 163].

Motivated by a user-centred design approach and Human Factors principles, this chapter presents three studies, representing the exploratory phase of this work (Figure 3.1). This includes formative research conducted to understand current clinical needs and practices of AD specialists, but also the opportunities offered by the technical innovation. Formative approaches are suitable to explore and inform the ways in which “work could be done” [20] for a novel system introduced in a complex healthcare scenario.

The first study aims to explore current clinical needs and practices of AD specialists, gather a first understanding of perceived potential of the EBM from a clinical perspective, and identify the context and early adopters that would be best placed to initially use the tool. Likely early adopters of the tool were identified with both psychiatrists and neurologists working in research-oriented clinical settings. The second study is an internet-based survey which reinforces and extends the initial findings, focusing on data use and access, and the specific indicators that inform clinical decisions on diagnosis and progression. The third study is a workshop involving technical experts, which aims to define the set of disease-progression models developed by the team and their potential in clinical practice from a technical perspective.
3.1 Study 1: Exploring Users’ Context

3.1.1 Aims

The first study was motivated by the need to understand opportunities for translation and adoption of an innovative EBM-based CDSS in healthcare from the early conceptual stage, before the development of prototypes and user testing. The aim is two-fold: (1) exploring adoption opportunities for CDSS in AD clinical care, and (2) testing a novel combination of methods to support this process. I focused on understanding current clinical needs and practices, and the potential for such a tool to be integrated into the setting, prior to its development. The user-centred approach was based on field observations and semi-structured interviews, analysed through workflow analysis, user profiles, and a design-reality gap model. The first two are common practice, whilst the latter provided added value in highlighting specific adoption needs. Moreover, as per few other studies [151, 91], clinical participants (in our case, AD specialists) were involved from an early stage of the development process, prior to any prototype development. This should shape the future design and translation process into healthcare to fit clinical needs and context, and promote explainability and trust from the outset.

3.1.2 Method

3.1.2.1 Approach

Methods selected in this work aim to understand users’ current needs and practices and identify possible early adopters of an EBM-based CDSS and contexts of use. This is done for a system that has not yet been developed. The focus is on the following three questions: (Q1) How is work currently done in AD clinical management?; (Q2) What are the characteristics and needs of front-line clinicians in AD?; (Q3) What are the perceived challenges and potential for adoption of the future tool? These findings will help defining the key implications for the intended tool to express its potential in AD clinical decision-making, and the early adopters and context that are best placed to initially benefit from it.

I applied three main methods in this study:
• **Workflow analysis.** Successful integration of innovative technologies in a real world setting requires adaptation to an existing workflow [133, 178]. Thus, I conducted a workflow analysis to understand the steps in the assessment of an AD patient. The sequence of tasks was generalised using a box-and-arrows representation.

• **User profiles.** I described users’ characteristics and needs through the definition of profiles that synthesise ethnographic research on a cohort of similar users [179].

• **Heeks’ Design-Reality Gap model (DRGm).** As described in Chapter 2.4, this semi-quantitative assessment has the advantage of being applied pre- and post-hoc. I adopted this method to assess how barriers and needs within the considered settings might influence the adoption of a possible system, before any design commitments.

### 3.1.2.2 Data Collection

Data collected for this study was generated by *field observations* and *semi-structured interviews*. To identify potential early adopters, the settings for observations and participant recruitment covered different options, with an eye on people oriented towards research and innovation. I conducted observations at Multi-Disciplinary Team (MDT) meetings. Semi-structured interviews involved neurologists and old age psychiatrists (recognised as front line specialists in the tasks for which the models are most likely to be beneficial), from hospitals and memory clinics. Ethical clearance to conduct the studies refers to Project ID UCLIC/1213/015 (Principal investigator Ann Blandford).

**Field observations.** I conducted a total of 6 observation sessions in MDT meetings held at the National Hospital for Neurology and Neurosurgery, Queen Square (London). I chose this setting to familiarise with the terminology and the issues discussed by clinicians, data mentioned, features of the disease, and the workflow. Approximately 10 to 15 medical professionals attended each meeting, including a
3.1. Study 1: Exploring Users’ Context

lead neuro-radiologist, neurologists, training students and others from related disciplines. At each meeting, 13 patients were discussed on average (min 6, max 22), within approximately one hour. Participants gave verbal informed consent to being observed. It was not possible to audio or video record the meeting, since patients’ names were read out, so data were recorded in the form of hand-written notes. Relevant information to record was identified through preliminary meetings with doctors, and iteratively adjusted as more meetings were attended. Field notes included: general workflow, questions asked by the audience, use of data and images, terminology and concepts describing the disease, and decision-making processes. Finally, in order to validate ambiguous information, specific clinical terms were reviewed by a neurologist not participating in the study.

Semi-structured interviews. I interviewed six AD clinical specialists: 5 neurologists and 1 psychiatrist. Participants belonged to first referral \( (n = 2) \) and secondary and tertiary referral \( (n = 4) \) centres, located in London \( (n = 5) \) and Belgium \( (n = 1) \). For privacy purposes and due to the small number of participants, affiliations are not mentioned here.

Participants were contacted through MDT meetings and word-of-mouth. Prior to the interview, participants were provided with information sheets, informed consent, and a background questionnaire, to track the variety of the sample (documentation is reported in Appendix B). Interviews had a dual purpose: to clarify insights from field observations, and explore the perceived value of predictive models (interview guide is reported in Appendix B.3). Questions related to the first point addressed both individual clinical practice and MDT sessions (inspired by the work of Lanceley et al. [180]). Regarding the second, participants were prompted to reflect on the EBM’s potential through a set of visual stimuli based on EBM visualisations used by the developers. These included one example of input screen (Figure 3.2) and six different examples of output screens (two examples are shown in Figure 3.3, all output stimuli are reported in Appendix B.4) with incremental information provided. The fixed content for all output screens was the sequence of stages generated by the EBM and the classification of the patient at a certain point of the scale. This
Chapter 3. Exploratory Phase

Figure 3.2: The input visual stimulus used in the semi-structured interviews (credits: icometrix early implementation).

represented the minimum information, as reported in the left example of Figure 3.3. Additional information was introduced gradually in the subsequent visuals, by reporting the list of markers becoming abnormal at each stage, the probability curve of a patient being classified at a particular stage, and the positional variance for each marker. The example on the right of Figure 3.3 represents the case with the highest complexity. These were used to stimulate discussion with participants, who were asked to comment on the visual stimuli and reflect on their understanding and applicability to the clinical scenario. I piloted the interviews with two neuro-radiologists and adjusted the script based on feedback. All interviews lasted between 40 and 70 minutes and all but two were recorded and transcribed. For non-recorded interviews, I collected detailed notes and checked the accuracy of information with the interviewees. A participation incentive was offered in the form of a free trial of the system once it becomes available.

3.1.2.3 Analysis

The methods listed and their order of application are the result of an exploratory process, whilst addressing our main questions. Firstly, the notes collected during field observations were used to generate a written report. This report grouped the
observed information around three different themes: main topics discussed, general patients discussion pattern, and clinical terminology. The content from the report informed the script for the semi-structured interview, which was structured in three main parts: (1) MDT meetings and the process of patients discussion; (2) Participant’s individual practice; (3) Introduction to EBM and discussion around its potential and limitations (Figure 3.4).

I first conducted an inductive thematic analysis on field notes and transcribed interviews (software used: NVivo 12): the analysis generated a set of 20 codes that were then grouped into 7 main themes. This first round of analysis revealed correspondence of themes with the factors reported in Gagnon’s work [142]. As Gagnon’s factors represent a more consolidated framework, they were then used as themes to run a second top-down analysis. Gagnon’s factors have sub-factors up to fourth-level; most of the sub-factors were covered by the analysis, except for 5 of 35 third-level factors and 16 of 32 fourth-level factors. This is due to the fact that interviews were not conducted based on this framework, and also because these factors aim to be generalisable, making it unlikely that a single case study would cover all of them. Outcomes from the thematic analysis were used to apply the methods described above, with the aim to address the main questions, in chronological order: workflow analysis contributed to the understanding of how work is currently done in AD management, user profiles gave insights into the characteristics and needs of front-line clinicians in AD, and Heeks’ Design-Reality Gap model (DRGm) assessed the perceived challenges and potential for adoption of the EBM.
technology. To provide more strength and guide the DRGm assessment’s process, Gagnon’s themes used for the deductive analysis were paired to Heek’s dimensions, as reported in Table 3.1.

### 3.1.3 Results

Results follow the set of questions formulated in Section 3.1.1, where each question is addressed by a particular method. Illustrative quotations are identified by participant number (P1, P2, ...) and clinical context (care-oriented: C-O or research and care: R+C) as described below, to maintain anonymity of participants.

#### 3.1.3.1 Current Clinical Practice and Needs

**Q1: How is work currently done in AD clinical practice?** I conducted a standard workflow analysis of AD current clinical practice, and in particular of AD patients’ diagnosis and prognosis. I produced a workflow for MDT discussion and one for individual clinical practice; their combination is represented in Figure 3.5. The process starts from availability of adequate data to support a clinical decision (e.g. diagnosis, referrals, treatment, additional investigations). Participants reported relying primarily on clinical history and symptoms, followed by cognitive assessment, and then biological and imaging analyses. The first barrier in the process (Figure 3.5, point 1) was identified in the challenge of gathering the breadth and quality of data to run the EBM. Typical issues on accessing and interpreting imaging scans

![Figure 3.4: Data collection scheme.](image)
Table 3.1: Gagnon’s factors organised by Heeks’ dimensions.

<table>
<thead>
<tr>
<th>Heeks’ dimensions</th>
<th>Gagnon’s factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>1. <strong>IT related factors</strong></td>
</tr>
<tr>
<td></td>
<td>1.2 Innovation characteristics</td>
</tr>
<tr>
<td></td>
<td>1.3 System reliability</td>
</tr>
<tr>
<td></td>
<td>1.6 Evidence regarding benefits</td>
</tr>
<tr>
<td></td>
<td>1.7 Validity of resources</td>
</tr>
<tr>
<td>Technology</td>
<td>1. <strong>IT related factors</strong></td>
</tr>
<tr>
<td></td>
<td>1.1 Design and technical concerns</td>
</tr>
<tr>
<td></td>
<td>1.3 System reliability</td>
</tr>
<tr>
<td>Process</td>
<td>2. <strong>Individual factors</strong></td>
</tr>
<tr>
<td></td>
<td>2.2 Attitude</td>
</tr>
<tr>
<td></td>
<td>4. <strong>Organisational environment</strong></td>
</tr>
<tr>
<td></td>
<td>4.1 Internal environment</td>
</tr>
<tr>
<td>Objectives and values</td>
<td>2. <strong>Individual factors</strong></td>
</tr>
<tr>
<td></td>
<td>2.2 Attitude</td>
</tr>
<tr>
<td></td>
<td>3. <strong>Human environment</strong></td>
</tr>
<tr>
<td></td>
<td>4. <strong>Organisational environment</strong></td>
</tr>
<tr>
<td></td>
<td>4.1 Internal environment</td>
</tr>
<tr>
<td>Staffing and skills</td>
<td>2. <strong>Individual factors</strong></td>
</tr>
<tr>
<td></td>
<td>2.1 Knowledge</td>
</tr>
<tr>
<td></td>
<td>4. <strong>Organisational environment</strong></td>
</tr>
<tr>
<td></td>
<td>4.1 Internal environment</td>
</tr>
<tr>
<td>Management</td>
<td>1. <strong>IT related factors</strong></td>
</tr>
<tr>
<td></td>
<td>1.5 Legal issues</td>
</tr>
<tr>
<td></td>
<td>1.9 Environmental issues</td>
</tr>
<tr>
<td></td>
<td>4. <strong>Organisational environment</strong></td>
</tr>
<tr>
<td>Other resources</td>
<td>1. <strong>IT related factors</strong></td>
</tr>
<tr>
<td></td>
<td>1.8 Cost issues</td>
</tr>
<tr>
<td></td>
<td>4. <strong>Organisational environment</strong></td>
</tr>
</tbody>
</table>
were: long waiting lists, delays in obtaining scans, frequent lack of radiologists’ reports or interpretation inaccuracies, and low quality images. However, clinicians belonging to research-oriented clinics reported having access to higher quality data, and reports from specialised radiologists. In the subsequent step, a provisional diagnosis is made, combining various cross-sectional data, and led by experience. However (point 2, Figure 3.5), in cases of critical diagnostic uncertainty, further investigations or MDT should be considered. The MDT meeting, on the right side of the map, amplifies the individual expertise and is recognised by all clinicians as a support to clinical uncertainty. The case discussion includes a precise request to the radiologist and to peers, usually to interpret or compare medical imaging data, and advice on further clinical action. The efficiency of MDTs (point 3, Figure 3.5) can be disrupted by lack of data clarity or difficulties in converging to a conclusion, leading again into the loop of the diagnosis decision node. When a provisional diagnosis is fulfilled, a prognosis might be attempted (point 4, Figure 3.5). Clinicians confirmed that they reason about prognosis, depending on their confidence in the current diagnosis. However, some interviewees focused more on how they communicate this information to patients, whilst others’ interest is in understanding the progression of the disease to more severe stages. The first group tend not to be too accurate in reporting the prognosis, to avoid giving misleading information. The progression of the disease largely depends on its past course and individual differences. One neurologist declared:

“Generally [the principle is that] the speed [at which it] has started will continue. Sometimes is quite difficult to define.” (P3, C-O)

Other clinicians, more interested in early detection of the disease and generally belonging to research-oriented clinics, placed more importance on their ability to detect critical conditions early on, and to monitor patients at risk until a diagnosis is made.

“The other fundamental point is progression. Understanding what percentage of MCI (mild cognitive impairments) will convert. Because if
there is a high probability that a patient will convert to MCI or AD, we would keep them.” (P6, R+C)

While MDT workflow is built on consistent responses from all participants, some discrepancies emerged with respect to the individual workflow. Therefore I investigated the variations in context for the clinical specialisations, to clarify how this can influence their needs and, consequently, the technology translation.

**Q2: What are the characteristics and needs of front-line clinicians in AD?** To answer this question, I developed three user profiles based on collected data and findings from the workflow analysis. Table 3.2 is an extract of profile contents highlighting convergent and divergent points. Results summarise clinical representative needs, context, and resources. Two main contexts could be identified, that are characterised by specific clinical needs, patient type, and availability of resources. These are referred to as Care-Oriented (C-O), and Research+Care (R+C). In C-O, including memory clinics, patients present at very different stages. The visit time is
limited, as is the data collection procedure, which mainly involves clinical and cognitive tests, while imaging is of variable quality and may or may not be informative for diagnosis and prognosis. R+C are more frequently involved in uncertain cases and early detection of the disease, until diagnostic criteria are met and the patient is discharged. Access to specialised data is facilitated and systematically recorded.

As per Table 3.2, I firstly explored whether, considering the same clinical professional role, but in two different contexts, differences emerged around needs, procedures and resources. Therefore, I compared the profiles of neurologists, based within two different contexts (Profile 1 in C-O and Profile 2 in R+C). Fundamental differences were found in the quality and availability of resources, characteristics of patients and clinical needs, more focused on supporting the disease path in profile 1, whilst more orientated to discern aspects of the disease in profile 2. I also explored the contrast between two different clinical professionals, a neurologist and a psychiatrist (Profiles 2 and 3), within a similar context (R+C). In this comparison, the availability of resources, time, goals, and needs of the team are aligned, regardless of the type of specialist involved. I found that the R+C vision of a clinical context is shaped by characteristics of the clinical centre’s director (who we will define as the champion), who might stimulate innovative ideas and a research-oriented mindset. Due to context classification being a result of the analysis and after the data collection was concluded, there was no C-O option for the psychiatrist profile, and only one psychiatrist was available at the time of data collection.

This analysis of user profiles gave a stronger hint as for who are the most suitable early adopters for the proposed tool, revealing that the type of specialisation (i.e. neurologist or psychiatrist) is not critical to the adoption of the EBM, provided that context and needs are aligned. I identified key factors that overcome these distinctions, like the presence of a champion who has a particular research-oriented vision and has access to particular facilities and resources.
### Table 3.2: Comparison of content for three user profiles.

<table>
<thead>
<tr>
<th>Description</th>
<th>Profile 1 Neurologist, C-O</th>
<th>Profile 2 Neurologist, R+C</th>
<th>Profile 3 Psychiatrist, R+C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resources</strong></td>
<td>Clinical data</td>
<td>Clinical data</td>
<td>Clinical data</td>
</tr>
<tr>
<td></td>
<td>Cognitive tests (basic)</td>
<td>Cognitive tests (complete)</td>
<td>Cognitive tests (complete)</td>
</tr>
<tr>
<td></td>
<td>Imaging (low quality, delays)</td>
<td>Imaging (high quality, delays)</td>
<td>Imaging (high quality, delays)</td>
</tr>
<tr>
<td><strong>Context characteristics</strong></td>
<td>Limited time and resources</td>
<td>Available time and resources</td>
<td>Available time and resources</td>
</tr>
<tr>
<td></td>
<td>Patients: all stages of disease</td>
<td>Patients: uncertain and rare cases</td>
<td>Patients: at early stages</td>
</tr>
<tr>
<td></td>
<td>If uncertainty: sent to later referral</td>
<td>Discharged at diagnosis</td>
<td>Discharged at diagnosis</td>
</tr>
<tr>
<td><strong>Needs</strong></td>
<td>Diagnosis</td>
<td>Diagnosis</td>
<td>Diagnosis</td>
</tr>
<tr>
<td></td>
<td>Patient management</td>
<td>Understand disease progression</td>
<td>Understand disease progression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Understand unclear and early cases</td>
<td>Understand unclear and early cases</td>
</tr>
</tbody>
</table>
3.1.3.2 Future Technology Adoption Assessment

I subsequently focused on framing the gap between the prospective technology and its clinical use, to assess the probability of failure or success. To meet this goal, I applied DRGm to the identified themes, particularly drawing on interview responses about individual practice and the reasoning around the visual stimuli (Figure 3.3).

Q3: What are the perceived challenges and potential for adoption of the tool?

Findings are here presented according to the DRGm dimensions.

Information. The most relevant themes here referred to the EBM’s benefit (defining stage and progression of the disease), reliability, and validity. The EBM would be beneficial to address the need for more detailed information, particularly on progression, expressed by both clinicians and patients.

“It’s a question most patients ask: how is this going [to progress], how is the evolution going to be, what is my prognosis? [...] And at this moment we don’t have any ways to predict the prognosis.” (P2, C-O)

On the other hand, a frequent concern referred to the validation and reliability of the EBM, it still being at an experimental stage. A specific concern was on the reliability of the biomarkers’ sequence generated by the model.

“Before I use such a tool, I would like to [...] set up a validation study like you do for such tools before starting to use them in research or clinical trials.” (P1, R+C)

What differentiated the two contexts is the availability of some biomarkers: in C-O clinics it is harder to access sophisticated data such as lumbar-puncture and comprehensive neuroimaging scans, whilst it is routine in R+C centres. Moreover, in R+C the majority of patients are referred for clinical uncertainty or difficult diagnosis, which shapes the information sought by the medical staff.

“It is easy to diagnose an evident AD. The real need is to understand more about this grey area, if we are keeping these patients.” (P6, R+C)
Technology. This dimension examined whether the current technical set-up of the workplace is suitable for the integration of the EBM. Given that the future CDSS can be easily implemented in desktop computers and that hospitals and clinics are already equipped with PCs, no additional technology would be needed. Notwithstanding this, data handling and interoperability would need to be carefully considered.

Process. This analysis was based on the process of data collection, the workflow, and the openness to additional steps that the EBM as a tool might require (Gagnon factor attitude). The heterogeneity of many biomarkers that are site-dependent (e.g. cognitive tests) or asset-dependent (fluids) can be an obstacle to the use of the EBM and to a scalable tool.

“Certainly MMSE (Mini-Mental State Examination - cognitive test) is very rough test but very commonly used, then yes, the test might depend on the site. I think cognitive tests used are very site dependent, unfortunately. That needs standardisation.” (P1, R+C)

Regarding imaging scans, the EBM uses volumetric data as input, which would not be available to clinicians. This will require an additional pipeline to the tool’s back-end system. Generally, R+C contexts have facilitated access to higher quality data and a workflow that will accommodate the future tool’s usage tasks, such as the systematic recording of data in a digital database. Once the diagnosis is reached, the patient is discharged. Thus, clinicians’ main interest is to follow them up in the early stages and during the eventual progression to more severe ones. Interestingly, the tool was envisioned as a useful resource to facilitate the discussion process in MDTs:

“In one MDT session a multitude of cases are discussed and that implies remembering numbers of all tests, eventually compared with previous scores and other information. It is hard to keep track of all information for each.” (P6, R+C)
Objectives and values. This dimension highlighted attitudes towards the prospective tool’s benefits and limitations, uncertainty, and interactions with patients. All interviewees affirmed they would require support in the diagnosis, with extra information on stages and progression. However, it is commonly recognised that the EBM would be primarily suitable for clinical trials and research settings.

“Where this can be very important is if you are going to do a trial for example or if you are going to pick any offered treatment for people in a particular stage, that might be helpful.” (P5, R+C)

The majority of clinicians were concerned about a precise numerical outcome for a stage, preferring a degree of uncertainty. This would allow them to combine the outcome from the EBM with their expertise. Regarding disease progression, clinicians would remain vague and hypothetical. The C-O context was more patient-focused, and concerned about how to communicate these contents to patients.

“We talk in vague terms about what to expect in the future. [...] we don’t tend to talk very much in time spans because that’s so difficult.” (P2, C-O); “if you put a scale up like that [...] and say right, you’ve just started having memory problems, it’s all very mild, that means you are at stage 5 out of 12, [the patient might reply] ‘well, I think you blind me, I am half-way gone!’, whereas we know that, although the disease may be present in their brain for 10 years, symptomatically they are at the beginning of that journey.” (P3, C-O)

Conversely, the R+C context was more disease-focused, highlighting the value of the EBM in advancing research, or as a support tool in daily practice.

“In this sense a visualisation would work. What we do now is doing that in our head, we compare with our experience, but it would be interesting to see how computers think” (P6, R+C)

Staffing and skills. Participants reported medium/high technical skills in using electronic medical software and inspecting MRI scans, given appropriate resources
and training. They found the proposed visual stimuli clear and intuitive to use, although the output was hard to interpret from a medical perspective.

**Management.** While most of the management issues are influenced by external factors (such as availability and quality of resources), the internal management structure of both C-O and R+C contexts is adequate to adopt a CDSS tool for the EBM, given regular quality checks and system assessment performed by a technician. Clinical champions with unique characteristics might be beneficial to encourage openness to innovation within the team, even in less suitable contexts. The implementation of the EBM could also address legal matters, as highlighted by one participant:

“[Staging] has a number of consequences with regard to reimbursement of medication and also to the driving capacities etcetera...” (P1, R+C)

**Other resources.** The most relevant factors within this dimension were cost and time-related issues. The CDSS should not have a high impact on additional tasks or time-on-task. However, gathering all required data and entering them in the tool will require additional steps that have to be carefully considered in the system’s design. Costs have not been explicitly assessed yet, excluding those already allocated for medical examinations.

Given that two main different contexts of adoption emerged from the previous analysis, I assessed both with DRGm, and the results are reported in Table 3.3. Each dimension provided an estimate of the actual gap between design and reality. The rating scale is described in Hawari and Heeks [181], where 0 indicates the absence of gap, 5 is a moderate degree of mismatch, and 10 represents a profound divergence between design and reality. Ratings were assigned and discussed with the research team. Single values are then combined to provide an overall score. According to the work of Hawari and Heeks [181], the total score can be predictive of failure or success. The context related to C-O received a score of 29 (“Project might fail totally, or might well be a partial failure, unless action is taken to close the design-reality gap” [181]), while the R+C scored 22 (“Project might be a partial failure” [181]).
Table 3.3: Assessment of Heeks’ dimensions for the two contexts. [*critical score when equal or higher than 5.]

<table>
<thead>
<tr>
<th>Heeks’ dimensions</th>
<th>Care-Oriented</th>
<th>Research+Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>5*</td>
<td>4</td>
</tr>
<tr>
<td>Technology</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Process</td>
<td>6*</td>
<td>3</td>
</tr>
<tr>
<td>Objectives and Values</td>
<td>7*</td>
<td>4</td>
</tr>
<tr>
<td>Staffing and Skills</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Management</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Other resources</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td><strong>22</strong></td>
</tr>
</tbody>
</table>

*Legend (likelihood as cause of failure): 0-2 very unlikely; 3-4 unlikely; 5 possible*; 6-8 likely*; 9-10 very likely*.

These results reflect a smaller design-reality gap in the R+C scenario, compared to the C-O one.

3.1.4 Conclusions

The findings supported this initial, exploratory work on framing the gap between the innovative technology and the user. This means first understanding the context of use, which was addressed through workflows and user profiles, then exploring facilitators and barriers that might contribute to the success or failure of the novel technology, based on the model proposed by Heeks and collaborators [132]. I found that research-oriented centres represented a narrower gap to adoption in this case, because early adopters would be advantaged to take on the tool thanks to their facilitated access to relevant data, and their need for early diagnosis and quantification of patients’ progression that better matches the tool’s capabilities. However, other internal and external factors can influence this transition, such as the validation and trust in the model, presence of a relevant champion, or characteristics and needs of the environment. The heterogeneity in clinical markers is also an apparent hurdle, with different centres using different tests, which suggests the requirement for the model to be adaptable to the specific set of clinical data available in a specific clinic. Generally, the tool is perceived as having potential, if the hurdles identified...
can be properly addressed. This study was affected by inevitable and acknowledged challenges within healthcare research [133]. From a practical perspective, this includes accessing clinical spaces, involving medical specialists and allocating time with them, recording protected data, or running any tests that can interfere with doctors’ routines. Therefore, parts of the data collection was recorded as notes, which may be more susceptible to transcription errors than audio recordings. Moreover, the sample was not extensive, due to difficulties in recruiting suitable participants and to their limited time availability, but also to the niche target audience considered in this study. However, this number was sufficient to identify challenges, as reported above. Similar studies [144, 130, 150] involve small data sets to test and learn from different frameworks’ applications, the sizes of which are comparable to this one. From a methodological perspective, the DRGm requires the analyst to assign quantitative scoring to each of the seven dimensions. Whilst there is an inherent subjectivity in quantifying qualitative results, the real value from this conceptual contribution lies in the process that led to those scoring [182]. The scoring is presented after the description of findings for each dimension, where the reader could find critical justification for the assigned numbers. Ultimately, scoring is here intended more as a tool for the analyst to think about which gaps are more difficult to address. I extensively used interviews for this study, which on one hand is more suitable for getting a rich input at this early stage of the development (given limited number of participants and the complex scenario), but on the other hand heavily depends on participants’ ability to articulate relevant gaps in the current workflow. In particular, it is hard for prospective users to assess the real impact of a future tool on their workflow, and how they would adapt to innovation (also known as the “task artefact cycle” [183]). This is also why end-users should be kept in the loop at all stages of the development process, as they will become increasingly aware of a tool’s potentials and use.
3.2 Study 2: Understanding Users’ Data Availability

3.2.1 Aims

The first study provided an initial understanding of the various factors playing a role in the considered scenario, the most relevant (but not in a particular order) being the presence of different clinical professionals, contexts, access to resources, and needs. However, I needed to consolidate the understanding of clinical data availability and use, given the importance of this aspect in the functioning of the model, as well as the clinical perspective of a future tool to support reasoning for staging and progression. Therefore, the specific aims of this study are: (1) to expand and reinforce the rationale behind whether and how clinicians make use of information on stages and progression; (2) to assess the availability and accessibility to resources, and how commonly used they are; (3) to gather clinicians’ opinions regarding the proposed model and future tool.

This second study is an online survey, and the choice of this data collection method was motivated by multiple factors. Firstly, it makes the data collection more feasible for the target population, given their limited time availability. Moreover, the previous study was limited to participants in two countries (UK, and Belgium), thus an online survey could reach out to other countries. Finally, the typical constrained responses of a questionnaire were more suitable to test some of the themes and issues that emerged in the previous study, although short open-ended questions were included to provide more granularity into some aspects of the research.

3.2.2 Method

The design of the online survey followed an iterative process, testing and improving each version as represented in Figure 3.6. A first draft was produced based on insights from Study 1 (section 3.1). The questions were initially focused on current practice and future vision on staging, prognosis, and potential of a diagnostic support tool. Version 1 was piloted with 15 clinical and research neurologists, recruited from EuroPOND partners, with the intention to keep their responses if the pilot was successful. Due to extensive changes to the questionnaire, this data was used as
3.2. Study 2: Understanding Users’ Data Availability

feedback to develop later versions of the survey.

Version 2 was then produced, and iterated based on internal and external feedback. The improvements were then incorporated into the Final version, such as including more open-ended questions, highlighting the focus on the early stages of the disease, and minor edits. Personal data of participants were not recorded, but participants could share their email contact if they wished to access a free trial of the system. The email list was alphabetically ordered and not linked to participants’ responses. The approximate time for completion of the survey was 15-20 minutes.

The recruitment strategy involved the participation of EuroPOND partners, clinical connections from partners, and word-of-mouth. The questionnaire was uploaded in the project website (http://europond.eu). Other participants were contacted, using the following criteria: clinicians working with AD and subtypes, with an interest in early detection of the disease. The questionnaire was disseminated through an online form, and responses recorded for each respondent. Data were analysed through descriptive statistics (mostly frequencies, given the large use of categorical data), and inductive thematic analysis for open-ended questions, using the software NVivo 12. Each respondent was recorded as S1 to S25. The full questionnaire is reported in Appendix C.

3.2.3 Results

First, the sample is described. Subsequently, findings are presented according to the three main sections explored through the survey: (1) results on disease staging; (2) results on disease progression; (3) perceived potential of predictive models.
3.2.3.1 Sample Description

Responses were collected from 25 participants. The majority of participants were neurologists \((n = 15)\), while the remaining were psychiatrists \((n = 2)\) or classified as other professions (neuropsychologist, neuroradiologist, or geriatricians; \(n = 8\)). The context of participants was classified according to results from Study 1 (section 3.1), as R+C \((n = 17)\), C-O \((n = 6)\), but 2 participants belonged to a purely research setting. These were embedded in the R+C category. Figure 21 shows the distribution of professions, according to context (A), and the country of work (B), to contextualise responses.

![Figure 3.7: Characterisation of sample. (A) Types of different clinical professionals (Neurologists, Psychiatrists, others), by context (R+C, C-O); (B) Respondents’ country of work.](image)

3.2.3.2 Results on disease staging

Questions regarding staging assessment aimed to explore current clinical criteria, data access, motivation to reason in terms of stages, and impact of early staging of patients. Within the R+C group, 16 participants confirmed they do stage their patients, and same applied for 5 respondents from the C-O context. Standardised diagnostic systems mentioned were the conventional classification mild/moderate/severe \((n = 8)\), while some respondents referred more precisely to the NIA-AA framework \((n = 4)\). Overall (21 respondents, 16 from R+C group), the main data used to stage patients is cognitive tests (100%), reported symptoms (90%), and patients’ history (90%), whilst only some R+C respondents incorporate
MRI and CSF data into their routine (see Figure 3.8).

Respondents agreed in considering patients’ staging as a key step in the assessment procedure, particularly at early stages, although this is much more relevant in R+C context (R+C average rating 4.4 on a 5-point Likert scale, versus 3.1 for C-O context). The rationale behind the importance of early staging was explored through a multiple choice question, where the majority (irrespective of the context) reported it being relevant to patients’ awareness (n = 15), treatment’s choice (n = 12), and legal purposes (n = 9). Four main reasons could be identified on supporting (or not) patients’ staging: (1) always needs to be considered, (2) it depends on various factors, (3) it depends on treatment, (4) valuable at early stages. In the R+C group, there was a stronger preference for staging being a relevant data for future treatment-related discoveries.

“The staging of prodromal disease will be most important because of the potential for secondary prevention strategies.” (S8, R+C); “Staging may influence symptomatic treatment choices in established disease (i.e. beyond prodromal stage).” (S2, R+C)

It is clear how various factors can play a role in adopting a staging approach, and in fact most responses considered this option:
“In the absence of disease-modifying treatments, it all depends on the patient’s choice or decision: the right to (not) know.” (S15, C-O); “Always relevant, only irrelevant if patient has another life threatening illness and poor life expectancy.” (S10, R+C)

On the other hand, accurate staging was not considered a relevant piece of information in later stages of the disease from three participants (S3, R+C; S5, R+C; S19, R+C). The process of assigning a specific stage is generally supported by a number of official and less official frameworks, and is an acknowledged element in patients’ assessment. Although staging is currently based mostly on clinical data, MRI and fluid biomarkers are key in determining the disease stage in R+C. This is the context where there was more interest towards early stages, as a proxy for understanding the potential of a new treatment. In some of the open-ended responses, both R+C and C-O highlighted the importance to consider individual patients’ choices as a right to know about their condition.

