

# Disentangling complex interventions

Evaluating the impact and value of Everyday Life Rehabilitation (ELR)

André Sjöberg



UMEÅ UNIVERSITY

This work is protected by the Swedish Copyright Act (Act 1960:729)  
ISBN: 978-91-8070-988-0 (print)  
ISBN: 978-91-8070-989-7 (pdf)  
ISSN: 0346-6612  
Umeå University Medical Dissertations New Series no. 2418  
Electronic version available at: <http://umu.diva-portal.org/>  
Printed by: Scandinavian Print Group, Hägersten

*"Knowledge [...] is not a gradual approach to the truth. It is rather an ever increasing ocean of mutually incompatible alternatives, each single theory, each fairy-tale, each myth that is part of the collection forcing the others into greater articulation and all of them contributing, via this process of competition, to the development of our consciousness. Nothing is ever settled, no view can ever be omitted from a comprehensive account. [...] Experts and laymen, professionals and dilettanti, truth-freaks and liars – they all are invited to participate in the contest and to make their contribution to the enrichment of our culture".*

*- Paul Feyerabend, Against Methods.*



Have you heard the tale of the Evaluation Monster...?

It does not live under your bed or in an enchanted forest. It is found in research papers, grant proposals, and policy documents. It has one simple rule: everything must be measured, for you see, the monster demands certainty. If you could see it, you would not find sharp teeth or claws. You would see yourself. For the monster is born from our discomfort with conditionality and variability.

From the moment of their conception, complex interventions resemble a ball of yarn, woven from threads from the very system in which they are implemented. This is where the monster becomes dangerous. In its pursuit of certainty, it turns to experiments as a way to cut into the yarn. The problem is not that rigour is wrong, but that threads are severed rather than traced.

Experiments do not unravel the entire ball of yarn, nor should they be expected to. They must be complemented by approaches that explain how effects emerge. However, as our evaluations shape real decision with real consequences it is also important to not forget the value of methodological rigour.

In this thesis, I invite you to join me in demystifying this so-called monster. What does it get right? Where does it fail? And how might we harness its power? Our goal is not to fully disentangle the yarn, but to trace enough threads to make precise and credible causal statements, while exposing the remainder honestly.

# Table of contents

- Table of contents ..... I
- Abstract .....IV
- Svensk sammanfattning.....VI
- Abbreviations .....VIII
- List of figures ..... X
- List of tables ..... XI
- List of original papers ..... XII
- Introduction ..... 1
  - Aim of the thesis ..... 3
  - Theoretical perspectives on evidence ..... 3
    - To evaluate or not to evaluate? .....5
    - The growth of translational approaches ..... 8
    - Resource allocation, preference and utility ..... 9
    - Bridging the implementation gap ..... 11
    - The heterogeneity revolution ..... 12
  - The case of Everyday Life Rehabilitation ..... 14
    - Similar recovery-oriented interventions..... 15
    - Personal recovery ..... 15
    - Recovery through meaningful everyday activities ..... 16
    - Person-centeredness ..... 17
    - Practice leadership ..... 18
    - The population: Serious mental illnesses ..... 18
    - The setting: Supported accommodation..... 19
    - The wider context: A decentralised welfare system ..... 21
- Methods ..... 24
  - Interventional procedures .....25
  - Randomised Controlled Trial (Study I & II) .....27
    - Recruitment process ..... 29
    - Randomisation and blinding ..... 29
    - Pragmatic trial..... 30
    - Internal pilot phase (Study I)..... 31

|   |    |
|---|----|
| Outcome measures.....   | 31 |
| Comparator .....  | 34 |
| Statistical methods.....  | 35 |
| Health economical evaluation (Study III).....                       | 36 |
| Model structure.....  | 36 |
| Identification and quantification of resource use.....              | 37 |
| Valuation of resource use.....                                      | 39 |
| Calculation of health benefits .....                                | 41 |
| Cost-effectiveness analysis.....                                    | 42 |
| Sensitivity analysis.....   | 42 |
| Budget impact analysis .....  | 46 |
| Process Evaluation (Study IV).....                                  | 47 |
| Data generation: Focus group interviews .....                       | 47 |
| Data analysis: Critical Discourse Analysis .....                    | 48 |
| Ad-hoc analysis of implementational outcomes .....                  | 49 |
| Ethical Considerations .....  | 51 |
| Let's get philosophical .....                                       | 52 |
| Findings .....  | 54 |
| Demographics.....   | 54 |
| Did ELR promote personal recovery? .....                            | 57 |
| Estimation of the level of exposure.....                            | 59 |
| Impact of transforming ReQoL to QALYs.....                          | 60 |
| Is the intervention cost-effective? .....                           | 64 |
| Estimated costs of the intervention .....                           | 64 |
| How sensitive are the conclusions? .....                            | 66 |
| Budget impact analysis .....  | 70 |
| Discourses among housing staff.....                                 | 71 |
| Contextual challenges in this setting .....                         | 72 |
| How were discourse influenced by context?.....                      | 73 |
| How did discourse influence the uptake of ELR?.....                 | 74 |
| Early implementation outcomes .....                                 | 76 |
| Acceptability: Conflation of practices .....                        | 76 |
| Adoption: High willingness, but lack of effort .....                | 77 |
| Appropriateness: Important, just not for their residents .....      | 77 |
| Feasibility: A hypothetical threat to organisational stability..... | 78 |
| Fidelity: Low adherence to the educational components .....         | 78 |
| Penetration: Unclear reach among participants.....                  | 79 |
| Discussion .....  | 80 |
| Why did ELR succeed when others failed? .....                       | 81 |
| The role of the occupational therapist .....                        | 81 |
| The importance of meaningful activities .....                       | 82 |

|  |     |
|--|-----|
| Explaining deviant cases and uncertainties .....         | 83  |
| Threats towards causal inference.....                    | 83  |
| Resistance and gatekeeping among staff .....             | 84  |
| Sustainability, vague leadership and accountability..... | 85  |
| Prioritising welfare services .....                      | 87  |
| Validity of ReQoL: Is a QALY always a QALY?.....         | 87  |
| What are we willing to pay for societal benefits?.....   | 88  |
| Future of EBM: The case for realism.....                 | 89  |
| Why ontology matters .....                               | 90  |
| Risky predictions and experiments.....                   | 90  |
| The role of qualitative research in evaluations .....    | 91  |
| Disentangling the evaluation, and its limitations .....  | 92  |
| For whom, under what circumstances, and how? .....       | 93  |
| Limitations of the studies .....                         | 95  |
| What about the evaluation monster?.....                  | 97  |
| Concluding remarks .....                                 | 98  |
| Acknowledgement.....                                     | 101 |
| References.....  | 102 |

# Abstract

**Background:** Despite growing interest among policymakers in comparative effectiveness and cost-effectiveness over the past two decades, remarkably few interventions used within the welfare sector can genuinely be described as “*evidence-based*”. However, even where evidence exists, significant questions remain about how well findings translate to real-world conditions and whether they are sufficient to inform decision-making. Such issues are further amplified when evaluating complex interventions, where experimental approaches alone are often insufficient, and need to be complemented by other methods.

This thesis forms part of a larger research project evaluating the intervention of Everyday Life Rehabilitation (ELR), a recovery- and activity-based intervention targeting persons with serious mental illness (SMI) living in supported accommodation. The thesis comprises four studies addressing the intervention’s effectiveness, cost-effectiveness, and early implementation. ELR involves residents, rehabilitation professionals, housing staff and their managers, aiming to promote personal recovery through engagement in meaningful everyday activities. The overall objective of this thesis is to evaluate ELR by integrating findings across four distinct analytical approaches, while contributing to the broader discussion on the role of evaluations in welfare settings.

**Methods:** The thesis revolves around a pragmatic cluster randomised controlled trial (RCT) evaluating the effectiveness and cost-effectiveness of ELR against treatment-as-usual in promoting personal recovery and quality of life among persons with SMI living in supported accommodation. Alongside the trial, data on resource use and early implementation have been collected, which are examined through a mixed-methods approach, drawing on health economics and implementation science.

Study I is an internal pilot study (19 units and 48 participants) evaluating the feasibility of the trial and intervention, assessing the outcome measures, and informing an updated power calculation. Study II is the full-scale RCT (60 units and 161 participants) evaluating the effectiveness of ELR on the primary outcome measure Recovering Quality of Life (ReQoL). Study III is a trial-based cost-utility analysis estimating the interventions impact on costs and quality adjusted life years (QALYs). Finally, Study IV is a process evaluation exploring housing staff’s (n = 21)

discourses in response to the early implementation process, based on seven focus group interviews conducted following the trial.

**Results:** The findings demonstrates that those who received ELR had significantly greater improvements in ReQoL scores at six months compared with the control (20.1, 95% CI: 15.8 to 24.4). This represents a strong and clinically meaningful effect, and high goal attainment were observed among those who received the intervention. In addition, analyses suggest that ELR has a high probability of being cost-effective, and that it is quite affordable. Despite positive findings, several implementational challenges were observed during the trial. Discourses among housing staff revealed a perceived sense of scarce resources and vague leadership, which constrained staff's agency and fostered apprehension towards residents own choices, leading them to employ discursive strategies to rationalise resistance and gatekeeping.

**Conclusions:** ELR has been found to not only be an effective intervention, but also a cost-effective one. These findings represent robust empirical evidence, which is used to illustrate how findings can be contextualised through translational approaches, to enable more precise causal statements. While effects are context-dependent, the findings suggest that meaningful improvements in resident's quality of life can be achieved with relatively modest efforts. ELR has been developed to fit well within the legal and organisational structures of supported accommodation in Sweden, and holds potential not only as an intervention, but as a means of integrating healthcare services, promoting resident's agency, and enhancing the overall quality of services.

# Svensk sammanfattning

Den som behöver hjälp av samhället ska känna sig trygg med att de insatser som erbjuds ska bygga på den bästa tillgängliga evidensen. Trots detta så är experimentella studier fortfarande en bristvara inom stora delar av välfärden. Att utvärdera komplexa interventioner innebär unika problem som ytterligare försvårar användandet av experimentella studiedesigner. Det har i praktiken lett till en låg användning av evidensbaserade interventioner, åtminstone sådana som har utvärderats i experiment av god kvalitet. För att kunna arbeta evidensbaserat krävs tillförlitliga underlag om både interventioners effekter och kostnader.

Denna avhandling är del av ett större utvärderingsprojekt av interventionen Everyday Life Rehabilitation (ELR). ELR riktar sig främst till personer med långvarig och omfattande psykisk sjukdom som bor i kommunala särskilda boenden. Dessa personer lever ofta ett stillasittande och passivt liv, trots att kommuner enligt lag är skyldiga att verka för att de får möjlighet att delta i samhällets gemenskap och att leva som andra. ELR är utformad för att leva upp till gällande lagstiftning genom att erbjuda god, långsiktig, och person-centrerad rehabilitering.

Det övergripande syftet med avhandlingen är att utvärdera ELR's effekt, kostnadseffektivitet samt implementering inom ramen för en pragmatisk kluster-randomiserad kontrollerad studie. Ambitionen är också att avhandlingen ska belysa de möjligheter och utmaningar som är förknippade med utvärdering av komplexa interventioner inom välfärdssektorn. Avhandlingen omfattar fyra studier: en intern pilotstudie, en pragmatisk kluster-randomiserad kontrollerad effektivitetsstudie, en hälsoekonomisk kostnadsnyttoanalys samt en processutvärdering av den tidiga implementeringsprocessen.

Resultaten visar att ELR, efter bara sex månader, haft en statistiskt och kliniskt signifikant effekt på boendes livskvalitet och återhämtning. Interventionen är också relativt billig, och det är sannolikt att den skulle vara kostnadseffektiv om den introduceras i praktiken. Dock observerades en del svårigheter vid kommunens implementering. Bland annat framkom att kommunerna ofta hade bristande struktur för implementeringsprocessen och vagt ledarskap, vilket ledde till fördröjning vid uppstarten av insatsen och i samverkan mellan boendepersonal och rehabiliteringspersonal. Till följd av detta anammade boendepersonal i vissa boenden en defensiv hållning jämt emot interventionen och brukarnas valda mål, vilket i vissa fall ledde till

konflikter. Trots att implementeringsprocessen inte alltid fungerade optimalt, uttryckte boendepersonal en önskan om mer struktur, där både ELR och en ökad samverkan med rehabiliteringspersonal sågs som önskvärda inslag.

Avhandlingens slutsats är att ELR är en effektiv insats med goda förutsättningar att vara kostnadseffektiv om den implementeras i svenska särskilda boenden för personer med allvarlig psykisk sjukdom. ELR bygger på ett personcentrerat arbetssätt, där resurser inom den kommunala hälso- och sjukvården tas tillvara för att erbjuda en god och integrerad rehabilitering till en ofta förbisedd brukargrupp. Både interventionen och denna utvärdering ligger även väl i linje med de krav som ställs i den nya socialtjänstlagen.

