CHAPTER 1

INTRODUCTION

Drug development typically consists of conducting multiple studies to evaluate a compound's efficacy, side-effect profile, and risk management strategies. The initial phase involves preclinical testing in vitro -primarily to confirm mechanistic target engagement, such as receptor binding and affinity- followed by animal studies to assess desired pharmacological effects, pharmacokinetic properties, and toxicology. Once preclinical findings sufficiently demonstrate both efficacy and safety, the compound can proceed to evaluation in humans.

Clinical trials are conducted in human subjects and are typically categorized into three types, as defined by the European Medicines Agency (EMA): human pharmacology, safety and efficacy, and special populations.[1] The earliest human study, known as a First-in-Human (FIH) trial, is designed to closely monitor safety and pharmacokinetics (i.e., absorption, distribution, metabolism, and excretion). Desired pharmacodynamic outcomes can also be evaluated during this stage.[2] To minimize confounding factors, these FIH trials are often conducted in healthy participants, thereby offering a clearer assessment of side effects unclouded by disease symptoms. Additionally, healthy subjects are typically more readily recruited, facilitating a faster start to clinical development. These studies in healthy volunteers are commonly referred to as Phase 1 studies.

Following favourable Phase 1 results—demonstrating safety and suitable pharmacokinetics—trials move on to the patient population. Early patient studies aim to identify an effective dose, evaluate safety in the target group, and verify the drug's underlying concept. Such trials are generally called Phase 2 studies. Larger, adequately powered studies that definitively test a drug's efficacy are referred to as Phase 3. By the time a drug reaches advanced clinical development, the combined outcomes of Phases 1, 2, and 3 must demonstrate both safety and efficacy, as well as a clear advantage over existing treatments.

To achieve this proof, the final stage before registration typically involves a trial in the intended patient population, using endpoints that drug regulators recognize as clinically meaningful. Known as clinical outcome assessments (COAs), these endpoints are often tied directly to the patient's quality of life. COAs may be patient-reported, observer-reported, clinician-rated, or part of a standardized performance measure.[3] For instance, in patients with epilepsy, a reduction in seizure frequency is widely accepted as a relevant outcome. Other examples include improved scores on daily-life questionnaires completed by asthma patients or increased walking distance in the 6-minute walk test for individuals with Duchenne muscular dystrophy.[3]

Progressing from a FIH trial to these large-scale registration studies is both lengthy and expensive. Moreover, the likelihood of success is relatively low—only around 10% of drugs entering clinical development ultimately achieve registration.[4] Although precise figures are infrequently disclosed, costs are estimated at around 500 million euros per approved drug, excluding expenses related to failed compounds.[5]

The clinical 'funnel' can be illustrated by the transition probabilities between phases—Phase 1 to Phase 2, Phase 2 to Phase 3, and Phase 3 to registration—with the majority of terminations occurring in Phase 2.[4] The relatively higher success rate in progressing from Phase 1 to Phase 2 is often attributed to the narrower focus of Phase 1 studies on safety and pharmacokinetics, as well as the solid pharmacological data collected from animal models.

In contrast, moving from Phase 2 to the large Phase 3 registration trials is less frequently successful. This juncture is considered critical because trial sizes and associated costs escalate significantly, making failures more financially consequential.[5] To mitigate risks, drug developers strive to gather robust evidence early in development and conduct smaller, proof-of-concept studies with sensitive assessments. These approaches help reduce the likelihood of unfavourable outcomes during registration trials.

One strategy aimed at curbing development costs is a phase-agnostic, question-based framework called Question-Based Drug Development (QBDD). This method emphasizes systematically asking the most critical—and riskiest—questions early. It encompasses six core questions spanning the drug's path through the human body, from 'Does the drug reach its site of action?' to 'What are the on-target and off-target pharmacological effects?' By directly translating these questions into study objectives and prioritizing them according to each drug's financial risks, early failures can be identified, thus reducing unnecessary later-stage expenditures.[6]

Even with QBDD principles, a considerable gap remains between initial patient studies and large, costly registration trials. Early patient studies (often termed Phase 2a) typically focus on target engagement—demonstrating that the drug modifies the disease's underlying pathophysiology.[7] Subsequent Phase 2b studies frequently resemble small-scale registration trials, employing the same COAs as in Phase 3. Unfortunately, these trials may be underpowered due to smaller sample sizes. The potential for false-positive (Type I) errors could lead to failure in Phase 3 and major financial repercussions, while false-negative (Type II) errors might cause premature abandonment of promising therapies.