### 3.2.3.3 Results on disease progression

Questions on disease progression followed a similar pattern to the staging section. All but one of the participants affirmed they do assess disease progression of their patients, making it a clear need for both groups. According to respondents, progression assessment is mostly based on cognitive tests’ results, patients’ symptoms, and patients’ history. Again, R+C specialists can rely on imaging data. All participants stated there are no standardised criteria to assess disease progression. Mostly, participants declared to consider speed of decline (n = 10) and patients’ complaints and history (n = 10) as the most sensitive features to detect progression patterns.

The importance of determining progression early in the disease course was more relevant to the R+C context (R+C average rating 4.3 on a 5-point Likert scale, compared to 3.5 for C-O context). The rationale for this statement related to supporting patients and families (66.7%) and their disease’s awareness (50%), in both contexts. Additionally, a good percentage of respondents in the R+C group referred to therapy plans (n = 6) and legal issues (n = 4). Two participants did not consider prognostic indications as relevant (n = 1, R+C; n = 1, C-O). Responses on the rel-
3.2. Study 2: Understanding Users’ Data Availability

evance (or not) to assess disease progression overlapped with those in the staging section, but did not follow a specific trend; moreover, responses from the C-O were not sufficient for a comparison between the contexts. Overall, the clinicians’ view on defining the progression at the early stages was affected by more uncertainty, due to lack of longitudinal data, and amount of patients that effectively visit the clinic at the prodromal stages.

3.2.3.4 Perceived potential of predictive models
The interest of participants on a disease staging and progression tool leaned towards a positive vote (R+C average rating 3.7 on a 5-point Likert scale, and C-O rating 4). Within 4 respondents having access to a similar tool, 3 of them considered it seemingly interesting. Results from this section were analysed in terms of implementability and perceived potential. Regarding implementability, data access and possible fit within the workflow were explored. As shown in Figure 3.9 on data access, all respondents indicated availability of cognitive tests across a consistent selection (A). MRI is more common amongst R+C, for a variety of sequences (B). Biological biomarkers, such as CSF (C) and blood or genetic tests (D), are the least accessed. Regarding the tool’s fit within the workflow, the majority of respondents would imagine themselves using it before the follow-up visit (n=16), with the two other relatively frequent scenarios being during patients’ consultation (n=8), and at early stages (n=7). Four participants responded they do not envision such scenario.

There was good alignment on the rationale for the perceived potential of the proposed tool: to support decision making (89%) — especially at early stages and uncertain cases (72%), data understanding and standardisation, e.g. in relation to workflow (56%), and communication with patients and families (61%). Decision making would be aided by “standardised evaluation and interpretation” (S3, R+C; S13, R+C; S15, C-O; S16, C-O) of AD-related indicators, and “facilitated data integration” (S2, R+C; S3, R+C). This would promote “objectivity in early diagnosis and intervention” (S22, R+C). Additionally, one respondent expressed the need of a tool that is “brief, inexpensive, and reliable” (S9, R+C). Particular attention was given to the communication with patients and families, as the tool “might provide
additional answers to patients and relatives” (S18, R+C), in a quantitative form (S13, R+C), such as a risk and prognosis score (S13, R+C). In fact, an early intervention would facilitate this process, given that “Prodromal Alzheimer’s disease (MCI with biomarker) are at risk of developing dementia, but they do NOT have dementia. Assessing their risk of conversion is important for taking care of these patients (memory clinics, drugs...)” (S20, R+C; S16, C-O), their carers (S10, R+C), and would “enable the patient to take care of the important life related issues” (S14, R+C; S16, C-O).

3.2.4 Conclusions

Study 2 extended the initial exploratory research. Staging and progression represented accepted clinical concepts in the selected population. A main difference between the two concepts lies in the criteria used: whilst there are recognised official criteria for assessing stages, this is not the case for progression. Moreover, the current assessment of staging and progression is largely influenced by the qualitative
nature of clinical work, data access (mostly cognitive tests, and patient’s history), and clinical needs. Therefore, two scenarios can be pictured for the presentation of staging and prediction information to future users. Preventative staging, based on currently adopted criteria, would facilitate not only early detection of the condition, but will also allow to compare and contrast different clinical cases thanks to the additional quantitative information. This would aid clinical decision making when assigning an intervention to a specific patient. Prediction, on the other hand, does not benefit from established criteria. However, a possible scenario that is suggested by participants, when a disease-modifying treatment will be available, is the ability to identify what course the disease will follow for a particular patient, allowing to plan ahead, personalise, and monitor their treatment more accurately. Having these resources at hand and embedded in the clinical workflow might also establish higher standards of care, saving time and resources.

A controversial topic is around the communication to patients, where concerns arise on whether this innovation could support patients’ knowledge and awareness or might, on the other hand, bring more confusion and frustration. Given that part of the recognised potential matches with expressed needs (such as decision support, future prevention of AD, data understanding, and standardisation of data interpretability), the tool can overall be currently seen as positively perceived by the sample of clinicians considered. However, some respondents represented an exception, and barriers for the adoption should be taken into account.

It is important to highlight that these outcomes cannot be currently generalised, due to the small number of respondents (particularly from C-O subgroup) and the fact that most of them are from partners and centres related to the project. Although specific recruitment strategies have been pursued (i.e. involving centres within and outside the partner centres, promoting dissemination, snowballing and back-snowballing), it proved challenging to involve the considered population. This could have been caused by the target users being such a narrow sample of the population, but also by the highly innovative scope of the project, which might find major support by clinicians that are familiar with the project’s objectives. Another
limitation is given by the type of study conducted: an online survey can be a good instrument to access a wider population, and to provide more quantitative and comparable insights within participants, but it is limited in the granularity of feedback.

Whilst more data should be collected to validate the insights from the first two studies, what was considered beneficial at this stage was to compare the clinical perspectives with the technical team’s capability, which is addressed in the third study.

### 3.3 Study 3: Defining Technical Potential

#### 3.3.1 Aims

The first two studies produced formative knowledge on AD specialists’ needs and practice, their use of data, and perspectives on the introduction of a CDSS built on a DPM to aid their managing and understanding of AD. However, two parties are involved in the process, and in the present study I focused on a better definition of the technical capability. This implies understanding the possible information that DPMs for AD could provide to clinicians. To this extent, I designed a workshop with the technical experts. This method was considered suitable for the current exploratory stage in the design process, to report initial insights on clinical needs and perspectives back to the technical team and narrow down the scope of the future tool. The advantages of a workshop are the possibility to gain input from many users in a relatively short time, and the richness and quality of the output that derives from participants’ discussion and idea stimulation. The aims of this study were to (1) understand the technical capability of DPMs for AD developed by the computational team; (2) define the technical team’s mental models on the importance of DPMs and their support to the clinical reasoning, how the team speaks about models, and how they would visualise them; (3) identify a set of potential design opportunities for the future CDSS.
3.3.2 Method

3.3.2.1 Participants and setting

Six post-doc researchers within or connected to the Centre for Medical Image Computing (CMIC), working on progression models of neurodegenerative disease, were contacted via email. Five of them accepted to take part on the chosen day. Most of them worked on AD; one works on Huntington Disease, but is familiar with AD data and models. The workshop took place on Thursday the 13th of December 2018 in the UCL Engineering building, between 10 am and 12 pm. A booklet was given to each participant, including a cover page with the agenda, the information sheet and consent form (see Appendix D.1), and a pen. A presentation was displayed on the screen, to guide participants through the workshop (see Appendix D.2). Other materials included: white paper sheets, big white paper sheets for conceptual maps, coloured markers, coloured post-it cards, three audio-recorders, notebook for the moderator. Refreshments were provided.

Figure 3.10: Agenda presented at the workshop.

3.3.2.2 Procedure

The workshop was planned to last 90 minutes. The structure of the meeting (Figure 3.10) was the following: (1) welcome and introduction, (2) generative activities, and (3) closing session. The welcome and introduction part lasted 10 minutes. Here, participants were given one minute each for a short presentation. The generative activities included an initial question (“How do you think disease progression
models can contribute and bring value to clinical decision making and patient management?”), allowing 10 minutes for discussion, and a set of three main activities. Activity 1 targeted technical capability: the group was asked to brainstorm, first individually and then at a group level, all the models they are working on and what information these models could provide to a research clinician working in AD. For the individual generation of contents (activity 1.1), participants were asked to write one model for each post-it card, reporting the data needed to run the model, and how it contributes to clinical practice. Consequently, the group was asked to discuss collectively about the contents created, and group them into categories (affinity diagram; activity 1.2). In activity 2, participants were split in two groups (of 2 and 3 participants), and each one chose few of the contents they generated on the board (activity 1.1 and 1.2). They were instructed to further discuss each category and relative models, particularly regarding data sources, clinical relevance, and possible visualisations. Activity 3 was again a group discussion, sharing the outcomes from activity 2. The closing session allocated 10 minutes to wrap-up and add final comments. The workshop was not pilot-tested, given the high specificity of the contents discussed.

The set of data collected for the workshop was composed of: audio recording of the entire session, audio recording of each group’s session, sketched and written material (post-it notes, affinity diagram map, notes from participants), and written notes from the moderator. All data was converted into a digital format and the session entirely transcribed. Participants are classified as W1 to W5. Inductive thematic analysis was conducted, inclusive of all data sources. All text material was coded by initially identifying nodes, pieces of content that address a specific concept, then grouping nodes into broader themes. The analysis, therefore, covered the following steps: (1) Bottom-up nodes generation; (2) Combination of nodes into themes; (3) Iteration of nodes and themes generation, until saturation (the point at which more coded text confirms current nodes and themes, and no new nodes are created); (4) Creation of a conceptual map.
3.3.3 Results

The thematic analysis generated a total of 20 nodes. These nodes were then grouped into 7 principal themes: technical motivation, perceived clinical motivation, clinical pathways, risk factor, accessibility and adoption, uncertainty, and barriers. The conceptual map reported in Figure 3.11 exemplifies the relationship between themes, that build around 3 areas of interest: (1) initial technical potential, (2) translational thinking, and (3) ideas generation. The initial technical potential represents the basis from which the following steps are built. The translational thinking and ideas generation influence each other iteratively, representing the reasoning process. This cycle is elicited by triggers (scenarios’ description and visualisation examples) on one side, and originates uncertainties and insights on possible barriers on the other. Findings are presented by area and theme.

**Figure 3.11:** Conceptual map representing the process of discussion during the workshop and exemplifying the relationship between areas and themes.

3.3.3.1 Technical potential

From activity 1.1 (to generate contents for all the models available), participants reported 8 items. These were classified according to the clinical need each model aims to address (activity 2). A schematic version of the map generated by the group and related contents are reported in Figure 3.12. The most central and interconnected clinical needs covered by the current items/models (in red) are: staging, prognosis,
stratification, and differential diagnosis. Disease understanding and biomarkers’ discovery relate more to clinical trials. The idea of data-driven workflow emerged from the thinking process: in fact, it is not supported by any specific models, but different models could contribute to this information (dotted arrows). Two themes having a particular influence on the technical potential are the technical motivation guiding models’ development, and the perceived clinical needs, meaning how these are seen and interpreted by participants.

Figure 3.12: Affinity diagram generated from Activity 1. Eight items are reported in the boxes (EBM = Event-Based Model; DEM = Differential Equation Model), whilst the clinical need they might address is reported in red.

**Technical motivation.** Disease progression models are generally data-driven, so they are inspired by the data available, and not directly by clinical need, as one participant explicitly mentioned. What motivates improvements to the models is the goal of making them “more robust” when used with a variety of datasets and conditions.

**Perceived clinical needs.** This theme referred mostly to staging, prognosis, treatment planning, subtyping, and decision support. The focus on providing temporal information through DPMs was particularly stressed. One interesting and important point was that the experts’ definition of “early stages” refers to initial biological
3.3. Study 3: Defining Technical Potential

markers’ abnormality, before observable symptoms’ onset, thus not reflecting clinical definition of “early stages” (i.e. patients would not present to clinicians before symptoms onset).

“I think that’s what they’re more interested in, trying to have model’s estimate of the time, an idea of the time between changes”; “this person has this percentage of [...] eventually getting AD, you might not be able to say how long you need until symptoms. Clinicians would want to know that.” (W1)

Opinions on whether clinicians prefer to interpret the disease considering a confidence interval or a precise number were discordant.

“Clinicians like that confidence interval [...] we think it will be in 4 and a half, they’re gonna say 3 to 6. So, the model or some of the models are ready, they’ve got that sort of built-in.” (W3); “[Clinicians] only need a number that’s like annual rate of change. If you keep going at this rate you’re gonna lose your function by 1 point a year, on this 50-point test.” (W1)

3.3.3.2 Translational Thinking and Ideas Generation

The translational thinking phase stimulated discussion around practical benefits that models could bring to the clinical setting, but also how models are likely to be accepted or implemented in the current clinical workflow. Practical thoughts about models’ optimisation to clinical purposes were also discussed.

Clinical pathways. One of the most discussed topics was around clinical pathways. The suggestion from participants was to combine existent models in order to produce data-driven guidance on the patient’s suggested care journey. According to participants, this application could have a unique impact on the clinical workflow, in terms of costs and clinical decision-making. This idea was defined by participants as data-driven clinical workflow.
“No one’s directly worked on that, but [...] I think there is definitely potential to do that, the kind of things that it could do that they suggest who to follow up, and when to follow them up, and [...] in what order you should collect different markers, depending on the disease stage and at what time you should collect these markers and what added value each marker would give you, and kind of suggest cost effectively.” (W2); “It will also give you a projection of if you collect these other marker then we expect that we will be able to tell you this much more... kind of precise information, and you can decide whether knowing about that projection is useful.” (W1)

The idea originated as a re-thinking of existent models, and was supported by imagining scenarios of application.

“Then you might say ok, you’re gonna need the CSF, or if someone [...] has really kind of classical AD therefore I can use only memory test and have a confident diagnosis, but someone might have kind of a mixture of symptoms you might do MRI or... so that type of things.” (W2)

**Risk factor.** Another theme of interest concerned risk factor, intended as a score within a confidence interval that informs the clinician on whether to undertake certain actions or not.

“The information, that would be probably one of the most important [...] was how we can provide information to the clinicians such as this person has a risk factor of this point disease stage, model stage, you need to see them again in 6 months.” (W1)

Participants imagined a scenario for the risk factor’s application in clinical practice, such as a system providing specific alerts, with flashing lights and notifications.
Accessibility and adoption. This theme was generated by discussion on visualisation options, thoughts around tool’s integration within the clinical workflow, and on translational skills of clinicians. Accessibility to the models’ contents could be promoted by visualisations that take into account analytical skills of clinical specialists.

“[Suitable visualisations are] the ones that require the least amount of data to build, to stage someone and that are closer to human intuition, so what clinicians tend to do is.. think either linearly or they binarise stuff... either you put 'yes' or 'no', or it's a linear prediction. That's it. So any model that can handle that, can be visualised in that way...”
(W1)

To facilitate clinicians’ accessibility to a set of predictive models, the group brainstormed the implementation of a framework that would include all different models with the possibility to plug in and out the chosen ones, case-by-case (W4). On adoption, participants recognised how models’ adoption in clinical practice is likely to start from a restricted group of clinicians and spreading as the models gain more trust from the wider user group. Moreover, they also expected future AD clinicians to have a translational set of skills, to be more prepared and keen on receiving such models.

“If we have lots of models that improve disease understanding and they start getting accepted, then neurologists just hear about it and "alright! it's sequence of events. Cool." [...] And the staging and stuff comes from trusting the models which have shown that they correlate with ... the whole thing will sort of... interrelate” (W1)

Uncertainty. Despite seeing the potentials and the applicability of models in clinical contexts, the technical team wondered whether the models would be needed.

“My perspective is that it might be less a question of "are the models ready?", it’s more a question of ”are they needed?” Like, do people, well.. what is the extra kind of... diagnosis and prognosis information that people need, given that there’s no treatment, I guess.” (W2)
Other unresolved questions from the team referred to the amount of uncertainty the models should be characterised of in order to meet clinical reasoning, and whether disease understanding from the computational point of view could be of any use or interest to them.

**Barriers.** This theme talks about the barriers to adoption as seen from the participants in the role of computational experts. Their comments on barriers are classified into clinical, technical, and translational. Clinical barriers were represented by clinicians trusting their instinct (W3), and making wider use of clinical and demographic indications to guide their reasoning, thus limiting the expected potential of DPMs.

“[Some other times we talked about] how much clinicians make decisions based only on demographic datasets… that seems to be most of the information? Cause in these cases that’s when that information of course it’s useful for what they see… they can’t apply models.” (W3)

From the technical perspective, data used for the models need to have specific characteristics.

“You need [a marker] that is dynamic at that point. Something that can help you to make predictions to accurate staging, and you can’t do that with something that’s not changing…” (W1)

Most of the barriers identified by the group, however, have a translational nature. This included discussion around visualisations and the need to simplify the model’s output, as well as clinical limitations in models’ understanding, which is expected to prevent models from being adopted and used in the short term.

### 3.3.4 Conclusions

The content that participants brought forward in the discussion clearly touched on specific clinical information needs that were encountered in the previous study, such as staging, stratification, or differential diagnosis. During the activities, the thinking process stimulated the generation of new ideas (such as the data-driven clinical
3.3. Study 3: Defining Technical Potential

workflow), but also identified adoption barriers and possible unresolved questions on how to leverage the potential of the models to improve clinical practice. The session was a good reflection on how the problem is framed from a computational perspective. In fact, the core engine in model development is data availability, particularly of those markers that are dynamic along the disease course, and how accurate the output is. During the discussion, it was interesting to observe that the technical experts shared a representation of how clinicians think and how they would prefer information to be displayed and presented. However, mismatches with the clinical approach can be identified, for example in the terminology, purpose, and use of data.

One of the aims of this study was to identify use cases that are seen from the technical team of experts as a relevant set of information to address important clinical needs in predictive models for AD. Those that were strongly discussed in the workshop are: (1) clinical pathways, (2) implementation of temporal information (in particular: time on a specific stage, and time of conversion to later stages), (3) risk scores (to represent confidence intervals and uncertain output), and (4) differential diagnosis or stratification.

This study presented some limitations. The selected set of participants were recruited from a specific research group. However, this aspect was justified by the innovative nature of the computational technology proposed and by limited access to such experts. The fact that participants knew each other and the respective work can be seen as a limitation, as they might have made use of tacit and shared knowledge, without clarifying every aspects of the discussion, but can also be interpreted as an advantage, that might have facilitated the thinking and creative phase of the session. Other limitations concerned the limited duration of the session, and the fact that the workshop was an isolated event, whereas a set of workshops might have been beneficial to produce more complete insights.
3.4 Discussion

The exploratory phase of this project was made up of three formative studies to define the current clinical practice and technical potential. The first study was focused on identifying the early adopters of a future EBM-based CDSS and the suitable context of use (addressed through workflows and user profiles). We found a narrower design-reality gap in the R+C context, which is better placed to initially adopt a future EBM-based support tool for AD, being mindful of the barriers highlighted by clinical experts. We have extended our understanding of clinical context through an online survey, looking at the use and access of data for specialised clinical contexts, the clinical conceptualisation of staging and progression, and the perceived potential for the use of data-driven predictive models in clinical practice. These findings shown how staging is mainly based on cognitive assessments, whilst progression is still a more uncertain concept, although it will be relevant when a disease-modifying treatment becomes available. The last study explored the technical perspective on the potential brought in by DPMs in AD, and in which way models could be leveraged to support and inform AD clinical practice.

3.4.1 Defining the Design-Reality Gap

These three studies not only reinforced the understanding of the subject matter from both technical and clinical sides, but most importantly they guided the exploration of possible dimensions that constitute this space. The main parameters identified are: the disease stage, and the environment. These parameters modulate the possibility to define the translation as possible or unlikely. Other important parameters identified are clinical needs, and availability of dynamic data. In the first case, the needs indicate the specific information required by a certain context dedicated to patients at a certain stage of the disease to efficiently support the clinical work. Satisfying the right needs should result in whether the tool would ultimately get adopted or not. The access to data is similarly determined by the context and the particular stage of the disease course. In this specific case, availability of suitable resources to run predictive models determines whether the technology would be used or not. In particular, clinicians expressed the concern around clinical tests being
3.4. Discussion

very specific for different centres, which might hinder the potential of the models to be more widely adopted.

![Diagram](image)

**Figure 3.13:** An approximate representation of the dimensions describing the technical-clinical gap, and the area that the future tool should cover to meet translational opportunities.

In Figure 3.13, these factors are represented on an imaginary timeline that refers to the course of the disease. Their coverage according to the timeline is represented as a green line, which is an approximate reference for when this factor is most relevant at a certain stage, and it is based on background research from Chapter 2 and the findings from the first three studies. At different stages, different data can be considered as dynamic, thus appropriate for the model’s purposes. However, the access to such data will depend on the context: specialised centres would have facilitated access to more sophisticated markers, because of their need for early detection of the disease and uncertain cases. The aim of this illustration was to identify the most appropriate end-user group. As Study 1 (section 3.1) suggests, the clinical specialisation is not as influential as the context, as long as the clinician belongs to a specialised research centre (R+C), with related needs and resources. Ultimately, clinical needs around early detection, staging, and progression are characteristic of preclinical to late MCI stages, whilst patient daily care is at the core of clinical
Table 3.4: Table summarising barriers and facilitators related to the exploratory phase of the project.

<table>
<thead>
<tr>
<th>Barriers</th>
<th>Facilitators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IT factors</strong></td>
<td></td>
</tr>
<tr>
<td>Does not use clinical terminology</td>
<td>Visualisation and team discussion</td>
</tr>
<tr>
<td>Lack of validity and reliability</td>
<td></td>
</tr>
<tr>
<td><strong>User factors</strong></td>
<td></td>
</tr>
<tr>
<td>Patient-focussed clinicians</td>
<td>Disease-focussed clinicians</td>
</tr>
<tr>
<td>Not properly fitting workflow</td>
<td>Need to assess early or uncertain cases</td>
</tr>
<tr>
<td></td>
<td>Familiarity with the EBM</td>
</tr>
<tr>
<td></td>
<td>Presence of a champion</td>
</tr>
<tr>
<td><strong>Contextual factors</strong></td>
<td></td>
</tr>
<tr>
<td>Lack of data quality</td>
<td>Availability of specialised data</td>
</tr>
<tr>
<td>C-O settings</td>
<td>R+C settings</td>
</tr>
<tr>
<td>Costs</td>
<td></td>
</tr>
</tbody>
</table>

scope towards the later stages. The fading tone of lines suggests that there is not a clear-cut and caution should be used in generalising this representation, given the heterogeneity of this condition and of clinical practice in different environments. The blue dotted line represents the span of factors best matching the future tools’ potential with clinical needs.

Having acquired a better understanding of these interrelated parameters, I will now discuss arguments around specific barriers and facilitators on the future adoption of such tool that can be identified from this formative study.

### 3.4.2 Opportunities for Adoption

In this work, the design-reality gap affects three main areas: IT factors, user factors, and contextual factors. IT factors are the elements of innovation proposed by the EBM and the tool; user factors are represented by the characteristics of the users, their expectations and beliefs; contextual factors indicate elements of the setting, resources, and external variables. Each of these areas contributes to define barriers and facilitators regarding the adoption of the tool considered in this study (Table 3.4).

Within IT factors, one barrier to adoption of a CDSS built on the EBM is the lack of validity and reliability. Whilst the EBM has proven potential, its evidence has only been tested with research datasets. To overcome this barrier, the EBM needs to be tested in a controlled healthcare environment, and replicate its results
based on a more diverse set of data. As for interpretability, a critical factor is the use of terminology. In this specific case, the EBM is programmed to detect very early stage AD, before any symptoms occur. However, ‘early stages’ in clinical terms are only defined when patients start to show visible symptoms, which would feature as mid-way through in the EBM staging. Therefore, this discrete numbering of stages would not suit current clinical terminology, and might also disrupt the communication with patients. What would facilitate the uptake of this CDSS is a clear and simple visualisation of the EBM output, designed based on the current clinical terminology and data used, and that could particularly support team discussion.

User factors include attitudes, meaning the perceived usefulness of the EBM, but also the benefit gained from it. Clinicians who focus more on the disease, because of their need to assess early or uncertain cases, were more open to the advantages brought by the EBM. Some study participants were already familiar with the model, and could see aspects of it that might improve clinical practice and early detection of at-risk individuals. Whilst specialists did not find the presentation of a precise stage number useful, they suggested the inclusion of a degree of uncertainty. The EBM does provide this information, and including it in the visualisations will supplement the clinician’s expertise.

Other implications around user factors are mostly interrelated with contextual ones, since users’ needs and practices are influenced by their context of work [91]. The EBM tool was judged useful for MDT meetings, to converge perspectives from multiple disciplines towards a consensus diagnosis. We also noted that the EBM is not currently applicable to all contexts. This can be defined as an early-late stage continuum, where contexts addressing early stages of AD (such as R+C) are focused on early detection, risk of conversion, and uncertain cases, whereas needs for late stage contexts do not properly match with the EBM purposes, as they focus more on accompanying patients through their daily-life, when the condition is evident. This is an important finding, as it highlights that the EBM is likely to be more beneficial for one context than another, which will influence subsequent design decisions. However, the presence of a champion can represent an exception to these
C-O contexts, e.g. by bringing a more systematic and specialised data collection approach in their team.

The exploratory studies determined that a research-oriented context is the most suitable to begin with, as the EBM will address a gap in the quantitative assessment of early stages of the disease and data-driven evolution, given multiple sources of biomarkers. However, it will be necessary to overcome current barriers to adoption related to technical, organisational and social assets. The most disruptive are the need for rigorous validation, adequate fit within context, and resource accessibility (particularly highlighted in Studies 1 and 2). Clarifying all these can contribute not only to guiding the research on the tool’s development, but also to a higher probability of translating the EBM into a tangible technology.
Chapter 4

Design Thinking and Testing

This chapter:

• Outlines the list of requirements defined from preliminary formative research;

• Describes the design thinking process, from narrowing down the chosen case study for this project, to describing the various design iterations and the rationale that led to the prototype used for testing;

• Reports on a study which tested the prototype with relevant clinicians (Study 4).

Preliminary results for Study 4 were presented at [C.3]. Study 4 has been submitted to [J.4].

The initial exploratory research provided the basis to understand the current clinical and technical scenarios, and defined needs that could inform the next stages in the tool’s design iterations. This chapter represents the design thinking and testing phase of the work (Figure 4.1). The first section describes the requirements derived from the exploratory research in Chapter 3, informing the next steps in the interface design process. The second section starts by referring to the different case studies proposed in the technical workshops and how, based on the research and requirements, the design thinking evolves until focusing on one concept that represents a narrower gap between clinical needs and technical potential. Subsequently, different iterations of possible interface designs are documented, until a first testable
option is reached. The last section describes the study in which the design concept is tested with the identified clinical early adopters.

4.1 Design Requirements

Various feedback and input were gathered in the exploration of needs and current practice from a clinical perspective (sections 3.1 and 3.2), and a better understanding of the technical potential was developed through the technical experts workshop (section 3.3). Altogether, these findings contribute to the definition of a set of requirements that represent more practical user-centred recommendations to guide the next steps in the design process and tool’s implementation.
Each individual requirement has been formulated similarly to the 3-part format of needs definition proposed by Nielsen Norman Group [184]: \([a \text{ user/item}] \text{ needs [be/do] something in order to accomplish [goal].}\) In this case, needs is replaced with \textit{should}. I grouped requirements by \textit{area} (Technical, User, and Organisation factors), each covering different \textit{dimensions}, defined by \textit{evaluation measures} (Table 4.1). This classification is not only supported by literature on design guidelines for CDSS ([18, 19, 20, 21], see section 2.4.3), but it also recalls the areas for which barriers and facilitators were defined in section 3.1. The full table of requirements (together with the second iteration presented in section 5.2) is reported in Appendix G.

The initial set of requirements pertaining to technical, user, and organisation factors are reported and described in detail in the following subsections. The parameters in the requirements table (e.g. Table 4.2) are the following:

- \textit{Dimension and evaluation measure}: specific subcategories;
- \textit{ID}: unique identifier for a requirement. The ID is built based on a tree-like structure: [category]-[dimension]-[n\_evaluation]-[n];
- \textit{Requirement}: the descriptive text of the requirement;
- \textit{Evidence from research}: which study within the exploratory phase in chapter 3 (Study 1: clinical interviews; Study 2: survey, Study 3: tech workshop) contributed to the requirement;
- \textit{Ref}: Literature supporting the requirement (Horsky et al., 2012 [18], Kilsdonk et al., 2017 [19], Miller et al., 2018 [20], Sutton et al., 2020 [21].)

### 4.1.1 Technical factors

Requirements related to technical factors cover the feedback on the EBM rationale and the future system’s functioning. Requirements are described along three dimensions: system quality, information quality, and information delivery.
# Chapter 4. Design Thinking and Testing

**Table 4.1:** Structure of requirements’ tables.

<table>
<thead>
<tr>
<th>Area</th>
<th>Dimension</th>
<th>Evaluation measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td>System quality</td>
<td>Data quality, Validation, Reference population</td>
</tr>
<tr>
<td></td>
<td>Information quality</td>
<td>Relevance, Completeness, Data entry methods, Reliability</td>
</tr>
<tr>
<td></td>
<td>Information delivery</td>
<td>Input, Output</td>
</tr>
<tr>
<td>User</td>
<td>System (intention to) use</td>
<td>Expectations and beliefs, Knowledge and expertise, Motivation</td>
</tr>
<tr>
<td></td>
<td>System understanding</td>
<td>Training, Interaction, Trust and uncertainty</td>
</tr>
<tr>
<td>Organisation</td>
<td>Structure</td>
<td>Workflow, Resources</td>
</tr>
<tr>
<td></td>
<td>Environment</td>
<td>Costing, Support</td>
</tr>
<tr>
<td></td>
<td>Interoperability</td>
<td>Standards, Third party access</td>
</tr>
</tbody>
</table>

## 4.1.1.1 System Quality

System quality includes the following items: *data quality, validation, and reference population* (Table 4.2). Generally speaking, system quality refers to the system performance. In the present case, given that the system is currently represented only by the algorithm, we refer to the EBM performance.

Requirements pertaining to data quality refer to the nature and characteristics of the markers used to feed and train the model. Inadequate data quality is acknowledged in the literature as one of the main causes of AI / ML inconsistencies [165, 166]. Here, to ensure EBM produces a reliable and unbiased output, data should generally be of high quality (IT-SQ-1-1). In particular, this means that data should be collected according to specified criteria (IT-SQ-1-3), and should be dynamic (IT-SQ-1-2), to allow longitudinal monitoring of changes related to the disease progression.
Clinicians mentioned the importance of conducting and reporting a validation of the model, as evidence of the system’s quality and its applicability to the clinical setting. The model has been validated against other neurodegenerative diseases [84, 69, 83, 185], and tested with hospital clinic data [2]. Validating the model on clinical practice data (IT-SQ-1-4) would consolidate the quality of the system for its use in real settings. Similarly, the reference population needs to be consistent with the population the system is intended to be used by (IT-SQ-1-5). Different countries or cultures can be characterised by different ranges of abnormality for each marker, and this has to be taken into account in the training phase and in the interpretation of the results. Therefore, the user should be able to set a specific reference population that will be used in the EBM training and generation of results.

4.1.1.2 Information Quality

Various measures account for information quality, such as relevance, completeness, data entry, and reliability (Table 4.3), all contributing to the information produced by the system as the output of the EBM. It is important to highlight how some of these measures can be subjective, as they come from the user’s perspective.

The fundamental requirement pertaining to relevance is that input markers
Table 4.3: Technical factors: information quality requirements.