# Abbreviations

EBM: Evidence-based Movement; Evidence-based Medicine.

RCT: Randomised controlled trials.

ELR: Everyday Life Rehabilitation.

SMI: Serious mental illnesses; Severe mental illnesses.

QALY: Quality-adjusted-life years

CFIR: Consolidated Framework for Implementation Research.

MRC: The Medical Research Council and National Institute for Health and Care Research

STAX-SA: Simple Taxonomy for Supported Accommodation

HSL: Swedish health care act; Hälso- och sjukvårdslagen.

SoL: Swedish social care act; Socialtjänstlagen.

LSS: Swedish disability act; Lag om stöd och services till vissa funktionshindrade.

PCRP: Person-centred recovery plan.

CHIME: Recovery framework; Connectedness, Hope, Identity, Meaning, Empowerment.

TAU: Treatment-as-usual.

ReQoL: Recovering Quality of Life.

RAS-DS: Recovery Assessment Scale—Domains & Stages.

GAS: Goal attainment scale.

COPM: Canadian Occupational Performance Measure.

ITT: Intention-to-treat.

MICE: Multiple imputations by chained equations.

CDA: Critical Discourse Analysis.

MD: Mean difference.

CI: Confidence interval.

ICER: Incremental cost-effectiveness ratio.

WTP: Willingness-to-pay.

# List of figures

|  |    |
|--|----|
| Figure 1 Ethical principles for priority setting.....                          | 9  |
| Figure 2 Illustration of the Swedish system, adopted from <sup>184</sup> ..... | 21 |
| Figure 3 Model structure .....   | 37 |
| Figure 5 Flowchart .....   | 55 |
| Figure 6 Distribution of ReQoL scores .....                                    | 58 |
| Figure 7 Distribution of RAS-DS scores .....                                   | 58 |
| Figure 8 Illustration of available process data.....                           | 59 |
| Figure 8 Distribution of utility scores.....                                   | 61 |
| Figure 9 Change of score in each item between baseline and 6 months            | 62 |
| Figure 10 Sensitivity analysis of items in the ReQoL-UI.....                   | 63 |
| Figure 11 Deterministic sensitivity analysis.....                              | 67 |
| Figure 12 Sensitivity analysis of effect retainment .....                      | 69 |
| Figure 13 Cost-effectiveness acceptability curve and plane.....                | 70 |
| Figure 14 Marginal costs per unit .....  | 71 |

# List of tables

|   |    |
|---|----|
| Table 1 Overview of thesis.....   | 2  |
| Table 2 Overview of the data material .....                             | 24 |
| Table 3 Overview of data collected in each of the waves .....           | 28 |
| Table 4 Mental health questions in ReQoL-20.....                        | 32 |
| Table 5 Identified and quantified cost components .....                 | 38 |
| Table 6 Hourly opportunity cost of different actors .....               | 40 |
| Table 7 Parameters in the deterministic sensitivity analysis.....       | 43 |
| Table 8 Parameters in the probabilistic sensitivity analysis.....       | 44 |
| Table 9 Proportion (%) using each service in the REFOCUS trial.....     | 45 |
| Table 10 Mean number of contacts among users in the REFOCUS trial       | 46 |
| Table 11 Parameters in the budget impact analysis .....                 | 46 |
| Table 12 Summary of focus group interviews .....                        | 48 |
| Table 13 Proctor's taxonomy of implementation outcomes .....            | 49 |
| Table 14 Summary of main findings for each study .....                  | 54 |
| Table 15 Demographics .....   | 56 |
| Table 16 Scores on the primary and secondary outcome .....              | 57 |
| Table 17 Ancillary analysis .....                                       | 59 |
| Table 18 Goal Attainment (GAS) .....                                    | 60 |
| Table 19 Estimated means QALY's gain for 24 months.....                 | 63 |
| Table 20 Base case analysis (n = 90) .....                              | 64 |
| Table 21 Unit costs for the intervention .....                          | 64 |
| Table 22 Total costs for the intervention .....                         | 65 |
| Table 23 Sensitivity analysis of other healthcare costs.....            | 68 |
| Table 24 Discursive representations found in the process evaluation ... | 72 |

# List of original papers

- I. Sjöberg, Liv & Lindström. “Recovery through engagement in meaningful everyday activities: An internal pilot study of Everyday life Rehabilitation (ELR)”. (Manuscript submitted).
- II. Sjöberg, Liv & Lindström. (2025). “Effect of Everyday Life Rehabilitation on recovering quality of life in individuals with serious mental illness in supported accommodation: a pragmatic cluster randomised controlled trial,” *BMJ Mental Health*, 28(1). <https://doi.org/10.1136/bmjment-2025-301757>.
- III. Sjöberg, Liv, Lindholm & Lindström. (2026). ”Evaluating the cost-effectiveness of Everyday Life Rehabilitation (ELR) for individuals with serious mental illness in supported accommodation,” *The European Journal of Health Economics*. <https://doi.org/10.1007/s10198-026-01905-8>
- IV. Sjöberg, Richter Sundberg & Lindström. “Housing Staff’s Discourses in response to the Implementation of Everyday Life Rehabilitation (ELR) in Supported Accommodation”. (Manuscript submitted).



# Introduction

There is a growing focus on the use of evidence-based interventions in both healthcare and the wider welfare sector <sup>1</sup>, including settings such as public health, social services, and education. These contexts require evaluations that address complex interventions and systems, capture a diverse range of outcomes, and consider issues of equity and rights. In addition, these evaluations are conducted in open systems, where “*irregularities in the flux of events*” make it difficult to isolate mechanisms and achieve experimental closure <sup>2</sup>. While this introduces substantial methodological challenges, it also raises questions about what counts as evidence, and how such evidence can, or should, be applied in practice <sup>3</sup>.

A new Social Services Act (2025:400) was recently implemented in Sweden<sup>4</sup>, further reinforcing evidence as a guiding principle for both practice and decision-making. Among other things, this reform establishes that social services should focus on enabling everyone to live a dignified life and experience wellbeing, and that all services should be informed by science. To ensure that research findings can be adapted and translated into reliable decision-making, it is worth revisiting the classical discussion on the value of experimental approaches, while examining their limitations. This is particularly pressing as many of the evaluation approaches now being applied in the welfare sector, originally were developed within the healthcare sector <sup>5</sup>. In the wider welfare sector, it is often more difficult to define relevant outcomes, and the influence of human agency and contextual factors tends to be more pronounced.

This thesis is based on an evaluation of the complex intervention Everyday Life Rehabilitation (ELR), which was developed for persons with serious mental illness (SMI) living in supported accommodation <sup>6–13</sup>. Due to the heterogeneous nature of SMI, the intervention’s content varies to match the wishes of residents and is designed to promote personal recovery through engagement in meaningful everyday activities. Supported accommodation in Sweden operates within a complex system that brings together multiple professional roles and sectors, while requiring practice to be integrated with the person’s everyday life. Despite being one of our most expensive services, one which also targets people with very severe conditions, it receives minimal attention in public discourse.

This work is conducted as part of a larger research project organised around a pragmatic cluster randomised controlled trial (RCT) evaluating

ELR. The role of this thesis is to bring together findings from four studies addressing different aspects of the evaluation process (see Table 1). The ambition is that this thesis also will demonstrate how pragmatic trials can be combined with more translational approaches, such as health economics and implementation science, to enrichen the analysis and generate more precise and nuanced causal statements.

*Table 1 Overview of thesis*

| Study   | Evaluation target    | Study objects |
|---|----------------------|---------------|
| <b>Study I:</b> Recovery through engagement in meaningful everyday activities: An internal pilot study of Everyday life Rehabilitation (ELR).   | Feasibility          | Residents     |
| <b>Study II:</b> The Effect of Everyday life Rehabilitation (ELR) on Recovering Quality of Life for those with Serious Mental Illness in Supported Housing: A Pragmatic Cluster Randomized Controlled Trial <sup>14</sup> . | Effectiveness        | Residents     |
| <b>Study III:</b> The cost-effectiveness of the recovery- and activity-oriented intervention Everyday Life Rehabilitation (ELR) for those with Serious Mental Illness living in Supported housing <sup>15</sup> .           | Cost-effectiveness   | Residents     |
| <b>Study IV:</b> Housing Staff's Discourses in response to the Implementation of Everyday Life Rehabilitation (ELR) in Supported Accommodation.   | Early Implementation | Housing staff |

The introduction is divided into two distinct sections. (1) Before turning to the specifics of the population, intervention, and setting, the theoretical perspective of evidence, its origins, and its close relationship to experimental knowledge will be explored. After that, the approaches used to contextualise findings from the trial will be introduced, namely health

economics and implementation science. (2) In the second section, the theoretical rationale of ELR, the situation in supported accommodation, the needs of persons with SMI, and the organisation of the Swedish welfare system will be further explored.

## Aim of the thesis

The overall aim of this thesis is to evaluate ELR by integrating findings across four distinct analytical approaches. By analysing effectiveness, cost-effectiveness, and early implementation of ELR, the thesis seeks to highlight both the challenges and potential value of evaluating complex interventions, while contributing to a broader discussion on the role of evaluation within welfare settings.

The specific research objectives in this thesis are:

1. To examine the hypothesis that ELR promotes personal recovery and quality of life through engagement in meaningful everyday activities, by integrating findings from the included studies.
2. To assess the potential value of the intervention in comparison with usual practice, by examining the evidence on incremental benefits, costs, and overall impact.
3. To explore housing staff's discourses in response to the six-month implementation of ELR within their practice and social context, and to map process data using Proctor's taxonomy to contextualise early implementation outcomes.

## Theoretical perspectives on evidence

Evaluation of complex interventions is a field that has evolved in recent decades, but taking methods from the lab into the open systems of society has proven to be more challenging than anticipated. The evidence-based movement (EBM) is a young paradigm, born in the 1990s with David Sackett and Gordon Guyatt who urged young doctors to adopt an "*enlightened scepticism*" toward their own practices <sup>16</sup>. This rejection of intuition, institutional truths, and mechanistic reasoning is telling for the early proponents of EBM <sup>17</sup>. Like a Kuhn-esque revolution, they challenged the status quo by questioning the role of authoritative experts.

*“The emphasis of EBM on scepticism and uncertainty – we will never be sure about the magnitude of the effects of our treatments or the power of our diagnostic tests – is central to the approach”<sup>16</sup>*

The movement developed as a response to a perceived lack of translational use of biomedical science in clinical practice<sup>18</sup>. In just 35 years, EBM has become a cornerstone of modern healthcare, and when the British Medical Journal asked its readers to rank the most important medical milestones since the 1840s, EBM placed 15th<sup>19</sup>. EBM assumes not all evidence is created equal, which has led to the RCT receiving special status as the gold standard for evaluating effect and causality. However, this hierarchical view of evidence has not gone unquestioned, and early on, EBM met a lot of resistance as it was seen as a threat to clinical freedom<sup>20</sup>. Today we see similar debates about the use of evidence-based practice, a concept which in this thesis is used interchangeably with EBM, when evaluating interventions in the wider welfare system<sup>21</sup>.

Critique against EBM tend to manifest as critique against the role and function of experimental designs<sup>22–25</sup>, especially in open systems, and maybe not without reason. In practice, randomisation is difficult to achieve, and blinding is often impractical or ethically difficult. In addition, evidence from non-randomised studies have been suggested to be able to approximate findings from randomised trials<sup>26–28</sup>. EBM is frequently being described as being too reductionist, valuing only what can be measured, leading to “*evaluation machines*”, or even an evaluation monster<sup>29</sup>. Although the number of published evaluations in Sweden have increased drastically since the 1990s, relatively few are conducted within the wider welfare sector<sup>30</sup>. In this sector, the use of evidence-based interventions remains uncommon, and the evidence for existing interventions tend to be weak<sup>31</sup>.

Social services represents a substantial part of the welfare system and target some of society’s most marginalised populations. The total regional budget in Sweden amounts to approximately 390 billion SEK, most of which is devoted to healthcare<sup>32</sup>. Municipalities, which are responsible for social services, have budgets roughly twice as large, with approximately 333 billion SEK allocated to care-related services<sup>33</sup>. Of this amount, 16% is directed toward healthcare, accounting for nearly one-third of Sweden’s total primary healthcare expenditure<sup>34</sup>. This financial distribution underscores the central role of municipalities in Sweden’s healthcare system, as they are responsible for much of the care that bridges healthcare and social services<sup>35</sup>, including services for the elderly

and individuals with functional impairments. This makes effective coordination between municipalities and regions essential.