Ideally, Phase 2b studies should incorporate endpoints that balance the mechanistic accuracy of Phase 2a measures with the real-world relevance of Phase 3 COAs. These 'bridging' endpoints are more precise and less variable than standard COAs -thus requiring fewer participants- yet are still closer to real-life outcomes than basic mechanistic readouts. In essence, they serve as an intermediary step between early proof-of-concept and full-scale clinical outcome assessments, aiding in a more reliable transition to successful Phase 3 trials.

BIOMARKERS

Biomarkers are objective measures of biological processes, states, or conditions that play a central role in evaluating safety and efficacy throughout drug development.[8] Typically classified into two main groups – 'safety biomarkers' and 'response biomarkers'- they support clinical decision-making by providing early and reliable indicators of a drug's performance.[8,9] An ideal biomarker should be safe, easy to measure, and cost-effective, while also meeting key technical requirements such as sensitivity, specificity, reproducibility, repeatability, and cross-species translatability.

Pharmacodynamic biomarkers quantify the biological response to a therapeutic intervention, providing direct evidence of target engagement and effect on the disease pathway. By reflecting the mechanism of action of a drug, these markers facilitate dose selection and enable monitoring of treatment efficacy. Pharmacodynamic biomarkers can help researchers bridge the gap between preclinical findings and human studies, as they deliver measurable endpoints that validate the compound's intended action early in clinical development. This is called proof-of-concept.

Surrogate endpoints, derived from validated biomarkers, serve as proxy measures that can predict or correlate with clinically meaningful outcomes, potentially accelerating drug development and reducing costs by providing an earlier readout of treatment efficacy or disease progression.[9] However, these endpoints are only referred to as surrogate endpoints when drug registration authorities formally accept their use in place of COAs that reflect patient symptoms and quality of life. Such acceptance requires extensive evidence demonstrating a strong correlation between the surrogate endpoint and the traditional clinical endpoints, ensuring that it reliably reflects the ultimate clinical benefit or risk. As a consequence, surrogate endpoints are seldom approved and rarely used. Nevertheless, in instances where a surrogate endpoint is thoroughly validated and mechanistically linked to the disease pathway, it can facilitate smaller, more efficient proof-of-concept studies without compromising scientific rigor.

COAs in drug registration studies evaluate clinically meaningful benefits, such as improvements in symptoms, function, or quality of life. Because COAs often rely on clinician-based evaluations, subjective patient-reported outcomes, or a combination of the two, they inherently introduce variability in the data. As a result, larger sample sizes are typically required in (phase 3) registration trials to ensure sufficient statistical power to detect true treatment effects.

Biomarkers used in the above-mentioned bridging trials should present a more real life setting and/or have a high relevance to the real-world context. The term often used to describe this characteristic is ecological validity.

ECOLOGICAL VALIDITY

The term ecological validity is frequently used in the field of (neuro)psychology and often confused with external validity and mundane realism.[10,11] Whereas external validity concerns the generalizability of findings to various populations, settings, and points in time, mundane realism relates specifically to how closely an experimental setting mirrors everyday life. Ecological validity may incorporate elements of mundane realism to determine whether the study's variables and conclusions are truly relevant and applicable to real-world contexts. By contrast, external validity extends beyond ecological validity by examining how well a study's results can be applied to a different target population. These differences are often misunderstood and can complicate the interpretation of the literature on ecological validity.[11]

Within this thesis, ecological validity refers to the degree to which biomarkers used in earlier phase trials can be generalized to the clinical and demographic conditions of the COAs used in registration trials, encompassing both trial settings and study populations. Although the use of the term ecological validity for this phenomenon is not common, the term is used across studies – spanning aviation research,[12] mild cognitive impairment,[13] Parkinson's disease,[14,15] and treatments involving benzodiazepines[16] or medication for opioid use disorder[17]- and reflects the pursuit of more realistic approaches to testing pharmacological interventions. Proposed strategies for enhancing ecological validity include integrating mobile phone data,[18] employing more detailed gait analyses,[19] and utilizing virtual reality.[20,21] However, the ecological validity of biomarkers in the early stages of drug development is rarely examined, limiting insights into biomarker potential and potentially affecting how results are interpreted.

RESEARCH OBJECTIVE AND STRUCTURE OF THIS THESIS

The aim of this thesis is to identify highly ecologically valid biomarkers for early phase clinical drug development and evaluate the ecological validity.

An example of translating clinical findings into measurements more relevant to real life is the study examining driving behaviour, described in **Chapter 2** of this thesis. This research focuses on biomarkers that assess the potential effects of pharmacological compounds on driving performance, as impaired driving carries significant safety risks. On-the-road driving tests, widely regarded as the gold standard with strong ecological validity, have traditionally been used to measure these effects.[22] However, these assessments are both time-consuming and expensive, as well as logistically complex to implement.