<table>
<thead>
<tr>
<th>Evaluation Measure</th>
<th>ID</th>
<th>Requirement</th>
<th>Evidence from research</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance</td>
<td>IT-IQ-1-1</td>
<td>Relevant cognitive tests should be included in the input set to improve the information quality</td>
<td>Clinical interviews</td>
<td>[19, 21]</td>
</tr>
<tr>
<td></td>
<td>IT-IQ-1-2</td>
<td>The heterogeneity of some markers should be considered in order to tailor the set of information used so to be relevant to the specific context/user</td>
<td>Clinical interviews</td>
<td>[19, 21]</td>
</tr>
<tr>
<td></td>
<td>IT-IQ-1-3</td>
<td>Only data pertaining to AD should be included in the input requirements to run the models, to promote relevance of data used and of output provided</td>
<td>Clinical interviews, Survey, Tech workshop</td>
<td>[19, 21]</td>
</tr>
<tr>
<td></td>
<td>IT-IQ-1-4</td>
<td>Terminology and scales used in the system’s interface should reflect clinical terminology to ensure the information provided is relevant to clinicians</td>
<td>Clinical interviews</td>
<td>[18]</td>
</tr>
<tr>
<td>Completeness</td>
<td>IT-IQ-2-1</td>
<td>Data used to train the model should be consistent for different patients to ensure a proper and informative output</td>
<td>Clinical interviews, Tech workshop</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>Data entry methods</td>
<td>IT-IQ-3</td>
<td>Volumetric data from the MRI should be automatised to facilitate data entry</td>
<td>Clinical interviews</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>Reliability</td>
<td>IT-IQ-4</td>
<td>The information provided as output from the model should be representative of the data given as input, the specific population, and resonate with clinical knowledge.</td>
<td>Clinical interviews</td>
<td>[19, 21]</td>
</tr>
</tbody>
</table>

should be typical of the AD condition (IT-IQ-1-3). It was also found in the clinical interviews and survey (section 3.1, 3.2) that the heterogeneity of various markers, particularly cognitive tests (IT-IQ-1-1, IT-IQ-1-2), could undermine the quality of the EBM’s output. Clinicians reported not being familiar with some of the cognitive markers used as input to the EBM, and absence of other tests that they normally use. The issues around heterogeneity, however, have an even greater impact, since different centres report using different tests to examine cognitive impairment. Running the EBM using a customised selection of tests for each centre would require an appropriate training dataset, validated on the target population. This approach does not seem to be realistic in terms of scalability and generalisability, and would go against the model’s aim to unify the way in which AD cross-sectional data (including clinical) is integrated at a large scale. Local, bespoke models would also require large datasets for each single centre, which is unrealistic. However, the local variations can still be accounted for, given that cognitive and behavioural test scores are standardised based on the relative population data. Additionally, the barrier of using different types of tests might be overcome through the concept of cognitive composites, presented in Section 2.1.4 [55]. Another important concern raised by
the clinical specialists referred to misalignment in terminology, and in particular the different use of “early stages” by clinicians and technical experts, as discussed in the previous chapter (IT-IQ-1-4). Therefore, the terminology used in the system should be clinically relevant.

To ensure completeness, data should be consistent across patients (IT-IQ-2-1). Even though the EBM can work with missing data if needed, consistent data would facilitate comparisons within and between patients. Interviewed clinicians reported the need to automate some data entry procedures to ensure that data quality is maintained in the process from collection to system output. This particularly refers to MRI volumetric segmentation and the need for this process (of segmentation and data input to the platform) to be automatised (IT-IQ-3). Finally, the requirement on reliability recalls concepts from data quality and validation, but this time focusing more closely on the output from the EBM, which should be representative of the input data and population, and should resonate with clinical knowledge (IT-IQ-4).

### 4.1.1.3 Information Delivery

The information delivery dimension concerns requirements on how the input and output interfaces should be designed to facilitate the interaction between the system and the users (Table 4.4).

The visual stimulus used in the clinical interviews for the input screen was well received, but participants suggested ways in which the process could be improved. Firstly, the steps to follow and type of data to input should be intuitive (IT-ID-1-1). Clinicians suggested grouping data entries according to the different type of markers, designing marker-specific data-entry fields or a drop-down menu (IT-ID-1-2). Demographic information of the patient should also appear in the input screen (IT-ID-1-3).

The output screen was more controversial. Generally, the output should not be too overwhelming to users (IT-ID-2-1). Clinicians commented on the need to have the available input data always visible to review in the output screen (IT-ID-2-2), and to be able to browse results from previous assessments, to facilitate comparisons (IT-ID-2-3). On the topic of terminology and how it should be adapted to clinical
<table>
<thead>
<tr>
<th>Evaluation measure</th>
<th>Requirement</th>
<th>Evidence from research</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>IT-ID-1-1</td>
<td>Steps required to input data in the platform should follow an intuitive logic to support a correct input of data</td>
<td>Clinical interviews [18]</td>
</tr>
<tr>
<td></td>
<td>IT-ID-1-2</td>
<td>Input of patient’s clinical information should be required to promote completeness of data</td>
<td>Clinical interviews [18]</td>
</tr>
<tr>
<td></td>
<td>IT-ID-1-3</td>
<td>Sections for different types of markers, with indications for value intervals and reference population, should be clear and separated to support a more intuitive input screen</td>
<td>Clinical interviews [18]</td>
</tr>
<tr>
<td>Output</td>
<td>IT-ID-2-1</td>
<td>The information presented on the screen should not be too overwhelming and dense to promote a better understanding and delivery of information</td>
<td>Clinical interviews [18]</td>
</tr>
<tr>
<td></td>
<td>IT-ID-2-2</td>
<td>The interface should always keep the available input visible to offer clinicians an overview of their available data</td>
<td>Clinical interviews [18]</td>
</tr>
<tr>
<td></td>
<td>IT-ID-2-3</td>
<td>The system’s output interface should allow to easily browse past assessments to facilitate comparison</td>
<td>Tech workshop [18]</td>
</tr>
<tr>
<td></td>
<td>IT-ID-2-4</td>
<td>The output should include a confidence interval to allow for clinical interpretation</td>
<td>Clinical interviews [18]</td>
</tr>
<tr>
<td></td>
<td>IT-ID-2-5</td>
<td>The scales used to represent the model output should reflect clinical current labelling and terminology, to facilitate the information delivery</td>
<td>Clinical interviews, Tech workshop [18]</td>
</tr>
</tbody>
</table>

Domain-knowledge, staging labels were described as not intuitive to read by clinicians, therefore a different way to represent the stages of progression needed to be used in the output screen (IT-ID-2-4). Finally, clinicians, felt uncomfortable with receiving a unique number as output, both from an interpretative perspective and for communication to patients. In this case, including a confidence interval in relation to the staging classification could enrich clinical reasoning, whilst leaving room for clinical interpretation (IT-ID-2-5).

### 4.1.2 User factors

The importance of human factors in the interaction with the technology is reflected in the requirements classified as system (intended) use and system understanding.

#### 4.1.2.1 System intended use

The exploratory research supported an initial understanding of the clinical perspective on the value and use of a EBM-based CDSS, and relevant feedback is integrated in the following measures: expectations and beliefs, knowledge and expertise, and motivation to its use (Table 4.5).
### Table 4.5: User factors: system’s intended use requirements.

<table>
<thead>
<tr>
<th>Evaluation measure</th>
<th>Requirements</th>
<th>Evidence from research</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expectations and beliefs</td>
<td>The system should provide information on stages and progression to support clinical decision making</td>
<td>Clinical interviews</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>UF-SI-1-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectations and beliefs</td>
<td>The system should provide information on early stages of the disease to allow a timely intervention</td>
<td>Clinical interviews</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>UF-SI-1-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge and expertise</td>
<td>The labels used in the output classification should reflect the knowledge and expertise pertaining to the clinical specialists to facilitate interpretation and understanding</td>
<td>Clinical interviews</td>
<td>[18, 19, 21]</td>
</tr>
<tr>
<td>UF-SI-2-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge and expertise</td>
<td>The system should not provide a unique numerical outcome for a stage to account for clinical interpretation</td>
<td>Clinical interviews</td>
<td>[18, 19, 21]</td>
</tr>
<tr>
<td>UF-SI-2-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>The system should integrate all markers in a unique visualisation to facilitate data understanding</td>
<td>Clinical interviews, Survey</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>UF-SI-3-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>The system should include quantitative data to support an objective understanding and interpretation of the condition</td>
<td>Clinical interviews, Survey</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>UF-SI-3-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>The system should support understanding of the progression for uncertain cases to support clinical management</td>
<td>Clinical interviews, Survey</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>UF-SI-3-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>The system should be used to facilitate the discussion process in MDTs</td>
<td>Clinical interviews</td>
<td>[19, 20, 21]</td>
</tr>
<tr>
<td>UF-SI-3-4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To be used and adopted, the system should align with users’ expectations and beliefs. According to earlier findings, clinicians would expect such system to provide patient-personalised information on the staging and progression of AD as an aid to their current, less quantitatively focused, clinical practice (UF-SI-1-1). More precisely, clinicians would expect such a system to support early and uncertain cases, which are the ones that are better placed to take advantage from such a data-driven system (UF-SI-1-2). On the other hand, it is important that the system reflects the clinical domain expertise and knowledge, by using clinically relevant labels for stages (UF-SI-2-1) as previously discussed, and that it is aimed at supporting clinical judgement, not replacing it. Therefore, the output should account for a range of probabilities, rather that providing a precise number of the stage the individual patient is at (UF-SI-2-2).

Despite the system not being developed and used yet, it was revealing to observe how positive clinicians were about its use in clinical practice. Clinicians reported seeing the value of the proposed tool in integrating different markers to provide a unique picture of the disease (UF-SI-3-1). Not only can this facilitate the understanding of possible progression trajectories for uncertain cases (UF-SI-3-2),
**Table 4.6: User factors: system understanding requirements.**

<table>
<thead>
<tr>
<th>Evaluation measure</th>
<th>ID</th>
<th>Requirement</th>
<th>Evidence from research</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>UF-SU-1</td>
<td>The clinical users should receive an adequate level of training and understanding of the model to use the system and be able to interpret its output</td>
<td>Clinical interviews, Tech workshop</td>
<td>[18, 19, 21]</td>
</tr>
<tr>
<td>Interaction</td>
<td>UF-SU-2</td>
<td>The system should include interactive features to allow clinical users’ on-demand exploration and investigation of information</td>
<td>Clinical interviews</td>
<td>[18]</td>
</tr>
<tr>
<td>Trust and uncertainty</td>
<td>UF-SU-3</td>
<td>The system should allow a degree of uncertainty, to account for clinical interpretation</td>
<td>Clinical interviews</td>
<td>[19, 21]</td>
</tr>
</tbody>
</table>

but it can consolidate clinical practice with quantitative information, where qualitative data are currently dominant (UF-SI-3-3). One particular setting where the system could find an interesting application is in MDT meetings, as a discussion facilitator (UF-SI-3-4).

### 4.1.2.2 System understanding

To facilitate the users’ understanding of the system, this group of requirements covers three main measures: training, interaction, and trust and uncertainty (Table 4.6).

Training would be required for the clinical users, to ensure the correct use of the new system and understanding of its nuances (UF-SU-1). However, the system should also include interactive features that allow exploration and customisation of the data and output provided (UF-SU-2). These features will also give clinicians more control over the output, rather than providing them with a prescriptive statement. Another, less interactive, way to achieve this is to maintain a degree of uncertainty in the EBM output on stages and progression, leaving room to clinical interpretation (UF-SU-3). This feature not only supports the ‘digital conversation’ between the user and the system, but is a trigger to trust, in that clinicians would struggle to trust a system that provides such a precise number out of a very complex and multifactorial condition.

### 4.1.3 Organisation Factors

Contextual elements play a fundamental role in the technical transition of innovation to clinical practice. Requirements extracted by exploratory research are grouped into three main dimensions: structure, environment, and interoperability.
4.1. Design Requirements

Table 4.7: Organisation factors: structure requirements.

<table>
<thead>
<tr>
<th>Evaluation measure</th>
<th>ID</th>
<th>Requirement</th>
<th>Evidence from research</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workflow</td>
<td>OR-S-1-1</td>
<td>The system should be seamlessly integrated within the workflow to facilitate adoption and use</td>
<td>Clinical interviews, Survey, Tech workshop</td>
<td>[18, 19, 21]</td>
</tr>
<tr>
<td>Resources</td>
<td>OR-S-2-1</td>
<td>The technical set-up of the Structure should accommodate the integration of the proposed system to promote seamless adoption and use</td>
<td>Clinical interviews</td>
<td>[18, 21]</td>
</tr>
<tr>
<td></td>
<td>OR-S-2-2</td>
<td>The Structure should ensure access to specific resources to run the model</td>
<td>Clinical interviews, Survey, Tech workshop</td>
<td>[19, 21]</td>
</tr>
</tbody>
</table>

4.1.3.1 Structure

Structure requirements are reported in Table 4.7. One of the key requirements informed by the exploratory studies is the importance of integrating the future tool within the current workflow, improving current clinical practice, without disrupting it (OR-S-1-1). The preliminary analysis of the current workflow (section 3.1.3.1) can, therefore, define opportunities and pain points for the future system. Similarly for resources, the adoption of such a system could be facilitated if the existent technical set-up is able to accommodate the integration of the new system (OR-S-2-1). Moreover, access to relevant data required for the functioning of the EBM should be ensured by the Structure (OR-S-2-1), an element of concern for clinicians, and particularly those belonging to less specialised settings.

4.1.3.2 Environment

Environment requirements are reported in Table 4.8, and focus specifically on costing and support measures. Whilst no costs other than a paid subscription are envisaged for such a system, it must be taken into account that to run the system other costs need to be evaluated (e.g. additional examinations) (OR-E-1). Moreover, a regular technical support needs to be provided as a preventative and safety requirement (OR-E-2).

4.1.3.3 Interoperability

Despite not being directly reported by clinicians and technical experts at this stage of the research, it is clear from literature the importance of designing and develop-
Table 4.8: Organisation factors: Environment requirements.

<table>
<thead>
<tr>
<th>Evaluation measure</th>
<th>ID</th>
<th>Requirements</th>
<th>Evidence from literature</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costing</td>
<td>OR-E-1</td>
<td>Costing should be evaluated to ensure adoption of the system</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td>Support</td>
<td>OR-E-2</td>
<td>A regular quality check and technical maintenance should be conducted to ensure the system is functioning correctly</td>
<td>Clinical interviews</td>
<td>[19, 21]</td>
</tr>
</tbody>
</table>

Table 4.9: Organisation factors: Interoperability requirements.

<table>
<thead>
<tr>
<th>Evaluation measure</th>
<th>ID</th>
<th>Requirements</th>
<th>Evidence from research</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standards</td>
<td>OR-I-1</td>
<td>The system should comply to official standards and be rated towards TRL to acknowledge its level of readiness to be used safely in clinical practice</td>
<td>-</td>
<td>[19, 21]</td>
</tr>
<tr>
<td>Third Party Access</td>
<td>OR-I-2</td>
<td>The system should facilitate access to relevant third parties and information sharing withing and between clinical centres, given the necessary precautions and permissions</td>
<td>-</td>
<td>[19, 21]</td>
</tr>
</tbody>
</table>

ing a system that is interoperable. An interoperable system is, in fact, expected to be developed based on methods or standards (OR-I-1) that facilitate and regulate the quality, safety, interaction, and exchange of information with other software, resources, and routines typical of other clinical and research centres. Such systems are often rated against the Technology Readiness Level (TRL) [186], originally developed by NASA and adopted EU-wide in 2014 to show “how far a technology is from being ready for use in its intended operational environment”. Third party access (OR-I-2) should also be regulated according to General Data Protection Regulation (GDPR) standards, and right to access clinically sensible information should be evaluated case-by-case.

4.1.4 Conclusions

In this section, I listed and described the requirements built on the formative research outcomes from Chapter 3 and from literature, according to a requirement matrix produced and reported in Appendix G. These requirements will inform and guide the design process, documented in the following section.
4.2 Design Process

The technical experts workshop (section 3.3) generated a set of possible case studies as ways in which the models can be translated for clinical AD use. Shown in Table 4.10, they are: data-driven clinical pathway, risk factor, differential diagnosis, and temporal information. Given the insights from the exploratory phase and the definition of requirements, this section expands and considers the different options, with the aim to narrow down the design problem and select the option that should be brought forward in the prototype design. It then describes the iterative process followed in the design of the CDSS prototype.

4.2.1 Exploring the Case Studies

The four case studies and their description are reported in Table 4.10. Here I explore each in turn: data-driven clinical pathway, risk factor, differential diagnosis, and temporal information.

One of the principal suggestions from the technical workshop was the data-driven clinical pathway. This concept entails the idea of leveraging data-driven models to produce a tool that reflects the evolution of the disease, so that it can provide information on the expected sequence of events, but also on the examinations that ought to be performed to improve the model’s prediction. In this form, the concept would address the clinical need of visualising the staging and progression of an individual patient, according to clinical labels, and of merging quantitative information from different sources in an integrated view of patient severity. From a technical perspective, this option reflects the current potential of the EBM in providing information on staging and forecast of future disease progression, including both individual and population data.

The idea for the risk score is to provide an index of the probability for a patient to progress to a more severe stage of the disease. Clinically, this data would be important in assessing the urgency of implementing a care plan to slow down the disease progression. In some cases, as reported in the clinical interviews, a risk score can guide the decision of discharging a patient that is not expected to progress to more severe stages. On the technical side, currently EBM does not produce a risk
Clinic Pathway

The representation of the patient evolution along time, including the clinical steps that the patient has gone through, the expected future events, and possible recommendations on preventative actions to take.

Risk score

A score with a confidence interval that informs the clinician on the risk of the patient to progress to more severe stages of the disease.

Differential diagnosis

The probability for a patient to differentiate between two or more conditions that could be typical of the patient’s symptoms.

Temporal information

The information about the duration of a stage and approximate time to conversion to a more severe stage.

Table 4.10: List and description of the four case studies derived from the technical experts’ workshop.

score value. Osuala & Arandjelovic [11] provided a set of visualisation examples to address the representation of this type of content in a healthcare setting. On the use and communication of a risk score, Van Maurik et al. [130] studied the needs around the use of this type of information and how it could be incorporated in the communication between clinicians and patients/caregivers.

Differential diagnosis is surely an unmet clinical need. The opportunity of excluding other conditions early on is both important and very difficult to implement at a technical level. The SuStAn model is seen as a viable starting point for developing a potential solution to this problem, as it is able to identify sub-stages within the same condition. A possible scenario for the use of differential diagnosis information in clinical practice would include information on how much an individual patient diagnosis resembles a particular subgroup and how much it differs from others. The work from Tolonen et al. [15] represents a good example of how differential diagnosis information could be communicated and displayed.

Lastly, there is an evident clinical need to understand the time a patient would remain in one disease stage or how long it takes for the patient to progress to the next stage of the disease. This has been mentioned as particularly relevant in ascertaining the pace of progression of one individual patient. On the technical side this feature had not been implemented at the time the workshop took place, but has since been developed into an early model [187]. Clinicians could take advantage of
this information to plan the next steps in the care journey, but also to facilitate the communication between them and patients.

4.2.2 Narrowing down the design problem

Predictive models considered here address some of the needs identified for AD clinical practice. My aim at this point was to narrow down the choices to the most immediate design problem, as an overlap between clinical needs and technical capability.

As previously discussed, different types of information can be provided by the models. These can be, but are not limited to:

- Data-driven pathway of the patient evolution;
- Quantitative interpretation of MRI scan;
- Integration of a multitude of AD-related markers into a single output;
- Risk score;
- Differential diagnosis;
- Time until conversion to more severe stages.

It was also discussed how other variables play a role in facilitating the clinical adoption of models, and specifically of the EBM. These refer particularly to the profile of early adopters, which are:

- Neurologists and psychiatrists, or other specialised staff who work in secondary and tertiary referrals or specialised research centres;
- Looking at early stages of the disease and difficult cases in order to disentangle the staging and progression of AD patients;
- Included in settings that can have facilitated access to specialised resources and data required for the models to work.
By considering clinical needs and the early adopters profile, whilst taking into account the current potential of the technology and particularly of EBM as the model (which has shown greater level of readiness in clinical practice), the concept of data-driven clinical pathway was eligible as the one being most immediately addressable both from a technical potential and clinical needs’ perspective. Clinical pathways have been used in healthcare to promote efficiency and quality of treatment approaches. There might be many definitions of pathways, but for this project it is referred to as proposed by Bettencourt-Silva et al. [188] as “an ordered set of patient-centric events and information relevant to a particular clinical condition”. As in their work, the data-driven pathway represented in this project was not intended as part of an intervention, but to describe, analyse, and evaluate the relevant clinical parameters of the condition over a period of time.

The reasons to not consider the other case studies in this project are now presented. All three other case studies were much needed information for clinical practitioners, but the required outputs were not yet supported by current models developed by the team. The differential diagnosis deserves a special mention, as SuStaIn could similarly address the understanding of subtyping within the AD condition. However, given the complexity of the translational and design problem being addressed here, the integration of a subtyping feature could be explored in the future, once the staging and progression design concepts have been studied and consolidated. Finally, it was shown how other initiatives have attempted at targeting either the differential diagnosis problem [15] or the risk score indication [130]. Therefore, addressing the problem of a data-driven clinical pathway represented a gap in this space that EBM seemed suitable to address, and for which there was clear clinical need. To highlight existing works and the issues they ought to address, Table 4.11 reports a comparative-competitive analysis of three tools (previously presented in section 2.3.3): PredictND [129], AmyloidRisk [189], and ADappt [130]. Eight parameters are considered here, with two of them (use of multiple AD markers and risk scores) being common to all three tools considered, whilst four of the investigated parameters, including a patient evolution curve, have not been implemented.
4.2. Design Process

This section describes the design concept brief for a clinical computer interface that provides quantitative data-driven information on the evolution and clinical pathway for individual Alzheimer’s Disease patients. This tool was intended to make use of a novel disease progression model (EBM) to support clinical practice in reasoning about patient pathways in a quantitative, data-driven manner. It should support clinical decision making and patient management, and promote a better understanding of the disease. Therefore, the tool was expected to:

- Showcase patient evolution according to marker changes with reference to a population;
- Provide a forecast of how the disease is expected to progress;
- Support clinical planning and reasoning around the assessments that would be relevant at each stage of the disease.

The tool was intended as to contribute to pivotal challenges in the study of data-driven model integration into clinical practice. The main ones are:

- To study the explainability (and possibilities for translation) of predictive models for AD in the clinical context;

\[
\begin{array}{|l|c|c|c|}
\hline
\text{Use of predictive models} & \text{PredictND} & \text{AmyloidRisk} & \text{ADappt} \\
\hline
\text{Provide future examination suggestions} & N & Y & Y \\
\hline
\text{Use of multiple AD markers} & Y & Y & Y \\
\hline
\text{Risk scores} & Y & Y & Y \\
\hline
\text{Temporal information} & N & N & N \\
\hline
\text{Differential diagnosis} & Y & N & N \\
\hline
\text{Patient pathway / evolution} & N & N & N \\
\hline
\text{Use of clinical data} & N & N & N \\
\hline
\end{array}
\]

Table 4.11: Comparative-competitive analysis of three tools that have developed an interface to display results from predictive models in relation to AD.

4.2.3 Design brief

This section describes the design concept brief for a clinical computer interface that provides quantitative data-driven information on the evolution and clinical pathway for individual Alzheimer’s Disease patients. This tool was intended to make use of a novel disease progression model (EBM) to support clinical practice in reasoning about patient pathways in a quantitative, data-driven manner. It should support clinical decision making and patient management, and promote a better understanding of the disease. Therefore, the tool was expected to:

- Showcase patient evolution according to marker changes with reference to a population;
- Provide a forecast of how the disease is expected to progress;
- Support clinical planning and reasoning around the assessments that would be relevant at each stage of the disease.

The tool was intended as to contribute to pivotal challenges in the study of data-driven model integration into clinical practice. The main ones are:

- To study the explainability (and possibilities for translation) of predictive models for AD in the clinical context;
• To understand the clinical relevance and intended use of introducing data-driven information on the evolution and pathway of patients;

• To develop a graphical tool to communicate data-driven insights effectively, that is designed on initial formative research on clinical needs and practice;

• To support, and not replace, the decision-making process of the clinician, by providing them with new ways of reasoning and combining data that facilitates their thinking.

To address these points, the following design aims were brought forward:

• Create a visualisation that can communicate the output of disease progression models in the form of patient-specific information to clinicians;

• Provide clinicians with a long-term quantitative picture of the entire patient’s disease evolution, in relation to a given population;

• Design an intuitive and easy to understand interface;

• Visually communicate future predictions on disease progression.

Out of the three main tools presented in the comparative-competitive analysis (Section 4.2.2), two in particular inspired the design statement. From PredictND [129], it was important to learn a way to display complex information from a disease staging and prediction model through a dashboard, that encapsulates information within multiple boxes or tiles. However, PredictND interface still looks very technical, and there is no reference to whether and how the users were part of the design process. ADappt [130] covers this aspect, describing in detail how clinicians and patients/caregivers had an active role in the design process, and the qualitative approach.

4.2.4 Design iterations

Based on the design brief, I produced various prototypes of the possible visualisation for the tool, from low- to mid-fidelity versions. In the following subsections,
I present the two main iterations preceding the final version, that was brought into user testing.

4.2.4.1 First Design Iteration

In the first iteration of the design, I focused on exploring possible visualisations for the input screen and the evolution pathway. The general set up was meant for a web-based scrollable dashboard that allows to keep a record of patients, input the required set of markers, and visualise the output from running the EBM.

The input screen (Figure 4.2 a) was organised as a form where markers’ values need to be filled for each tab. Different tabs contain data entries for the different types of information or markers (demographic data, imaging, CSF, clinical), as per IT-ID-1-3 (Table 4.4). The imaging section presented a visual clue, instructing the user to upload the MRI scan (see IT-IQ-3, Table 4.3). Consistently with previous icometrix work [190], volumetric data was expected to be automatically extracted from the MRI file, with no need from the user to input those values manually in the system.

For the output screen (Figure 4.2 b) reporting the EBM results and primarily
the pathway visualisation, the concept was that of a scrollable dashboard made up of various tiles. These were: patient ID data; overview of input data (as per IT-ID-2-2, Table 4.4); main pathway output; interactive timeline bar to navigate between visits (as per IT-ID-2-3, Table 4.4); spider graphs for each biomarkers’ type; colour coded MRI highlighting areas of significant volume change between visits. The timeline bar and the spider graph were inspired by the work from Osuala et al. [191].

Regarding the **pathway representation**, a visualisation that could support the individual patient evolution over time was the Sankey Diagram. One example was provided by CareFlow [192] and Outflow [125], previously mentioned in the background chapter (Figure 2.8). **Sankey diagrams** display flow relationships over a sequence of discrete ordinal stages, where stages are represented as parallel displays. Another type of visualisation for the individual patient’s progression, which was suggested in the technical workshop, was survival curves [193]. **Survival curves**, also called Kaplan-Meier curves, show the probability of an event along a predefined time interval. Both these ideas were integrated in the first sketches of the main output tile (Figure 4.3). On the $x$-axis, I reported the stages as a numerical discrete sequence, whilst on the $y$-axis I reported the sequence of markers as generated by the EBM, starting from the top rather than the base of the axes, to allow the visualisation of a downward curve. Two main paths were represented: a red one, which indicates the sequence of progression that would characterise the population, whilst the blue represents the pathway of the individual patient. Other semi-transparent red paths could represent other sub-groups, which resemble less the individual characteristic progression pathway. The yellow dot was the current stage, surrounded by a shape representing a confidence interval. Forecast for the progression of the disease was displayed with dashed lines. The background visual recalled the positional variance diagram represented in the EBM [9].

This first design iteration was informally discussed with the technical and quality team at the industry partner icometrix. This visualisation presented some advantages: an intuitive interface for the conceptual display of progression communicated a sense of a trend that is progressing and highlighted the individual patient’s out-
4.2. Design Process

4.2.4.2 Second Design Iteration

The aims for the second iteration of the prototype were: to generate a clearer output representing the pathway; to produce a fully digital version; and to design additional content other than the pathway representation. The second iteration was designed as a fully digital wireframe of the input and the output screens (Figure 4.4).

The input screen (Figure 4.4, left) was kept as in the first iteration, but upgraded to higher fidelity. The output (Figure 4.4, right) included slightly different tiles: *input overview*; the pathway output name was changed to *evolution* as better conveying the concept of the individual progression along time; *signatures*, which included the spider graphs for each type of marker. The spider graphs reported the probability of abnormality for each marker, with the aim to provide more transparency to the data in the evolution tile.

The evolution graph reported the population trend as a pink descending curve, whilst the individual patient was represented with a white line, and a white dot
referring to the current assessment. The timeline at the top marked the various
stages or the clinical journey (visit, follow-up, predicted progression).

Whilst the evolution visualisation was now cleaner, it lost details on the pop-
ulation and the clinical labelling. One main issue was that the patient trend ap-
peared continuous, which is somewhat misleading since assessments are discrete.
Moreover, the use of incremental numbers for the stages was still not aligned with
clinical knowledge (see IT-IQ-1-4, Table 4.3; UF-SI-2-1, Table 4.5). Overall, the
patient evolution remained unclear with this visualisation, therefore a new iteration
focused on this issue more specifically.

4.2.5 Final version: icompass

The final prototype became the product icompass. The current version of icom-
pass is an interactive wireframe, made up of input (Figure 4.5) and output (Figure
4.6) screens of three sequential assessments: baseline, one year, and three years
follow-up. The input screens are mainly composed of two modules (Figure 4.5, a):
patient information on top (New Patient module) and markers on the bottom (Input
Data module), with the last one including various tabs for different type of markers
(imaging, tests, etc.). The user can get additional information on input markers us-
ing the button on top of the active field (Figure 4.5, b). A button on the lower-right
corner leads to the output screen. The output has three main modules (Figure 4.6). The input at the top allows the user to explore the input markers available. The evolution module is the main output from the EBM. The visualisation of patient evolution is inspired by the positional variance diagrams used in the EBM, whereas the patient’s trajectory recalls survival curves. The \textit{x}-axis is the patient’s timeline with age and the \textit{y}-axis is the sequence of markers as generated from EBM, starting from the top. The background curves represent the diagnostic population distribution according to age and normal/abnormal markers, with AD in blue and MCI in yellow. This type of visualisation recalls stacked area charts, useful to show the change over time of quantitative values. An example is provided by Yau [194].

To mark each visit, the patient is represented as a white dot, according to age (\textit{x}-axis) and the number of abnormal markers (\textit{y}-axis). Patient information at each visit can be accessed interactively by clicking on the corresponding dot (Figure 4.7, a). Once multiple visits are recorded, the evolution and predicted trajectories in time (\textit{x}-axis) and disease stage (\textit{y}-axis) are visualised (Figure 4.6). The evolution trajectory is indicated with a solid line connecting visits (dots), whilst the predicted trajectory, which is the patient expected progression, is indicated with a dashed trajectory connecting visits (dots) and the probability range, visualised as a white shadow. Another feature is the timeline bar on the top, allowing the user to browse back and forth along the output timeline to check previous data (Figure 4.7, b).

The last module of the output screen includes three sections for each of the main markers (cognitive, imaging, CSF), and visualises the probability of abnormality of each marker in a spider-graph (Figure 4.6). This module provides additional information on the classification of each marker as normal/abnormal, in an attempt to provide more transparency to the model.

\subsection*{4.2.6 Conclusions}

Icompass was identified by the technical and quality team as a prototype that could be used for a first design concept testing to gather clinical feedback on the overall visualisation, potential, limitations, and intended use of this tool.

This does not exclude the fact that some elements of the design were still con-
Figure 4.5: Example of (a) icompass input screen and (b) details call-out for one of the cognitive tests.

Considered trade-offs, balancing the effort of designing a tool that is clinically relevant, according to the criteria listed in the design brief (Section 4.2.3), with the model assumptions. In particular, the so-called design tensions [161] were identified around:
Figure 4.6: Example of icompass output screen.

- **The patient trend**: the evolution graph provides information on the patient progression along time. However, it is designed to always be a monotonic downward trend, which is an assumption inherited by the EBM, but does not fully describe the real world.

- **Sequence of markers**: although patients could follow different progression pathways, the model assumes that there is a main sequence of biomarkers becoming abnormal along the disease progression, with certain probabilities. Having tested ways of embedding the positional variance diagram in the evolution visualisation to represent those probabilities, it was concluded that it
made the interface more complex.

- **Stage and age**: The EBM provides a staging system that is not based on currently used clinical criteria. Although I could have mapped EBM stages onto more fine-grained clinical stages, I decided that the exact staging is such an individualised outcome, and suggestive of a diagnosis. To preserve the nature of this tool as a decision-support system, rather than diagnostic, I decided to represent the timeline with the patient age, providing the reference population classification in the background, and leaving to clinicians the task of integrating these elements into their preferred clinical staging system.

The next section will describe the clinical testing of icompass.

![Figure 4.7: Details of icompass output screen.](image)

### 4.3 Study 4: Design Testing

#### 4.3.1 Aims

In this study, the design concept of icompass was tested with previously identified end-users: neurologists and psychiatrists working in secondary and tertiary AD care. The aim was to evaluate the potential of the proposed design concept from a clinical perspective, and particularly to explore: impact (perceived value and intended use); user needs addressed by the tool; and integration within the clinical workflow. The motivation for addressing these questions on an early version of the prototype was to avoid misunderstanding and faulty assumptions at a later stage of the product development, while reinforcing correctly identified requirements.
4.3. Method

4.3.2 Task

A think aloud task was designed to stimulate the interaction of clinicians with icompass. To introduce participants to the tool, a brief video was shown. Additional details on the EBM or on the tool functioning were not provided in the introduction, to test participants’ understanding of the tool and areas of concern. Moreover, the video ensured a consistent introduction for all participants. However, a brief explanation on EBM was additionally given to participants if they struggled during the task or requested it.