In this thesis, an evaluation of the intervention ELR will function as a case illustrating how pragmatic experiments, in the tradition of EBM, can be combined with more translation approaches to address the challenge of evaluating complex interventions in the wider welfare sector. This is even when they are implemented and evaluated in the messy open systems of society. The included studies represent individual parts of a larger research project, which will be combined in this thesis to provide an extensive account of both the intervention and evaluation. However, before we continue, it might be good to pause and reflect: Why do we even evaluate in the first place?

### To evaluate or not to evaluate?

Our appetite for evaluations seems endless and the number of published scientific articles is at an all-time high. Over seven million papers are published each year, up from about one million a year in the 1980s <sup>36</sup>. More than two papers per minute are published in the PubMed database alone <sup>37</sup>. Everything is expected to be evaluated: effects, costs, feasibility, sustainability, equity, adverse events, culture, law, ethics, user experience, and much more. Even if a lot of research is produced, much of it goes to waste. Research waste refers to low-quality publications that offers little value to clinicians and decision-makers <sup>38</sup>. It has been estimated that as much as 85% of medical research can be considered wasted <sup>39</sup>. This is not a novel issue, and already in the 1990s, Doug Altman warned about the poor state of medical research <sup>40</sup>. There are many causes for research waste, including flawed methods, misaligned funding models, weak use of previous research, biased reporting, and human error.

*“We need less research, better research, and research done for the right reasons” – Doug Altman <sup>40</sup>*

Alongside the issue of research waste, a replication crisis has emerged, a phenomenon in which replicated studies tend to report weaker effects than the original studies <sup>41</sup>. Replication is a core principle of causal inference, and the replication crisis points to structural flaws in our current research and publication practices. As the publishing sector has expanded, so have the need for systematic reviews and meta-analyses, which have increased nearly 20-fold between 2000 and 2019 <sup>42</sup>. Given this rapid development, one might forget that meta-analysis was once dismissed as an exercise in “*mega-silliness*” <sup>43</sup>. Systematic reviews are

primarily a form of data collection, not analysis, one which takes inspiration by the bias-reducing principles of RCTs. Even though systematic reviews are powerful, they rely entirely on the quality of the included studies. Problems like p-hacking, HARKing (hypothesising after results are known), and cherry-picking occur in both primary studies and systematic reviews <sup>44, 45</sup>.

*What does “being evidence-based” really mean?*

Historically, EBM emerged from several different schools of thought <sup>46</sup>: John Wennberg’s data-driven outcomes research movement with focus on registries and probabilistic knowledge; David Sackett and Gordon Guyatt’s clinical epidemiology advocating RCTs and critical appraisal; and the HTA movement, with Gene Glass and Thomas Chalmers, promoting systematic reviews and meta-analyses. Despite these diverse origins, the core principle is quite straightforward: practice should be based on research of adequate quality, or as it is commonly framed, the “*best available evidence*”.

The term EBM is commonly used to refer to evidence-based medicine, reflecting its historical origins in healthcare. Since then, EBM has spread to several other fields, where it is often referred to as evidence-based practice, but the two terms are frequently used interchangeably <sup>47</sup>. Although subtle differences may exist, this thesis treats them as equivalent and uses EBM as a broad umbrella term for the whole evidence movement. The three philosophical pillars of EBM are typically described as the best available evidence, clinical experience and practitioner expertise, and patient values and circumstances <sup>18</sup>.

A source of confusion is the dual use of the term EBM, the first being the evaluation of interventions (the research process), and the second being critical appraisal of material (the decision-making process) <sup>48</sup>. A too large focus on the research process has created controversy and, sometimes, misconceptions on what classifies as evidence, and how it is valued by this movement. EBM has been described as a “*cookbook approach*”, which diminishes practitioners autonomy and fosters dogmatic thinking <sup>49</sup>.

Realigning the definition with the decision-making process, and the original writings, brings the approach back in line with EBM’s core principle: an “*conscientious, explicit and judicious use of current best evidence in making decisions about the care of individuals*” <sup>48, 50</sup>. In practice, the debate has never really been about whether to rely on evidence or not, but rather what qualifies as sufficient evidence <sup>51</sup>. This is

important, as the normative principles of the movement, such as being critical consumers of research, reflecting on one's current practices, and relying on the most trustworthy information, are rarely disputed.

*“Obviously, medicine should be evidence-based. The issues lie in the details: what exactly counts as evidence? Do certain kinds of evidence carry more weight than others? (And if so, why?) And how exactly should medicine be based on evidence?” - John Worrall <sup>51</sup>*

While proponents of EBM argue that experiments are more reliable than alternative designs, this does not mean other types of knowledge are irrelevant, only that experimental evidence takes precedence over observational evidence <sup>52</sup>. This hierarchy of evidence is often illustrated using “*evidence pyramids*”, with recent adaptations emphasising the complementary nature of different approaches <sup>53</sup>. Traditionally, such hierarchies’ position practitioners’ expertise as an internal source of evidence, typically regarded as lower in quality. By contrast, more contemporary frameworks treat expertise as external, which then has to be appropriately integrated with the other pillars <sup>54</sup>. The challenge, then, is deciding what to do when expertise, preferences, and evidence point to different courses of action. As decision-making inherently is an expression of power, transparency of inputs, assumptions, and priorities are key aspects of balancing these three pillars <sup>55</sup>.

### *The comparative nature of causal inference*

So, what makes experimental designs special? When discussing effects, we are always talking about a difference. To determine if something has changed, we must know what it was before, which requires data from at least two time points, ideally before and after an exposure <sup>56</sup>. This ensures the Bradford Hill’s criteria of temporality, where causes must come before effects <sup>57</sup>. However, this is not the only difference that matters, an effect should also be relative, meaning that it must be compared against something <sup>58</sup>. Ideally, for causal inference, we would like to compare what happened with what would have happened without the intervention. This is known as counterfactual reasoning, a Latin term meaning approximately “*in contrary to fact*” <sup>56</sup>. With this perspective, an effect represents the numeric estimation of the difference between what is, and what could have been <sup>59</sup>.

Since time travels are currently unavailable, it is impossible to directly observe counterfactual outcomes. Instead, experiments can be used to mimic it. By splitting a sample into equivalent groups, one with the

intervention and one without, we create a counterfactual scenario. If all but the exposure is equal, the difference in outcomes between the groups is assumed to reflect the true effect. Even though effects are relative, we like to talk about them in absolute terms, as if an intervention is effective or not <sup>60</sup>. It is important to remember that an effect is measured for a specific population, in a specific setting, on a specific outcome, during a specific time, which means evidence is not necessarily generalisable to other similar populations and contexts.

## The growth of translational approaches

One major problem with experiments is their lack of external validity, referring to limited generalisability and applicability in real-world practice <sup>61</sup>. While this limitation is important to acknowledge, it is mainly a problem when findings are intended to inform decision-making; for hypothesis testing, it poses far less of an issue. However, as you may have noted, this thesis is presented as an example of how experimental research can be augmented with more translational approaches, for which the ultimate goal is in fact practical usability.

Translational research represents a subdiscipline of applied research where the objective is to bridge the gap between research and practice. Traditionally, the term has been used to refer to “*bench to bedside*” research, where findings are “*translated*” from basic research, such as animal or lab studies, to clinical practice. However, the field is larger than that, and it can be classified into four phases (T1-T4): Translation to humans, translation to patients, translation to practice, and translation to communities <sup>62</sup>.

In this thesis, two disciplines are applied that were both developed to address the fact that decision-making requires fundamentally different types of information than what is generated by experiments: health economics and implementation science. Both fall within the later phases of translational research, with implementation science mapped to T3 <sup>63</sup> and cost-effectiveness research to T4 <sup>62</sup>. The term translational research is therefore used in this thesis as an umbrella term for these disciplines, despite the translation into practice not being the primary objective.

Findings from experimental studies often translate poorly to clinical practice <sup>64</sup>. While such studies typically offer high internal validity, they are commonly described as measuring *efficiency*, how the intervention performs under ideal conditions, rather than *effectiveness*, how it performs in real-world settings <sup>65</sup>. Decision-makers must be able to assess whether

an intervention will be successful in their specific setting, information that is often not provided by experimental studies <sup>66</sup>. Establishing internal validity through an RCT represents only one part of predicting how mechanisms will operate in different settings <sup>67</sup>.

Experiments are controlled environments, which can be thought of as small, shielded worlds, where researchers attempt to control all variables that might influence outcomes <sup>68</sup>. Within these small worlds any heterogeneous factor becomes a potential source of distortion, often leading to extensive standardisation of content and procedures. This comes at a cost, as it undermines external validity. The gap between efficiency and effectiveness is largely driven by unrepresentativeness, where the artificial procedures or eligibility criteria used in an study fail to reflect the target population or real-world setting <sup>69</sup>. Some suggest that interventions should be evaluated as they occur in practice, meaning more “*out of control*”, rather than in idealised experimental conditions <sup>70</sup>. This creates a sort of continuum between pragmatic and explanatory trials, where pragmatic approaches trades internal validity for feasibility and enhanced external validity <sup>71</sup>.

## Resource allocation, preference and utility

Since 1997, three ethical principles have guided priority setting in Swedish healthcare <sup>72</sup>: the principle of human dignity, the principle of needs and solidarity, and the principle of cost-effectiveness (see Figure 1). These principles are to be considered hierarchical, with human dignity as the overarching principle. It has been suggested that the same, or similar, principles should apply to resource allocation across the welfare sector <sup>73</sup> as well. As such, discussions around health economics and equity have become central perspectives in most Swedish evaluation projects.

- **The principle of human dignity:** Addresses factors that should not determine priorities for care, such as age, gender and income, effectively functioning as an anti-discriminatory clausal.
- **The principle of needs and solidarity:** Holds that those with the greatest need should receive more resources, emphasizing equitable care rather than the maximization of health
- **The principle of cost-effectiveness:** Establishes that when choosing between different services one should strive for a reasonable relationship between costs and effects.

*Figure 1 Ethical principles for priority setting*

### *Theoretical perspectives of health economics*

The objective of health economics is to provide an ethical framework to make statements about which states of our society are preferable to others <sup>74</sup>. This means that focus is on the desirability of outcomes, rather than economic variables. The framework is built on key tenets, such as the utility principle, individual sovereignty and consequentialism. Utility is considered a numeric representation of preference, and can be thought of as “*a quantity that an individual should maximize or that society should help him to maximize*” <sup>75</sup>.

- **The Utility Principle:** Individuals are able to rationally maximise their welfare by choosing preferred options.
- **Individual Sovereignty:** Individuals themselves are the best judges of what contributes to their utility.
- **Consequentialism:** Utility is derived from outcomes, rather than the process themselves or intentions.

One of the critiques of this classical approach, often called “*welfarism*”, is its individualistic focus on utility. In contrast, the economic perspective adopted in this thesis is called “*extra-welfarism*”, which means that the welfarism framework is extended beyond individual utility, to also include other characteristics <sup>74</sup>, such as health, function or capacity. This reflects a semi-paternalistic view of resource allocation, and is needed as maximising preferences tend to produce suboptimal health outcomes due to market failures in healthcare <sup>76</sup>.

In healthcare, the appropriate choice of maximand has naturally been health, which is operationalised as quality-adjusted life years, or QALYs. With this, health is conceptualised as an entity, of which individuals have in varying amounts, allowing for comparisons between different health states <sup>77</sup>. The idea is that everyone prefers more health to less health, making health the appropriate maximand. This makes health maximisation as a policy goal in itself, assuming that individuals derive similar utility from the same health states <sup>78</sup>.

### *Quality-adjusted-life years (QALYs)*

Health economists are not interested in the symptoms alleviated by interventions, but rather what impact this have on the individual’s life, and their utility. Quality of life is a common outcome measurement in both clinical trials and health economic evaluations. It represents an individual’s “*perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their*

*goals, expectations, standards and concerns*”<sup>79</sup>. To allow for comparisons between interventions and populations, health economists has developed QALYs, a composite outcome based on the number of life years combined with the aggregated utility during the same period.

One of the main strengths of QALYs is that it creates a common currency for comparing interventions between populations and sectors. To be able to do this type of analysis, a core assumption is that “*a QALY is a QALY is a QALY*”, meaning that it does not matter who gains the QALY or how it is acquired, they are always valued the same<sup>80</sup>. Utility weights are estimated using preference-based instruments, based on discrete-choice or time-trade of experiments, the most famous being the EuroQoL 5 Dimension (EQ-5D)<sup>81</sup>. As with all patient-reported outcomes, these types of measures are inherently subjective by nature.

One of the problem with many preference-based measures is that they have poor sensitivity for individuals with SMI<sup>82–84</sup>. For example, EQ-5D is based on five generic domains (mobility, self-care, usual activities, pain and discomfort, and anxiety and depression), some of which are not directly impacted by mental disorders. This creates a “*ceiling-effect*”, where the instrument is not capable of distinguishing between health states that are close to full health<sup>85</sup>. This increases the risk of underestimating disease burden, as well as the potential benefit of treatment for these individuals.