This study proposes using a driving simulator as an intermediate tool that bridges the gap between fundamental, laboratory-based evaluations of aspects of driving behaviour and the real-world on-the-road tests mandated by regulatory bodies such as the FDA.[23] While simulators inherently exhibit lower ecological validity than on-road driving tests, they offer higher ecological validity than psychomotor tests focusing on hand–eye coordination. Through the use of sleep deprivation to induce impaired driving performance, the study investigates how well results from these methods translate to one another. This tiered approach provides a structured pathway for assessing medication-related effects on driving before advancing to on-the-road trials.

Another bridging study between the clinical research unit and a real-life setting, described in **Chapter 3** of this thesis, uses a biomarker that evaluates a potential increase in fall risk. A commonly employed biomarker of postural stability in early-phase clinical trials is body sway, measured while individuals stand still with eyes closed. The total horizontal sway over a defined period is compiled into a single endpoint, and an increase in this metric generally signifies diminished stability, assumed to be associated with a greater likelihood of falling. However, actual fall risk is not directly measured, introducing uncertainty about potential safety concerns in later-phase drug development. An outcome measure that can be assessed in healthy volunteers but has higher ecological validity can be expected to have stronger predictive value for real-world fall risk in future trials.

In this chapter, we aimed to validate the previously described Interactive Walkway (IWW) for evaluating drug-induced effects. In the study described, we induced a mild fall risk by administering a benzodiazepine to healthy older adults and comparing its effects to a more selective hypnotic agent and placebo. The IWW offers higher ecological validity and greater specificity to fall risk compared to the body sway test. Additionally, the Timed Up and Go (TUG) test was included – a brief walking assessment in which participants stand from a chair, walk three meters, return, and sit down again. Finally, we assessed the ecological validity of each task within the context of the study findings.

Chapter 4 describes our efforts to develop a more ecologically valid method for quantifying muscle handgrip strength. Although grip strength is routinely used as an objective measure of muscular function in clinical trials and is often abnormal in patient populations, it does not adequately capture the functional outcomes relevant to patients' daily lives. An increase in grip strength alone is insufficient for regulatory approval; regulatory agencies require COAs, such as the 6-Minute Walk Test (6MWT), survival metrics, or timed chair-stand tests, which typically involve larger patient populations to detect clinically meaningful changes.

Emerging tools like PowerJar offer a better balance between quantifiability and ecological validity. Opening a jar, for instance, demands more than just hand grip and can also reflect the challenges posed by a unilateral impairment. PowerJar provides a more functional, contextually relevant measure of muscle strength than traditional grip strength, while offering higher resolution – and thus greater precision- than tests like the 6MWT. By more closely mirroring real-world tasks, PowerJar may serve as a more meaningful tool for assessing and demonstrating clinically relevant improvements in strength.

Chapters 5 and **6** of this thesis describe our initial efforts to establish a pain model that incorporates an affective component into a purely nociceptive task. Traditional pain assessments, which focus primarily on nociceptive processes,[24] may overlook drug effects that alter pain perception via emotional modulation. Anxiolytics, for instance, could be beneficial for pain management but are not typically evaluated with biomarkers that capture the emotional response component in healthy volunteers. By integrating an affective dimension into an evoked pain task, we enhance its ecological validity.

In many chronic pain conditions, the emotional response is believed to significantly influence the perception of pain.[25,26] To replicate this aspect in healthy volunteers, we increased the emotional response to pain through a virtual reality simulation that displays skin damage at the site of the painful stimulus. This novel addition to the nociceptive testing battery was initially examined for repeatability and validated using patient-reported outcome measures. Subsequently, a first clinical trial was conducted with a drug known to suppress emotional responses, advancing our understanding of how the affective component interacts with pain modulation.

Finally, **Chapter 7** presents the findings from these studies and discusses the evaluation of the biomarkers. Current literature on frameworks for assessing ecological validity is reviewed, and a new structured framework is proposed. This framework enables a standardized approach to quantifying the ecological validity of both existing and novel biomarkers, thereby promoting more effective use of biomarkers in early-phase clinical drug development and enhancing overall efficiency in the drug development process.

REFERENCES

1 European Medicines Agency. ICH guideline E8 (R1) on general considerations for clinical studies [Internet]. 2021. Available from: www.ema.europa.eu/contact

2 Cohen AF, Burggraaf J, Van Gerven JMA, Moerland M, Groeneveld GJ. The use of biomarkers in human pharmacology (Phase I) studies. Vol. 55, Annual Review of Pharmacology and Toxicology. Annual Reviews Inc.; 2015. p. 55–74.