The general task was divided in three sub-tasks, one for each of the three sequential assessments, and involved clinicians using the interface mock-up and attempting to establish the following information: (1) what is the clinical status of the patient; (2) what is the suggested care plan; (3) whether the patient is progressing or not (only for second and third assessments). During the task, participants were encouraged to talk, verbalising any concerns, clinical reasoning, and possible conclusions. Instructions were given before each sub-task, to clarify and re-iterate the information requested. After completing all sub-tasks, participants were asked to complete a survey, which included the evaluation on a 5-point Likert scale of the following questions: (1) How clear was the information presented to you in the interface; (2) How confident were you in assessing the patient; (3) How confident were you in the care suggestions you came up with; (4) How confident were you in making a prognosis?

4.3.2.2 Participants

Participants were recruited amongst clinical contacts within hospitals in UK and Belgium. A total of 17 clinical experts took part in this study, 14 based in UK and 3 in Belgium. Amongst those, 9 were Neurologists, 7 Psychiatrists, and 1 Senior Nurse. In the analysis, the Nurse was included in the Psychiatrists group, for similarity of feedback and because they worked in a Psychiatric context. The level of training varied across the sample, with 8 medical doctors, 3 Professors, 2
clinical fellows, 3 trainees and 1 ‘other’ (the Senior Nurse). They worked either in secondary (n = 9) or tertiary referral (n = 8).

4.3.2.3 Procedure
The study could be completed in person or remotely. The remote option was included to both accommodate the busy schedules of clinicians, and to reach out to participants in Belgium. Initially, participants read the instructions and completed the consent form (see Appendices E.1, E.2) and an anonymous background survey (see Appendix E.3), including questions on their level and context of work. At this point, the audio recording started. Participants were presented with a short demo of the tool, asked whether they had further question, then introduced to the task. Further instructions were given after each assessment. Once the three sub-tasks were completed, participants filled in the post-task metrics. Finally, a semi-structured interview was conducted (see Appendix E.4). Questions addressed potential benefits and downsides of the tool within the clinical environment, how they imagined using this tool within their current clinical workflow, and possible influence of the tool on their decision-making. The sessions lasted between 40 minutes and 1 hour. All participants agreed to be audio-recorded and for data to be used for research purposes. Ethics for this study have been approved and are registered under reference UCLIC/1819/006/BlandfordProgrammeEthics.

4.3.2.4 Data and Analysis
The data collected was in the form of audio recordings, background survey, and task metrics. Audio-recordings were transcribed, anonymising any references to places or people. Quantitative descriptive analysis was performed on the numerical and quantifiable data. Inductive thematic analysis was conducted on transcribed data from the task and the interviews. In the first round of thematic analysis, 10 analysis codes were identified pertaining to concepts discussed and to requirements to improve the current version of the tool. A re-analysis produced five core themes, which will be described in the results section.
4.3.3 Results

Results are presented at three different levels: task assessment, thematic analysis, and Concept-based Analysis of Surface and Structural Misfits (CASSM).

4.3.3.1 Task Assessment

Task data was analysed according to the match between clinical judgement and the model’s output, for each of the sub-tasks. As shown in Table 4.12, there were three conditions for each sub-task: the clinician agrees, does not agree, and not sure. The number of participants for each condition was counted. For the two follow ups, there were cases in which the clinicians generally agreed with the tool, but they would judge the patient as being worse, i.e. diagnosis of AD pathology.

<table>
<thead>
<tr>
<th>SUBTASK</th>
<th>Baseline</th>
<th>Follow-up 1</th>
<th>Follow-up 2</th>
<th>TOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>10 (58.8%)</td>
<td>14* (82.3%)</td>
<td>14* (82.3%)</td>
<td>38 (74.5%)</td>
</tr>
<tr>
<td></td>
<td>N = 5; P = 5</td>
<td>N = 8; P = 6</td>
<td>N = 8; P = 6</td>
<td>N = 8; P = 6</td>
</tr>
<tr>
<td></td>
<td>*(4 exp worse)</td>
<td>*(4 exp worse)</td>
<td>*(4 exp worse)</td>
<td>*(4 exp worse)</td>
</tr>
<tr>
<td>Not sure</td>
<td>3 (17.6%)</td>
<td>2 (11.8%)</td>
<td>1 (5.9%)</td>
<td>6 (11.8%)</td>
</tr>
<tr>
<td></td>
<td>N = 3; P = 0</td>
<td>N = 1; P = 1</td>
<td>N = 1; P = 0</td>
<td>N = 1; P = 0</td>
</tr>
<tr>
<td>No</td>
<td>4 (23.5%)</td>
<td>1 (5.9%)</td>
<td>1 (11.8%)</td>
<td>7 (13.7%)</td>
</tr>
<tr>
<td></td>
<td>N = 1; P = 3</td>
<td>N = 0; P = 1</td>
<td>N = 0; P = 2</td>
<td>N = 0; P = 2</td>
</tr>
</tbody>
</table>

Table 4.12: Clinical agreement with icompass output for the three sub-tasks.

In total, 74.5% across all three cases agreed with the tool’s judgement on the patient status. Amongst them, 23.5% would have assessed the mock patient as being worse. I further explored how decisions changed over the three assessments and according to the two types of experts (Figure 4.8, a-c). At baseline there was some discordance in each of the two groups (Figure 4.8, a), visible as the number of participants split across agreeing with icompass output (B.Y n = 10), not agreeing (B.N n = 4), and not sure (B.NA n = 3). Although the majority agreed with the system, the agreement became more consistent in the last two follow-ups (FU1 and FU2). This might be due to the fact that in the last two assessments the cognitive tests started to show abnormality. Cognitive tests, being the most commonly used marker in clinical practice, might have improved the confidence of clinicians in
identifying the worsening status of the patient. Looking at neurologists and psychiatrists separately (Figure 4.8, b-c), there was not a clear trend at baseline. However, in the last two assessments, neurologists were more consistently aligned with the tool’s output, whilst psychiatrists disagreed more. This might be due to neurologists generally relying more on data, whilst psychiatrists are more focused on the clinical impression and behaviour assessment of patients.

![Sankey diagrams representing changes in clinical decision along the three sub-tasks, a) collectively, and for b) neurologists and c) psychiatrists separately.](image)

**Figure 4.8:** Sankey diagrams representing changes in clinical decision along the three sub-tasks, a) collectively, and for b) neurologists and c) psychiatrists separately.

In reporting metrics related to the task (Figure 4.9), clinicians provided a positive rating for the clarity of information presented in the interface, with comments saying that it was intuitive and well presented. For the metrics related to the confidence on three task-related questions (assessing the status of the patient, suggesting a care plan, and making a prognosis), scores are more uncertain/moderate. This was
justified by all participants as a lack of preliminary patient information.

![Scores](image)

**Figure 4.9:** Scores are related to judgements on a 5-point Likert scale for questions linked to the task.

### 4.3.3.2 Thematic Analysis

I identified five overarching themes from both the think-aloud tasks and the semi-structured interviews: *clinical impression, mental model, intended use, uncovered needs,* and *trust.*

**Clinical impression.** The first theme highlights the impact clinical impression has on the initial sense-making of a clinical condition, and the “*risk to rely too much on numbers, and not talk sufficiently to the patients to judge things like anxiety*” (P15).

A common issue with scoring, indeed, is that it is affected by various factors. E.g. “*If you are working in an area [where] social or educational background is lower than in an area like here for example, you could have MMSE scores lower than someone with Alzheimer’s disease*” (P16).

Despite the support provided by quantitative instruments, what really matters as an outcome to a clinical professional is making sure the patient copes with daily life: “*the thing about MCI versus AD as diagnosis is that they are hinged upon a person’s function in that context. So you wouldn’t make a diagnosis of Alzheimer’s as opposed to MCI of the AD type if a person is still functioning in the context*” (P4)
The lack of clinical history made the task too “artificial” to them, given its relevance in decision-making. Clinical impression is what drives the care management plan. This does not mean support tools are not contemplated in this reality, but acknowledges that their role is to inform better practice.

**Mental model.** The interaction with the tool stimulated clinicians to build a mental model of how the tool works. Three features particularly triggered their reasoning: status of the biomarker, forecast, and supporting evidence.

Regarding the first, clinicians tried to make sense of the pattern of normal/abnormal biomarkers and relate them back to the possible clinical status of the patient: “So the red ones means that they’re are abnormal and that’s why MCI would probably be the prudent diagnosis because tau positive, amyloid-beta positive even though the cognitive scores are still okay” (P9).

They also discussed the sequence of biomarkers, and although they stated this did not reflect the pathway for all patients, they considered it as a guide in providing evidence for deterioration. “What I was not getting quite right is that this is sequential so your model has something implicit” (P16); “Looks like they’re deteriorating clearly so you can see even on the graph. It’s just going down because you get more supporting biomarkers, you got the ADAS [AD Assessment Scale, cognitive test] and now even the MRI [imaging] is more abnormal” (P9).

Clinicians tried to find justifications to support the tool’s output, perhaps linked to age and coping strategies of the patient (“I think it’s getting a bit worse, but if we are seeing it according to the age of the patient, it’s quite normal. So he’s still in the MCI phase which can be due to the age instead of something else” (P1)), or assumptions on the speed of progression (“There was a big drop with a one year intro, but actually, there’s not that much over the two years. So maybe it is just in that kind of limbic predominant [note: a form of dementia affecting specifically the limbic system]” (P6)).

**Intended use.** The interaction with the tool and the interviews stimulated participants to consider how they would use the tool in the real world and whether it
would add value. Comments on intended use were grouped into three main areas, here described and reported in Table 4.13.

1. **Facilitates work.** Generally, the tool was judged as providing a nice visualisation and overview of the patient’s status. This promoted immediate understanding, whether it’s a known patient or a new one. “*I think it gives you a very clear, comprehensive overview of all the biomarkers integrated into one model because now we have them separately and we have to integrate it ourselves*” (P8). Some clinicians stated this tool does what they already do in their head, therefore it would alleviate the cognitive load caused by browsing too much data simultaneously. At an individual level, participants imagined themselves using the tool to prepare for the follow-up visits, making the process much faster: “*It would be quicker to evaluate the patient before them coming in for feedback. It would save me time because at the moment I have to open a file, look at the MRI results, then the cognitive tests, open the computer to look at bloods. With your tool, I’ll have all in one place*” (P9). Icompass would also bring advantages in collaborative settings like MDT meetings. The tool would facilitate discussion and interaction, and reduce the load caused by an often overwhelming number of cases, as confirmed by all participants. Similarly, it would facilitate the exchange of opinions between colleagues “*When I’m supervising residents, I see patients that were followed up by my colleagues. And then it would be very helpful to me to have such an overview of what the disease has been like for a certain patient even if I’m not intimately familiar*” (P10). Although comments were expected on possible workflow disruption caused by entering data in the system, this was not mentioned as a barrier, with two clinicians stating that digitisation of data is already performed by junior staff or trainees.

2. **Educational/insightful.** A substantial group of comments judged icompass as a valuable tool to educate or better understand the disease. Regarding the latter, this refers to enhanced insights on detecting early stages or addressing difficult and uncertain cases: “*It could be useful when it isn’t obviously clear from numbers that there is a decline. Just by placing the numerical output of an image registration and of a psychometric registration side by side and giving you some kind of global
decline measure” (P3). Relevant insights were also provided by prediction and recommendations: “What this model is doing, is trying to tell clinicians something that they don’t know. And what clinicians don’t know from a good assessment is probably about the significance of this for the future? Who do we need to follow up and who do we not need to follow up” (P17). Comments on educational advantages referred to sharing knowledge with trainees, and only to some extent with patients, as a support in their planning and lifestyle.

3. Specialist use. Despite some potential to educate patients, the majority of clinicians stated the tool should be intended for clinical use, and might represent a source of misunderstanding and additional stress if used with patients “It’s useful for clinicians, definitely. The language can be difficult to explain to patients, some people might get worried. I’m not sure” (P4).

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Intended use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facilitates work</td>
<td>General: - Overview</td>
</tr>
<tr>
<td></td>
<td>- Keep record</td>
</tr>
<tr>
<td></td>
<td>- Prepare for visit</td>
</tr>
<tr>
<td></td>
<td>- Not seen patient before</td>
</tr>
<tr>
<td></td>
<td>- MDT</td>
</tr>
<tr>
<td></td>
<td>- Opinion from peers</td>
</tr>
<tr>
<td></td>
<td>Individual: - Early stages</td>
</tr>
<tr>
<td></td>
<td>- Difficult cases</td>
</tr>
<tr>
<td></td>
<td>- Understand progression / trend</td>
</tr>
<tr>
<td></td>
<td>- For trainees</td>
</tr>
<tr>
<td></td>
<td>- External consultations</td>
</tr>
<tr>
<td></td>
<td>- Educate / prepare</td>
</tr>
<tr>
<td></td>
<td>Collaborative: - Not for communication to patients</td>
</tr>
<tr>
<td>Educational /</td>
<td>Understand disease: - Overview</td>
</tr>
<tr>
<td>Insightful</td>
<td>For peers: - Early stages</td>
</tr>
<tr>
<td></td>
<td>- Difficult cases</td>
</tr>
<tr>
<td></td>
<td>- Understand progression / trend</td>
</tr>
<tr>
<td></td>
<td>For patients: - Not for communication to patients</td>
</tr>
<tr>
<td>Specialists use</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.13: Overview of intended use of icompass from clinicians, grouped into areas of advantage.

Uncovered needs. Participants recognised the tool could overcome many current clinical needs. Two sources of uncovered needs were identified: aspects of clinical practice that the tool does not currently acknowledge, and features that clinicians wish could be implemented in the future.
Regarding the first, clinicians wished the tool could report more context about the patient (e.g., clinical history, current medication). This comment might have been triggered by the fact that this information was not given at all, whereas in a realistic situation clinicians would have this information available and the tool would be a way to enhance information on the disease evolution. Despite this, integrating notes on clinical history in the system could address the potential recognised by clinicians to have all information in one place, giving them the opportunity to browse both numerical and historical data. Another important issue is that different centres use their own clinical tests, and more behavioural assessments, which are key for them to diagnose patients and plan for care. Most of them were not familiar with ADAS-cog, and they wondered whether this tool would require them to adopt different tests.

The interaction of clinicians with icompass triggered some brainstorming around possible features that would make their approach more quantifiable and support a better understanding of the condition. They mainly referred to a more precise timescale of progression, stratification of patients, and forecasting. “Predictions is difficult to know at present, but it would help us for the future. Because we’re notoriously bad at predicting anything. I think if you get a lot of these visual scales, you will develop pattern recognition” (P15). Clinicians saw the real value of such a tool once a disease-modifying treatment is available: “Once we get to the point of having drugs that work, this stuff is going to be critical. Because then you’re looking at biomarkers in the pre-clinical phase. That’s going to be a totally different scenario” (P7).

**Trust.** The theme of trust was also extensively discussed, with comments that could be grouped in three main areas: data quality, algorithm, and uncertainty.

Doubts on data quality could relate to the fact that many tests were not familiar to participants. Moreover, the volumetric data extracted from the MRI was reported as a set of numbers; these were meaningless to clinicians who would have liked to see the original images or at least a written report. Other questions touched on cutoff scores, which is often known to depend on the specific normative population, so
that it is less generalisable across different centres, and the training dataset.

Trust related to the algorithm itself was critical. The immediate concern was on the predefined sequence of biomarkers, which would not be applicable to the understanding of rare cases. One suggestion to overcome this issue was providing the option to explore different evolution graphs for different types of biomarkers. Moreover, clinicians wanted to know more about the EBM validation and scientific evidence from a clinical perspective. This information was not extensively provided before the testing, but would be reported in a user manual, and references to relevant publications would be included in the tool as well.

These considerations converged to the theme of uncertainty. Whilst there is the desire to provide more precise data, some clinicians did not see quantification as always helpful: “It can be misleading, I mean, we have to work with some level of uncertainty. And when people ask about progression, we are very cautious in terms of saying very strong statements about what’s coming in future, because we know there’s some huge variability between patients” (P16). This statement suggests that the quantification provided by the EBM might play a role in clinical reasoning, and to compare and contrast various cases more objectively. However, clinicians also need to take into account the possible imprecise nature of the algorithm, and leave room for uncertainty when it comes to care plan and communication to patients.

4.3.3.3 CASSM analysis

Given the findings, I wanted to further explore how the proposed concept for icompass supports the clinicians’ intended use, and where it does not. The CASSM framework was chosen as an evaluation method to better define areas of misalignment between the concept of the tool and clinical practice, thanks to its proven suitability at the early stages of a design solution and its focus on concepts rather than processes. The initial evaluation identified concepts pertaining to the user, system, and interface, and each of them is judged as being present, absent, or difficult to ascertain. Each concept was then classified into entities and attributes, and labelled according to the extent of them being created/deleted (for entities) and set/changed (for attributes). There is a “surface misfit” when concepts that are salient to the user
are not well represented in the system, and vice versa.

Through the CASSM analysis, six entities and 31 attributes were defined, as per Figure 4.10. For each one, I judged whether they were present/difficult/absent at the user, system, and interface level. Then, I assessed the extent to which they can be created/deleted or set/changed. The analysis was discussed with the lead author of the framework [144].

Three classes of surface misfits were identified (Table 4.14). User concepts not represented by the system can be easily detected as all those items that are absent for I (interface) and S (system), such as availability of clinical history, choice of specific tests, issues on evolution and patient status. Most of these items have already been discussed and they refer to aspects of current clinical practice that are not covered by icompass. System concepts the user has to know about are, for example, the underlying functioning of the EBM, or characteristics of the training dataset or normative scores. Knowledge on this can facilitate trust and uptake of the technology, as it would give more context to the outcome. Lastly, user/system concepts that are similar but not identical refers to classifying patients at early/late stages (for clinicians early stages is when patients start to show symptoms, which is middle stages for the system), and again about the sequence of biomarkers.

The presented results allow to see what aspects of the design-reality gap are still problematic and should be addressed and evaluated in the next iteration of the prototype. ‘User concepts not represented by the System’ represent the clinicians’ expectations, according to their standard procedures: elements of the tool that they do not see, but expect to see, or would like to see in the future. These are mainly about the quality or clinical relevance of data presented. The second two areas of misfits represent the gap in the explainability: concepts proposed by the tool that are not familiar to clinicians, not part of standard clinical practice, or simply assumptions of the model. This refers to the two critical areas in the translation of AI into clinical practice presented in Section 2.4.4: data quality and explainability. Some of these items will need design and technical improvements, whilst others might need to be classified as models trade-offs, that might be overcome when a
better model becomes available and narrows down the design-reality gap.

### 4.3.3.4 Conclusions

Icompass visualisation and concept was generally welcomed by clinicians. The aims of this study were to look at the tool’s potential impact, how it could address
4.3. Study 4: Design Testing

<table>
<thead>
<tr>
<th>Areas of surface misfits</th>
<th>Items</th>
</tr>
</thead>
</table>
| U concepts not represented by S | - Clinical history  
- Customised tests  
- Evolution due to treatment  
- Patient can improve  
- Clinical impression  
- Differential diagnosis  
- Assessment based on cognitive domains |
| S concepts U has to know about | - EBM  
- Sequence of biomarkers  
- Patient conversion  
- Time to conversion  
- Imaging  
- Cut-off scores |
| U/S concepts similar but not identical | - Early / late stages  
- Sequence of biomarkers |

Table 4.14: Overview of surface misfits (U = user; S = system).

user needs, and integration within the workflow. These aspects are inevitably linked, as we assume that a tool is perceived as valuable when addressing users’ needs. Figure 4.11 illustrates how the tool could be integrated within the clinical workflow, based on the findings from this study. The graph is composed of four different levels, according to phases in the user journey (patient interaction, decision making, data collection/analysis, and peers discussion). The orange dots represent where icompass is expected to be used by participants. Icompass serves both individual and collective purposes, particularly providing a general overview and supporting individual and group reasoning, contrast and comparison, and communication with other specialists. Miller et al. [20] highlighted how clinical decision-making is shared and iterative, and CDSS design should therefore support these aspects of practice. However, comments also referred to the overarching potential effect of icompass on the workflow. For example, the tool is expected to speed up work and ease cognitive workload. These can be considered strengths of the current version, that would enable a smoother adoption in clinical practice. This was extensively discussed by Medlock et al. [195], but, to date, little research has covered the
effects of CDSS on user and workflow, making this work a contribution towards closing this gap.

**Figure 4.11:** Clinical workflow, representing current practice and possible icompass integration (marked with an orange circle).

Clinicians recognised the tool will not represent a substitute for clinical impression. In fact, the purpose of CDSS is about supporting and enhancing knowledge [195]. This is reflected in the other main expected use of icompass as an educational tool. Predictive algorithms provide insights on data faster and more accurately than the human brain is able to, but particularly in healthcare, clinicians need to be able to account for the nuances and idiosyncrasies of each individual patient. This is well represented in comments from clinical impression and mental model themes. Smith & Koppel [145] highlighted how relevant the triangulation of patient-clinician-IT is and how imbalances can lead to issues. In line with this model, our study highlights how the three actors in this triangulation interact in a 3-step process. The first step is the information stream from the patient to the clinician, who generates a first impression of the patient’s condition. The second step is the clinical interaction with the IT, querying the system and building a mental model of its output. The final
4.3. Study 4: Design Testing

Step is the clinician feeding-back to the patient information that has been enhanced after the interaction with IT. Overall, this process accounts for the added value of CDSS in clinical practice, by supporting the “naturalistic human information process” [20].

Ultimately, the extent to which clinicians will adopt and rely on the system heavily depends on trust. In our study, major concerns were around the quality of data, assumptions of the algorithm, and the coverage of borderline cases. These are common concerns for AI algorithms in healthcare, as reported in Cai et al. [174], where they further suggest interaction design solutions to increase confidence and transparency.

The main limitation of this study is that the patient’s scenario resulted in being too artificial, with no active input of data from the users (data were already provided in the interface) and clinical history. However, our aim for this early version of the prototype was not to conduct a full usability study, but to test the design concept, how clinicians interpreted the information provided, and potential use in the clinical setting.

The next iteration of the prototype should strive to increase trust and adoption. This should be achieved by testing the prototype within the real practice scenario. Moreover, findings from uncovered needs and misalignments represent directions to work on together with the technical team, defining a strategy based on technical feasibility and clinical relevance for new design and features of the tool.

4.3.4 Discussion

This chapter presented the design process, from the insights generated from the formative research, to the development and testing of a first prototype representing the concept of the CDSS built for the EBM, called icompass.

The first section described the requirements that were extracted by initial research and literature, and how, in the second section, these influenced the process of narrowing down the design problem to the concept of a data-driven clinical pathway. Through various iterations of the conceptual design of the prototype, with consistent reference to the requirements previously defined, icompass was finally
identified as a suitable prototype to be tested with the clinical early adopters.

Therefore, the third section of this chapter tested the design concept for icompass, a new CDSS for Alzheimer’s Disease specialist settings providing information on the current and expected evolution of an individual patient. At a general level, it was demonstrated that adopting a user-centred design approach early in the development of such tools can uncover important understanding on how the end users (in this case clinicians) intend to use an innovative system, and therefore how to design a tool that has better chances to be used in the real scenario, with minimal resources investment. In particular, the study focused on important aspects of the technology translation to clinical practice, often neglected in CDSS research and cause of failure in adopting such systems: integration within clinical workflow and clinical mental model of these tools. In this specific case, icompass showed potential in facilitating the work and reducing mental effort, but it is also a great educational resource. Finally, I illustrated how this tool can be integrated in the clinical workflow, according to the users’ intended use, both at an individual and collaborative perspective.

The next chapter reports these results back to the technical team, where findings are discussed from a technical perspective and a second iteration of requirements is generated.
Chapter 5

Design Review

This chapter:

• Presents a workshop study (Study 5) with the technical experts that is meant to revise and build on the insights from icompass testing with clinicians (Section 4.3);

• Elaborates on a second iteration of the requirements presented in section 4.2, consolidated and extended by the clinical testing in section 4.3 and the workshop study presented in this chapter (Study 5, section 5.1.

Preliminary results from 5.1 were included in [C.4].

The conceptual testing of icompass, described at the end of the previous chapter, generated novel insights on the prototype’s potential use, limitations, and possible improvements, all of which suggested directions for the next iteration of the prototype. A necessary step at this point was to report the user testing results back to the technical team, in order to reflect upon the testing outcomes and generate directions and practical actions for the next iteration of the prototype. This chapter forms the design review and recommendations part of the thesis roadmap (Figure 5.1): a first part presents a workshop study with four groups of technical experts discussing key insights from the clinical testing, and a second part where the set of requirements formulated in Chapter 4 is updated, based on the last two studies.
5.1 Study 5: Technical workshop

5.1.1 Aims

The set of technical workshops aimed to build on insights generated by clinical testing of the icompass prototype. Specifically, aims were to:

- Address misfits that were identified in the user testing;
- Evaluate clinical needs currently not covered by the tool;
- Define strategy and opportunities to fit the tool into the workflow.

5.1.2 Method

5.1.2.1 Participants

The workshop is a methodology that was adopted in section 3.3. However, whilst that study was intended to understand the model’s potential and therefore included
only disease progression modelling experts, this time I targeted also other technical experts that have a role in the development process, from the model development to the system’s transition into clinical practice. Participants were recruited from the Progression Of Neurodegenerative Disease group in the UCL Centre for Medical Image Computing (CMIC) and icometrix (Leuven, Belgium). To ensure context-specificity, each participant was assigned to one or more of the following four categories of technical expertise:

- Models experts (CMIC)
- Clinical evidence and applications team (icometrix)
- Front-end developers (icometrix)
- Regulation and quality experts (icometrix)

5.1.2.2 The Activities

Three activities were planned to focus on the key outcomes of the clinical testing with icompass, and closely reflected the aims of this study:

- Misalignments between users and the system;
- Clinical needs;
- User journey.

Activity 1 targeted aspects of the tool where the mental model of clinicians and the tool’s characteristics were not aligned (e.g. a predefined sequence of biomarkers or the concept of stages). This activity was structured as a guided brainstorming activity to address the identified misalignments. I produced a set of seven cards, each one stating a different clinical/tool misalignment. Each card had a title and a short description, as per example in Figure 5.2; the full list of cards is reported in Appendix F.2. In this activity, participants were asked to individually add as many comments as possible to the presented cards, in the form of post-it notes, for a given time. After collecting comments, I moderated the discussion among the participants using the following questions as prompts:
• Amongst all cards, which one do you think represents the biggest gap / misalignment?

• Which of the misalignments do you think is the most easily achievable in terms of bridging the gap?

• Why do you think clinicians have this perspective for this particular card?

**Figure 5.2:** Example of card used for the misalignment activity.

*Activity 2* was a guided discussion around uncovered needs mentioned by clinicians as a result of the interaction with the tool. Uncovered needs refer to both (1) something that they have available in their current practice, but the tool does not provide, and (2) something that they wish the tool could do but it does not at the moment. This activity was conducted as a group, with a card-sorting task. Participants were provided with a set of 13 cards, 6 pertaining to the needs clinicians currently have (2 cards on need for more context and 2 cards on customised markers), and 7 cards on what clinicians wished the tool could do in the future (2 cards on: time indications, level of detail in the graph, stratification and forecast; 1 card on future value). The group of participants was asked to sort all cards that were
relevant to them into a bi-dimensional grid according to feasibility (high, low, x-axis) and relevance (high, low, y-axis). The list of cards used and a short description is reported in Appendix F.3.

Activity 3 was built on the user journey map presented in Chapter 4.3 (Figure 4.11). The purpose of this activity was to immerse participants in the user journey flow, discussing possible uses and context of use for the tool signposted on the map, as well as possible barriers, technical/regulatory issues or alternative options. This activity was a guided brainstorming. The map of the user journey was presented to participants, highlighting when the tool was intended to be used by clinicians, how, and possible advantages. Each participant was firstly asked to individually stick post-it notes on the map, based on where they saw barriers, alternative opportunities, technical issues, or suggestions. This was followed by a group discussion, to elaborate on the rationale behind all contributions, and compare and contrast participants’ perspectives.

5.1.2.3 Materials and data collection

The workshop was conducted virtually, due to restrictions in travelling and gathering caused by the SARS-CoV-2 outbreak. To run the workshop, I used the collaborative platform MURAL (https://www.mural.co). The main screen was organised as a big board, where the three activities where arranged in different areas (Appendix F.1). At the top, an introduction section reported the agenda of activities and provided the QR code to a background questionnaire (Appendix F.5) for the participants. The background questionnaire was a brief 3-minute set of questions, to ascertain the level of technical knowledge of each group asking about their level of expertise in AD clinical practice, DPM development, CDSS and related quality and regulations, and software development. In the board, each activity area was marked with different colour shades, the title and a short description. No additional instructions were reported in the virtual board, as I was guiding the session verbally. For the calls, the video-conferencing platform MS Teams (GDPR compliant) was used.

The session was audio recorded and transcribed verbatim, anonymising any
references to participants and sensitive information. In summary, the data collected was composed of:

- Audio-recording of each session;
- Background survey for each participant;
- Written materials produced through MURAL;
- Moderator notes from the session.

Ethics for this study have been approved and are registered under reference UCLIC/1819/006/BlandfordProgrammeEthics.

5.1.2.4 The Workshop Plan

The four workshops were conducted in April 2020, and were composed of: (1) a presentation of findings from the clinical testing to the two bigger groups (CMIC and icometrix), and (2) the four main workshops (one with each subgroup of experts) with the following agenda:

- Introduction (5 minutes);
- Activity 1: Addressing misfits (15 minutes);
- Activity 2: Addressing uncovered needs (15 minutes);
- Activity 3: Addressing workflow (20 minutes);
- Closing (5 minutes).

The presentation of findings was planned as a separate session to the main workshop, and included a 30-minute presentation of the main results from the clinical testing of the interface and a quick demo of the interface. This was intended to provide contextual knowledge to the technical teams and familiarise them with the insights and content prior to the workshop.

The main workshop lasted approximately one hour. Once participants connected to the call, they were asked to register to MURAL. An introduction to the
platform and its functionalities was given, and participants were asked to follow the link in the “welcome” area of the virtual board to complete the consent form (see Appendix F.4) and the background survey (see Appendix F.5). Then, they were briefed on the overall activities. Each activity was timed with the integrated timer tool of the interface.

5.1.3 Results

5.1.3.1 Sample description and background

In total, 13 technical experts took part in the workshops across all four sessions. As shown in Table 5.1, only one participant constituted the Quality and Regulation (QR) group, whilst Clinical Application (CA), Model Developers (MD) and Front-end Developers (FD) were composed of four participants each. The background knowledge on the core topics of the workshop is consistent with each group’s expertise, with a specific mention for CA that has a more even knowledge across all areas.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Knowledge of AD and clinical needs</th>
<th>Knowledge of CDSS</th>
<th>Knowledge of DPM</th>
<th>Knowledge of system development</th>
<th>Knowledge of quality and regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>QR</td>
<td>1</td>
<td>Poor</td>
<td>Avg</td>
<td>Poor</td>
<td>Poor</td>
<td>Excellent</td>
</tr>
<tr>
<td>CA</td>
<td>4</td>
<td>Avg</td>
<td>Good</td>
<td>Good</td>
<td>Avg</td>
<td>Good</td>
</tr>
<tr>
<td>MD</td>
<td>4</td>
<td>Avg</td>
<td>Excellent</td>
<td>Good</td>
<td>Avg</td>
<td>Avg</td>
</tr>
<tr>
<td>FD</td>
<td>4</td>
<td>Poor</td>
<td>Avg</td>
<td>Excellent</td>
<td>Avg</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Sample and background knowledge description

5.1.3.2 Analysis

Two rounds of analysis were conducted. Given the differences in expertise for the four groups, each workshop was initially analysed independently to fully understand each group’s perspective. Coding spreadsheets were used for all the written content in the MURAL board, whilst the software NVivo was used for the analysis of transcripts. Here, the analysis codes created for Activity 1 mirrored the cards used for the activity, the codes for Activity 2 reflected the items that participants had to sort, and the codes for Activity 3 represented themes identified from the
discussion during the activity.

In the second round of analysis, the aim was to compare and contrast the contributions from different groups to each activity. Therefore data was analysed by activity, and in particular looking at recurrent themes and specific roles played by different groups. A new MURAL board was created to collect post-it notes from all the groups in one single space, with colour codes (orange = QR; green = CA; blue = MD; pink = FD). All data (post-it notes, coded text from NVivo, and other metadata from the MURAL board) was recorded in an excel spreadsheet, to facilitate filtering and comparison of contents. Quotes are reported in the text, labelled as [group_participantID] when it refers to a verbal comment, and labelled as [group] when it refers to a written contribution (i.e. written comment or post-it note).