## Bridging the implementation gap

Implementation science explores how the uptake of evidence-based interventions into routine practice can be improved<sup>86</sup>. The mechanisms that influence implementational outcomes are not necessarily the same as the ones impacting clinical outcomes. Process evaluations are a practical application of implementation science and explores the factors and processes responsible for observed outcomes, caused by both interventional components and implementational strategies<sup>87</sup>. Early implementation science has been described as “*an expensive version of trial-and-error*”, with limited attention being paid to the theory behind the experiments<sup>88</sup>. As a result, considerable effort has been devoted to identifying and classifying different types of theoretical approaches<sup>1</sup>, such as process models, determinants frameworks, classic theories, implementation theories and evaluation frameworks<sup>89</sup>.

---

<sup>1</sup> Sometimes referred to as TMFs, standing for theories, models and frameworks

A massive number of theories exists, one of the most common being the Consolidated Framework for Implementation Research (CFIR) <sup>90</sup>. This is a determinant framework comprising of 37 constructs across five major domains: the intervention, the inner and outer setting, the individuals involved, and the process by which implementation is accomplished <sup>91</sup>. While determinant frameworks map out implementational barriers and facilitators, evaluation frameworks specify what aspects should be used to determine implementational success. One of these are Proctor's taxonomy, which consists of eight implementational outcomes: acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability <sup>92</sup>. These outcomes are considered to be dependent variables which can be used to hypothesise on the impact of implementational outcomes on other types of outcomes <sup>93</sup>.

Although all these theories seek to establish a common, standardised language, in practice they often manifest as “*structured lists of disconnected items*” <sup>94</sup>. As a result, the field would benefit from critically analysing existing theories, rather than developing new ones <sup>95</sup>. Implementation science can be seen as a natural evolution of EBM, emerging in response to the slow and inconsistent uptake of evidence-based interventions. Studies estimate that it can take up to 17 years for research evidence to be integrated into practice <sup>96</sup>. Highlighting the urgent need to prioritise the usability of research findings, and to consider the scalability of interventions, when introducing them to a system.

## The heterogeneity revolution

Karl Popper (1902 – 1994) is perhaps the most renowned philosopher of science of our time, and asserted falsification as the demarcation criteria for distinguishing science from pseudo-science. According to Popper, a hypothesis gains empirical credibility when it has been tested through risky predictions <sup>97</sup>. This paradigm was developed as a critique to both logical positivism as well as historicists like Hegel and Marx who relied on historical trends to make law-like claims <sup>98</sup>. The rapid growth of “*neo-inductive*” approaches, which rely on data-driven predictive models using big data, has also renewed this longstanding debate <sup>99</sup>.

Over the past century, experiments have become not only a cornerstone for establishing causality, but a model of administration <sup>100</sup>. This view can be exemplified by Donald Campbell (1916 - 1996), who closely followed in Popper's footsteps, and portrayed researchers as the “*methodologists for the experiment society*” <sup>101</sup>. Even if the field has made significant progress,

it is still likely far from the utopian society envisioned by Campbell. As Ray Pawson, another prominent evaluator, described it: “*The field has reached industrial proportions but remains feudal in its capacity to create change*”<sup>102</sup>. As a result, many contemporary evaluators are turning towards more theory-driven and explanatory approaches, including realist evaluations, program theories, process evaluations, logic models, and theory-of-change models<sup>103</sup>.

This growing emphasis on contextual factors and theoretical reasoning has been described as a “*heterogeneity revolution*”, suggesting that the replication crisis stems from insufficient attention to the natural variation of effects across populations and contexts<sup>104</sup>. Many trials rely on broad causal statements and fail to specify the conditions necessary to translate those claims into practice<sup>105</sup>. Validity tends to be treated as an empirical issue, which is challenging, as assessments of external validity often require assumptions on unobserved phenomena<sup>106</sup>. In contrast, theory-driven approaches aim to provide clear descriptions of the desired outcomes, the activities required to achieve them, and explanations on how these activities are expected to produce the outcomes<sup>107</sup>.

#### *What is complex in complex interventions?*

In this thesis, ELR is being described as a complex intervention. It is not uncommon for interventions to be characterised as complex simply because they involve many moving parts or components. For example, the Medical Research Councils (MRC) guidelines for evaluating complex interventions stated that “*the more difficult it is to define the active ingredient, the more likely it is that we are dealing with a complex intervention*”<sup>70</sup>. In practice, it is challenging to define active ingredients, is for example *small talk* between housing staff and residents included, or is it only formal actions performed during the visit?<sup>107</sup>.

“... receiving all the right parts in the right dose and order may have insufficient value for understanding effects or the notion of complexity invoked by the use of the phrase complex interventions”<sup>108</sup>

Hawe et al (2004) suggest an alternative system-based definition, which is borrowed from the field of complexity science, where complexity is defined as when: “*systems display behavioural phenomena that are completely inexplicable by any conventional analysis of the systems’ constituent parts*”<sup>70</sup>. Interventions are then mainly seen as “*disruptions*” that affect and integrate into these complex systems<sup>108</sup>. Even if causation in open systems are complex, it does not mean that the mechanisms lack

distinct causal powers <sup>109</sup>. However, it may require a larger emphasis on the contextual properties of the system in which the intervention is implemented.

## The case of Everyday Life Rehabilitation

Now, we have arrived at the second section, where the specific intervention evaluated in this thesis will be examined in more detail. Everyday Life Rehabilitation (ELR) is a person-centred, recovery- and activity oriented, intervention developed in different phases since 2006 by my supervisor Dr Maria Lindström <sup>6-10</sup>. It was developed as a response to the lack of meaningful activities and agency among persons with SMI living in Swedish supported accommodation, and preliminary outcomes has previously been evaluated in an uncontrolled feasibility trial <sup>6-13</sup>.

The intervention involves residents, housing staff, housing managers, and rehabilitation professionals. These rehabilitation professionals refer to both occupational therapists and physiotherapists, who are employed by the municipalities (see “Healthcare in Swedish municipalities”). Although ELR primarily targets occupational therapists, certain activities may involve shared or transferred responsibility with physiotherapists. For ease of reference, the term occupational therapists will be used throughout this thesis, even if physiotherapists may also be involved.

The intervention operates at three levels. At the praxis level, it provides individually tailored, long-term, outreach-based, and recovery-oriented rehabilitation and activity engagement for residents. At the practitioner level, it aims to strengthen staff practices through targeted education, and an emphasis on rehabilitation and resident autonomy. At the organisation level, it promotes the importance of long-term, structured rehabilitation to improve overall service quality, while fostering collaboration among residents, housing staff, housing managers, and occupational therapists. The strategies used in the intervention will be discussed in more detail later, and at this stage the focus is on the change processes they are intended to support through the preparation, change and maintenance phase. However, broadly the main changes introduced are as follows:

1. Delivery of a brief web-based educational package for all staff.
2. Deeper integration of occupational therapists into usual practices.
3. Collaborative methods for staff, managers, occupational therapists and residents.
4. Development of a person-centred recovery plan (PCRP).

5. Exploration and training in real-life activities of resident's choice.
6. Meetings with a focus on collegial learning and collaboration.

## Similar recovery-oriented interventions

There exists two similar interventions which have been evaluated in an RCT: REFOCUS and CARE<sup>110, 111</sup>. REFOCUS have also been studied in combination with the PULSAR program in Australia<sup>112</sup>. The results from these evaluations have been mixed, and a recent systematic review concluded that there currently are inclusive evidence on the effectiveness of recovery-oriented staff training interventions<sup>113</sup>. The structure of these interventions is quite similar, and all of them attempt to provide a forum to discuss participants wishes and goals, which then are formulated into some form of recovery plan. However, despite similarities, it is important to acknowledge the differences between these approaches.

One of the most apparent differences are the scope and format of the staff education. ELR relies on a brief web-based educational package (approximately 3.5 hours), while REFOCUS and CARE provides a more comprehensive educational package. In REFOCUS, staff are offered 3 x 4 hours in a classroom setting, with additional coaching and reflection sessions. The material focuses on implementing three working practices: holistic/person-centred, strength assessment and goal orientation. CARE has a similar structure, with seven training sessions (3 full days theory meeting, 4 half-days "*training on the job*"), while highlighting the relationships with the client, strength assessment, formulated wishes and goals and formulating and executing a recovery worksheet.

The second large difference is the organisational setting of these interventions. ELR is implemented in Swedish supported accommodation and involves two distinct groups of staff: housing staff and occupational therapists. CARE has also been implemented in supported accommodation, but in the Netherlands, where the teams consist of social workers and nurses. REFOCUS on the other hand was implemented by multidisciplinary community-based mental health teams in the UK. The utilisation of occupational therapists in ELR fits well in the legal frameworks of Sweden and provides additional competency when developing the PCRPs and the gradual increase in the intensity and difficulty of the recovery-oriented activity training in real-life challenges.

## Personal recovery

ELR classifies as a recovery-oriented intervention, meaning that it is founded on the belief that individuals with SMI can lead fulfilling lives

despite experiencing clinical symptoms. Modern healthcare has largely been developed within a biomedical paradigm, where the effectiveness of treatments is measured by their impact on symptom manifestations <sup>114</sup>. This paradigm frames clinical recovery as being free from disease, pain, and suffering <sup>115</sup>. In everyday usage, the term recovery denotes regaining control of something previously lost, taken, or diminished <sup>116</sup>, implying a return to a normative state of normality <sup>117</sup>.

In contrast, the concept of “*personal recovery*” emerged from the civil rights movements of the 1960s, when former psychiatric patients spoke out against violations of their civil rights <sup>118</sup>. The most well-known definition of personal recovery is from William Anthony who in 1993 defined it as: “*a deeply personal, unique process of changing one’s attitudes, values, feelings, skills, or roles, and involves the development of new meaning and purpose in one’s life*” <sup>119</sup>. This definition brought new optimism on the possibility for individuals with SMI to lead meaningful lives, despite the presence of severe symptoms.

Personal recovery was further operationalised by Leamy et al when they developed the CHIME framework based on a systematic review and narrative synthesis <sup>120</sup>. CHIME is a mnemonic representing five key elements of personal recovery: Connectedness, Hope, Identity, Meaning, and Empowerment. Personal recovery is understood as both an outcome and a process: it is an outcome in which the individual achieves a state of recovery, and a process in which the individual takes active control over their recovery. This whole movement rejects paternalism, and questions the assumption of provider expertise, instead emphasising a balance between positive risk-taking and the duty to treat <sup>121, 122</sup>.

## Recovery through meaningful everyday activities

The theoretical rationale of ELR is that persons can progress towards personal recovery by engaging in self-chosen and meaningful everyday activities performed out in the real world (“*Recovery through Activity*”). Residents in supported accommodation rarely engages in activities just for the sake of pleasure, even though this foster a positive sense of doing <sup>123</sup>. When individuals are supported in doing what matters to them, they can regain a sense of meaning, identity and agentic self, transforming them from passive recipients of care to active participants in their own lives <sup>6, 8, 11, 124</sup>.

For persons with SMI, their condition limits not only their functions, but also their motivation and ability to choose how to spend their time. It is

essential for the recovery process that residents are allowed to participate in societal life as an part of it, rather than as a deviation from it <sup>125</sup>. Research has shown low level of agreement between resident's perception of their own goals and needs and those expressed by staff <sup>126</sup>. Even when chosen activities might appear insignificant, it is important to recognise that residents must be able to participate in society like ordinary citizens, with all that this entails <sup>127</sup>. It is precisely because these activities are so common in our own lives that it is difficult to notice their significance.

What ELR try to accomplish is creating space for meaningful activities in the usual practices, which allows relationships to develop naturally between residents and practitioners <sup>128</sup>. Nevertheless, it is important to remember that staff do not hold the key to recovery, they are simply there to provide recovery triggers and believe in the resident <sup>119</sup>. By prioritising self-chosen activities, ELR attempts to support residents in rebuilding a positive sense of identity, regain personal agency, and find meaning in their daily lives.

## Person-centeredness

ELR adopts a person-centred approach to motivate residents to take agency over their own lives. To do this, it combines ideas from different theoretical traditions, together with new components, ranging from shared decision-making, goal setting, therapeutic alliance and negotiation of expectations. Person-centeredness places the individual in the centre when talking about health and meaning and activities in their daily life. It emerged as a response to what was perceived as reductionist and dehumanising models of healthcare <sup>129</sup>. Person-centeredness, patient rights, and increased participation are explicit policy objectives in Swedish healthcare <sup>130, 131</sup>.