3 US Food and Drug Administration. Clinical outcome assessment (COA) compendium [Internet]. 2021. Available from: https: //www.fda.gov/regulatory-

4 Thomas D, Chancellor D, Micklus A, LaFever S, Hay M, Chaudhuri S, et al. Clinical Development Success Rates and Contributing Factors 2011–2020. 2021.

5 Sertkaya A, Beleche T, Jessup A, Sommers BD. Costs of Drug Development and Research and Development Intensity in the US, 2000-2018. JAMA Netw Open. 2024 Jun 28;7(6).

6 de Visser SJ, Cohen AF, Kenter MJH. Integrating scientific considerations into R&D project valuation. Nat Biotechnol. 2020 Jan 1;38(1): 14–8.

7 Torres-Saavedra PA, Winter KA. An Overview of Phase 2 Clinical Trial Designs. Int J Radiat Oncol Biol Phys. 2022 Jan 1;112(1): 22–9.

8 Califf RM. Biomarker definitions and their applications. Exp Biol Med. 2018 Feb 1;243(3): 213–21.

9 US Food and Drug Administration. BEST (Biomarkers, EndpointS, and other Tools) Resource. 2016.

10 Beechey T. Ecological Validity, External Validity, and Mundane Realism in Hearing Science. Ear Hear. 2022 Sep 1;43(5): 1395–401.

11 Hammond KR. Ecological Validity: Then and Now. 1998.

12 Ehlert AM, Wilson PB. Stimulant Use as a Fatigue Countermeasure in Aviation. Aerosp Med Hum Perform. 2021;92(3): 190–200.

13 Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: Systematic review. J Med Internet Res. 2019 Aug 1;21(8).

14 Foreman KB, Addison O, Kim HS, Dibble LE. Testing balance and fall risk in persons with Parkinson disease, an argument for ecologically valid testing. Parkinsonism Relat Disord. 2011 Mar;17(3): 166–71.

15 Ramsperger R, Meckler S, Heger T, van Uem J, Hucker S, Braatz U, et al. Continuous leg dyskinesia assessment in Parkinson's disease – clinical validity and ecological effect. Parkinsonism Relat Disord. 2016 May 1;26: 41–6.

16 Curran HV. Psychopharmacology Benzodiazepines, memory and mood: a review. Psychopharmacology (Berl). 1991;105–6.

17 Marsch LA, Chen CH, Adams SR, Asyyed A, Does MB, Hassanpour S, et al. The Feasibility and Utility of Harnessing Digital Health to Understand Clinical Trajectories in Medication Treatment for Opioid Use Disorder: D-TECT Study Design and Methodological Considerations. Front Psychiatry. 2022 Apr 29;13.

18 Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: An explorative study. J Med Internet Res. 2016 Mar 1;18(3).

19 Ilg W, Seemann J, Giese M, Traschütz A, Schöls L, Timmann D, et al. Real-life gait assessment in degenerative cerebellar ataxia: Toward ecologically valid biomarkers. Neurology. 2020 Sep 1;95(9): E1199–210.

20 Segawa T, Baudry T, Bourla A, Blanc JV, Peretti CS, Mouchabac S, et al. Virtual Reality (VR) in Assessment and Treatment of Addictive Disorders: A Systematic Review. Front Neurosci. 2020 Jan 10;13.

21 Höller Y, Höhn C, Schwimmbeck F, Plancher G, Trinka E. Effects of Antiepileptic Drug Tapering on Episodic Memory as Measured by Virtual Reality Tests. Front Neurol. 2020 Feb 20;11.

22 Verster J, Roth T. Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP). Int J Gen Med. 2011;359.

23 FDA-CDER. Evaluating drug effects on the ability to operate a motor vehicle – guidance for industry [internet]. 2017. Available from: https: //www.Fda.Gov/drugs/guidancecomplianceregulatoryinformation/guidances/default.Htm

24 Siebenga PS, van Amerongen G, Okkerse P, Denney WS, Dua P, Butt RP, et al. Reproducibility of a battery of human evoked pain models to detect pharmacological effects of analgesic drugs. Eur J Pain [Internet]. 2019/02/23. 2019; Available from: https: //onlinelibrary.wiley.com/doi/pdf/10.1002/ejp.1379

25 Lumley MA, Cohen JL, Borszcz GS, Cano A, Radcliffe AM, Porter LS, et al. Pain and emotion: A biopsychosocial review of recent research. J Clin Psychol. 2011;67(9): 942–68.

26 Ruscheweyh R, Nees F, Marziniak M, Evers S, Flor H, Knecht S. Pain Catastrophizing and Pain-related Emotions. Clin J Pain. 2011;27(7): 578–86.