5.1.3.3 Findings from Activity 1

Figure 5.3 is only illustrative of the overall contribution in the form of colours spreading across the seven cards. The two groups who contributed more to the activity were Clinical Application (CA, in green) and Model Developers (MD, in light blue). Both of them stated that the “clinical history” card would be the most difficult to address, mainly due to GDPR regulations, but also to the fact that data entered might vary across patients causing data inconsistencies, e.g. having notes for some patients but not for others. All other contributions and feedback touched mainly on four topics, as identified from the analysis: supporting evidence and user manual, supporting decisions rather than being prescriptive, implementation barriers, and system features suggestions.

Supporting evidence and user manual. QR, MD, and CA stressed the fact that supporting evidence, such as references to relevant publications, will enhance trust and provide knowledge on how the model works. In particular this refers to the sequence of biomarkers and their normal and abnormal cut-off scores. The suggestion from these three groups is to provide a user manual that clarifies terminology (QR) and provides publications references (CA, MD), particularly on the model functioning and assumptions (MD).
Supporting decisions rather than being prescriptive. The CA and QR groups particularly commented on how the tool should support the clinical decision making process, rather than being prescriptive on the diagnosis and prognosis.

“A tool should provide more insight into the patient, not a yes or no diagnosis, that one maybe it’s not too useful, but to provide insights into the status and prediction of the patient is... seems very useful. […] I think you really should take into account that these tools are helping the clinicians and not taking over some kind of decision making or something” [CA_01]

The QR participant highlighted, from a regulatory perspective, that despite the way in which the output information is implemented in the tool, it is fundamental that the tool does not make conclusions, e.g. on the fact that the patient progress is determined by a specific treatment.

“If we want to include information about interventions, we need to be careful about the conclusions/claims we are making out of the data e.g. ’the medication worked’ is a very risky claim to make” [QR].
Implementation barriers. The MD team was particularly focused on the implementation barriers that could be caused by the model’s assumptions. In particular, this was relevant for: including information on the evolution due to treatment, as the EBM is not designed for this purpose, and providing clear cut-off scores for normal/abnormal biomarkers, given that the EBM does not use thresholds as it is a probabilistic model.

“The tool is not designed for this, the model even assumes that no disease modifying AD intervention exists (which is true)” [MD].

System features suggestion. The FD team contributed from a purely interface implementation perspective, in two ways. On one side, they were referring to features they had already implemented for another of their products, trying to think of how previous work could be applied to this scenario. This particularly refers to the possibility of recording Patient Reported Outcomes (PROs), meaning that the patients could digitally enter functional data independently (e.g. on a patient-facing application) and use this data as a feed to the clinical tool.

“I think the bigger question is, how you will [deliver] it to the... to the public. Will it be still a companion AD... how do you label it, because that has some effects on the decision that we have to make on how we implement it” [FD_01].

FD participants suggested possible features that could be implemented, such as the patient-facing app just mentioned, or reminders and incentives for clinicians to complete or plan for an upcoming assessment.

In summary, the technical experts suggested various solutions to address the clinicians/tool misalignments listed in Activity 1. A user manual could address many of the suggestions here, including: indicate the supporting evidence (Clinical Application group) for the specific model’s assumptions, state its limitations (Model Developers group), clarify the intended use as being a support to the treatment, and not a diagnostic tool (Quality and Regulations participant). Some additional
5.1. Study 5: Technical workshop

Functionalities were suggested (Frontend Developers) on how to facilitate repeated assessments and recording of patient data.

5.1.3.4 Findings from Activity 2

Findings from Activity 2 revealed the technical perspective on clinical current and future needs, and whether there were consistent similarities or notable contrasts across groups. Figure 5.4 maps all groups’ contributions into one grid that is built on the two dimensions of \textit{feasibility} and \textit{relevance}. It is immediately visible how very few items are located in the bottom-half of the grid, therefore with low relevance, but they are spread along the whole feasibility spectrum. This indicates how the majority of cards are considered of high relevance for all teams. Moreover, it is noticeable how again the majority of the contribution came from CA and MD groups, with fewer cards sorted by QR and FD.

In the bottom half of the grid, the \textbf{Quadrant [Low feasibility, Low relevance]} can be considered as containing all needs classifiable as “out of scope”. This area comprises cards representing needs that are less likely to be brought to a second stage of the prototype development. Here, the MD group allocated those needs judged as being out of scope for the model, such as including treatment-related factors in the interface outcome.

\textit{“The model is completely agnostic to that [evolution due to treatment]. We can’t tell if someone’s taking medications and that’s helping their depression and therefore helping their cognitive scores and therefore helping their model stage. But a clinician would definitely take that into account. So, from a technical expert perspective, I think that’s completely irrelevant and completely not feasible”} [MD_P01].

The \textbf{Quadrant [High feasibility, Low relevance]} is home to those needs that can be considered “less of a priority”, easier to address, but not considered urgent. There is again little contribution to this area of the grid, and the cards seem in disagreement. In particular, the CA group raised the point that, although easier to implement, features like “MRI images of reports” and “Separate evolution for type
Figure 5.4: Summary of the post-it notes from all groups contributing to Activity 2 on sorting current and future clinical needs based on the two parameters feasibility and relevance. The figure has illustrative purposes only, to visually communicate the contribution of different groups by looking at the colours’ distribution.

of markers” might complicate the interface, despite being two of the most requested features by clinicians.

“It’s feasible, because basically you will just enhance this timeline just showing each biomarker separately. If it is really relevant, I’m not sure... You complicate the app much” [CA_01].

In the top half of the grid, the Quadrant [Low feasibility, High relevance] comprises cards that could represent important features to address, but are currently difficult to implement for various reasons. Therefore these are high priority technical issues, that would require time and resources to address, and success is not guaranteed. Most of the cards in this quadrant belong to the CA group. In particular, cards reported here refer to items that touch on predicting the course of the
disease (evolution due to treatment, forecast, conversion risk) or that are linked to model-related requirements (e.g. need of certain type or amount of data to successfully work).

“Yeah, I think forecasting years, that’s something that I’m working on right now, is highly relevant, but it’s not so feasible. I think it’s, it’s very difficult. It depends also on... certain factors and it’s a trade-off, because if you can achieve higher accuracy to forecast, the problem is that the clinician will not always be able anymore to understand what the model is learning from, because the models will get more complex and more complex, they will get higher accuracy, but the interpretation will drop” [CA_02].

Lastly, the Quadrant [High feasibility, High relevance] hosts items judged as having a greater chance of being addressed, as they are both relevant from a technical perspective, but also easily achievable. The majority of needs reported in this area refer to tracking the course of the disease (e.g. tracking, subtyping, separate evolutions). This suggests that providing advanced statistical information on the current status of the disease is a more consolidated and achievable aim, whereas predictions are recognised as being more uncertain and challenging, due to the complexity of the model and its assumptions, but also uncertainty of the disease itself. The FD group deserves a specific comment, as the majority of their contributions are clustered in this quadrant. In fact, as long as the model works and is validated, the FD experts considered most of the items as totally feasible, with different levels of relevance, but still distributed in the top-half of the grid.

“So I would say if this kind of information comes out of the model, and it’s currently displayed on a graph, just splitting it up is a no brainer, but only if it’s available from the model” [FD_02].

In conclusion, Activity 2 allowed to prioritise relevant clinical needs to address, according to various technical constraints. In particular, tracking and forecasting-related needs are the most relevant, although the latter is less feasible than the former and would require much more data and computational resources. Some other
needs might require less computational effort and resources, but it is important they do not complicate the interface.

5.1.3.5 Findings from Activity 3

Overall, all the groups found the clinical journey meaningful and easy to understand. Findings for the Activity 3 can be summarised according to four predominant themes: data sharing, workflow integration, suggested features, and regulations and instructions.

**Data sharing.** The majority of feedback on data sharing referred to *sharing within the clinical team, third party sharing, and anonymisation*. The CA group commented on how data sharing between different experts would be fundamental to a better understanding of the models and decision making; however anonymisation is required when sharing with other experts, as per previous experience of the FD group. Another debated topic is on data sharing *with patients*. The CA group reported previous experience of patients feeling discouraged at following their downward trends in the disease progress, whilst others found it educational.

“The feedback from the core team [of that project] was that they explicitly wanted this [option for patient to see their progress] to be disabled because it can be very discouraging for the patient themselves. […] And also don’t forget that lots of patients are very interested in how these models work. So I think it’s great to have this, this option” [CA_03].

**Workflow integration.** The main comments on the integration of icompass into the clinical workflow came from CA and FD. In particular, they talked about the importance of establishing the data needed to run the model, how data are smoothly passed from one hospital or referral to the other, and whether this data is entered automatically in the system or someone has to enter it manually. Specifically on the MDT discussion, the CA group suggested that an IT expert should be part of the meetings, to facilitate the use and interpretation of the tool.

“I think for the MDT, computer experts and some people that have more
knowledge on the technical side can also provide some insight like ‘hey, why did this give this output and can I trust it?’ so that you also have this human interaction from somebody who is an expert in the field” [CA,02].

**Suggested features.** System functionality ideas were suggested by all groups to facilitate or enhance the integration of the tool in clinical practice, as was partly done in Activity 1. The QR participant focused on how clinicians need an easy way to extract a patient report from the system. The CA and QR groups suggested how icompass could be used as a tool to inform what relevant data should be collected in the next steps of the care pathway, in a way to optimise costs and resources.

“It might even be like: Look, you’ve collected this data, but the predictions might be more robust if you also collect this other type of data. So consider, consider making these extra examinations” [QR].

Both CA and FD group suggested how patient-reported tests, such as behavioural or well-being questionnaires, might be collected independently by a patient-facing app and fed-back into the clinical system. The MD group, instead, was interested in the clinical perception of the model performance and suggested a way for clinicians to provide regular feedback on how good the results from the model were, so to improve the model.

**Regulation and instructions.** The QR participant commented on the need for the flow to be smooth, informed by regulation principles and with clear guidance for clinicians on the steps where the tool should be used. To this purpose, the QR expert proposed the production of a user manual that can support and guide the understanding of the model’s nuances and assumptions. The QR participant also suggested that a risk assessment should be conducted and warnings should be reported in the user manual.

“That is something very important. So, for the overall process or the overall chart you’re showing here, for each step, we do a risk analysis. So we have now three interactions with the tool, and for those three
interactions, we need to analyse which risks are related to those. […]

For some results we need to show a warning slide [in the manual]” [QR].

In conclusion, the three activities produced technical insight/feedback on the clinical input that can be grouped into overarching themes and a set of new requirements, that — together with the clinical feedback — will be incorporated in a second iteration of the requirements. Whilst requirements will be defined and discussed in the second section of this chapter, the overarching themes are now discussed.

5.1.4 Discussion

5.1.4.1 Overarching themes

The findings can be grouped into a set of overarching themes: icompass as an educational tool; implementation barriers, leveraging the current model potential; and open questions (Table 5.2).

<table>
<thead>
<tr>
<th>Overarching theme</th>
<th>Minor themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>icompass as an educational tool</td>
<td>- Support rather than replace</td>
</tr>
<tr>
<td></td>
<td>- Trust and evidence</td>
</tr>
<tr>
<td></td>
<td>- Future potential</td>
</tr>
<tr>
<td>Implementation barriers</td>
<td>- Model assumptions and limitations</td>
</tr>
<tr>
<td></td>
<td>- Data used for training</td>
</tr>
<tr>
<td></td>
<td>- IT implementation of the interface</td>
</tr>
<tr>
<td>Leveraging the current tool’s potential</td>
<td>- Possible features suggested</td>
</tr>
<tr>
<td></td>
<td>- Comments on the flow</td>
</tr>
<tr>
<td></td>
<td>- Interoperability and data sharing</td>
</tr>
<tr>
<td>Open questions</td>
<td>- Communication with patients</td>
</tr>
<tr>
<td></td>
<td>- Clarity and interpretability</td>
</tr>
</tbody>
</table>

Table 5.2: Overarching themes across all activities and relative findings.

icompass as an educational tool. It is a shared perspective that icompass in its current form has more potential as an educational tool, and it is not ready yet as a consolidated clinical device, given the lack of validation of EBM in clinical practice. Nonetheless, the clinical and technical feedback confirmed the potential of this tool, in particular to aid clinical practice and the decision making process along the
patient pathway. To this purpose, clear evidence from the scientific community and clinical validation is needed to earn clinicians’ trust in the tool. Ultimately, such a tool should leave room for clinical interpretation, so that it does not replace clinical judgement (e.g. through claims on the effect of medications or coming to diagnostic/prognostic conclusions), but it supports and advises clinicians by augmenting features in the data that would not be perceivable otherwise. When such a tool is expected to play a fundamental role is when a disease-modifying treatment will be available.

**Implementation barriers.** The current study identified technical constraints in meeting some of the clinical needs. The primary barrier was represented by the model assumptions, such as the established sequence of markers to become abnormal. Data required to run the model is also a barrier to a wider applicability. For example, customisation of the input data would need big datasets including that data and consistent longitudinal data points to train the model. However, it is part of the model’s aim to be scalable in the future, and creating bespoke models or tools for every single clinic would go against this principle. The potential solution to this barrier, then, might not be represented by tailored models, but a more standardised approach to the use of clinical data, considering cognitive composites or digital markers for cognitive tests. Additional barriers were represented by the IT implementation, such as developing customisable features that may complicate the interface.

**Leveraging the current tool’s potential.** The representation of the clinical journey triggered participants to see how the tool could be embedded in the clinical ecosystem and where possible gaps were. Participants in fact contributed with comments on how to optimise the flow, e.g. using the tool’s forecast as a suggestion for the most relevant next examinations to perform. Other comments referred to data sharing, in particular how existing data is transferred between clinical centres when a patient is referred, and how different clinicians could access data within the tool to discuss a particular case. Moreover, all groups suggested various features, e.g. a downloadable PDF report to keep for the records (QR), or a way in which clinicians
can provide a feedback to the system on the quality of its predictions (MD).

**Open questions.** Some open questions are identified by all groups. One of these was if and how the tool might facilitate the communication between patients and clinicians, to support the care plan in shared decision-making process. Although ambitious and relevant, this was seen as unlikely, given the domain-knowledge required to not only use the interface, but also to interpret the results. Additionally, this information can trigger distress and fear in patients. Another open question referred to the clarity in which the results are proposed to clinicians, and what is the level of interpretability and transparency of the model, which will influence clinical trust in the system.

In conclusion, this study provided a deeper understanding of the technical constraints and possibilities moving forward in the tool’s development process. These findings and those from the clinical testing (section 4.3 generated material for the consolidation and expansion of the original set of requirements 4.1.

## 5.2 Updated Requirements

This section preserves the original structure for requirements used in section 4.1 (Technical factors, User factors, and Organisation factors), and indicates those requirements that have been consolidated by the design testing phase (clinical testing: section 4.3, and technical review: section 5.1) as well as new requirements. This section will present the updated content for each category of the structure, through revisited versions of the tables in section 4.1. The content of the updated tables of requirements is contained in Appendix G.

### 5.2.1 Technical Factors

The majority of technical requirements were validated by both clinical testing and technical experts’ review. Regarding **system quality**, the two measures that were consolidated in the studies were *validation* and *reference population*. The first [IT-SQ-2] was addressed in particular by the clinical application (CA) group, when discussing the possibility to include new tests [CA_03]. Details of the reference population [IT-SQ-3] was a frequent question from clinicians, as based on this in-
5.2. Updated Requirements

formation they could evaluate how much the tool’s outcomes are applicable to their specific patients.

On information quality, clinicians confirmed many of the previously defined requirements, with some new entries. Clinicians stressed the need to be able to use tests that are more familiar to them, e.g. Addenbrooke’s Cognitive Examination (ACE), instead of tests that are less used in clinical practice (e.g. ADAS-cog) [IT-IQ-1-1], although we have elaborated on the technical challenges and the limitations to model scalability that this might represent. Therefore, this requirement should be addressed through alternative solutions, e.g., considering each test’s validation data on a specific population, cognitive composites, or different means to collect cognitive behavioural data, such as digital markers. Moreover, to facilitate clinical practice and completeness of data, clinical notes should also be included in the patient-related information [IT-IQ-2-2].

New requirements were introduced on information delivery. Many clinicians expected the MRI images or the radiologist report to be accessible [IT-ID-1-5]. All input queries, moreover, should indicate measure units and cut-off scores, to facilitate the process of data entry and understanding [IT-ID-1-4]. Technical experts mentioned the importance for the clinical data to be GDPR compliant and be consistent, if that data is used as part of the clinical record. Regarding the system output, one new requirement from clinicians suggested visualising the evolution trajectories for each type of marker [IT-ID-2-6], which, particularly for complex cases, should facilitate the understanding of the nuances in the patient-individual progression. Technical experts, instead, highlighted the importance for the system to not make any clinical claim, as this would go against regulation for CDSS [IT-ID-2-7].

The dimension constraints was added to the IT requirements, given the input received from the technical experts. The new requirements [IT-C-1] to [IT-C-3] refer to limitations currently characterising the models, such as it being agnostic to disease modifying treatments or the impossibility to add new, customised markers if training data is not available.
5.2.2 User Factors

Nearly all requirements stated in the first version (section 4.1) were confirmed by clinical testing and technical experts’ workshop studies. On the system intended use, the technical experts suggested that one of the key expectations from the system is to inform clinicians on the next most relevant examinations to optimise both the system performance, but also the use of clinical resources [UF-SI-1-3]. New requirements were added on Knowledge and expertise from both clinicians and technical experts. [UF-SI-2-3] stresses the importance of clinical impression as the key factor in interpreting the outcome from the system, and [UF-SI-2-5] mentions how cognitive scoring is often affected by patients’ individual and background factors that might lead to misinterpretation. A relevant requirement from technical experts highlights the main purpose of a CDSS, which is to be intended as an aid and not as a diagnostic tool, therefore it should not be prescriptive [UF-SI-2-4]. Lastly, new requirements on motivation reflect the findings discussed on the clinical intended use in 4.3, such as preparation for the follow-up appointment [UF-SI-3-5] or the limitations in using the tool to communicate with patients [UF-SI-3-7].

System understanding touches on the need of producing a user manual [UF-SU-1-2], which should report system functioning, risks, limitations, and supporting evidence. Lastly, interactive features are seen by clinicians as a way to allow customised data exploration, possibly promoting understanding and trust in the system [UF-SU-3-2], although it might be complex to implement them in the tool from a frontend development perspective.

5.2.3 Organisation Factors

Most of the organisation factors’ requirements were consolidated in the design testing and review research. The majority of new requirements are attributed to the structure, and in particular to the workflow. In fact, it is highlighted how the workflow should account for one person entering the data into the system [OR-S-1-2]. It was also suggested that a technical expert should take part in the MDT meetings to support clinicians in the interpretation and use of the system [OR-S-1-3]. Regarding resources, the technical team mentioned issues they often encounter in
transferring information from one hospital or medical centre to another. Given the fact that this system is particularly focused on secondary and tertiary referrals, it would be fundamental that previously collected patient information are smoothly shared and automatically uploaded in the system, according to GDPR regulations and with a suitable format [OR-S-2-3]. Overall, the consolidated and the new requirements represent an important step towards the future implementation of the system. Firstly, information quality and delivery is stressed as a key factor for clinicians to build trust in the system and its output. Secondly, the intended use of such a system, accounting for inevitable technical constraints, should inform the best way towards adoption. Lastly, aspects on the integration within the workflow and the production of a user manual have been further stressed by both clinicians and technical experts.

5.3 Discussion

The updated requirements for the prototype, based on Study 4 (section 4.3 and Study 5 (section 5.1), included generic and specific considerations for the next phase of the development process. This section discusses some clear directions for the next stage in the implementation of icompass.

The design concept of icompass was well received by the technical teams. They saw its potential in enhancing clinical understanding of the multitude of available biomarkers and their expected evolution over time, in supporting MDT discussions, and possibly informing the most relevant assessments to consider next in the patient journey. However, icompass is currently seen more as an educational tool. This is due to the EBM assumptions, the fact that the EBM still needs to be validated more extensively with real clinical data, and limitations around data quality. Many of these challenges are those classified with higher relevance and lower feasibility, particularly around disease progression forecast and customisation of input markers (e.g. personalised cognitive tests). A way to acknowledge these limitations in the use of the tool is by writing a user manual for icompass, that points the users to relevant publications, as well as a thorough description of the model’s limitations.
and assumptions. This is also declared by Sutton et al [21] as an evidence-based mitigation strategy for diagnostic support systems.

From a data visualisation and interaction design perspective, specific comments from both clinicians and technical teams suggested improvements in the way in which complex information is conveyed. Despite concerns on the sequence of markers, both groups considered the visualisation trend, the use of patient’s age as a timeline, and the forecasts intuitive and clear. One particularly discussed feature was to provide different options for the evolution graph, such as visualising the disease evolution trend for types of markers, in addition to the general overview for all markers. This would support clinicians in the exploration of data as an aid to the decision-making process, and is a common rule in information visualisation design, as predicated in the visualisation mantra from Shneiderman [196]: “overview first and details on demand”. Other relevant features should be considered, such as inspecting clinicians’ notes and MRI reports written by the radiologist, and in general what has been classified by the technical team as highly relevant and feasible.

Data visualisation is only one aspect of the human factors perspective. Another relevant aspect highlighted in the technical workshops and included in the second iteration of requirements is the role of CDSS as an aid to clinicians, without prescriptive purposes. Whilst this was not explored in depth in the first iteration of the design, feedback from the research demonstrated how relevant this is for the next iterations, as stated in Sutton et al.’s guideline [21]: “avoid prescriptiveness in system design. Evaluate system impact on an ongoing basis”. As the authors further mentioned, the fact that clinicians can use such a tool in a useful way for them to adhere to guidelines or inform their practice, without feeling their autonomy ‘jeopardised’, is a way for them to build trust in the system, therefore be more likely to use it and accept it.

Another key aspect, which represents the bridge between human factors and the organisation ecosystem, is how well the technology fits within the workflow. This is a concept that has been explored since the initial understanding of user needs (section 3.1), and that has become more tangible in this study. Acceptance and
adoption are heavily influenced by how well the tool can be integrated in the system, ranging from data sharing across hospitals, compliance to hospital procedures and standards, up to assigning a person who is in charge of manually entering the data in the system. Integrating a new tool should not disrupt the flow, but optimise it. This implies that a robust evaluation of risks and tasks needs to be conducted, as well as an assessment of how seamlessly the tool integrates with existing electronic health records.

The next chapter will merge these findings with those in previous research chapters, and outline in detail the contribution from this work, together with additional implications, limitations, and future work.
Chapter 6

Discussion, Limitations, Future Work

This chapter:

- Summarises the findings from previous chapters, contextualising them in relation to the background literature;
- Clarifies the three main contributions to knowledge;
- Discusses limitations and future work.

Parts of this chapter’s content is published in [J.2, J.3]. Methodological contributions were presented in [C.4]. Ethical consideration and explainability were partially presented in [W.1].

6.1 Summary of Research Findings

Disease Progression Models (DPMs) have the potential to enhance aspects of clinical practice towards a more quantitative and integrated understanding of AD [71]. This could support current clinical practice, which is predominantly qualitative and commensurately variable, due to the overwhelming quantity of data available. However, a big barrier lies in the transferability of DPMs to clinical practice. Such specialised computing applications often fail when introduced into practice, as it may succeed in meeting computational challenges, but not be clinically relevant [174]. One of the core reasons for this is that there is a lack of human factors considerations, with innovative technologies being introduced in the clinical scenario without
evaluation and consideration of end-users’ needs, workflow and practice, all influential elements along the development process [133]. Indeed, the motivation for this work was to identify translational opportunities and directions to exploit a DPM for AD in clinical practice. In particular, I was interested in whether, and under what circumstances, a DPM could be valuable to end users, and to frame an approach to facilitate this translation.

To address this challenge, I adopted a user-centred design approach. Prior to introducing a new technology in the clinical setting, I needed to understand the current clinical needs and practice of AD specialists. As previous studies reported [133], this should be the first step when building tools that are designed with the end-user in mind, but very little is said on what would be a suitable combination of methods. I proposed a combination of HCI methods and analysis that allowed me to explore the clinicians’ context, their needs, and their perspective in using a DPM-based tool in the future to aid clinical practice. I conducted field observations, semi-structured interviews, and an online survey (sections 3.1 and 3.2). Results from these studies clarified the main aspects of the clinical workflow in AD practice, as well as facilitators and barriers in the uptake of DPMs in clinical practice. It was clear that more specialised centres would be better prepared to adopt a model such as the EBM, due to their access to relevant data and their need for early diagnosis, and that the target users were predominantly neurologists and psychiatrists in secondary and tertiary healthcare. In fact, the context for the future adoption of a CDSS was more important than the specialisation of the users. I also needed to explore the technical potential of the EBM and the perspective of the technical team developing DPMs on their possible clinical exploitation in practice (section 3.3). The experts proposed various models and combinations of them, all addressing possible different case studies, such as models to provide information on the clinical pathways, on the temporal information, or on the risk to progress to later stages of the condition. From this initial formative research, I could address the first aim stated in Chapter 1, to “frame the gap between technical capability and clinical application”, and identify where the tool could initially best fit, according to disease stage, context, data used, and clin-
6.1. Summary of Research Findings

Through a design thinking process (sections 4.1, 4.2), I narrowed down the design problem to embed the EBM into a tool that could provide clinicians with information on the disease evolution and progression of individual patients, given a set of markers. The design was built on a set of requirements defined from the initial formative research, and principles of data visualisation literature. The final version of the tool, named icompass, was developed after a cycle of design iterations and was tested with relevant clinicians (section 4.3), in a task-based think-aloud study, followed by a semi-structured interview. Clinicians provided invaluable feedback not only on the flow of the tool, their interpretation of data displayed, and on the clarity overall, but most importantly on their intended use. This last point clarified even further what was the perceived potential of such a tool and gave indications on how and at which steps the tool could potentially be integrated in the clinical workflow. Therefore, the main outcomes from this study were: the definition of misalignments between clinical mental model and how the tool was set to work, various needs that originated from the clinical interaction with the tool, and a more solid understanding of the clinical workflow and how the tool could play a role at specific steps in the clinical journey. Moreover, at this point other aims stated in Chapter 1 were addressed: “develop an understanding of the design concept of a tool that allows clinicians access to predictive models for Alzheimer’s Disease” and “create appropriate low and high fidelity design prototypes, enabling users to provide feedback”. Finally, I discussed the feedback from clinicians with different teams of technical experts that will be involved in the next steps of the tool’s development, to understand the implications of clinical input, and to define a strategy for the tool’s integration within the clinical scenario (section 5.1). The clinical and technical feedback provided material to consolidate and expand the set of requirements in section 4.2, which altogether better defined the role of icompass’ use in clinical practice, its limitations, and the next steps in its development. Overall, this project has provided three key contributions to knowledge, which will be described in the next sections and are: contribution to clinical adoption of CDSS for AD; design contribution; and methodological contribution.
6.2 Contribution to Clinical Adoption of CDSS for AD

In this project, I studied opportunities and barriers to the adoption of a CDSS specifically intended for AD clinical practice. As part of the EuroPOND project, I focussed on a tool built on the EBM, with the long term goal of using this research to inform the design and development of a CDSS that is usable, fit for purpose, and adoptable. Technical papers on the EBM anticipate its potential clinical utility as providing a new understanding of the disease to help clinicians put measurements into perspective, providing more accurate prognosis and a staging system that may guide treatment, care choices, and communication with patients [9]. However, previous work has been technically focused, neglecting the need to clarify whether these tools are realistic and, if so, how we should best package and present relevant information to exploit its potential.

This work addresses an unmet need in the burgeoning field of CDSS for healthcare [175, 91, 197, 198, 199] and this is the first time that opportunities for adoption are explored for a DPM-based CDSS to be used by AD specialists.

<table>
<thead>
<tr>
<th>Requirements for Adoption</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technical factors</strong></td>
</tr>
<tr>
<td>- IT terminology adapted to clinical setting</td>
</tr>
<tr>
<td>- Data reliability and validity</td>
</tr>
<tr>
<td>- Interpretability</td>
</tr>
<tr>
<td><strong>User factors</strong></td>
</tr>
<tr>
<td>- System perceived as useful</td>
</tr>
<tr>
<td>- System meets critical needs</td>
</tr>
<tr>
<td>- System should fit with workflow</td>
</tr>
<tr>
<td>- Team / champion promotion IT</td>
</tr>
<tr>
<td><strong>Contextual factors</strong></td>
</tr>
<tr>
<td>- Availability of resources</td>
</tr>
<tr>
<td>- Specialised setting</td>
</tr>
<tr>
<td>- Costing support</td>
</tr>
</tbody>
</table>

**Table 6.1:** Table summarising requirements for adoption of CDSS for AD.

Some of the reported findings can be generalised to CDSS developed for AD, but built on different underlying algorithms. As for the rest of the project, and in line with previous literature [18, 21], these discussion points are presented accord-
6.2. Contribution to Clinical Adoption of CDSS for AD

ing to the usual three main areas of IT, User Factors, and Contextual Factors, as summarised in Table 6.1.

From the IT perspective, clinicians need transparency on CDSS output to feel confident in adopting them. This touches on key topics about explainability and trustworthiness of AI systems in healthcare [162, 200, 163]. Healthcare is a particularly challenging space, where the demand for explanation is much higher compared to other applications [173], and where explanations depend more profoundly on domain knowledge and the needs of the specific context. Within healthcare, AD brings additional challenges given by its heterogeneity, which affects the range of explanations that models have to account for. Various data sharing initiatives in AD aim to generate a common platform that could guide towards data harmonisation, as a more solid base for computational models to be developed and deployed [59, 86, 3].

From the User Factors’ perspective, such as stakeholders’ needs and beliefs [91], involving users early in the design process allows timely understanding of the benefits and value of including innovations in the health space, with less resources spent in the run up to the fully functioning product. Stakeholder engagement represents a support in defining the intended use of CDSS. In fact, it is not only about translating the outcome from an algorithm into a tool, but more about translating a concept (the EBM) between two very different fields. To translate a concept, each side’s beliefs, mental models, and terminology need to be taken into account. One might ask how can this “black-box” be more transparent for clinicians? Human Factors contribution to the answer is absolutely critical. For instance, Cai et al. [174] demonstrated that instead of improving the performance of imperfect algorithms, researchers could work on improving users’ interaction with the system. This will stimulate users to build appropriate mental models through the interaction with the algorithm, and therefore gather those explanations in a progressive enquiry process.

It was often stressed the importance of technology fitting within the current workflow in a seamless and interoperative way [91, 142]. This refers to the applicability or the extent to which the innovation might be suitable for a particular clinical
situation (workflow, setting, needs, type of patient, and resources). One barrier to adoption is the time taken to input data in the system and who would be completing that task. Since most clinicians make use of written notes and do not record patients’ data in datasets, this additional step might limit the integration of the tool into clinical routine. Other contextual factors include those external variables that do not depend on the user or the technology [147], such as facilitated access to data, data quality and heterogeneity, costing, and technical resources.

The present work aligns with these threads. Only few studies, such as [151, 91], involved clinical experts in the development of a CDSS, prior to an actual tool being developed. This approach allowed me to define promising strategies for the successful future adoption of this particular predictive model in a CDSS for AD.

One finding that was common throughout the research and all levels of specialisation is the importance of cognitive tests in clinical practice. Multiple clinicians reported concerns over the cognitive tests that are included in the EBM as not reflecting standard clinical practice. The application and development of cognitive tests is a key aspect of clinical research into improving diagnosis, characterisation of samples and longitudinal change, outcome measures including composite scores, and for validating potential AD biomarkers and data-driven subtypes [54, 201]. Therefore, computational scientists using cognitive tests data from big datasets to train their models need to be aware of their heterogeneous and multifactorial nature, especially when the results have implications for clinical practice. The work published in [J.3] [3] was inspired by this outcome and aimed to present common pitfalls and promote best practices for data-driven computational analyses of cognitive measures to maximise their value in the global efforts to understand and manage AD. It highlights key challenges through examples using cognitive tests commonly available in open access AD datasets.

6.3 Contribution to Design

This project makes a design claim on a possible solution to represent and deliver a DPM for AD to the clinical audience. Chapter 4 illustrated how the design was
6.3. Contribution to Design

guided by literature, formative research, and existing products. From a literature perspective, as reported in 2.4.3, previous work provided guidelines for the design of CDSS that are mainly centred around the system’s quality, user factors, and context [18, 19, 20, 21]. These guidelines set a scalable ground truth on what is important to consider for the implementation of other CDSS, but also give awareness on the design tensions to be considered in such a complex design space [161]. Based on formative research with the clinical users and technical experts (Chapter 3), I have outlined the potential case studies for which there was a strong need or interest, with the aim to narrow down the scope for this work. Tools have been developed to visualise some of the contents that other models for AD could deliver, for example AmyloidRisk [189] and ADappt [130], representing visualisations for risk score models, or PredictND [129], focused on a differential diagnosis model.