The concept originate from Edith Balint, when she in 1969 wrote about patient-centred care and the importance of understanding patients as human beings <sup>132</sup>. Since then, the concept has evolved, and today we often talk about person-centred care. The idea is that the patient should be transformed from a passive recipient of care, to an active person who have both responsibilities and rights in all situations which concerns her <sup>133</sup>. How to apply person-centeredness in practice is not always self-explanatory, but a recent framework <sup>134</sup> outlines that person-centred care, can be interpreted as follows:

1. Cultivating communication
2. Respectful and compassionate care and rehabilitation

3. Engaging patients in managing their care and rehabilitation
4. Integration of care and rehabilitation

The first domain involves creating space for listening to the individual and sharing information so that they have enough information to make rational decisions. The second domain positions the individual as the expert on their own health, highlighting the need to be sensitive to their preferences and values. The third domain is to let individuals be involved in co-designing plans through shared decision-making and goal setting. The last domain is about how the system itself should be integrated, and that information should be shared smoothly between actors across the whole continuum of care.

### Practice leadership

ELR includes monthly reflection sessions, led by housing manager, to promote engagement and quality in efforts among housing staff. These are based on pre-written reflection questions and takes inspiration from practice leadership <sup>135</sup>. This is a style of management which has developed from research in supported accommodation for those with intellectual disabilities. It aims at focusing staff's attention on providing support that enable the people they serve to have a good quality of life <sup>135</sup>. It has been described as containing the following five domains:

1. Focus on the quality of life among residents and how staff can support this.
2. Allocating and organising staff to provide support when and how residents need and want it.
3. Coaching staff to deliver support by spending time with them, providing feedback and modelling good practices.
4. Supervising the practice of each staff individually.
5. Facilitating teamwork, share information, and ensure constancy through regular team meetings.

### The population: Serious mental illnesses

Many services provided by Swedish municipalities act as a safety net for our most marginalised populations, those who are furthest from the idealised image of a citizen <sup>136</sup>. ELR specifically targets persons with SMI living in supported accommodation, which are among the most marginalised groups in our society. Their symptoms restrict them from engaging in society and have significant negative effects on their physical, cognitive, social, and overall health <sup>137–139</sup>. SMI are among the leading

contributors to human suffering globally and remain a major cause of lost potential and productivity <sup>140–142</sup>.

SMI is an umbrella term for psychiatric disorders characterised by their persistence and severely disabling effects on multiple domains of functioning <sup>143</sup>. While there is no formal definition, schizophrenia, schizoaffective disorder, and bipolar affective disorder are commonly included diseases <sup>144</sup>. This population is far from homogeneous, showing substantial variation within and between groups, and the definition of this population varies widely <sup>145–147</sup>. The population in Swedish supported accommodation includes persons with multiple diagnosis, severe neuropsychiatric disorder or dual diagnosis involving substance abuse.

Emerging early in neurodevelopment phases of life and peaking in late adolescence <sup>148</sup>, these disorders have long been regarded as chronic and progressively deteriorating <sup>149</sup>. Nevertheless, research suggest that nearly a fourth of those with schizophrenia, which is one of the disorders with the worst illness trajectory, are able to recover over a twenty-year period <sup>150</sup>. However, individuals with SMI still have a higher risk of somatic diseases, and their average life expectancy is 15 years shorter for men and 12 years shorter for women compared with the general population <sup>151</sup>.

## The setting: Supported accommodation

Thirty years ago, those with severe manifestations of SMI would probably reside in one of the many large mental institutions. Today they live in a similar fashion, but in smaller facilities in the lesser-known margins of our communities <sup>152</sup>. In Sweden, a major psychiatric reform was implemented in 1995 that led to the closure of the mental institutions, and instead redirected patients to community-based care. The aim of the reform was that those with SMI would receive service, support and care in open and normalised forms in their own communities <sup>153</sup>. The number of available beds in psychiatry decreased drastically <sup>154</sup>, followed by an increased demand for outpatient treatment.

In Sweden, access to social services is determined through a needs assessment, favouring services in the individual's own residence (ordinary housing), which effectively transforms the home into a site of professional intervention <sup>155</sup>. Although most Swedes rates their mental health as good, the number of individuals receiving psychiatric care is increasing <sup>156</sup>. It is estimated that approximately 60 000 individuals with SMI receive interventions from social services <sup>157</sup>. For those with severe manifestations, services in ordinary housing is often not enough. For this

population, various forms of “*group homes*” are available, which in this thesis will be referred to as supported accommodation. The most common type of supported accommodation in Sweden is congregate housing with staff available on-site 24/7, providing high levels of support, and placing limited emphasis on transition to more independent living. This type of unit corresponds to type-1 units as classified by the STAX-SA taxonomy<sup>158</sup>. Hybrid models are also relatively common, usually consisting of some type of satellite housing, which is an individual housing arrangement located near housing staff.

Those living in supported accommodation can be described as a low-volume, high-needs population<sup>159</sup>, which even if being relatively few requires extensive care from both healthcare and social services<sup>159, 160</sup>. While only representing 9% of the SMI population, those living in supported accommodation in Sweden accounts for 38% of the total municipal expenditure on services for people with SMI<sup>161</sup>. Research on housing satisfaction among persons with SMI is limited, but one small study found that residents in both ordinary- and supported accommodation were generally satisfied with their arrangement<sup>162</sup>. Despite having poorer health and lower function, those living in supported accommodation have even rated their quality of life higher than those living in ordinary housing<sup>163</sup>.

Studies from Denmark and Sweden have found that residents in supported accommodation struggle to assert self-determination and agency, making it challenging for them to realise their needs and preferences<sup>164–166</sup>. Shared common areas can also create a sense of enforced togetherness<sup>167</sup>, with most aspects of residents’ daily lives occurring together with housing staff or other residents<sup>168</sup>. Individuals with SMI are in general at risk of living sedentary, passive, and socially isolated lives<sup>169, 170</sup>. Residents in supported accommodation often spend a lot of their time alone in their homes with limited opportunities for meaningful engagement<sup>123, 165, 167</sup>. When activities are available, they tend to be brief and lack clear sense of purpose<sup>168, 171</sup>.

Housing staff in supported accommodation have low levels of education and limited relevant training<sup>172, 173</sup>. The fact that housing staff often have such low education is problematic, as the population have such complex needs. For example, in Denmark persons living in supported accommodation had higher mortality rates than matched psychiatric patients<sup>174</sup>. In addition, managerial support is limited due to prioritisation of administrative responsibilities<sup>6, 175</sup>. The fragmented division between health and social services in Sweden further complicates

the delivery of integrated care <sup>176, 177</sup>. As a result, housing staff often lack the necessary conditions to effectively enable residents to engage in meaningful activities <sup>178</sup>.

The delivery of high quality services are not only limited by the contextual challenges, but it is also unclear how staff ideally should work with their residents. Rehabilitation efforts provided by occupational therapists are typically short-term and focused on prescribing technical aids, with long-term, individualised efforts rarely being offered <sup>179, 180</sup>. Healthcare services in supported accommodation are often focused on nursing and distribution of pharmaceuticals, and there is quite low ambition when it comes to more structured interventions.

### The wider context: A decentralised welfare system

Sweden has a highly decentralised welfare system, with the state, regions (n = 21) and municipalities (n = 290) being responsible for different segments of the health system (see Figure 2) <sup>181</sup>. Both regions and municipalities have considerable autonomy over service planning and delivery, as well as independent taxing rights <sup>35</sup>. Regions have the formal responsibility for the healthcare sector, while municipalities are responsible for care for specific populations. However, this fragmentation makes service coordination challenging, contributes to health inequalities, and makes collaborative efforts between professional groups difficult <sup>182, 183</sup>.

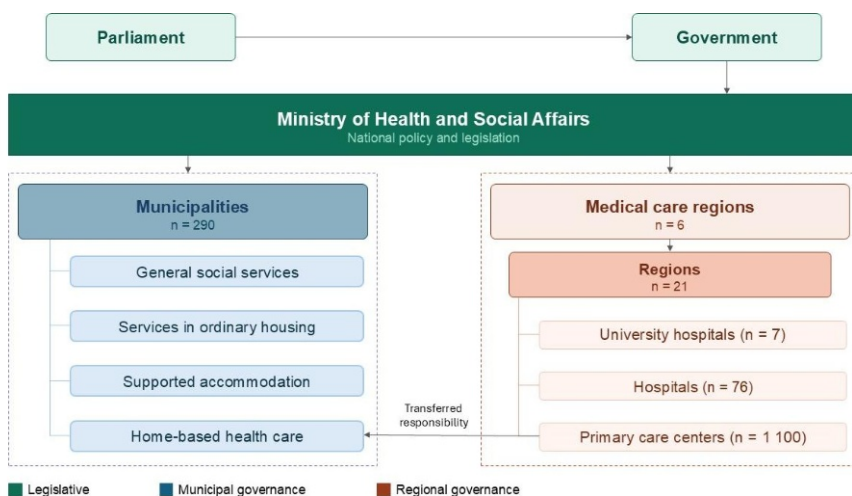


Figure 2 Illustration of the Swedish system, adopted from <sup>184</sup>

Sweden currently have one of the lowest number of hospital beds per capita in Europe <sup>185</sup>, and there have been a general shift from inpatient to outpatient care. This transition towards community-based services aligns with the national “*Good and Accessible Healthcare*” reform, which seeks to move away from a hospital-centred model by strengthening primary care as the first point of contact, while fostering integrated collaboration between regional and municipal providers <sup>34</sup>. This shift is partly being facilitated by the gradual transfer of healthcare responsibilities to municipality-driven welfare services <sup>186</sup>. Municipalities provide almost 40% of primary healthcare in Sweden <sup>34</sup>, and serve populations with increasingly complex and extensive needs. Yet the formal healthcare system continue to have limited knowledge of the municipal context in which the care is delivered <sup>187</sup>.

### *Social services for those with SMI*

There exists two major legislations dictating the rights to social services for individuals with SMI in Sweden: The Social Care Act (SoL; in Swedish “*Socialtjänstlagen*”) <sup>188</sup> and The Disability Act (LSS; in Swedish “*Lag om stöd och services till vissa funktionshindrade*”) <sup>189</sup>. As supported accommodation are provided under both legislations, it is not always clear under which legislation individuals are entitled to services. The scope of these legislations differs slightly, with LSS emphasising citizenship and rights, and SoL, obligations in the social contract between individuals and the state <sup>190</sup>. LSS targets specifically people with significant and long-term functional disabilities, and often applies to those with intellectual disabilities, autism, or permanent functional disabilities <sup>189</sup>. Still, the circumstances under which individuals are entitled to services can be important, as illustrated by how LSS ensures a “*good standard of living*”, where SoL only ensures a “*reasonable standard of living*”.

### *Healthcare in Swedish municipalities*

Swedish municipalities provides a large proportion of home-based healthcare care, and in 2021, approximately 380 000 individuals received such services <sup>191</sup>. As municipalities are not allowed to hire physicians, municipal healthcare services, are provided by nurses, physiotherapists, and occupational therapists. Several laws regulate healthcare-related services, but municipalities responsibilities for providing healthcare are primarily established under the Healthcare Act (HSL; in Swedish “*Hälso och sjukvårdslagen*”) <sup>192</sup>. Since residents receive social services under SoL/LSS alongside healthcare under HSL, staff must navigate overlapping legal frameworks, whose boundaries are often unclear, and for residents of little importance <sup>180</sup>. Even though municipalities are

required to provide basic healthcare and opportunities for community participation for residents in supported accommodation <sup>188, 189, 192</sup>, access to structured, long-term rehabilitation remains limited <sup>193</sup>.

In supported accommodation, healthcare is mostly delivered by unlicensed housing staff, on delegation by licensed healthcare professionals. Delegation refers to the transfer of responsibility for a specific task from a licensed healthcare professional to unlicensed staff, where the delegator retains accountability for the outcome <sup>194</sup>. This is common in Swedish social services, and studies show that up to 95% of staff in ordinary housing perform delegated tasks <sup>195</sup>, which comprise up to 14% of their working hours <sup>196</sup>. It is also possible to receive healthcare as self-care (in Swedish: “*Egenvård*”), which refers to healthcare tasks that licensed professionals have assessed an individual can perform independently, or with support from others. These “*others*” may include family or staff providing services outside of the scope of HSL, such as housing staff working under SoL or LSS. When assisting with self-care, staff are not classified as healthcare staff and are therefore not bound by HSL, as they would be when carrying out delegated healthcare tasks <sup>197</sup>.

# Methods

This project is based on an evaluation following the updated MRC guidelines for evaluation of complex interventions <sup>198</sup>. These guidelines emphasise the importance of continuous development of the program theory, the use of pilot and feasibility trials, as well as the inclusion of cost- and implementation aspects. In this chapter, I will outline the most important aspects of the methods, sometimes keeping the discussion at a relatively high level of abstraction. For more detailed descriptions, I urge you to read the individual studies. Study II are reported in the original publication in the *British Medical Journal Mental Health* <sup>14</sup>, and Study III in the *European Journal of Health Economics* <sup>15</sup>.

The data used in this thesis is part of a larger research project, which was conceptualised and organised by the principal investigator Dr Maria Lindström. Please note that not all studies produced through this larger project is included in this thesis. This thesis revolves around a pragmatic cluster RCT evaluating the effectiveness and cost-effectiveness of ELR compared to treatment-as-usual (TAU). The data used in the thesis have been either collected through this RCT, or from follow up focus group interviews with housing staff who participated in the trial (see Table 2).

*Table 2 Overview of the data material*

| Sub study | Main data material used   |
|-----------|---|
| Study I   | Internal pilot with a subset of participants from Study II (n = 48) |
| Study II  | Full-scale RCT, including participants in Study I (n = 161)         |
| Study III | Health economic evaluation based on Study II (n = 161)              |
| Study IV  | Interviews with housing staff participating in Study II (n = 21)    |

It is worth mentioning that the conceptualisation of the larger project, including the trial, was planned and organised by the principal investigator (ML), together with the project statistician (PL) and a senior health economist (LL). Data was mainly collected using external blinded assessors and data administration was facilitated by two university administrators (UN & KJ). I want to stress that everything in this thesis has been performed in close collaboration, support, guidance and supervision by the principal investigator (ML), as well as the co-authors of the individual studies (PL, LL & LRS).

The methodological chapter is divided into six sections, where we start by looking at the interventional components of ELR, which is followed by three sections describing the methods of the individual papers. Note that information on Study I has been combined with Study II, as it is an internal pilot and only represents a subset of the full-scale RCT. The chapter concludes with two broader sections addressing the ethical considerations and philosophical framework of this thesis.

## Interventional procedures

ELR is a recovery-oriented, activity-based, and person-centred intervention, where residents' own wishes and desires are used to create a person-centred recovery plan, or in short PCRP. The idea is that this plan will be used as a tool by the occupational therapist, housing staff, and resident, to engage in meaningful everyday activities in a real-world context. The intervention also provides education, information and manuals, to all staff who works at the unit, as well as worksheets to residents. In addition, ELR includes components and tools for practice leadership, collaboration, and collegial learning.

### *The interventions three phases*

The intervention is delivered through three phases: a preparation phase, change phase, and maintenance phase. The preparation phase starts already when residents receive information brochures of ELR, encouraging them to reflect on what activities they would like to engage with. When the intervention begins, the change phase is initiated, during which residents work collaboratively with an occupational therapist over six months to build alliance, establish motivation, identify interests and wishes, and set goals. Throughout this process, selected goals and activities are outlined in a PCRP, which is developed collaboratively during weekly sessions with the occupational therapist.

Residents explore and train these real-life activities together with the occupational therapist, who may delegate specific tasks to housing staff as needed. When residents have achieved their goals, or the intervention period is over, the maintenance phase is initiated, during which the occupational therapist and resident together evaluate the changes achieved, any further rehabilitation needs, and plan how to maintain progress. It is then assessed whether this can be managed as self-care, with or without support from housing staff, which also imply a transfer of responsibility from healthcare to social care.

### *Educational material and support to staff*

Before the start of the intervention, both occupational therapists and housing staff, are offered a brief recover-oriented web-based educational package including key components of ELR. The material consists of 10 pre-recorded videos introducing both theoretical and practical aspects of the intervention. It is estimated to be watched during work hours and take around 3.5 hours in total. Manuals and tools are also shared with housing staff, occupational therapists, and housing managers. Both housing staff and occupational therapists were expected to have partaken in the educational package before the trial started.

In addition, the intervention includes components promoting collegial learning and intersectoral collaboration. Firstly, when a resident agree to participate in the intervention, staff and occupational therapists are expected to participate in a one-hour collaborative meeting, to review expectations, attitudes and roles. Secondly, during the intervention period, housing managers are expected to lead 15-minute monthly reflection sessions to promote collaborative learning and to address quality and practical issues related to the intervention, taking inspiration from practice leadership<sup>135</sup>. Pre-written reflection questions for these sessions were distributed in advance.

### *Integrated long-term rehabilitative efforts*

The intervention revolves around weekly individual rehabilitation sessions between the resident and an occupational therapist. Each session was estimated to last 30 minutes per week, with an additional 30 minutes being allocated for administration and travels. During the initial meetings, residents are encouraged to express their interests, preferences and priorities, supported by motivation-strengthening approaches. Goals and expectations are then collaboratively negotiated using a person-centred mapping approach.

Through this shared decision-making process, a PCRP is developed, outlining goals, planned activities and strategies for the coming months of activity exploration and training in real-life context. Depending on the content of the PCRP, specific parts can be supported by either the occupational therapist or the housing staff, who integrate the strategies into their usual practices. These long-term outreach-based rehabilitative efforts are in most municipalities quite a large shift from how they are currently working. But the provision of integrated and structured long-term healthcare services is consistent with existing legal frameworks and ongoing policy discourse in Sweden.

### *Engagement in meaningful everyday activities*

During the intervention, both the occupational therapist and housing staff, is expected to provide continuous support in line with the CHIME-framework <sup>120</sup>. Goals and activities are selected by residents themselves, resulting in a wide variety of activities, varying levels of ambition, and differing amounts of time devoted to activities. In the internal pilot the most commonly chosen activities were some forms of physical activity, outdoors activity, social activity, or hobby. It was also common with more generic goals related to self-care, such as learning to cook, being outside more, or getting into better physical shape.

It is important to note that there are limited data available on the extent of these activities, especially those which was performed together with housing staff. For Study III, it was conservatively assumed that time spent by housing staff on activities increased linearly from zero to 30 minutes per week, with occupational therapists assumed to keep meeting with residents weekly (30 + 30 min). The rationale for the linear increase is that occupational therapists can delegate more responsibility to staff over time. As housing staff work on-site, they are not expected to spend any additional time on administration or travels.

### *Trial-based components*

To build implementational readiness, maintain fidelity, and address potential challenges, the trial incorporated some components that were not a formal part of ELR. Following the internal pilot, it was observed that many municipalities lacked routines for collaboration with the occupational therapist, and that many organisations was not prepared to implement the intervention in practice. For following waves, municipality representatives were therefore required to attend preparatory workshops organised by the principal investigator, each lasting approximately two hours. During the trial, all occupational therapists were offered three group-based coaching sessions with the principal investigator, each lasting approximately 1.5 hours. During the first session, they were introduced to the instruments used in the PCRPs and the data collection process, and in the following two sessions, they received collegial coaching with occupational therapists from other municipalities.

## Randomised Controlled Trial (Study I & II)

The RCT was initiated in 2021 following the publication of a pre-registered study protocol on ClinicalTrials.gov (NCT05056415) <sup>12</sup>. It is

designed as a pragmatic cluster RCT with two parallel arms and was conducted in northern and central Sweden. Data collection was completed in June 2024, and a total of 161 residents across 60 housing units from 16 municipalities participated in the trial.

Participating housing units were randomised to either the intervention arm, who received ELR, or the control arm, who received TAU. The study is designed as a superiority trial, meaning that all statistical tests evaluate the null hypothesis that the two study arms are equal. Residents in the control were also placed on a waitlist to receive the intervention after the trial was completed. When invited to participate, all residents received an offer of “*personalised rehabilitation period of six months with a possible waiting list of six month to start*”<sup>12</sup>, which provide partial blinding and increased motivation to participate.

For practical reasons, the trial used an adaptive design, by applying an internal pilot and stepwise recruitment. Eligible municipalities, units, and participants were recruited and randomised in “*waves*”. These recruitment waves were conducted annually between 2021 and 2024, with the number of participating residents per wave detailed in Table 3. Data were collected at baseline and six months by blinded external assessors. Written process diaries were also submitted by occupational therapists in the intervention arm. As these documents formed part of the intervention, they are not available for those in the control.

*Table 3 Overview of data collected in each of the waves*

| Wave          | n (ELR/TAU) | Completed (ELR/TAU) | Process diaries* |
|---------------|-------------|---------------------|------------------|
| Wave 1 (2021) | 33 (16/17)  | 23 (11/12)          | 11               |
| Wave 2 (2022) | 15 (10/5)   | 9 (5/4)             | 5                |
| Wave 3 (2023) | 81 (44/37)  | 66 (37/29)          | 42               |
| Wave 4 (2024) | 32 (20/12)  | 24 (15/9)           | 15               |
| Total         | 161 (90/71) | 122 (68/54)         | 73               |

\* *Process diaries were only collected for the ELR group.*

## Recruitment process

Before each wave, new municipalities were approached and asked to participate in the trial. Initially, all municipalities within a 270 km radius of the city of Umeå were eligible. However, to ensure adequate sample size the recruitment area was later increased to include municipalities in Northern and Central Sweden. Before randomisation occurred, participating municipality representatives sent in lists with eligible units. These lists were used to randomise each housing unit to either the intervention or control arm.

Around two months prior to the intervention, residents in the participating units received invitation brochures and oral information about the trial. After reflection, they could either post their written informed consent to the data administrators (UN & KJ) or attend to a meeting with the blinded assessors to discuss any questions and sign the informed consent form. It was the housing units themselves who were responsible for distributing invitations to eligible residents.

To be included in the study, participants had to be adults aged 18 years and older, with a diagnosis of SMI or severe neuropsychiatric disability, and currently living in supported accommodation. Exclusion criteria were individuals with dementia or severe intellectual disabilities, those unable to communicate effectively in Swedish, as well as individuals presenting acute psychosis or suicidal risk.

## Randomisation and blinding

Given that housing staff are part of delivering the intervention, randomising individual residents would have created a considerable risk of contamination. For this reason, a pragmatic cluster randomisation approach was adopted, using each housing unit as the units of randomisation. Although cluster randomisation reduces contamination, it requires larger sample sizes, is more resource-intensive, and increases the risk of selection bias<sup>199</sup>.

The randomisation sequence was computer-generated by an independent statistician (HH), separately for each wave, and was stratified by municipality, targeting a 1:1 allocating ratio within each municipality. This was partly done for practical reasons, as it would be unfeasible for municipalities if too many housing units were assigned to the intervention arm. One exception was made for municipalities with only one eligible unit, which were allocated to the intervention. As the number of

participants within each unit naturally varies, the ratio of participating residents was also not expected to be fully balanced.

Blinding was conducted in three steps. Data-collection was facilitated by 14 blinded external assessors who visited each of the residents on-site and were available when residents filled out the outcome measurements. As residents were not informed of their allocation, they are considered to be blinded at baseline. However, once informed of their start date, it was theoretically possible for residents to figure out their allocation, potentially compromising one of the blinding. Data were coded, managed and stored by two university administrators (UN & KJ), who were not involved in the intervention or analysis.

Coded data remained blinded to researchers until all analyses were completed, and were stored in a locked, fire-safe box, to which only the administrators had access. Data entry, data cleaning and initial statistical analysis were performed by me, which formally also was blinded. However, due to the nature of the data, it was difficult to ensure adequate blinding. The project statistician (PL), who performed the final analysis, remained blinded during the whole process. In addition, a third statistician, not involved in the project or a co-author, independently conducted the statistical analysis for safety and quality control purposes.

## Pragmatic trial

The objective of a pragmatic trial is to “*determine the effectiveness of an intervention in a real world setting to inform clinical decision-making*”<sup>200</sup>. Particular attention is therefore paid to ensure the population and context is as similar as possible to usual practices. This means that participating organisations often have large influence of the delivery of the intervention. Although clear instructions were provided in the trial, relatively few measures were implemented to ensure fidelity and exposure. Given the challenges of translating findings from highly controlled settings, trials must attempt to balance internal validity with external validity, to ensure results are applicable in real-world contexts.

The Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) can be used to assess trials across ten key domains to determine whether a trial is more pragmatic or explanatory<sup>201</sup>. By doing this, we found that this trial is considered predominantly pragmatic. Eligibility criteria were minimal and focused on safety, with recruitment facilitated by housing staff, as it would occur in practice. The content of the intervention was entirely determined by residents and occupational therapists, and no

exclusion criteria were applied to staff. Although an external assessor visited residents to support data collection, which deviates slightly from usual practice, outcomes were based on self-reported measures. Adherence was not monitored, and the primary analysis followed an intention-to-treat approach including all randomised participants.