The specifics of the information provided by the EBM on AD staging and forecast creates the opportunity to tackle a different design challenge not previously addressed. With icompass, I specifically contributed in designing the experience of visualising quantitative data-driven information on the evolution and clinical pathway of an individual AD patient. Although previous works have explored the possibility of visualising care pathways information for different purposes and conditions [188], this is the first time that such information is visualised for AD, as a support to clinical decision making. Various visualisation strategies were considered in the initial prototype, based on the first set of requirements (Chapter 4) and established visualisation methods, balancing a trade-off between interpretability of the model and adaptation to existing clinical ways of visualising patient’s data. There is no doubt that incorporating a predictive model way of thinking in the qualitative and individualised current clinical practice for AD would represent a change from how the disease is currently discussed.

To stimulate a positive and constructive clinical behaviour change in the tool’s uptake, the key strategy was to follow a rigorous user-centred design approach, introducing novel features with a design that could facilitate clinicians in grasping the characteristics of the model and information presented. More specifically, there
were three main challenges that represented a novelty in clinical thinking. Firstly, stages in the EBM are determined by an irreversible, incrementally-numbered sequence of markers becoming abnormal. However, the design of icompass maintains the standard clinical labels to classify the probable status of the patient, and uses the patient’s age as a timeline, instead of using the sequence of markers as stages, as in the initial visualisation of the EBM. A second novelty brought by the EBM in the way AD is currently discussed is to combine a set of markers in an integrated description of the patient’s status. This, together with the quantitative contribution of the EBM to current clinical practice, have the potential to extract patients’ key information from big, heterogeneous datasets, and enhance data that might be overlooked if taken in isolation. From a data visualisation perspective, this unfamiliar information from the EBM was represented through survival curves, more familiar to clinical specialists to indicate patient’s evolution. A third challenge was to communicate the model-estimated uncertainty, both for the disease forecasts and for the potential model-based classification of patients into diagnostic labels. Having said this, the tool is not intended to be a replacement for clinical expertise and knowledge, therefore it is not designed to make a decision on the specific diagnostic label. The proposed solution takes the clinically familiar form of a normative model where a patient is assessed within the context of a population-based model. The confidence interval reported in the forecast is another way to provide an estimation of the trend, without being prescriptive on the exact progression of a patient.

The main validation of clinical acceptance of the design concept came from the clinical testing (Chapter 4.3), and particularly the findings on intended use. Given the design concept, clinicians were able to identify how icompass would add value to specific steps in their workflow, and how it would facilitate their data understanding as well as peer discussion and planning. This might not come without risks (e.g. disruption of workflow given required interaction with other systems), but holds great advantages as reported by Sutton et al. [21].
6.4 Contribution to Methodology

The overarching challenge here was on how to translate computational innovations to clinical adoption. Specifically, computational approaches can contribute to inform clinical practice for chronic conditions (such as neurodegenerative diseases) with advanced quantitative insights. The primary barrier is that such approaches often originate from a technical opportunity, and not a direct clinical need, which makes the translation to clinical use difficult, if not impossible [133]. Failures in the translation of technologies to clinical practice are a known challenge [133] and there is no current consensus on an optimal approach to promote successful adoption and integration to the healthcare environment.

This project brings a methodological contribution on how to address this challenge. I propose here a conceptual framework which builds on human factors and user-centred design research methodology. This framework (Figure 6.1) is characterised by theoretical (top, green) and practical (bottom, yellow) areas, and three phases (definition, concept testing, and iterative integration). When a technical opportunity arises [a], a matching clinical need must be identified, in the particular context [b]. This helps to frame the technical-clinical gap and detect a common scope for a new technology to be implemented [c]. Next, the technical innovation moves from a theoretical perspective to addressing more practical issues, e.g., algorithm improvements and design of a first prototype [d], which is tested at a conceptual level with clinicians to understand the tool’s potential and intended use [e]. This informs the concept testing node, where results from testing and adoption strategies are addressed [f], leading to a technical improvement of the tool, towards better integration into real-world situations [g]. The new prototype is then tested in a controlled clinical setting [h], where results inform the Technology Readiness Level [i]. The process iterates back to [g].

The approach adopted in developing icompass amounts to a specific case study of the general framework above. icompass is a tool using data-driven models for Alzheimer’s Disease as a support to clinical decision-making, and we ran workshops to define these models’ potential [a]. Clinical needs, practices, and the variety
of settings and resource access were studied by means of field observations, interviews, and surveys [b]. These findings allowed to frame the technical-clinical gap [c], identifying the technical features that were better suited for clinical use (data-driven clinical evolution), and research-oriented clinicians as the best placed to initially adopt the technology [3]. I designed a prototype [d], tested it with clinicians [e], and reported the outcomes to technical teams through a series of workshops, to address the feedback from the clinical testing [f]. This approach has promoted a thorough understanding of the clinical users, context, their perceptions of the potential of AD predictive models and the intended use, with minimal investment of technical resources in the first stages of prototype development. As per diagram, the next steps should envision increasing the prototype fidelity [g], with interleaved clinical testing [h], each time refining the design, algorithm, and strategies to facilitate the tool’s integration within clinical practice [i].

Whilst it is appreciated that other applications are characterised by different scopes and features, it is intended that this framework is sufficiently general to be applicable to other medical conditions and computational solutions.
Figure 6.1: Framework to translate computational, data-driven approaches to clinical tools.
6.5 Limitations

Limitations were reported for each study in the relevant chapters. However, there are some common and additional considerations.

Firstly, this work made extensive use of interviews and discussion groups. On one hand these methods were considered more suitable for getting a rich input at this early stage of the tool’s development (given the limited pool of participants and the complex scenario), but on the other hand they heavily depend on participants’ ability to articulate relevant gaps in the current workflow. In particular, it is hard for prospective users to assess the real impact of a future tool on their workflow, and how they would adapt to innovation (also known as the “task artefact cycle” [183]). This is also why end-users should be kept in the loop at all stages of the development process, as they will become increasingly aware of a tool’s potentials and use. Consequently, the findings generated by this work are, necessarily, based predominantly on qualitative research. Whilst I adopted a rigorous approach in the application of qualitative research standards and all the findings were reviewed by the supervision team, it is possible that the coding and interpretation of the data could have led to alternative conclusions or design implications if they were conducted by a different researcher.

A second challenge that had an influence on data was the limited access to facilities and resources as well as limited availability of clinical specialists. Particularly when it comes to healthcare, a number of factors come into play to restrict the breadth of options available. My access to clinical work and decision-making was limited to specialists’ verbal reports and to some observations of MDT meetings where only written notes not including patients’ data could be recorded. Therefore clinical specialists were my main source of information, but were themselves difficult to get hold of, given their limited time availability and continuous change in schedule. As the research focus narrowed to specialists working in secondary and tertiary referrals for AD, preferably neurologists and psychiatrists, the range of options was reduced and so was the number of participants that could be recruited for each study. It is worth mentioning that, although clinicians accepted to be
terviewed motivated by their contribution to research and no other rewards, all of them were enthusiastic to bring their unique perspective.

Additional limitations are due to the technical characteristics of the project and the design-reality gap. Models are still not validated in clinical practice, making it hard to build that trust that is sought from the clinical side, who see the technical opportunity still more as a “black-box” rather than a realistic and tangible perspective. For this reason, and due to the fact that the true potential of the models might find its full capacity only once a disease modifying treatment will be available, it is acknowledged that this innovation is at its early stages.

Although acknowledging that the best opportunities in technology adoption and diffusion of innovations are enhanced by targeting early adopters first as a necessary step towards a wider adoption and scalability of the tool, this represents a bias in itself. All these considerations contribute to the fact that the intended gap between technical opportunities and clinical adoption resulted in being wider and more intricate than initially expected, thus requiring more resources and time in understanding and clarifying the challenges to overcome those barriers.

Finally, the Covid-19 outbreak, which happened in the very last steps of this research, impeded the final consolidation of the research findings with the industry partner icometrix, which was set to generate a better definition of the next iteration and development of the icompass prototype on the company side. As these were unprecedented times, it was not possible to find a suitable time to have further discussions on future work.

### 6.6 Ethical Considerations and Explainability

Although this work sets promising expectations for the use of a CDSS and predictive models in the AD care space, the recent discourse on AI-related ethics and explainability in healthcare has highlighted a whole new set of challenges.

Big-data approaches and improved computational power have the potential to disentangle causes and treatments for AD, integrating large-scale multivariate datasets in a way that would not be achievable by small-scale approaches or by
relying on clinical expertise. Data in AI serves different purposes, that can be classified into training, performance and databank, and patient treatment. Data used for training is often subject to bias, that can perpetuate or amplify the models’ outcome [202]. As such, the predictions generated can fail if applied to a different population, with outcomes ranging from discriminatory or racial, to clinical under-estimation of threat. This poses the question of whether it is preferable to use a well-controlled research dataset for training, or a real-world one. Both approaches are approximation of the reality, with the first potentially enhancing bias, and the second potentially missing edge-cases [203]. For training datasets, it is also important to consider the discourse around data ownership, privacy, and pricing, particularly when data is not formally recorded through non-clinical sources. Performance and databank indicate datasets used to assess the accuracy, sensitivity, specificity, and determinability. Even for an algorithm performing close to perfection, a whole set of other real-world related factors can disrupt the outcome, such as human error or low quality input data. Performance data continuously evaluates that the model performs as intended, and the users continue to benefit from the system without being harmed. Data for patient treatment encapsulates all the risks related to the other two methods, in addition to the range of privacy and security requirements. This involves data collection and anonymisation procedures, transfer of data from hospitals to platforms, and prevention from identifying the individual patient’s entries by third parties. Terms and conditions related to how the system manages patients’ data needs to be established, such as the limit of time patients’ records will be stored, deleted, inspected, modified, and treated. Current regulations are inappropriate for new technologies handling elaborate large-scale data. Some examples refer to ensure control over a higher granularity of data (which current consent forms cannot cover), or consent for longitudinal projects, where the cognitive condition of the patient might decline. This compromises public trust and willingness to share data for research. Additionally, it has deleterious effects on privacy and security. Some of the more sophisticated phenotypical data can be traced even after anonymisation, and data breaches are a quite frequent phenomenon [204]. Encryption methods
6.6. Ethical Considerations and Explainability

should be improved to guarantee privacy and security of patients’ data. In relation to the specific case of predictive models, such a system would require clinical guidelines to regulate when it is meaningful to make use of the system, and this might depend on the stage of the disease (i.e. whether it adds value to be used at later stages or it is only disruptive). It is paramount to consider patients’ right (not) to know and to be informed about procedures and outcomes deriving from the use of computational approaches, taking into account a person-centred perspective. Having considered all the ethical challenges, it is also worth mentioning that disease prediction approaches, if characterised by low risk, can represent a benefit to society in different ways. Big data can bring *volume, variety and velocity* [205], set a standard of care to refer to, or be used for prevention.

Technical and digital infrastructures are in need of improvements, to support the computational demand required by advances in data analytics, and the need for big-data repositories [206]. Specifically for this project, the system that would allow clinicians to make use of disease progression models should run on appropriately protected systems, and should be supported by the available hardware and software. To promote the reliability of outputs, quality controls should be constantly performed by a specialised team, and information on the systems’ criteria should be available for clinicians to consult.

Finally, the adoption of data-driven models in healthcare, many of which may be considered “black-box” in nature, received a number of criticisms [162, 207, 173]. General concerns regarding interpretability of machine learning and artificial intelligence algorithms are arguably particularly relevant in clinical applications, where results can influence clinical decisions and health outcomes and present unique ethical challenges [173]. Interpretability touches on all stages of the development and use of these models, including the dataset used, the explainability of the models’ decision, and the interpretation of results according to domain knowledge. Understanding of models’ decisions is of particular importance in establishing replicability and generalisability of results. While limitations in understanding are often contextualised within trade-offs between their explainability and performance, there
are increasing efforts to explain model decisions and results, for example based on presenting model features with observed behavioural data [208]. Regarding interpretation based on domain knowledge, nominally significant results do not necessarily constitute clinically meaningful or informative findings at the population, group- or individual-level. Within this space, HCI contribution is set to be absolutely crucial. Including clinicians from the very early stages of the development of a tool that uses AI algorithms will make sure that the models’ outcome is not only validated, but also clinically relevant, and that the design of such tools support clinicians in building a mental model of these algorithms, through visualisation and interactive features [174, 163].

6.7 Future Work

The insights produced by this work contribute with the foundation knowledge needed to confidently achieve the long-term goal: to develop a tool that brings a novel predictive model developed by the EuroPOND project [1] into AD clinical practice, and that is tailored around clinicians’ needs, beliefs, and expected use. Based on the current progress from this project, further iterations and engagement with a greater number of clinicians and multiple centres is desirable to ensure this innovation can scale.

The future steps envisioned for this work map onto the framework defined in Figure 6.1: the development of an improved working prototype of icompass [g], that could possibly be integrated in the IT system and use clinical data of at least one centre, so that it can be tested in the real world setting [h], and the output from this testing to be used for next level improvements [i]. The table in Appendix H.1 outlines these steps more in detail, referring to four potential Phases.

In Phase 1, icompass design and technical specifications need to be iterated based on the outcome from research. Two main elements should be considered at this stage to promote even greater adoption in the long run: EHR integration and digital cognitive testing. The first feature would promote the adoption and seamless integration of icompass in the clinical stream of work, potentially being able to
6.7. Future Work

fetch data from existing patient records to feed into the icompass platform. The integra-
tion and interoperability required for this step opens up a range of challenges, such as understanding the Application Programming Interfaces (APIs), or how the data is protected, sent, shared, and stored. The expected interaction would see clinicians being able to access the full medical record of one patient, including previous clinical notes, together with the decision support provided by icompass. As for the digital cognitive testing, one strategy could be a partnership with a provider such as CANTAB (https://www.cambridgecognition.com/cantab). Adjusting the models to work with a digital cognitive assessment would have a number of benefits. From a clinical point of view, such tests are language independent and can be completed independently by the patient or with the support of the caregiver, therefore streamlining the process on the care team end. From a technical point of view, digital assessments would automatically feed data into the system, removing manual data recording, thus saving time, resources, and reducing error. From a data-related point of view, digital assessments is a revolutionary approach that, if adopted by clinics, will allow data consistency and comparability, as well as facilitated standardisation. However, this would require models developers to program and train the model so that it can run with the updated cognitive markers, and web developers to integrate the system of gathering, recording, and storing data into the icompass platform.

The design and user research team would be in charge of exploring how the current platform can be adapted to these features, from a data visualisation and user flow perspective. Phase 1 would also involve the Quality and regulation team, on a first analysis of risks (a good initial method can be Preliminary Hazard Analysis - PHA [209]) and requirements for eventual medical device classification regulations.

Phase 2 would involve the testing of a fully functioning system in a clinical setting, but with a controlled set of data and study conditions. This step is advised to make sure there is further testing, before resources are invested into a real-world implementation. Two studies are suggested for this Phase (see Table in Appendix H.2). A usability study would focus on the general efficiency of the interface, and would involve approximately 15 to 20 target clinicians completing a specific task
of patients evaluation at various stages. Methods used to record data can be: eye tracking recording, to analyse the exploration of the screen by means of heatmap and gaze analysis; data logged by the system (such as clicks, back and forth actions, action repetition, time in between different actions, successful/unsuccessful actions); time on task and visit preparation time; a set of questionnaires and interviews. The second study would be observational, involving on-field documentation of 8 to 10 clinicians interacting with icompass during a daily preparation for a set of appointments and during MDTs, whilst the researcher is asking them contextualised questions. This study can be concluded with questionnaires and semi-structured interviews. Reports from these studies would inform the subsequent iteration and identify potential centres to involve in the real-world testing phase.

Phase 3 would require the whole team of designers, developers, and quality managers to review the insights from Phase 2, and build a release of icompass that is suitable to work with data for those clinics that would be involved in the real-world testing. Additionally, the quality and regulation team should perform a more in-depth risk analysis (e.g., adopting principles from ISO 14971 standard [210] and Failure Mode and Effects Analysis - FMEA [211]), and work on a user manual.

In Phase 4, icompass could be tested with 1 or 2 clinical centres over a period of time (e.g., 4 months), as proposed in Appendix H.2. During this time, various data sources are monitored, to make sure everything is running smoothly (e.g., log data, observations and note-taking on site, regular surveys), and a weekly meeting should be held with the appointed clinical responsible. At the end of this period, a workshop with the clinical and technical teams might be helpful to discuss and review the experience. This Phase might also produce new inputs for a subsequent iteration and release of icompass.

Afterwards, icompass might be ready for an initial launch (provided necessary regulations are in place), initially for a limited number of centres, including those that will take part in the final testing. Continuous use will provide data for improvement and iterations. This process is only one example of a CDSS for AD; more application areas (e.g. other neurodegenerative conditions or different DPMs)
should be explored, to consolidate the proposed requirements for the translation and adoption of CDSS for AD.
Chapter 7

Conclusions

Alzheimer’s Disease (AD) is an overwhelming condition for the patients, their families, and the healthcare system, yet not well understood and for which there is no disease modifying treatment available [22]. The need to understand more about AD characteristics stimulated the systematic collection of a variety of data (clinical, imaging, biological), called markers, that can detect and provide information on the status of the disease [24], as well as the development of algorithms that can provide a better understanding and insights into this data [71].

In this project, I developed a framework that exemplifies how these algorithms can be translated for clinical use. I demonstrated the process on a particular computational model, the EBM, as part of the EuroPOND Horizon 2020 consortium [1] and in collaboration with the industry partner icometrix. Through the various studies, I addressed the aims stated in the Introduction (Chapter 1), as follows:

- I framed the gap between technical capability and clinical application, based on formative research with AD specialists and the models’ experts;

- I explored a visualisation strategy to allow clinicians access to the EBM purpose and outcomes;

- I created low and high fidelity design prototypes, to bring the concept into research. This served to gather clinical and technical input and test the efficiency of the design concept;

- I defined a set of recommendations in two different iterations, one based on
formative research and one based on the testing of an early design concept of the CDSS with clinicians. This result should serve model developers, by enhancing the utility and relevance of the model, as well as system developers, to implement a system that is built on user needs and requirements, and that clinicians see as embedded in their workflow.

Overall, the project made three main contributions to knowledge. The first is on studying and assessing the potential for clinical adoption of a DPM-based CDSS developed for Alzheimer’s Disease. Through a set of formative studies with clinical specialists as well as technical experts, I identified the IT, user factors, and context characteristics that would be best suited to adopt such tool in the first place, with the vision for scalability in the future. Based on these outcomes, I defined a set of implications for adoption that can be applied to other CDSS developed in the context of AD.

The second is a design contribution. I explored a visual way of communicating the complexity and novelty of the EBM to clinical specialists, in a way that is understandable and clear to them, but at the same time brings a positive and constructive improvement to their current clinical practice. I gathered their feedback and that from the technical experts, and represented it in a set of requirements that are set to drive the development of a fully functioning prototype for the clinical tool.

The third is a methodological contribution. There is no previous guidance on how to best translate and incorporate advanced computational knowledge that has been developed to improve clinical practice, in the target setting. Starting from the assumption that the end users need to be included in the loop since the early stages of the development, I proposed a framework on a novel combination of HCI and user-centred design methods, as well as a sequence of steps involving clinical specialists and technical experts. This framework could be generalisable to other computational applications that struggle to be adopted and used in the healthcare setting, despite the great potential they hold.

In conclusion, disease progression models that are developed to advance our understanding and management of Alzheimer’s Disease will only serve limited
goals if not leveraged to be used in the clinical setting. For this and similar challenges, the translation and the way in which such concepts are shaped around the actual clinical purpose, will potentially have an impact on AD and many other clinical conditions. This is only possible when a multidisciplinary effort is implemented, taking into account each party’s needs, potentials, and striving to have a positive impact on the delivery of better healthcare.
Bibliography


[7] Ivana Rubino, Bruce A Swanson, Samantha B Haeberlein, Johan Luthman, and Philipp von Rosenstiel. Modifying the course of


[78] Hubert M Fonteijn, Marc Modat, Matthew J Clarkson, Josephine Barnes, Manja Lehmann, Nicola Z Hobbs, Rachael I Seahill, Sarah J Tabrizi, Se-


[84] Arman Eshaghi, Ferran Prados, Wallace J Brownlee, Daniel R Altmann, Carmen Tur, M Jorge Cardoso, Floriana De Angelis, Steven H van de Pavert, Ni-


[151] Jennifer L Martin, Daniel J Clark, Stephen P Morgan, John A Crowe, and Elizabeth Murphy. A user-centred approach to requirements elicitation in


Appendix A

Background resources

A.1 Main cognitive tests for Alzheimer’s Disease
Table 1 - Overview on a set of commonly used cognitive tests for AD.

<table>
<thead>
<tr>
<th>Test</th>
<th>Short Description</th>
<th>Subscales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s Disease Assessment Scale – cognitive (ADAS-cog)(^1)</td>
<td>Brief cognitive battery, with items intended to evaluate different cognitive domains (in particular memory, language, praxis, and orientation).</td>
<td>13</td>
</tr>
<tr>
<td>Mini-Mental State Examination (MMSE)(^2)</td>
<td>Brief screening and staging instrument for mild cognitive dysfunction. Mainly evaluating orientation, attention, working and short term memory, language, and constructional praxis.</td>
<td>10</td>
</tr>
<tr>
<td>Montreal Cognitive Assessment (MoCA)(^3)</td>
<td>Brief screening instrument assessing various cognitive domains</td>
<td>11</td>
</tr>
<tr>
<td>Ray Auditory Verbal Learning Task (RAVLT)(^4)</td>
<td>Test of episodic memory, immediate recall or delayed recall</td>
<td>9</td>
</tr>
<tr>
<td>Boston Naming Test (BNT)(^5)</td>
<td>Visual naming test</td>
<td>-</td>
</tr>
<tr>
<td>Category Fluency (CF)(^6)</td>
<td>Word fluency test according to a specified category of names, to test semantic memory and word recall.</td>
<td>-</td>
</tr>
<tr>
<td>Clock Drawing Test (CDT)(^7)</td>
<td>Test of visuo-constructional abilities</td>
<td>-</td>
</tr>
<tr>
<td>Trail Making Test (TMT) A / B(^8)</td>
<td>Test of visual attention, psychomotor speed and task switching, in two parts.</td>
<td>2</td>
</tr>
<tr>
<td>Logical Memory (LM) I-II(^9)</td>
<td>Learning test of a short story, immediate and delayed.</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2 - ADAS-cog: subscales and description
<table>
<thead>
<tr>
<th>Item</th>
<th>Subscale</th>
<th>Task</th>
<th>Range</th>
<th>Scoring</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Word Recall</td>
<td>Three trials to learn as many high-frequency words</td>
<td>[0-10]</td>
<td>Mean N of recalled words in the three trials</td>
<td>Memory</td>
</tr>
<tr>
<td>Q2</td>
<td>Commands</td>
<td>Performance of 5 separate commands involving 1 to 5 steps per command</td>
<td>[0-5]</td>
<td>Steps correctly performed</td>
<td>Language</td>
</tr>
<tr>
<td>Q3</td>
<td>Constructional Praxis</td>
<td>Copy of 4 geometric forms with increasing level of difficulty (circle, overlapping rectangles, diamond, cube).</td>
<td>[0-5]</td>
<td>Score is based on the number of correctly copied forms</td>
<td>Visuospatial functions, Executive functions</td>
</tr>
<tr>
<td>Q4</td>
<td>Delayed Word Recall</td>
<td>Free recall of the word list in Q1, approximately 5 minutes after trial 3 of Q1</td>
<td>[0-10]</td>
<td>Mean N of recalled words in the three trials</td>
<td>Memory</td>
</tr>
<tr>
<td>Q5</td>
<td>Naming Task</td>
<td>1. Name 12 drawn objects with high/medium/low frequency; 2. Name the fingers of the dominant hand</td>
<td>[0-5]</td>
<td>5 ranges of scores, with 0 being the best performance</td>
<td>Language</td>
</tr>
<tr>
<td>Q6</td>
<td>Ideational Praxis</td>
<td>Performance of a sequence of actions participants would do in order to send a letter to themselves (fold letter, put in envelope, seal, address and put a stamp on it).</td>
<td>[0-5]</td>
<td>Scoring based on correctly performed components of the task</td>
<td>Apraxia</td>
</tr>
<tr>
<td>Q7</td>
<td>Orientation</td>
<td>Questions regarding orientation to time and place</td>
<td>[0-8]</td>
<td>Number of correct responses</td>
<td>Orientation</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Item</th>
<th>Subscale</th>
<th>Task</th>
<th>Range</th>
<th>Scoring</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q8</td>
<td>Word Recognition</td>
<td>After reading a list of 12 words, participants have to recognise these words from a new list were including both old and new words. It comprises three trials</td>
<td>[0-12]</td>
<td>Mean N of recalled words in the three trials</td>
<td>Memory</td>
</tr>
<tr>
<td>Q9</td>
<td>Remembering Test Instructions</td>
<td>Examiner’s assessment of the participant’s ability to remember test instructions</td>
<td>[0-5]</td>
<td>Examiner’s assessment (0 = no difficulties; 1 = very mild; 2 = mild; 3 = moderate; 4 = moderately severe; 5 = severe)</td>
<td></td>
</tr>
<tr>
<td>Q10</td>
<td>Comprehension</td>
<td>Examiner’s assessment of the participant’s ability to understand speech</td>
<td>[0-5]</td>
<td>Examiner’s assessment</td>
<td></td>
</tr>
<tr>
<td>Q11</td>
<td>Word Finding Difficulty</td>
<td>Examiner’s assessment of the participant’s ability in word-finding</td>
<td>[0-5]</td>
<td>Examiner’s assessment</td>
<td></td>
</tr>
<tr>
<td>Q12</td>
<td>Spoken Language Ability</td>
<td>Examiner’s assessment of the participant’s ability to communicate verbally</td>
<td>[0-5]</td>
<td>Examiner’s assessment</td>
<td></td>
</tr>
<tr>
<td>Q13</td>
<td>Number Cancellation</td>
<td>Selective cancellation of target numbers on a page (timed: 45 seconds)</td>
<td>[0-40]</td>
<td>[N of correctly crossed items in 45 seconds] – [N of incorrectly crossed items] – [reminders]</td>
<td>Visuospatial/Attention</td>
</tr>
<tr>
<td>Item</td>
<td>Subscale</td>
<td>Task</td>
<td>Range [min - max]</td>
<td>Scoring</td>
<td>Domain</td>
</tr>
<tr>
<td>------</td>
<td>-------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------</td>
<td>----------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Q1</td>
<td>Orientation</td>
<td>Questions to assess patient's orientation in space and time</td>
<td>[0-10]</td>
<td>1 point for each correct answer</td>
<td>Orientation</td>
</tr>
<tr>
<td>Q2</td>
<td>Immediate Recall</td>
<td>Three common words are read and patients have to repeat them in the same order</td>
<td>[0-3]</td>
<td>Number of correct words in the correct position</td>
<td>Memory</td>
</tr>
<tr>
<td>Q3</td>
<td>Attention</td>
<td>Patients are asked to subtract 7 from 100 for 5 times. Alternatively, they are asked to spell a word of 5 letters backward</td>
<td>[0-5]</td>
<td>Scoring is given by the number of digits within the patients response that has the most ascending digits.</td>
<td>Attention / Executive functions</td>
</tr>
<tr>
<td>Q4</td>
<td>Delayed Recall</td>
<td>Recall the three words in Q2</td>
<td>[0-3]</td>
<td>Number of correct words in the correct position</td>
<td>Memory</td>
</tr>
<tr>
<td>Q5</td>
<td>Naming Task</td>
<td>Patients are asked to name a wristwatch and a pencil, when shown to them</td>
<td>[0-2]</td>
<td>Number of correct responses</td>
<td>Language</td>
</tr>
<tr>
<td>Q6</td>
<td>Command</td>
<td>Patients have to execute a command that implies a sequence of 3 actions</td>
<td>[0-3]</td>
<td>1 point for each correctly performed action</td>
<td>Executive functions, Language</td>
</tr>
<tr>
<td>Q7</td>
<td>Repetition</td>
<td>The examiner reads a sentence and patients have to repeat it</td>
<td>[0-1]</td>
<td>1 point when the sentence is correctly repeated</td>
<td>Language</td>
</tr>
<tr>
<td>Q8</td>
<td>Reading</td>
<td>Patients have to read a sentence shown on a paper sheet and perform the action</td>
<td>[0-1]</td>
<td>1 point if the instructions are completely and correctly performed</td>
<td>Language</td>
</tr>
<tr>
<td>Q9</td>
<td>Writing</td>
<td>Patients are asked to write a sentence they choose</td>
<td>[0-1]</td>
<td>1 point for a semantically and grammatically correct sentence</td>
<td>Language</td>
</tr>
</tbody>
</table>

Continued on next page
Table 3: MMSE: subscales and description (continued from previous page)

<table>
<thead>
<tr>
<th>Item</th>
<th>Subscale</th>
<th>Task</th>
<th>Range</th>
<th>Scoring</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q10</td>
<td>Construction</td>
<td>Patients have to copy the design of two intersecting pentagons</td>
<td>[0-1]</td>
<td>The figures should have five angles each and intersect to score 1 point</td>
<td>Visuospatial</td>
</tr>
<tr>
<td>Item</td>
<td>Subscale</td>
<td>Task</td>
<td>Range [min - max]</td>
<td>Scoring</td>
<td>Domain</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>-----------------------------------------------------------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Q1</td>
<td>Trails</td>
<td>Draw a line (1) connecting numbers in ascending order (2) alternating one number and one letter in ascending order</td>
<td>1 / 0</td>
<td>Correct if the sequence is correct / autocorrected</td>
<td>Executive functions</td>
</tr>
<tr>
<td>Q2</td>
<td>Cube</td>
<td>Three common words are read and patients have to repeat them in the same order</td>
<td>1 / 0</td>
<td>Specific criteria to assess this item</td>
<td>Visuospatial</td>
</tr>
<tr>
<td>Q3</td>
<td>Clock</td>
<td>Draw a clock with all numbers, then set the time to 11:10</td>
<td>1 / 0</td>
<td>Criteria: contour, numbers, hands</td>
<td>Visuospatial</td>
</tr>
<tr>
<td>Q4</td>
<td>Naming</td>
<td>Tell the name of drawn animals</td>
<td>[0-3]</td>
<td>Number of correct items</td>
<td>Executive functions</td>
</tr>
<tr>
<td>Q5</td>
<td>Memory Immediate</td>
<td>Read a list of 5 works and recall them afterwards</td>
<td>[0-5]</td>
<td>Number of words correctly recalled at the first trial</td>
<td>Language</td>
</tr>
<tr>
<td>Q6</td>
<td>1. Digit Span Forward (attention)</td>
<td>Repeat a sequence of 5 digits (1 digit/sec)</td>
<td>[0-1]</td>
<td>1 point if performed correctly</td>
<td>Executive functions</td>
</tr>
<tr>
<td></td>
<td>2. Digit Span Backwards (attention)</td>
<td>Repeat a sequence of 3 digits backwards (1 digit/sec)</td>
<td>[0-1]</td>
<td>1 point if performed correctly</td>
<td>Executive functions</td>
</tr>
<tr>
<td></td>
<td>3. Letters and tapping (vigilance)</td>
<td>The examiner reads a list of letters, the participant has to tap anytime the letter A is read</td>
<td>[0-1]</td>
<td>No point if more than 2 errors</td>
<td>Executive functions</td>
</tr>
<tr>
<td></td>
<td>4. Serial 7s (vigilance)</td>
<td>The patient is asked to subtract 7 from 100 and other 7 until the stop command</td>
<td>[0-3]</td>
<td>4 or 5 correct = 3pt; 2/3 correct = 2pt; 1 correct = 1pt</td>
<td>Executive functions</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Item</th>
<th>Subscale</th>
<th>Task</th>
<th>Range</th>
<th>Scoring</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q7</td>
<td>Sentence Repetition</td>
<td>The examiner reads a sentence and patients have to repeat it (2 sentences)</td>
<td>[0-2]</td>
<td>1 point correctly repeated sentence</td>
<td>Language</td>
</tr>
<tr>
<td>Q8</td>
<td>Verbal Fluency</td>
<td>Name the maximum amount of works beginning with F (time: 1 min)</td>
<td>N</td>
<td>Total number of words, minus any intrusions, duplication, or other errors</td>
<td>Phonemic fluency</td>
</tr>
<tr>
<td>Q9</td>
<td>Abstraction</td>
<td>Tell how two items are alike</td>
<td>[0-2]</td>
<td>1 trial and 2 counted trials</td>
<td>Semantics</td>
</tr>
<tr>
<td>Q10</td>
<td>Delayed Recall</td>
<td>Recall the words in Q5</td>
<td>[0-5]</td>
<td>Criteria: recall with no cue, category cue, multiple choice cue</td>
<td>Memory</td>
</tr>
<tr>
<td>Q11</td>
<td>Orientation</td>
<td>Questions to assess the orientation of patients in space and time</td>
<td>[0-6]</td>
<td>Number of correct responses</td>
<td>Orientation</td>
</tr>
<tr>
<td>Item</td>
<td>Subscale</td>
<td>Task</td>
<td>Range [min - max]</td>
<td>Scoring</td>
<td>Domain</td>
</tr>
<tr>
<td>------</td>
<td>------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>1</td>
<td>List A: Trial I</td>
<td>The examiner reads a list of 15 words and the participant has to recall as many as possible</td>
<td>[0 - 15]</td>
<td>Number of recalled words (number of intrusions is also counted)</td>
<td>Episodic memory</td>
</tr>
<tr>
<td>2 to 5</td>
<td>Trial II to V</td>
<td>The examiner reads the same list of 15 words and the participant has to recall them</td>
<td>[0 - 15]</td>
<td>(as above)</td>
<td>Episodic memory</td>
</tr>
<tr>
<td>6</td>
<td>List B</td>
<td>The examiner reads a new list of 15 words and the participant has to recall as many as possible</td>
<td>[0 - 15]</td>
<td>(as above)</td>
<td>Episodic memory</td>
</tr>
<tr>
<td>7</td>
<td>Trial VI</td>
<td>The participant has to recall as many words as possible from List A</td>
<td>[0 - 15]</td>
<td>Number of correct items</td>
<td>Episodic memory</td>
</tr>
<tr>
<td>8</td>
<td>30 min delayed</td>
<td>After 30 minutes, the participant has to recall as many words as possible from List A</td>
<td>[0 - 15]</td>
<td>(as above)</td>
<td>Episodic memory</td>
</tr>
<tr>
<td>9</td>
<td>Recognition</td>
<td>The participant has to recognise as many words as possible from a list mixing words from List A with new words</td>
<td>[0 - 15]</td>
<td>Number of recognised words (number of intrusions is also counted)</td>
<td>Episodic memory</td>
</tr>
<tr>
<td>Item</td>
<td>Subscale</td>
<td>Task</td>
<td>Range [min - max]</td>
<td>Scoring</td>
<td>Domain</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>BNT</td>
<td>Boston Naming test</td>
<td>The patient has to name 30 images&lt;br&gt;The patient has to name as many items from an assigned category as possible (animals, vegetables, time: 60 sec each)</td>
<td>[0 - 30]</td>
<td>Number of correctly recognised items (sum of uncued + semantic + phonemic cues)&lt;br&gt;Number of correct responses</td>
<td>Language</td>
</tr>
<tr>
<td>CF</td>
<td>Word fluency test</td>
<td>The patient has to name as many items from an assigned category as possible (animals, vegetables, time: 60 sec each)</td>
<td>[SUM]</td>
<td></td>
<td>Semantic fluency</td>
</tr>
<tr>
<td>CDT</td>
<td>Constructional ability</td>
<td>The patient is asked to (1) draw a clock and the time; (2) copy a clock and the time</td>
<td>[0 - 5]</td>
<td>Score according to the number of criteria met&lt;br&gt;Number of correct responses</td>
<td>Executive functions, visuospatial functions</td>
</tr>
<tr>
<td>LM I/II</td>
<td>Episodic memory</td>
<td>I. The examiner reads a story and the patient has to recall it.&lt;br&gt;II. Delayed recall of the story</td>
<td>[0 - 25]</td>
<td>Number of chunks of information correctly recalled</td>
<td>Episodic memory</td>
</tr>
<tr>
<td>TMT A/B</td>
<td>Trail Making Test</td>
<td>A. Connect numbered circles in ascending numerical order&lt;br&gt;B. Connect circles with numbers and letters in alternate order</td>
<td>[time] N omissions N commissions</td>
<td>Time to complete the task (A. max 150 sec; B. max 300 sec)&lt;br&gt;Number of commissions or omission errors</td>
<td>Executive functions, visuospatial functions</td>
</tr>
</tbody>
</table>
REFERENCES

A.2 Predictive models selection and assumptions

Figure A.1: Flowchart representing example (A) research questions and steps regarding (B) test selection, (C) quality control and standardisation, and (D) computational/statistical methods. Example research questions (A) correspond closely to test selection (B), while subsequent processes outlined in steps C and D are broadly relevant across questions and tests. Example measures are reported in italics and section headings underlined. Dashed arrows indicate revisiting steps, for example, revisiting test selection owing to missing data [3]
Appendix B

Study 1: Research Material

B.1 Information
Information Sheet for Participants in Research Studies

You will be given a copy of this information sheet.