## Internal pilot phase (Study I)

In the study protocol, it was originally stated that the first wave was intended to be used in an internal pilot, with the purpose of investigating any needs for adaptations, and to update sample size calculation before continuing with the full-scale RCT <sup>12</sup>. According to the MRC-guidelines, pilots are an important phase when evaluating complex interventions, as it provides an opportunity to adjust the study protocol and solve potential issues <sup>198</sup>. Study I was designed an internal pilot, meaning that the data used in it are included in the full-scale RCT. This design is commonly used when there are uncertainties regarding recruitment, randomisation and attrition <sup>202</sup>.

Since the early waves coincided with the Covid-19 pandemic and the start of the war in Ukraine, recruitment of municipalities proved more challenging than originally anticipated. In addition, several municipalities reported problems recruiting and retaining housing managers. Consequently, it was decided to extend the pilot to include the first and second wave, as well as participants in the control group of the first wave who received the delayed intervention as part of the waitlist.

Initial sample size calculations, in the statistical analysis plan, estimated a target sample size of 35 housing units per study arm. Following the sample size recalculation in the internal pilot, it was determined that the targeted statistical power was unattainable given the observed recruitment and attrition rates. To address this, a fourth wave was implemented, the geographical recruitment area was widened, and the targeted detectable effect size were adjusted from five to ten points on the primary outcome. Adding 28 units to the 17 units included in Study I were estimated to yield 80% power for detecting a difference of 8.8 points. While this increases the risk of type-2 errors, 10-points have been suggested to be a minimum important difference for the primary outcome <sup>203</sup>, ensuring clinical relevance of the findings.

## Outcome measures

As the primary outcome in the RCT, we used an alternative preference-based measure called Recovering Quality of Life (ReQoL), which was

developed specifically for individuals with SMI. The instrument was developed in the UK and comes in a 10- and 20-item version, where we used the 20-item version (ReQoL-20). ReQoL is a concise and reliable tool for assessing quality of life in persons with SMI <sup>204</sup>, and covers seven domains: meaningful activity, relationships, autonomy, hope, self-perception, well-being, and physical health. Scores range from 0 to 80, with 0 representing the lowest quality of life and 80 the highest <sup>205</sup>. The questionnaire has been translated and linguistically validated in Swedish before the start of the trial <sup>206</sup>.

The first ten questions in the 20-item version are identical to those in the 10-item version, meaning that scores from the two versions are considered comparable. ReQoL-20 consists of 20 mental health items and one for physical health. For the physical item, respondents are asked to describe their physical health (problems with pain, mobility, difficulties caring for yourself or feeling physically unwell) over the last week using scores ranging from “*No problems*” to “*Very severe problems*”. For the mental health items, respondents are asked to describe their thoughts, feeling and activities over the last week (see Table 4) using scores ranging from “*None of the time*” to “*Most or all of the time*”. Positive responses are rated as 4, while negative responses are rated as 0.

*Table 4 Mental health questions in ReQoL-20*

|  |
|--|
| 1. I found it difficult to get started with everyday tasks |
| 2. I felt able to trust others                             |
| 3. I felt unable to cope                                   |
| 4. I could do the things I wanted to do                    |
| 5. I felt happy  |
| 6. I thought my life was not worth living                  |
| 7. I enjoyed what I did                                    |
| 8. I felt hopeful about my future                          |
| 9. I felt lonely   |
| 10. I felt confident in myself                             |
| 11. I did things I found rewarding                         |
| 12. I avoided things I needed to do                        |

|                                    |
|------------------------------------|
| 13. I felt irritated               |
| 14. I felt like a failure          |
| 15. I felt in control of my life   |
| 16. I felt terrified               |
| 17. I felt anxious                 |
| 18. I had problems with my sleep   |
| 19. I felt calm                    |
| 20. I found it hard to concentrate |

*Secondary outcome*

As a secondary outcome, the Swedish translation of the Recovery Assessment Scale - Domains & Stages (RAS-DS) was used, which is a measure of self-reported recovery and daily functioning <sup>207</sup>. The instrument was developed in Australia and has been translated, into at least, 18 languages <sup>208</sup>. It is considered to be a valid and reliable self-report measure of service user-defined recovery. The instrument consists of 38 items, ranging over four domains: valuing activities, looking forward, mastering illness, and social belonging, which is rated on a Likert scale from “*untrue*” to “*completely true*” <sup>209</sup>. The scores are added up for all items, which produces a total raw score between 0 and 152 <sup>207</sup>.

*Written process diaries*

For residents in the intervention arm, occupational therapists filled out written process diaries, which were used both as a clinical tool and a process document. These include estimates on goal attainment using the Goal Attainment Scale (GAS), performance and satisfaction with everyday activities using the Canadian Occupational Performance Measure (COPM), self-rated contentment scores (1 to 10), adherence rates and free-text writing. As these process diaries are only available for the intervention arm, they are in this thesis primarily used to contextualise other findings.

GAS is a method developed by Kiresuk and Sherman in 1968 <sup>210</sup>. It is based on first establishing a criterion for what is an “*successful outcome*” for the specific participant <sup>211</sup>. The outcome is then scored on a 5-point scale depending on the degree of attainment for each goal. If they achieve the expected level, they receive a score of 0. Positive scores (+1 or +2) are

given when they achieve it more than expected, while negative scores (-1 or -2) are assigned when they achieve somewhat less than expected.

COPM is a person-centred outcome measure developed in the 1990s, and is commonly used in occupational therapy <sup>212</sup>. It begins with a semi-structured interview to ask participants to identify routines in daily life related to self-care, productivity and leisure. Participants are then asked to prioritise the most important and challenging daily tasks. These are then rated using a score between 1 and 10 based on the participants self-rated level of performance and satisfaction with each problem.

### *Proxies for exposure*

As this is a pragmatic trial, there is limited possibilities to influence adherence to the intervention in practice. This means that we mostly have to rely on crude proxy indicators for intervention exposure. From the web-based educational package it is possible to determine the total number of views on each module, but there is no way to connect these to a specific viewer or unit. In the process diaries, occupational therapists were responsible for reporting adherence to the intervention and the overall process in free-text format. In Study II, a per protocol analysis were performed, based on residents who participated in at least 70% of the weekly rehabilitation sessions with the occupational therapist. As the interventional hypothesis assumes that increased engagement in meaningful everyday activities promote personal recovery, we would expect a dose-response relationship, based on the exposure to these weekly sessions. Therefore, in this thesis, this per-protocol analysis is used as a proxy for intervention exposure. In a similar fashion, goal attainment can be interpreted as an indication of “adequate exposure”, and GAS scores are therefore also used as a proxy for adherence.

### Comparator

Units allocated to the control continued their current practices without any standardised instructions, other than distributing the initial informational brochures to residents. TAU include a wide range of residential services and psychosocial support provided by housing staff, which vary depending on individual approaches, staff commitment, and unit norms. Before starting the intervention, each municipality were asked to describe their TAU. To ensure fairness, a wait-list design was used, ensuring that the control would receive access to the intervention after the study period.

## Statistical methods

A statistical analysis plan was pre-published at ClinicalTrials.gov (NCT05056415), detailing the statistical methods used for sample size recalculations, handling of missing data, and for analysing outcomes.

### *Sample size recalculations*

The study was originally designed to detect a difference of 5 points on the ReQoL scale at 80% power. This was based on assumed standard deviation of 10, in line with previous studies<sup>204</sup> and a guesstimated intraclass correlation at housing unit level of 0.1. In Study I, a mixed effects model was fitted with post ReQoL as dependent variable using the R function *lmer* from the R package *lme4*<sup>213</sup>. Study arm and baseline scores were included as independent fixed effects, and housing unit was included as a random effect. The *simR* package was then used to investigate the power of the initially planned sample size. The interim analysis was performed by an independent and blinded statistician (SV), otherwise not involved in the study.

### *Primary and secondary outcomes*

To assess differences in ReQoL a mixed-effect model were used, with the 6-month follow-up measurement as outcome, and unit as a random effect to account for the cluster design. Study arm and baseline scores were included as fixed effects. Baseline ReQoL was adjusted for, using both the individual ReQoL scores and the average baseline score within the corresponding housing unit to avoid cross-linking bias. Group difference in RAS-DS was analysed using the corresponding mixed effect model. GAS utilises a 5-point scale ranging from -2 to +2, with negative scores indicating less change than expected. Scores were transformed into a total T-score, considering an expected intercorrelation between goals of 0.3<sup>211</sup>. Following the GAS manual, T-scores were rescaled to have a mean of 50 and a standard deviation of 10. Reported T-scores are unweighted and do not include any weighting for goal importance or difficulty.

### *Missing data*

For both Study II and III, missing values were imputed for ReQoL-20, ReQoL-UI, and RAS-DS. To handle missing data, we used a modified intention-to-treat (ITT) approach, excluding municipalities and units that withdrew from the trial prior to the recruitment of residents. First, participants with two or fewer missing values on ReQoL had values imputed using the mean of completed items, following the scoring manual<sup>205</sup>. Secondly, the remaining missing values were imputed using Multiple

imputations by chained equations (MICE) with the mice package in R <sup>214</sup>. Partial mean matching from 10 possible donors was used to generate 30 imputed data sets. For RAS-DS, we calculated the domain mean and used these as predictors to impute total RAS-DS scores.

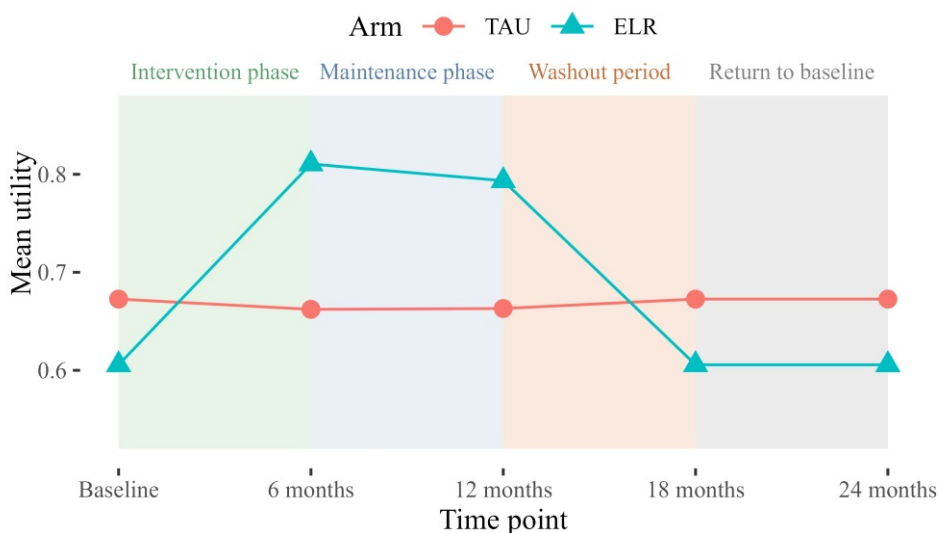
## Health economical evaluation (Study III)

Study III is a trial-based health economical cost-utility analysis based on data collected during the RCT. It adopts a pseudo-societal perspective, by considering the interventions impact on municipalities, regions and residents. However, due to data limitations, many of these analyses are conducted as sensitivity analyses, meaning that the final base case mostly reflects a municipal payer perspective. In the absence of long-term data, the analysis employs a relatively short time horizon of 24 months, with costs and benefits being extrapolated accordingly. In the sensitivity analysis, the impact of this extrapolation is explored by altering the rate at which benefits are retained.

### Model structure

The health economic model is structured around four phases, which is illustrated in Figure 3 using unadjusted utility data. During the first phase (baseline to 6 months), residents are simply assumed to incur costs and benefits as observed during the RCT. To account for attrition, it was assumed that residents who dropped out did so at the six-month mark. From that point onward, they were assigned costs and utility at the same rate as in the control.

During the second phase (6 to 12 months), benefits were extrapolated using data from a previous uncontrolled, feasibility study of ELR <sup>9</sup>. Although this study did not report QALYs, we estimated the median proportion of benefits retained 6 months post-intervention across all outcomes. This showed that participants had retained 92% of the benefits gained during the intervention at follow-up. During the third phase, residents were assumed to return to baseline over a wash-out period of six months, followed by an additional six months at the baseline phase.



*Figure 3 Model structure*

## Identification and quantification of resource use

During Study III, a total of eight intervention components were identified and used to estimate the cost of the intervention (see Table 5). Costs were categorised into four distinct types: start-up costs, intervention costs, maintenance cost, and trial-driven costs. Start-up costs refer to those incurred prior to the start of the intervention. Intervention costs represent those incurred during the six-month intervention period. Maintenance cost is those assumed to be incurred as part of the maintenance phase of ELR. Lastly, trial-driven costs are those that are not formally part of the intervention but were incurred to ensure feasibility during the trial.