Title of Project: Research on current clinical practice and novel visualization methods for Neurodegenerative Diseases.

This study has been approved by UCLIC Research Department’s Ethics Chair [Project ID No]: UCLIC/1213/015

Name, Address and Contact Details of Investigators:
Principal investigator: Professor Ann Blandford

We would like to invite you to participate in this MRes/PhD project overseen by researchers at UCL. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide whether you want to take part, it is important for you to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

We are aiming to understand and observe neurologists’ individual clinical practice and decision-making for Neurodegenerative Diseases (Multiple Sclerosis and Alzheimer’s Disease). Moreover, we are interested in getting initial insights on possible graphical representations of novel statistical models that estimate the progression of these diseases. This research is an initial step in planning the development of a computer interface that can support clinicians in their management and diagnostic process. An interview will be conducted individually, and will last around 30 min. The interview will be audio-recorded, for research purposes. Written notes will also be taken. The aim of the interview is to gain a clearer understanding of processes, data usage and decision-making in the MDT scenario and in the individual clinical practice. In the second part of the interview, graphical data representation will be presented and you will be asked to answer some questions and provide your professional insight. You will not be asked personally sensitive information, and can refuse to answer any questions that you don’t wish to.

All data will be handled according to the Data Protection Act 1998 and will be kept anonymous. Only the researcher and her supervisors will have access to your anonymised data. With your permission, anonymised quotations from your interview may be used in reports about the study.

It is up to you to decide whether or not to take part. If you decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason.
Informed Consent Form for Participants in Research Studies

(This form is to be completed independently by the participant after reading the Information Sheet and/or having listened to an explanation about the research.)

Title of Project: **Research on current clinical practice and novel visualization methods for Neurodegenerative Diseases.**

This study has been approved by UCLIC Research Department’s Ethics Chair

[Project ID No]: UCLIC/1213/015

<table>
<thead>
<tr>
<th>Participant's Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>I ……………………………………………………………………………</td>
</tr>
<tr>
<td>agree that I have</td>
</tr>
<tr>
<td>▪ read the information sheet;</td>
</tr>
<tr>
<td>▪ had the opportunity to ask questions and discuss the study;</td>
</tr>
<tr>
<td>▪ received satisfactory answers to all my questions or have been advised of an individual to contact for answers to pertinent questions about the research and my rights as a participant and whom to contact if I have any concerns.</td>
</tr>
<tr>
<td>▪ I understand that notes will be taken during the interview.</td>
</tr>
<tr>
<td>▪ I understand that my interview will be audio recorded and I am aware of and consent to the analysis of the recordings.</td>
</tr>
</tbody>
</table>

I understand that I am free to withdraw from the study without penalty if I so wish. I understand that I consent to the processing of my personal information for the purposes of this study only. I understand that any such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.

Signed: ____________________  Date: ________________

<table>
<thead>
<tr>
<th>Investigator's Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>I ……………………………………………………………………………</td>
</tr>
<tr>
<td>confirm that I have carefully explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable).</td>
</tr>
</tbody>
</table>

Signed: ____________________  Date: ________________
B.2 Background Survey
Background Questionnaire

This questionnaire is for us to track and understand details on our participants.

All data will be handled according to the Data Protection Act 1998 and will be kept anonymous. Only the researcher and her supervisors will have access to your anonymised data. With your permission, anonymised quotations from your interview may be used in reports about the study.

1. Age: _____
2. Gender: M F Other
3. Studies/Specialization: ________________________________________________
4. Years of practice: ____________________________________________________
5. Job Location (main): _________________________________________________
6. Estimated skills in browsing and interpreting MRIs (and other related imaging techniques):
   1 = no expertise (I have never checked scans myself, I do not know software used and how to start)
   2 = very little expertise (I can hardly browse and interpret scans)
   3 = moderate expertise (I can browse and interpret scans, with help from an expert)
   4 = good expertise (I am quite independent in browsing and interpreting scans)
   5 = expert (I am totally independent in browsing and interpreting scans)
7. Use of medical interfaces (electronic health records, imaging platforms, databases, and similar):
   1 = never
   2 = very little use
   3 = moderate use
   4 = frequent use
   5 = high frequency use
8. Attendance at MDT meetings (on average): __________ per month
B.3 Interview Guide
SEMISTRUCTURED INTERVIEW – GUIDE

Duration and material
- The interview should not last longer that 1 h (around 30-40 min)
- Consent Form (with consent to record)
- Background Form
- Audio Recorder
- Notes (the researcher)
- Graphical presentation to show (from the laptop)

Opening the interview
- Present myself and the work I am doing
I am Maura, PhD student at UCL. My research topic is to work on interfaces that can aid clinical practice, providing computerised tools to reason about the disease progression.

Today I am interviewing you in order to get some clarifications on MDT meetings, on the individual clinical practice of a neurologist/psychiatrist and to discuss some possible future way of presenting clinical information on neurodegenerative progression.

Background questionnaire

Part 1: MDT meetings
These questions are oriented at understanding the meaning MDT has for the individual neurologist. It is also to counterproof the observations.

- What does MULTIDISCIPLINARY DECISION MAKING mean for you? (Prompts: does it means there are formal opportunities to communicate with other staff; can you see it resulting in improved patient care? Has it done anything to change traditional ways of decision-making?)
- How would you characterize the discussion during MDT meetings? (Prompts: Laboured, lively, everyone has a voice, etc).
- What are the criteria in deciding which of your clinical cases need to be discussed in the MDT?
- What are you usually asking/getting feedback for from the radiologist in MDT?
- How would you describe the procedure in discussing a case in MDT?
- What are the elements that improve clinical decision making or provide benefits for the patient?

Part 2: Specialist’s individual practice
These questions are oriented at understanding what / how the individual clinician uses information about one patient and makes decisions / manages the clinical case.

In the last question, the concept of ‘progression’ is suggested and initially inspected with the interviewee.
- Do you personally visit and get in contact with your patients or the patients you study?
- Could you list all the clinical information you need for making a provisional diagnosis in AD? (Prompt: can you think of a particular recent case? Patient’s history, cognitive tests, CSF, location/number of lesions)
- In what order and how do you use these indicators with the purpose of making a provisional diagnosis? (Prompt: can you think of two recent cases? A new one and a case where the disease is much further advanced - please describe the process and your eventual concerns).
- Do you inspect the MRI scans? What information do you gather from the scan and how do they help you in the understanding of the disease?
- What information do you gather from the radiologist’s report and how do they help you in the understanding of the disease?
- (If there are cases that are not discussed in the MDT) Do you discuss other cases with any colleagues? Could you think back to a couple of recent cases you didn’t bring to an MDT meeting and describe the decision you took about each of those, and what information you based that decision on?
- What information do you consider to make a prognosis on how the disease can progress? Is the concept of ‘progression’ useful for the management of patient and the disease itself?

Part 3: Novelties
This part of the interview introduces the novelty of considering the ‘progression’ in AD, through the presentation of visual stimuli. The interviewee is asked to try to understand what information is conveyed and if it can be effective to his/her work. The researcher is interested in the language used during this ‘think aloud’, the concept proposed by the clinician and where in the classical workflow such future tool would be effective.

- Introduction to the EBM (slides).
- Researcher: Now I will show you different graphical outcomes that inform on the stage at which the patient is classified, using some novel algorithms that measure the progression of his/her disease.
  SLIDE 1 TO 6
  o Would this be useful?
  o What information do you get from this image?
  o Would you need more information for understanding what you need from a picture like that?
  o Where/when would you like to have this information in your routine practice?
  o What are the differences between the usual information you are provided with?

Closing the interview
This part of the interview is meant as a closure and eventual additional comments from the respondent.

- Thank you for your discussion on this work.
- Would you like to add anything related to how MDT meeting are important in your individual practice?
- Would you like to add anything related to your usual routine, use of data or management of the disease?
- Would you like to add anything related to the graphical alternatives and options we discussed about?

- Thank you for your collaboration!
B.4 Visual Stimuli
Appendix C

Study 2: Research Material

C.1 Online survey
University College London (UCL) and icometrix are currently involved in a European project (www.europond.eu) to investigate the interest of clinicians in automated software for assessing and monitoring patient progression in early stages and diagnostic uncertainty of Alzheimer's Disease (AD). This system aims to give information on staging and progression of the disease, given clinical scores, MRI scans, blood and cerebrospinal fluid markers as input.

In this questionnaire we are exploring the current clinical workflow, in particular, how and whether clinicians make use of information on stages and progression of the disease, and how the proposed system might fit within current workflow.

Your feedback would be invaluable to help us developing a graphical tool that can truly address your needs in clinical practice and be an added resource to make your workflow more efficient and quantifiable.

Thank you very much for your time and input!

* Required

Respondent details

1. What title reflects your current professional activities the most? *

   Mark only one oval.
   - Radiologist
   - Neurologist
   - Neuroradiologist
   - Psychiatrist
   - Other: ____________________________

2. Please, specify your level of clinical training (e.g. senior doctor, junior, trainee)

   ____________________________

3. If applicable, please select your specific field of expertise

   Mark only one oval.
   - Alzheimer's Disease
   - Vascular Dementia
   - Frontotemporal Dementia
   - Posterior Cortical Atrophy
   - Other: ____________________________
4. What environment do you work in (mark all appropriate)? *
   Check all that apply.
   □ Community mental health
   □ Memory clinics
   □ Specialist cognitive clinics (e.g. tertiary referral centre)
   □ Research (e.g clinical trials or observational research)
   □ General medicine/Care of the elderly
   □ Other: ________________________________

5. What country do you work in? *

1. Questions on disease staging
   These following sections cover questions about disease staging. By this we mean the approximate assessment of where the patient is located along the disease path, at the time of the evaluation.

6. Do you assess the disease stage of your patients? (E.g. do you define your Alzheimer patients according to whether mild, moderate or severe) *
   Mark only one oval.
   □ Yes
   □ No  Skip to question 10.

7. What kind of stage systems for AD do you know? *

8. What staging systems do you use? *
9. What sources of data are more relevant to you in assessing the STAGE of disease?  
(Select all that apply)  
Check all that apply.  

- [ ] Patient/Informant history  
- [ ] Symptoms  
- [ ] Cognitive tests  
- [ ] Cerebrospinal fluid biomarkers  
- [ ] Blood test  
- [ ] MRI  
- [ ] Other: ____________________________________________________________

1. Questions on disease staging

10. How important is it to stage patients in the prodromal phase of the disease?  
*Mark only one oval.*  

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not important</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Very important</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

11. Why is it relevant, if so?  
*Check all that apply.*  

- [ ] For legal purposes (driving licence, carer, …)  
- [ ] To choose the appropriate treatment  
- [ ] For patients' awareness  
- [ ] I do NOT think it is relevant  
- [ ] Other: ____________________________________________________________

12. When might it (not) be more relevant to assess the stage of a patient?

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

2. Questions on disease progression

These following sections cover questions about disease progression. By this we mean the approximate forecast on how the disease will develop along time, given the present conditions.
13. Do you assess the disease progression of your patients? *
Mark only one oval.

☐ Yes
☐ No  Skip to question 16.

2. Questions on disease progression

14. What sources of data are more relevant to you in assessing the PROGRESSION of disease? (Select all that apply)
Check all that apply.

☐ Patient/Informant history
☐ Symptoms
☐ Cognitive tests
☐ Cerebrospinal fluid biomarkers
☐ Blood test
☐ MRI
☐ Other:

2. Questions on disease progression

15. What criteria guide you the most in assessing the probable progression of a patient?
(e.g. information from data, personal experience, speed of decline, complaints of patient or caregiver, ...) *

2. Questions on disease progression

16. How important is it to assess disease progression in the prodromal phase of the disease? *
Mark only one oval.

1 2 3 4 5
Not important  ☐  ☐  ☐  ☐  ☐  Very important
17. **Why is it relevant, if so?**

*Check all that apply.*

- [ ] For legal purposes (driving licence, carer, …)
- [ ] For planning the therapy
- [ ] For patients’ awareness
- [ ] To help patients and their families to plan for the future
- [ ] I do NOT think it is relevant
- [ ] Other: ________________________________

18. **When might it (not) be more relevant to estimate the progression of a patient?**

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

3. **Disease staging and progression tool**

19. **Would you be interested in an automated tool for assessing and monitoring patient progression based on multiple sources of data (cognitive tests, CSF values, blood test, MRI volumetric data)?**  

*Mark only one oval.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not interested</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

20. **Do you already use/have access to any automated systems to combine and visualise different sources of clinical data?**  

*Mark only one oval.*

- [ ] Yes
- [ ] No

**Which of the following data would be available to you?**

Select all options that apply.
21. **Cognitive tests**  
*Check all that apply.*

- Mini-Mental State Examination (MMSE)
- Alzheimer's Disease Assessment Scale (ADAS)
- Rey Auditory Verbal Learning Test (RAVLT)
- Montreal Cognitive Assessment (MoCA)
- Addenbrooke's Cognitive Examination (ACE-III)
- **Other:**

22. **Cerebrospinal fluid analysis**  
*Check all that apply.*

- Amyloid $\beta$ 1-40 / 1-42
- Total Tau protein
- Phosphorylated Tau (P-tau181)
- **Other:**

23. **Blood test**  
*Mark only one oval.*

- Yes *
- No

24. **Blood test: If yes, please specify which**

25. **MRI**  
*Check all that apply.*

- 2D T1-weighted scan (slice thickness < 3mm)
- 3D T1-weighted scan
- 2D FLAIR
- 3D FLAIR
- T2-weighted scan

26. **Genetic test**  
*Mark only one oval.*

- Yes
- No

3. **Disease staging and progression tool**
27. Do you think an automated tool for assessing and monitoring disease progression in dementia patients can bring value to clinical practice? *
   Mark only one oval.
   ○ Yes
   ○ No
   * Skip to question 31.

3. Disease staging and progression tool

28. Where do you see might be the value brought by the proposed tool in clinical practice? *
   Check all that apply.
   [ ] Improve the structure of the current working routine
   [ ] Patient's knowledge
   [ ] Communication between clinician and patient
   [ ] Support the clinician's decisions
   [ ] Speed up data inspection
   [ ] Early stages and uncertain cases
   [ ] Mid/late stages
   [ ] Other: _______________________________________

29. Can you briefly motivate your choices?
   __________________________________________________
   __________________________________________________
   __________________________________________________
   __________________________________________________

30. Which are the needs that this tool might support in your daily work?
   __________________________________________________
   __________________________________________________
   __________________________________________________
   __________________________________________________

3. Disease staging and progression tool
31. When do you imagine yourself using the proposed tool? *  
   Check all that apply.
   - During patient consultations
   - Before the follow-up visit
   - Only in the early stages of the disease
   - Only at the first meeting with the patient
   - Only at follow-up
   - I do not imagine using it
   - Other: __________________________________________

3. Disease staging and progression tool

32. Would a disease progression score/stage be useful information to communicate to patients? *  
   Mark only one oval.
   - Yes
   - No
   - Maybe

33. Briefly explain why it would (not or might) be useful *

Thank you for your contribution!

34. Would you be interested in using a free tryout of the system once it becomes available? *  
   Mark only one oval.
   - Yes Skip to question 35.
   - No Stop filling out this form.

Thank you for expressing your interest!
Please provide your e-mail address below, so you can be notified once a free tryout of the system is available.

35. E-mail address
Appendix D

Study 3: Research Material

D.1 Information
Information Sheet for Participants in Research Studies

You will be given a copy of this information sheet.

Title of Project: **Workshop with technical experts on disease progression models**

This study has been approved by UCLIC Research Department's Ethics Chair

[Project ID No]: UCLIC/1617/004/Staff Blandford HFDH

Name, Address and Contact Details of Investigators:

Principal investigator: Professor Ann Blandford

We would like to invite you to participate in this workshop for experts in disease progression models, as part of a PhD project overseen by researchers at UCL. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide whether you want to take part, it is important for you to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

We are aiming to stimulate discussion and contents generation among experts in developing disease progression models for Alzheimer’s Disease, regarding the technical capabilities and possibilities in the clinical translation of this knowledge. This study is part of a project on the development of a computer interface that can support clinicians in their management and diagnostic process, through the use of predictive models. The workshop will last approximately 90 minutes and will involve open discussion, interaction, and idea generation within the group. The session will be audio-recorded for research purposes, and other recording systems will be involved (written notes, written material generated from the discussion, and pictures of the material). You will not be asked personally sensitive information, and can refuse to answer any questions that you don’t wish to.

All data will be handled according to the Data Protection Act 1998 and will be kept anonymous. Only the researcher and her supervisors will have access to your anonymised data. With your permission, anonymised quotations from your interview may be used in reports about the study.

It is up to you to decide whether or not to take part. If you decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason.
# Informed Consent Form for Participants in Research Studies

(This form is to be completed independently by the participant after reading the Information Sheet and/or having listened to an explanation about the research.)

**Title of Project:** Workshop with technical experts on disease progression models

This study has been approved by UCLIC Research Department’s Ethics Chair

[Project ID No]: UCLIC/1617/004/Staff Blandford HFDH

## Participant’s Statement

I …………………………………………………………………………………………………

agree that I have

- read the information sheet;
- had the opportunity to ask questions and discuss the study;
- received satisfactory answers to all my questions or have been advised of an individual to contact for answers to pertinent questions about the research and my rights as a participant and whom to contact if I have any concerns.

- I understand that my participation will be audio recorded and I am aware of and consent to the analysis of the recordings.

I understand that I am free to withdraw from the study without penalty if I so wish. I understand that I consent to the processing of my personal information for the purposes of this study only. I understand that any such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.

Signed: Date:

## Investigator’s Statement

I ………………………………………………………………………………………………

confirm that I have carefully explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable).

Signed: Date:
D.2 Session
WORKSHOP WITH TECHNICAL EXPERTS

AGENDA

- Welcome and introduction
- Idea generation activities
  - Initial questions
  - Activity 1: exploring technical capabilities
  - Activity 2: potential of disease progression models
  - Activity 3: show and tell
- Closing
introduction

- aims of the workshop
- group agreements

- raise a forefinger when you wish to contribute to the discussion
- raise both forefingers if your point is a direct response, this way you will skip the queue
- wave a hand if you agree with an opinion

introduction

- presentations

Each one presents him/herself and their work.
idea generation

‣ initial question

How do you think disease progression models can contribute and bring value to clinical decision making and patient management?

idea generation

‣ activity 1: Exploring technical capabilities

Models on the table

‣ short description of main output
‣ what input needed
‣ what ‘clinical need’ is targeting
idea generation

› activity 1: Exploring technical capabilities

| Models on the board | 15' |

idea generation

› activity 2: Potential of disease progression models

| In groups | 20' |

- Are the available models a suitable information for the content identified?
  - If yes, how? Try to imagine how clinicians might be using this information
  - If no, are there any improvements or modifications to the models that can target the identified clinical need?
- Try to sketch how you think this information might be visualised by clinicians
- Think about why this information should be important for clinicians
idea generation

› activity 3: Show and tell

Group discussion

› What the model does
› Why important information for clinicians
› Possible visualisations
› Possible integration in clinical workflow

closing

› group discussion

› anything that has not been covered?
› wrap-up
THANK YOU!
Appendix E

Study 4: Research Material

E.1 Information
You are being invited to take part in my Ph.D. research project. Before you decide, it is important for you to understand why this study is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the project’s purpose?
The aim of this project is to better understand clinicians’ needs and practices with Alzheimer’s Disease patients at early stages, and to test a novel digital tool to help with managing that process.

Why have I been chosen?
You have been invited to participate because you are:
- A clinician working with Alzheimer’s Disease patients
- Aged 18 or over
- Able to communicate effectively in English, and do not consider yourself to be a vulnerable adult.
- Able to give informed consent.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up that point.

What will happen to me if I take part?
You will be invited to participate in a short testing of the tool, that will involve you making some provisional assessments of an ideal patient, based on the tool’s usage. We will then ask you to rate your experience through some metrics, and ask some questions related to the use of this tool in clinical practice. This test should take approximately 40 minutes. You will be asked to fill in an online survey about your current clinical practice, which will take 5 minutes to complete. Reasonable travel expenses will be reimbursed, and you will get free trial of the tool once it becomes available.
6. Will I be recorded and how will the recorded media be used?
With your permission, the interviews will be audio recorded. Transcriptions of the audio recording will be used only for analysis and for illustration in conference presentations and lectures. The audio recordings will be deleted once they have been transcribed and any identifying information will be removed during transcription.

7. What are the possible disadvantages and risks of taking part?
No disadvantages or risks of taking part have been identified. In the unlikely event that participating causes you any distress, you are free to withdraw, to discuss concerns with the researcher or the Principal Investigator.

8. What are the possible benefits of taking part?
While there are no immediate benefits to you from taking part, you will contribute in shaping the design and concept of this tool. We hope that you will find the study interesting and that it will help you to reflect on how predictive models for Alzheimer’s Disease can soon become a support in current clinical practice. We aim to share our findings with developers so that they may help to inform the future design of the tool.

9. What if something goes wrong?
If you have any concerns with the conduct of this study, please raise them in the first instance with Professor Ann Blandford (a.blandford@ucl.ac.uk). If your concerns are not addressed to your satisfaction, then you may contact the Chair of the UCL Research Ethics Committee – ethics@ucl.ac.uk

10. Will my taking part in this project be kept confidential?
All the information that we collect about you during the course of the research will be kept confidential, subject to legal constraints and professional guidelines. You will not be identifiable in any ensuing reports or publications.

11. What will happen to the results of the research project?
This study is for my Ph.D. project, and the findings will be reported in my dissertation. Depending on the findings, my supervisor and I may also publish the results in a journal or conference paper. If you would like to receive a copy of these, let me know and I will send it to as soon as the documents will be available. Anonymised data will be stored securely for five years, and may be reviewed in subsequent studies that have a related focus.

12. Local Data Protection Privacy Notice
The controller for this project will be University College London (UCL). The UCL Data Protection Officer provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk
The only personal information retained will be a copy of your informed consent and your chosen contact details if you wish to be informed of the outcome of this study. These will be held securely and separately from the anonymised data that you provide for the study.
Further information on how UCL uses participant information can be found at https://www.ucl.ac.uk/legal-services/privacy/ucl-general-research-participant-privacy-notice
The lawful basis that would be used to process your personal data will be performance of a task in the public interest.

16. Contact for further information
Contact details for me and my supervisor are provided at the top of this sheet; feel free to contact either of us if you have queries or concerns.

Thank you for reading this information sheet and for considering taking part in this study.
E.2 Consent
CONSENT FORM FOR ICOMPASS PROTOTYPE TESTING

Please complete this form after reading the Information Sheet or listening to an explanation of the study.

**Title of Study:** icompass – prototype exploratory test  
**Department:** UCLIC  
**Researcher(s):** Maura Bellio  
**Principal Researcher:** Ann Blandford  
**UCL Data Protection Officer:** Spenser Crouch

This study has been approved by the UCLIC Research Ethics Committee:
- Project ID number: UCLIC/1819/006/BlandfordProgrammeEthics
- Data Protection Number: Z6364106/2019/07/128

Thank you for considering taking part in this research. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to.

I confirm that I understand that by ticking each box below I am consenting to this element of the study. I understand that unticked boxes means that I DO NOT consent to that part of the study. I understand that by not giving consent for certain elements, I may be deemed ineligible for the study.

<table>
<thead>
<tr>
<th></th>
<th>Tick</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me and to ask questions which have been answered to my satisfaction. I agree to take part in the testing and interview.</td>
</tr>
<tr>
<td>2.</td>
<td>I understand that data will be anonymised and that it will not be possible to link my personal data (consent, contact details) with the study data. I understand that according to data protection legislation, ‘public task’ will be the lawful basis for processing.</td>
</tr>
<tr>
<td>3.</td>
<td>I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified. Data gathered in this study will be stored anonymously and securely. It will not be possible to identify me in any publications.</td>
</tr>
<tr>
<td>4.</td>
<td>I understand that the information I have submitted will be published as a report and I wish to receive a copy of it. Yes/No</td>
</tr>
<tr>
<td>5.</td>
<td>I consent to my interview being audio recorded and understand that the recordings will be destroyed following transcription.</td>
</tr>
<tr>
<td>6.</td>
<td>I am aware of who I should contact if I wish to lodge a complaint.</td>
</tr>
</tbody>
</table>

_________________________  ___________________________  ___________________________
Name of participant  Date  Signature

_________________________  ___________________________  ___________________________
Researcher  Date  Signature
E.3 Background survey
Study 4 - background questions

Start of Block: Block 1

Q1.1 The following survey collects some information on your current job, expertise, and clinical practice. This data will be used to explore the variety of our sample, and no names will be retained or matched to the survey's responses.

Thank you for your time and input!

End of Block: Block 1

Start of Block: Background questions

Q2.1 Please tell us about you:

Country of work:

Q2.2 Your specialisation (select all that apply)

- Neurologist (1)
- Psychiatrist (2)
- Neuropsychologist (3)
- Neuroradiologist (4)
- Geriatrician (5)
- Other (6) ________________________________________________
Q2.3 Level of training (select the most representative)

- Medical Doctor (1)
- Researcher / Professor (2)
- Clinical fellow (3)
- Trainee (4)
- Other (5) ________________________________________________

Q2.4 Working environment (select the most representative)

- First/self referral centres (1)
- Secondary referral (e.g. memory clinics) (2)
- Tertiary referral (e.g. specialised centre) (4)
- Research (e.g. clinical trial) (3)

Q2.5 Field of expertise (select all that apply)

- Dementia (1)
- Multiple Sclerosis (2)
- Parkinson's Disease (3)
- Other (4) _________________________________

End of Block: Background questions
E.4 Testing and interview guide
Study 4: mock-up exploratory tests

Aims
The aim of this fourth study is to collect feedback on the early development of the clinical mock-up.

Motivation
At this stage, we have reached a good understanding of the early adopters that will be involved in the use of the future tool, their access to data, and their needs. We integrated recommendations from the end-users, technical perspective, and literature, in the development of an interface that allows the exploration and interaction with the EBM output. This mock-up has gone through three iterations, based on input from the technical side. At this stage, we are presenting it to the end-users (neurologists and psychiatrists working in Alzheimer’s Disease, in specialist centres). However, we are still defining potential solutions in the visualisation and representation of models through the interface.

Through this first exploratory test on the mock-up with the end-users, we aim to evaluate the potential of the design concept proposed, and address the following questions:

1. **Impact**: what do the users think about the proposed tool? What value might it add to the classical clinical diagnosis?
2. **User needs**: are our identified user requirements correct? Have we misunderstood any requirements? Does it have value for the user?
3. **Fit within clinical workflow**: is the user positive in using it? How do they see it fitting within the clinical workflow?

By addressing these questions on an early version of the prototype, we avoid misunderstanding and faulty assumptions at a later stage of the product development, while reinforcing correctly identified requirements. This study being exploratory, we cannot make assumptions on results. We can, however, expect that the majority of our feedback previously collected on users’ needs and practice will be confirmed, while we expect to receive new feedback on the visual interface and its fit within the workflow.

Method

**Procedure**

**Consent form, info sheet (5 min)**
The participant is given the information sheet to read and a consent form to sign (vocal agreement if online).

**Background questions (1 min)**
These questions are filled in an online questionnaire. The questions are the same that were planned for the new short survey. **AIM**: tracking background details of the participant (country, specialisation, level of training, environment of work, field of expertise, data used).
Demo of the tool (1 min)
This is done by using the video promo I developed for AAIC conference. It maintains the amount of information equal for all participants, while giving an introduction to the tool.

Introduce the task (2 min)
Here I introduce the task the participant is asked to perform. The task involves the clinician using the interface mock-up, and attempt at providing the following information:
- Status of the patient (CN, MCI, AD – early/late to define it better)
- What is the suggested care plan?
- Is the patient progressing or not? (from the second assessment on)
The task will be performed in three parts, following the three different time points of the patient.

Script for part 1:
“Now I will present you 3 sequential assessments of the same patient on the interface. We will start with the first one. Here, while using the tool, I will ask you to tell me what you think the current status of the patient is, according to commonly used diagnostic criteria, and what you would suggest as clinical plan (medication, plans for next assessment, and anything else you think it is valuable to add). Please, think aloud while doing the task, even about any concerns or actions you are about to perform, or anything that comes to your mind while doing the task. Try to respond quickly, as you have a limited amount of time, but reasonably.”

Script for part 2
“Now we will move on to the second assessment for this patient. As you did previously, while using the interface, please tell me what you think the current status of the patient is, according to commonly used diagnostic criteria, and what you would suggest as clinical action (medication, plans for next assessment). In this case, also review whether the patient is progressing or not, and if yes, how is the progression expected to be. I ask you to think aloud while doing the task. Again, try to respond quickly but reasonably.”

Script for part 3
“Now we will move on to the third assessment for this patient. Again, while using the interface, please tell me what you think the current status of the patient is, according to commonly used diagnostic criteria, and what you would suggest as clinical action (medication, plans for next assessment). As you did previously, I ask you to review whether the patient is progressing or not, and if yes, how is the progression expected to be. Again, please think aloud while doing the task and try to respond quickly but reasonably.”

Performing the task (15’)
Screen recording.
- Instructions part1
Participants to perform part 1 of the task
- Instructions part 2
- Participants to perform part 2 of the task
- Instructions part 3
- Participants to perform part 3 of the task

Testing metrics (3 min)
After the full task is completed, the participant is given a link to an online form with the following metrics on 5-point Likert scales:

1. How clear was the information presented to you in the interface?  
   [Extremely unclear – extremely clear]
2. How confident were you in assessing the patient?  
   [Not confident at all – extremely confident]
3. How confident were you in the care suggestions you came up with?  
   [Not confident at all – extremely confident]
4. How confident were you in making a prognosis?  
   [Not confident at all – extremely confident]

Semi-structured interview (20 min)
The participant is asked questions related to: the role of the tool in their clinical practice, fitness within the workflow, and on comparing the tool to the list of data.

Questions:
- What do you think is the potential benefit of the tool to be used in clinical practice, in addition to what is done currently?
  - What could be also the downsides?
  - Do you have any suggestions for improving the visualisation or interface / interaction?
  - Was there other information about the patient that you would expect to see, or that you need to be confident in your assessment?
- Can you see a role for this in your clinical practice?
  - What is your clinical practice like?
  - How might it fit within the workflow?
  - Investigate also uses/value in multidisciplinary discussion (collaborative setting)
  - Investigate also uses/value in communicating to patients
- Do you think the evolution graph might have influenced your judgement on the patient assessment and prognosis? In which way?
- Do you have any other comments on the prototype?

STOP RECORDING

Thanks and contacts
Ask the participant if there are other people available for the interview, and if they are happy with us keeping their contact for testing the working prototype.
Appendix F

Study 5: Research Material

F.1 Activities overview
Welcome!

Please fill this form

**Items**

- Currently have: more context
- Currently have: customised markers
- Wish: time
- Wish: level of detail in the graph
- Wish: stratification and forecast
- Wish: future value

**Clinical history**

- Current medications
- or interventions
- Functional tests
- MRI images of reports

Their own functional tests

Most common cog markers

Timeline with intervals between assessments

Forecast in years

Separate evolutions for type of markers

Customisable Subtypes

Conversion risk

Evolution due to treatment or intervention

**High relevance**

- Low relevance

**High feasibility**

- Low feasibility

---

**AGENDA**

**ACTIVITY 1: MISALIGNMENTS**

**CARD 1: CLINICAL HISTORY**

Clinicians make decisions based on patients' history and clinical impression. While they normally use notes, this information is not available from the tool.

**CARD 2: FUNCTIONAL TESTS**

Clinical decision is ultimately about how the patient can cope with daily life. Thus, functional and behavioural tests are a fundamental indicator and usually collected data.

**CARD 3: EVOLUTION DUE TO TREATMENT OR INTERVENTION**

Clinicians evaluate progress also based on particular intervention (cognitive, behavioural, etc) or start of a medication. This is not tracked in the interface.

**CARD 4: SEQUENCE OF BIOMARKERS**

The sequence of biomarkers becoming abnormal will be unique for every patient. Although the tool provides a data-driven sequence of expected progression, clinicians need to know more about how this has been generated.

**CARD 5: NORMAL/ABNORMAL BIOMARKERS**

Biomarkers' cut off is not always explicit and can depend on the specific patients' population. It is not clear from the interface what is the intended threshold for a biomarker to be considered abnormal.

**CARD 6: REPEATED ASSESSMENTS**

The tool assumes a set of assessments is repeated regularly. However, mostly cognitive assessments are repeated, perhaps not at the same time point, and only if the patient is not too deteriorated.

**CARD 7: CLINICAL LABELS AND STAGES**

Clinicians use more precise labels than AD and MCI. They would discriminate early and late MCI, or preclinical stage. The current status of the interface is less fine grained.

---

Write your comment here

ACTIVITY 2: UNCOVERED NEEDS

Questions
- Among all cards, which one do you think represents the biggest gap/misalignment?
- Which of the misfits do you think is the most easily achievable in terms of bridging the gap?
- Why do you think clinicians have this perspective for this particular card?

Double-click to create a note

**ACTIVITY 3: USER JOURNEY**

Double-click to add a sticky note

How would you address this misalignment to make the tool more adapted to clinicians expectations and needs?

How do you classify each item in relation to relevance and feasibility?

Discuss as a group.

Thanks for your collaboration!

This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1) and by the industrial partner Icometrix (Leuven, Belgium {https://icometrix.com/).

This work is part of the EuroPOND initiative, funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 666992.
F.2  Activity 1: cards
## List of Cards

<table>
<thead>
<tr>
<th>Card 1</th>
<th>Clinical History: Clinicians make decisions based on patients history and clinical impression. While they normally use notes, this information is not available from the tool.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card 2</td>
<td>Functional tests: Clinical decision is ultimately about how the patient can cope with daily life. Thus, functional and behavioural tests are a fundamental indicator and a data usually collected which is not currently represented in the tool.</td>
</tr>
<tr>
<td>Card 3</td>
<td>Evolution due to treatment or intervention: Clinicians evaluate progress also based on particular interventions (cognitive, behavioural, etc) or start of a medication. This is not tracked in the interface.</td>
</tr>
<tr>
<td>Card 4</td>
<td>Sequence of biomarkers: The sequence of biomarkers becoming abnormal will be unique for every patient. Although the tool provides a data-driven sequence of expected progression, clinicians need to know more about how this has been generated.</td>
</tr>
<tr>
<td>Card 5</td>
<td>Normal / Abnormal biomarkers: Biomarkers’ cut off is not always explicit and might depend on the specific patients’ population. It is not clear from the interface what is the intended threshold for a biomarker to be considered abnormal.</td>
</tr>
<tr>
<td>Card 6</td>
<td>Repeated assessments: The tool assumes a set of assessments is repeated regularly. However, mostly cognitive assessments are repeated, perhaps not at the same time point, and only if the patient is not too deteriorated.</td>
</tr>
<tr>
<td>Card 7</td>
<td>Clinical labels and stages: Clinicians use more precise labels than AD and MCI. They would discriminate early and late MCI, or preclinical stage. The current status of the interface is less fine grained.</td>
</tr>
</tbody>
</table>
F.3  Activity 2: cards
### Needs based on current clinical practice, not present in the tool

**More context**

<table>
<thead>
<tr>
<th>Needs based on current clinical practice, not present in the tool</th>
<th>More context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical history</td>
<td>Clinicians normally have access to patients’ clinical history and having this information available would be fundamental to support the interpretation of icompass outputs.</td>
</tr>
<tr>
<td>Current medications or interventions</td>
<td>Clinicians can judge the progression of the condition also on the basis of medication or interventions. This information is currently not part of the outcome.</td>
</tr>
<tr>
<td>Functional tests</td>
<td>Functional tests, such as behaviour questionnaires, or activities of daily living, are a fundamental clinical indicator of the patient’s progress, which is currently not included in icompass.</td>
</tr>
<tr>
<td>MRI images of reports</td>
<td>The current version of the tool only extracts volumetric data. However, clinicians would normally have access to the radiologist report or to the images, a fundamental resource for them to contextualise the imaging data.</td>
</tr>
</tbody>
</table>

**Customised markers**

<table>
<thead>
<tr>
<th>Needs based on current clinical practice, not present in the tool</th>
<th>Customised markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Their own functional tests</td>
<td>Each clinician/centre has a personalised set of functional tests they consistently use with patients. Not only functional tests are currently not part of the model, but in case they will be in the future, they would also need to account for heterogeneity.</td>
</tr>
<tr>
<td>Most common cognitive markers</td>
<td>Each clinician/centre has a personalised set of cognitive tests they consistently use with patients. The current tool works with a limited and rather not-representative set of cognitive tests.</td>
</tr>
</tbody>
</table>

### Needs clinicians wish the tool could have in the future

**Time information**

<table>
<thead>
<tr>
<th>Needs based on current clinical practice, not present in the tool</th>
<th>Time information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeline with intervals between assessment</td>
<td>Clinicians would need a tool that could indicate the amount of time between consecutive assessments.</td>
</tr>
<tr>
<td>Forecast in years</td>
<td>Clinicians would need a tool that can indicate the specific years to patients’ conversion to more severe stages.</td>
</tr>
</tbody>
</table>

**Level of detail in the graph**

<table>
<thead>
<tr>
<th>Needs based on current clinical practice, not present in the tool</th>
<th>Level of detail in the graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separate evolutions for type of markers</td>
<td>Clinicians would need a tool that not only merges the trajectory of all markers into a single output graph, but to also be able to visualise trajectories for different types of markers.</td>
</tr>
<tr>
<td>Customisable</td>
<td>Clinicians wish they could personalise the output graph in different ways, for example the types of markers included in the output screen if any of them is not relevant to their practice, or the level of data granularity such as for time.</td>
</tr>
</tbody>
</table>

**Stratification and forecast**

<table>
<thead>
<tr>
<th>Needs based on current clinical practice, not present in the tool</th>
<th>Stratification and forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtypes</td>
<td>Clinicians would need to be able to see different subtypes of the disease progression, given that patients could follow different disease courses. One single pathway of progression is felt as reductive.</td>
</tr>
<tr>
<td>Conversion risk</td>
<td>Clinicians wish to access a score that represents the risk for the patient to progress to more severe stages of the disease.</td>
</tr>
</tbody>
</table>

**Future value**

<table>
<thead>
<tr>
<th>Needs based on current clinical practice, not present in the tool</th>
<th>Future value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evolution due to treatment</td>
<td>Clinicians wish the tool could integrate treatment or medication data into the evolution graph.</td>
</tr>
</tbody>
</table>
F.4 Information and consent
YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of Study: icompass – technical workshops
Department: UCLIC
Researcher(s): Maura Bellio
Principal Researcher: Ann Blandford

1. Invitation Paragraph
You are being invited to take part in my Ph.D. research project. Before you decide, it is important for you to understand why this study is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

2. What is the project’s purpose?
The aim of this project is to explore, discuss, and produce actionable insights with technical teams involved in the development of predictive models or the clinical tool that has been initially tested with clinicians. The workshop is designed based on the outcomes from the end users, whilst the outcomes from this study will be informative for the next steps in the tool’s development.

3. Why have I been chosen?
You have been invited to participate because you are:
- Involved in the methods, regulatory, or technical side of the tool’s development (disease progression models’ experts, front-end developers, clinical evidence team, regulatory requirements and quality team)
- Aged 18 or over
- Able to communicate effectively in English, and do not consider yourself to be a vulnerable adult.
- Able to give informed consent.

4. Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. You can withdraw at any time without giving a reason. If you decide to withdraw you will be asked what you wish to happen to the data you have provided up that point.

5. What will happen to me if I take part?
The workshop will last approximately 60 minutes and will involve three different group activities, addressing different aspects of clinicians perspective and feedback on the tool. You will not be asked personally sensitive information, and can refuse to answer any questions that you don’t wish to. There will be a follow-up session to inform participants about results from this study.
6. Will I be recorded and how will the recorded media be used?
With your permission, the session will be audio recorded. Transcriptions of the audio recording will be used only for analysis and for illustration in conference presentations and lectures. The audio recordings will be deleted once they have been transcribed and any identifying information will be removed during transcription.

7. What are the possible disadvantages and risks of taking part?
No disadvantages or risks of taking part have been identified. In the unlikely event that participating causes you any distress, you are free to withdraw, to discuss concerns with the researcher or the Principal Investigator.

8. What are the possible benefits of taking part?
While there are no immediate benefits to you from taking part, you will contribute in shaping the design and concept of this tool. We hope that this study will help you reflect on how predictive models for Alzheimer’s Disease can soon become a support in current clinical practice. We aim to share our findings afterwards, to inform the future design of the tool.

9. What if something goes wrong?
If you have any concerns with the conduct of this study, please raise them in the first instance with Professor Ann Blandford (a.blandford@ucl.ac.uk). If your concerns are not addressed to your satisfaction, then you may contact the Chair of the UCL Research Ethics Committee – ethics@ucl.ac.uk

10. Will my taking part in this project be kept confidential?
All the information that we collect about you during the course of the research will be kept confidential, subject to legal constraints and professional guidelines. You will not be identifiable in any ensuing reports or publications.

11. What will happen to the results of the research project?
This study is for my Ph.D. project, and the findings will be reported in my dissertation. Depending on the findings, my supervisor and I may also publish the results in a journal or conference paper. If you would like to receive a copy of these, let me know and I will send it as soon as the documents will be available. Anonymised data will be stored securely for five years, and may be reviewed in subsequent studies that have a related focus.

12. Local Data Protection Privacy Notice

Notice:
The controller for this project will be University College London (UCL). The UCL Data Protection Officer provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk

This ‘local’ privacy notice sets out the information that applies to this particular study. Further information on how UCL uses participant information can be found in our ‘general’ privacy notice:

For participants in research studies, click here

The information that is required to be provided to participants under data protection legislation (GDPR and DPA 2018) is provided across both the ‘local’ and ‘general’ privacy notices.

The lawful basis that will be used to process your personal data are: ‘Public task’ for personal data.

Your personal data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data you provide we will undertake this, and will endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, or if you would like to contact us about your rights, please contact UCL in the first instance at data-protection@ucl.ac.uk.
13. Contact for further information
Contact details for me and my supervisor are provided at the top of this sheet; feel free to contact either of us if you have queries or concerns.

Thank you for reading this information sheet and for considering taking part in this study.
CONSENT FORM FOR: ICOMPASS – TECHNICAL WORKSHOPS
Please complete this form after reading the Information Sheet or listening to an explanation of the study.

Title of Study: icompass – technical workshops
Department: UCLIC
Researcher(s): Maura Bellio
Principal Researcher: Ann Blandford
UCL Data Protection Officer: Alexandra Potts
This study has been approved by the UCLIC Research Ethics Committee: Project ID number: UCLIC/1819/006/BlandfordProgrammeEthics Data Protection Number: ZE384106/2020/03/123

Thank you for considering taking part in this research. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to.

I confirm that I understand that by ticking each box below I am consenting to this element of the study. I understand that unticked boxes mean that I DO NOT consent to that part of the study. I understand that by not giving consent for certain elements, I may be deemed ineligible for the study.

|   |   
|---|---
| 1. | I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me and to ask questions which have been answered to my satisfaction. I agree to take part in the testing and interview.  
|   | Tick  
| 2. | I understand that data will be anonymised and that it will not be possible to link my personal data (consent, contact details) with the study data. I understand that according to data protection legislation, ‘public task’ will be the lawful basis for processing.  
|   |   
| 3. | I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified. Data gathered in this study will be stored anonymously and securely. It will not be possible to identify me in any publications.  
|   |   
| 4. | I understand that the information I have submitted will be published as a report and I wish to receive a copy of it. Yes/No  
|   |   
| 5. | I consent to my interview being audio recorded and understand that the recordings will be destroyed following transcription.  
|   |   
| 6. | I am aware of who I should contact if I wish to lodge a complaint.  
|   |   

Name of participant __________________________ Date __________________________ Signature __________________________
Researcher __________________________ Date __________________________ Signature __________________________
F.5 Background survey
icompass workshop

Start of Block: Background questions

Q15 The following questions are for us to better understand your background and your area of expertise.

Q2.2 Your job title

________________________________________________________________

Q2.4 Location

________________________________________________________________

Q2.3 Years in this position

________________________________________________________________

Q2.5 Previous occupation/expertise

________________________________________________________________

End of Block: Background questions

Start of Block: Expertise
Q16 How much knowledge do you have on how Alzheimer's Disease (AD) clinical specialists work in practice?

- Excellent
- Good
- Average
- Poor
- None

Q47 How much knowledge do you have on clinical needs and practice in managing AD?

- Excellent
- Good
- Average
- Poor
- None

Q49 How much knowledge do you have around clinical decision-support tools?

- Excellent
- Good
- Average
- Poor
- None
Q51 How much knowledge do you have on predictive models for neurodegenerative disease, how they work and their output?

- Excellent
- Good
- Average
- Poor
- None

Q53 How much knowledge do you have on system development (front-end and back-end)?

- Excellent
- Good
- Average
- Poor
- None

Q55 How much knowledge do you have on quality and regulations around the use of AI systems in healthcare, and the integration of AI tools in the clinical scenario?

- Excellent
- Good
- Average
- Poor
- None
Appendix G

Requirements
Requirements for the initial set and the first iteration are presented in the following tables. Tables are divided by category: Technical factors (Table 1), User factors (Table 2), Organisation factors (Table 3).

The parameters in the table are the following:

- **Dimension and evaluation measure**: specific subcategories;
- **ID**: unique identifier for a requirement. The ID is built based on a tree-like structure: [category]-[dimension]-[n_evaluation]-[n];
- **Requirement**: the descriptive text of the requirement;
- **Evidence – Research**: which study within the exploratory phase (Chapter 3) contributed to the requirement;
- **Evidence – Testing**: which study within the testing and review phase (Chapter 4.3, 5.1) contributed to the requirement;
- **Ref**: Literature supporting the requirement (a = Horsky et al., 2012, b = Kilsdonk et al., 2017, c = Miller et al., 2018, d = Sutton et al., 2020.)

The iteration of requirements is highlighted by two factors. The **Evidence – Testing** column reports which study in the testing or review phase confirmed a previously defined requirement. When new requirements were introduced as a consequence of the testing or review phase, that row is highlighted in grey.
1. Technical factors' requirements.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>System quality</td>
<td>Data quality</td>
<td>IT-SQ-1-1</td>
<td>Data used to train the model should be of high-quality standards to ensure system quality.</td>
<td>Clinical interviews, Tech workshop</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-SQ-1-2</td>
<td>Data used to train the models should be dynamic to allow the models to run correctly.</td>
<td>Tech workshop</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-SQ-1-3</td>
<td>Data should be collected according to specified requirements to ensure quality.</td>
<td>Tech workshop</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td>Validation</td>
<td></td>
<td>IT-SQ-2</td>
<td>Model output should be validated for other applications and with real-world data to assess the quality of the system.</td>
<td>Clinical interviews, Tech experts workshop</td>
<td></td>
<td>a, b, d</td>
</tr>
<tr>
<td>Reference population</td>
<td></td>
<td>IT-SQ-3</td>
<td>The population used for the training should be clarified and consistent with the specific intended scenario to be representative of the context.</td>
<td>Clinical interviews</td>
<td></td>
<td>a, b, d</td>
</tr>
<tr>
<td>Information Quality</td>
<td>Relevance</td>
<td>IT-IQ-1-1</td>
<td>Relevant cognitive tests should be included in the input set to improve the information quality.</td>
<td>Clinical interviews</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-IQ-1-2</td>
<td>The heterogeneity of some markers should be considered in order to tailor the set of information used so to be relevant to the specific context/user.</td>
<td>Clinical interviews</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-IQ-1-3</td>
<td>Only data pertaining to AD should be included in the input requirements to run the models, to promote relevance of data used and of output provided.</td>
<td>Clinical interviews, Survey, Tech workshop</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-IQ-1-4</td>
<td>Terminology and scales used in the system’s interface should reflect clinical terminology to ensure the information provided is relevant to clinicians.</td>
<td>Clinical interviews</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
<td>IT-IQ-2-1</td>
<td>Data used to train the model should be consistent for different patients to ensure a proper and informative output.</td>
<td>Clinical interviews, Tech workshop</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-IQ-2-2</td>
<td>Data required should include clinical notes, to facilitate clinical practice by having all information in one place.</td>
<td>Clinical interviews</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td>Information delivery</td>
<td>Input</td>
<td>IT-ID-1-1</td>
<td>Steps required to input data in the platform should follow an intuitive logic to support a correct input of data</td>
<td>Clinical interviews</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-ID-1-2</td>
<td>Input of patient’s clinical information into the system should be required to promote completeness of data</td>
<td>Clinical interviews</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-ID-1-3</td>
<td>Sections for different types of markers, with indications for value intervals and reference population, should be clear and separate to support a more intuitive input screen</td>
<td>Clinical interviews</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-ID-1-4</td>
<td>Each input query should be complemented by measure units and cut-off scores to promote better understanding of data and facilitate clinical data entry</td>
<td>Clinical testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-ID-1-5</td>
<td>The volumetric data from the MRI automatically entered should also give clinicians a view on the radiologist report or the images to facilitate input understanding</td>
<td>Clinical testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT-ID-1-6</td>
<td>The data requested as clinical information input should comply to GDPR regulations, and be consistent and relevant across patients to represent a meaningful data entry</td>
<td>Clinical testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td>IT-ID-2-1</td>
<td>The information presented on the screen should not be too overwhelming and dense to promote a better understanding and delivery of information</td>
<td>Clinical interviews</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IT-ID-2-2</td>
<td>The interface should always keep the available input visible to offer clinicians an overview of their available data</td>
<td>Clinical interviews</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT-ID-2-3</td>
<td>The system’s output interface should allow to easily browse past assessments to facilitate comparison</td>
<td>Tech workshop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT-ID-2-4</td>
<td>The output should include a confidence interval to allow for clinical interpretation</td>
<td>Clinical interviews</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT-ID-2-5</td>
<td>The scales used to represent the model output should reflect clinical current labelling and terminology, to facilitate the information delivery</td>
<td>Clinical interviews, Tech workshop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT-ID-2-6</td>
<td>The system should provide the option to visualise the evolution for every type of marker to facilitate the understanding of individual marker’s trajectories</td>
<td>Clinical testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT-ID-2-7</td>
<td>The system should not make clinical claims or come to clinical conclusions to comply to regulatory requirements for CDSS</td>
<td>Tech experts workshop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Constraints**

| IT-C-1 | The system should currently not be intended to account for disease-modifying treatments | Tech experts workshop |
| IT-C-2 | The range and type of customisable features should be carefully considered and evaluated with user testing and understanding of the system. Customisation might currently be intended for the way in which the information is displayed, to facilitate output understanding, interpretability and trust from the user. In order to include a customised marker (i.e. cognitive or behavioural), enough training data for a specific population should be available for the model to provide reliable outcomes. | Tech experts workshop |
| IT-C-3 |  | Tech experts workshop |
## 2. User factors' requirements.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>System (intended) use</td>
<td>Expectations and beliefs</td>
<td>UF-SI-1-1</td>
<td>The system should provide information on stages and progression to support clinical decision making</td>
<td>Clinical interviews</td>
<td>Clinical testing, Tech experts workshop</td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SI-1-2</td>
<td>The system should provide information on early stages of the disease to allow a timely intervention</td>
<td>Clinical interviews, Tech workshop</td>
<td></td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SI-1-3</td>
<td>The system should suggest to clinicians the most relevant examinations to consider to optimise the system results and provide more robust outcomes</td>
<td>Tech experts workshop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge and expertise</td>
<td></td>
<td>UF-SI-2-1</td>
<td>The labels used in the output classification should reflect the knowledge and expertise pertaining to the clinical specialists to facilitate interpretation and understanding</td>
<td>Clinical interviews</td>
<td>Clinical testing, Tech experts workshop</td>
<td>a, b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SI-2-2</td>
<td>The system should not provide a unique numerical outcome for a stage to account for clinical interpretation</td>
<td>Clinical interviews</td>
<td></td>
<td>a, b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SI-2-3</td>
<td>The system should acknowledge the importance of the clinical impression as a relevant factor in the results interpretation to facilitate clinical adoption and use</td>
<td>Clinical testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SI-2-4</td>
<td>The system should aid the clinical decision-making process without being prescriptive, to work as a support to clinical practice and to not overtake the role of clinicians</td>
<td>Tech experts workshop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SI-2-5</td>
<td>The cognitive scoring should take into account the nuances of a specific population or environment to be a valid indicator of patient cognition.</td>
<td>Clinical testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
<td>UF-SI-3-1</td>
<td>The system should integrate all markers in a unique visualisation to facilitate data understanding</td>
<td>Clinical interviews, Survey</td>
<td>Clinical testing</td>
<td>b, d</td>
</tr>
<tr>
<td>System understanding</td>
<td>Training</td>
<td>UF-SU-1-1</td>
<td>The system should include quantitative data to support an objective understanding and interpretation of the condition</td>
<td>Clinical interviews, Survey</td>
<td>Clinical testing</td>
<td>b, d</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>---------------------------</td>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SU-1-2</td>
<td>The system should support understanding of the progression for uncertain cases to support clinical management</td>
<td>Clinical interviews, Survey</td>
<td>Clinical testing</td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SU-1-3</td>
<td>The system should be used to facilitate the discussion process in MDTs</td>
<td>Clinical interviews</td>
<td>Clinical testing</td>
<td>b, c, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SU-1-4</td>
<td>The system should facilitate clinical individual preparation for follow-up visit to support patient care</td>
<td>Clinical interviews</td>
<td>Clinical testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SU-1-5</td>
<td>The system should be intended to facilitate the exchange of opinions between colleagues</td>
<td>Clinical testing</td>
<td>Clinical testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UF-SU-1-6</td>
<td>The system at its current concept should not be used to communicate to patients to prevent them from being discouraged or misunderstanding specialised domain-knowledge</td>
<td>Clinical testing, Tech experts workshop</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| System understanding | Interaction | UF-SU-2 | The system should include interactive features to allow clinical users’ on-demand exploration and investigation of information | Clinical interviews | Clinical testing | a |

<table>
<thead>
<tr>
<th>System understanding</th>
<th>Trust and uncertainty</th>
<th>UF-SU-3-1</th>
<th>The system should allow a degree of uncertainty, to account for clinical interpretation</th>
<th>Clinical interviews</th>
<th>Clinical testing</th>
<th>b, d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>UF-SU-3-2</td>
<td>The system should allow for interactive data exploration and customisation for what/how data is visualised, to account for trust and interpretability</td>
<td>Clinical testing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Organisation factors’ requirements.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>Workflow</td>
<td>OR-S-1-1</td>
<td>The system should be seamlessly integrated within the workflow to facilitate adoption and use</td>
<td>Clinical interviews, Surveys, Tech workshop</td>
<td>Clinical testing, Tech experts workshop</td>
<td>a, b,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR-S-1-2</td>
<td>The workflow should account for a person who enters relevant data in a digital form to make the system work</td>
<td>Clinical testing, Tech experts workshop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR-S-1-3</td>
<td>A technical expert should be part of MDTs to support the discussion on the system’s characteristics and model’s output</td>
<td>Clinical testing, Tech experts workshop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resources</td>
<td></td>
<td>OR-S-2-1</td>
<td>The technical set-up of the Structure should accommodate the integration of the proposed system to promote a seamless adoption and use</td>
<td>Clinical interviews</td>
<td></td>
<td>a, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR-S-2-2</td>
<td>The Structure should ensure access to specific resources to run the model</td>
<td>Clinical interviews, Surveys, Tech workshop</td>
<td>Clinical testing, Tech experts workshop</td>
<td>b, d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR-S-2-3</td>
<td>The Structure/clinical setting should have access to existent data in a GDPR compliant and regulated manner, to acknowledge previously recorded data</td>
<td>Clinical testing, Tech experts workshop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td>Costing</td>
<td>OR-E-1</td>
<td>Costing should be evaluated to ensure adoption of the system</td>
<td></td>
<td>Tech experts workshop</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>Support</td>
<td>OR-E-2</td>
<td>A regular quality check and technical maintainance should be conducted to ensure the system is functioning correctly</td>
<td>Clinical interviews</td>
<td>Clinical testing</td>
<td>b, d</td>
</tr>
<tr>
<td>Interoperability</td>
<td>Standards</td>
<td>OR-I-1-1</td>
<td>The system should comply to official standards and be rated towards TRL to acknowledge its level of readiness to be used safely in clinical practice</td>
<td>Clinical interviews, Survey</td>
<td>Clinical testing</td>
<td>b, d</td>
</tr>
<tr>
<td>OR-I-1-2</td>
<td>A manual should include information of the relevant standards and limitations of the system to be safely used in the clinical environment</td>
<td>Clinical interviews, Survey</td>
<td>Clinical testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Third party access</td>
<td>OR-I-2</td>
<td>The system should facilitate access to relevant third parties and information sharing within and between clinical centres, given the necessary precautions and permissions</td>
<td>Clinical interviews, Survey</td>
<td>b, d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix H

Future work proposal

H.1 Suggested Phases
<table>
<thead>
<tr>
<th>Phase</th>
<th>Teams</th>
<th>Tasks</th>
</tr>
</thead>
</table>
| **Icompass iteration: design and technical implementation** | Design team<br>User research team | • Integrate icompass in a system of patients’ records  
• Explore notes recording functionalities  
• Explore integration of digital cognitive tests |
| (Estimated time: 6-8 months)                      | Technical team (model developers, front-end and back-end developers) | • Plan for the integration with EHR system (e.g., Epic) and technical requirements  
• Plan for the handling of data in the backend  
• Plan for the integration of digital cognitive tests in the system and in the model |
| Quality and regulation team                                                  | • Initial risk assessment plan (PHA)  
• Initial exploration of requirements for medical device classification |
| **Controlled clinical testing**                       | User research team                          | • Design the studies  
• Apply for ethics  
• Study 1: usability study (+ pilot)  
• Study 2: observational study (+ pilot)  
• Data analysis  
• Presentation of results  
• Identification of clinics for real-world testing |
| (Estimated time: 4 months + NHS ethics)            | Design team                                      | • Address eventual design feedback from testing                       |
| **Icompass iteration: design and technical implementation** | Technical team | • Adaptation of the system to the clinical setting of choice for the real-world testing |
| (Estimated time: 8 months)                          | Quality and regulation team                   | • Risk assessment (FMEA)  
• Eventual work on medical device classification  
• Production of manual |
| **Real world testing** (1/2 clinics)                 | User research team                           | • Design the study  
• Apply for ethics  
• Study 3: observational and user experience research (+ pilot)  
• Data analysis |
| (Estimated time: 4 months of data collection)       |                                            |                                                                      |
| + 2 months for design and analysis + NHS ethics) | • Presentation of results  
• Implications for launch (eventual new release?) |
H.2 Suggested Future Studies
<table>
<thead>
<tr>
<th>Phase 2: Usability Study</th>
<th>Phase 2: Observational study</th>
<th>Phase 4: Real-world testing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aim</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation of the general usability of the interface and interaction flow</td>
<td>Observe the clinical interaction with icompass, eventual obstacles, uncertainties, and use</td>
<td>Study the integration of icompass in the real-world scenario, and eventual hurdles, over a period of 4 months (1 or 2 clinical centres)</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Eye tracking data (heatmaps and gaze analysis)</td>
<td>• Assigned task (e.g., preparation for appointments and MDTs discussion using icompass)</td>
<td>• Weekly report from the clinical representatives</td>
</tr>
<tr>
<td>• Log data (e.g., timing, clicks)</td>
<td>• Observation on site</td>
<td>• Observation on site</td>
</tr>
<tr>
<td>• Time on task</td>
<td>• Note taking</td>
<td>• Log data</td>
</tr>
<tr>
<td>• Visit preparation time</td>
<td>• Contextual inquiry</td>
<td>• Note taking</td>
</tr>
<tr>
<td>• Questionnaires on usability</td>
<td>• Questionnaire and semi-structured interview at the end</td>
<td>• Regular questionnaires and semi-structured interview</td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 15-20$</td>
<td>$N = 8-10$</td>
<td>$N = 1-2$ clinical centres</td>
</tr>
<tr>
<td>• Neurologists, neuropsychologists, Psychiatrists working in specialised memory clinics with AD patients</td>
<td>• Neurologists, neuropsychologists, Psychiatrists working in specialised memory clinics with AD patients</td>
<td>• Neurologists, neuropsychologists, Psychiatrists working in specialised memory clinics with AD patients</td>
</tr>
<tr>
<td>• Various levels of experience</td>
<td>• Various levels of experience</td>
<td>• Various levels of experience</td>
</tr>
</tbody>
</table>