The number of residents was operationalised as the 90 residents who were allocated to the intervention in the RCT. According to process data, approximately 20 occupational therapists were involved in the trial. However, there are no available data of the number of participating housing staff. Instead, the number of views ( $n = 1\ 676$ ) on the ten educational modules in the web-based educational package was used to estimate the number of housing staff. Assuming all occupational therapists viewed all modules, that only eligible actors accessed the material, and no modules were not re-watched, the number of views suggests that 147.6 housing staff participated in the intervention. The average number of actors per unit was derived by dividing the estimated total number of actors across the 33 units in the intervention arm.

*Table 5 Identified and quantified cost components*

| Identified components   | Quantification of components  |
|---|---|
| <i>Start-up costs</i>   |   |
| Web-based education for all staff and occupational therapists               | Assumed that all occupational therapists and housing staff completed the modules within the designated 3.5-hour period.   |
| Collaborative start-up meeting for all staff                                | Assumed that all housing staff per unit, their housing manager and one occupational therapist participated in a 1-hour meeting per unit.                                  |
| <i>Interventional costs</i>   |   |
| Weekly rehabilitation sessions for residents with an occupational therapist | Assumed that one occupational therapist performed a 30 min session with each resident every week, including, 30 min extra time for administration and travels every week. |
| Monthly reflection sessions with the housing manager                        | Assumed that all housing staff and their housing manager spent 15 min of a monthly meeting to discuss the pre-written reflection questions.                               |
| Activities for residents together with housing staff                        | Assumed that housing staff spent no additional time on activities at baseline, with engagement increasing linearly to 30 minutes per week per resident over six months.   |
| <i>Maintenance costs</i>  |   |
| Activities for residents together with housing staff                        | Assumed that housing staff continued engaging with each resident for 30 min per week throughout the entire maintenance phase.   |

|   |  |
|---|--|
| Education for new employees                           | Assumed that each unit hired new staff at the average annual municipal turnover rate of 13% <sup>215</sup> , all of whom were assumed to complete the 3.5-hour educational package.                |
| <i>Trial-based costs</i>                              |  |
| Supervision meetings with the occupational therapists | Assumed that each municipality were offered three coaching sessions with the principal investigator of the trial, in which all occupational therapists participated in the 1.5-hour long sessions. |
| Introductory workshop for upper management            | Assumed that five municipal representatives participated in one of the ten introductory workshops, which lasted two hours.   |
| Development of the educational material               | Assumed that the principal investigator spent 3.5 hours recording the educational material, followed by six hours of preparation and editing.  |

## Valuation of resource use

The value of each component was estimated using a micro-costing approach based on resource use during the RCT. Utilisation of other healthcare services were not captured by the trial and were therefore not included in the base-case, but its hypothetical impact is explored in the sensitivity analyses.

All costs are reported in Swedish kronor (SEK), with 1 SEK equivalent to 0.091 EUR <sup>216</sup>, and adjusted to 2024 price levels using the national consumer price index <sup>217</sup>. We applied a human capital approach using the mean market salary of the participating professions to estimate the value of allocated labour time (see Table 6). Salaries were obtained from national registries by Statistics Sweden <sup>218</sup>, and adjusted to reflect total labour costs by applying a national payroll tax rate of 31.42%.

Table 6 Hourly opportunity cost of different actors

| Professions            | Salaries               |
|------------------------|------------------------|
| Housing Staff          | 269 SEK <sup>218</sup> |
| Occupational therapist | 309 SEK <sup>218</sup> |
| Housing manager        | 410 SEK <sup>218</sup> |
| Project Investigator   | 443 SEK <sup>218</sup> |
| Upper management       | 522 SEK <sup>218</sup> |
| Residents              | 138 SEK <sup>219</sup> |

By claiming a pseudo-societal perspective, value of time must not only be considered for providers, but also for residents <sup>220</sup>. While individuals with SMI have lower productivity than the general population, their time still holds economic value. Engaging in activities also provide an “*enjoyment value*” <sup>221</sup>. However, as no robust estimate of the value of resident’s time are available, it is not suitable to be included in the base-case. Instead, a conservative willingness-to-accept estimate of 138 SEK per hour of leisure time, based on a Dutch study of the general population <sup>219</sup>, is used in the deterministic sensitivity analysis.

Allocating staff resources to a single resident inherently generates an opportunity cost for other residents. This consideration is particularly important given the lack of quality-of-life data for those living at the supported accommodation, but who is not participating in the trial. Spending time with residents are part of housing staff’s job description, and other time not spent with residents are spent on housekeeping or administrative duties. While ELR could be considered “*better*” spent time, by allocating a unit of time to one-on-one activities with a single resident, that time is still inherently unavailable to other residents.

Assuming an equal distribution of staff resources, the monetary value of one hour of one-on-one time equals staff wages (269 SEK) divided by the average number of residents per unit ( $n = 8$ ), resulting in approximately 33 SEK per resident. By summing the potential reduction in service for the remaining residents ( $n - 1$ ), the opportunity cost of reallocating one hour of staff time to a single resident is approximately 236 SEK. For the

other professions, it was assumed that their market salary represented the true opportunity cost.

## Calculation of health benefits

The ReQoL instrument allows for conversion of recovery outcomes into QALYs through their utility index (ReQoL-UI) <sup>222</sup>. Seven of the twenty items in the instrument are used to categorise each respondent into one of 78 125 unique health states. The score floor of ReQoL-UI is -0.195, with a score of 1 representing perfect health. Since there is no country-specific preference-based value set for Sweden, the UK tariff was applied.

QALYs were calculated using total area under the curve based on patient-level data <sup>223</sup>. For each time period the QALYs was calculated (Equation 1) assuming a linear change in utility over time <sup>224</sup>. In this equation,  $n$  is the number of time-points during the study period and  $Q_t$  is the individual utility obtained at the  $t^{th}$  measurement.  $T$  represents the number of time units in a year (12 months). As the interval length of this model is 6 months, the delta ( $\delta$ ) for all intervals is estimated as 0.5, representing the percentage of a year covered ( $\delta = \frac{6-0}{12}$ ). This delta value is used to re-scale the average QALY for each interval, which is then summed across all individuals.

$$QALY = \sum_{t=0}^n \frac{(Q_1 + Q_{t+1})}{2} \cdot \frac{(T_{t+1} - T_t)}{T} \quad (1)$$

To assess the effect of treatment on the total discounted QALYs, while adjusting for baseline differences, a linear regression model was used <sup>223</sup>. Initially the *lmer* function from the *lme4* package in R was used to fit a mixed-effects model, including a random intercept for unit <sup>213</sup>. But while fitting the model we encountered issues with matrix singularity due to the small variance of random effect. As a result, it was decided to switch to a linear regression model to ensure robust estimation. The final model included 6-month QALYs as dependent variable, and study arm and baseline QALYs as fixed-effect independent variables.

Estimated marginal means were used to adjust for baseline utility scores, holding them constant at their mean value, using the *emmeans* package <sup>225</sup>. For handling missing data, multiple imputation using the *mice* package in R was used <sup>214</sup>. Imputation was performed using 30 imputations with a donor pool of 10. Although it is not possible to rule out that the data are

missing not at random (MNAR), the assumption of missing at random (MAR) was considered reasonable based on the observed data structure.

## Cost-effectiveness analysis

Cost-effectiveness was assessed by estimating the incremental cost-effectiveness ratio (ICER), which is calculated by dividing the difference in costs ( $\Delta$  Costs) by the difference in benefits ( $\Delta$  QALYs) between the intervention and the control. To determine whether the intervention is cost-effective, the ICER is compared against a predefined willingness-to-pay (WTP) threshold of 350 000 SEK per QALY.

As there are no formal WTP-thresholds in Sweden, especially for valuing benefits in the municipal sector, we had to rely on indirect evidence from the healthcare sector. In the base-case, we rely on empirical evidence showing that the average revealed preference for pharmaceuticals in Sweden is approximately 350 000 SEK per QALY<sup>226</sup>. A threshold of 350 000 SEK per QALY lies within the range classified by the Swedish National Board of Health and Welfare as a “*modest cost per QALY*” (100 000–500 000 SEK)<sup>227</sup>.

In this thesis, findings are also presented as net health benefit, calculated as  $\Delta$  QALYs –  $\Delta$  Costs / WTP, and net monetary benefits, calculated as  $\Delta$  QALYs  $\times$  WTP –  $\Delta$  Costs<sup>228</sup>. Where positive values indicate that the intervention is cost-effective at the specified threshold. The WTP was also used to quantify the potential opportunity cost of the intervention. This was calculated as  $(\Delta$  Costs / WTP) /  $\Delta$  QALYs, representing the QALYs displaced elsewhere in the system for each QALY gained due to the resources required to implement the intervention<sup>229</sup>.

## Sensitivity analysis

In Study III, both a deterministic sensitivity analysis and a probabilistic sensitivity analysis were performed. A deterministic sensitivity analysis is conducted by altering a single parameter in the model one at a time, while probabilistic sensitivity analysis is done by simultaneously varying all parameters across an assigned distribution. There is almost always one or a few parameters that “*drive*” the model, meaning that small adjustment in these critical assumptions can have a major impact on the results<sup>230</sup>. In this sense, sensitivity analyses can be used like a modern application of Occam’s razor, helping to distinguish the most critical assumptions from those of lesser importance<sup>231</sup>.

### *Deterministic sensitivity analysis*

For the deterministic sensitivity analysis, fourteen parameters were systematically adjusted across their plausible ranges (see Table 7). Uncertainty is commonly presented using confidence intervals, where a narrower interval is generally preferred because it excludes a greater range of false values <sup>232</sup>. Accordingly, for QALYs, the confidence interval was used directly as the plausible range. For the economic variables, such as salaries and time-use, an arbitrary range of  $\pm 20\%$  were applied. For attrition rate, staff per unit, turnover rate, and other health care costs, a non-specific value were applied by halving or doubling the base-case estimates. Some parameters, such as activity intensity and inclusion of resident's time, are binary variables that effectively function as on/off switches in the model.

*Table 7 Parameters in the deterministic sensitivity analysis*

| Variable                                 | Min    | Base   | Max    |
|--|--------|--------|--------|
| Opportunity cost for HS                  | 0      | 236    | 236    |
| Activity intensity by HS                 | 0.000* | 1.000* | 1.000* |
| Turnover rate among staff                | 0.000  | 0.065  | 0.130  |
| Housing staff per unit                   | 2.472  | 4.473  | 6.473  |
| Time-used by intervention ( $\pm 20\%$ ) | 0.800* | 1.000* | 1.200* |
| Salaries ( $\pm 20\%$ )                  | 0.800* | 1.000* | 1.200* |
| Time-used by maintenance ( $\pm 20\%$ )  | 0.800* | 1.000* | 1.200* |
| Inclusion of resident's time             | 0.000* | 0.000* | 1.000* |
| Effect retention (12mo)                  | 0.000  | 0.917  | 1.000  |
| Intervention duration                    | 12.000 | 12.000 | 24.000 |
| Discount rate                            | 0.950  | 0.970  | 1.000  |
| QALYs gained (ELR)                       | 1.400  | 1.430  | 1.460  |
| QALYs gained (TAU)                       | 1.230  | 1.260  | 1.300  |
| Attrition rate                           | 0.142  | 0.242  | 0.342  |

*\* Represents a multiplicative factor rather than the raw value*

### *Probabilistic sensitivity analysis*

In the probabilistic sensitivity analysis, 10 000 Monte Carlo simulations was used to assess the impact of uncertainty in model parameters. Using this approach, all parameters are treated as random variables associated with a specific probability distribution <sup>233</sup>. Twelve parameters were included in the probabilistic sensitivity analysis (see Table 8). For very uncertain assumptions, such as housing staff per unit and other health care costs, a uniform distribution was applied. As these values are assigned equal probability, their inclusion is negligible, other than showing the range between the most and least optimal scenarios.

*Table 8 Parameters in the probabilistic sensitivity analysis*

| Parameter                           | Parameters                               | Distribution        |
|-------------------------------------|--|---------------------|
| QALYs gained (ELR)                  | Mean = 1.423; sd = 0.015                 | Normal              |
| QALYs gained (TAU)                  | Mean = 1.262; sd = 0.017                 | Normal              |
| Time-used by intervention<br>(±20%) | Mean = 1; sd = 0.1<br>(Bound 0.8 to 1.2) | Truncated<br>Normal |
| Time-used by maintenance<br>(±20%)  | Mean = 1; sd = 0.1<br>(Bound 0.8 to 1.2) | Truncated<br>Normal |
| Salaries (±20%)                     | Mean = 1; sd = 0.1<br>(Bound 0.8 to 1.2) | Truncated<br>Normal |
| Value of resident's time            | $\alpha = 8; \beta = 2$                  | Beta                |
| Effect retention (12mo)             | $\alpha = 9.167; \beta = 0.833$          | Beta                |
| Attrition rate                      | $\alpha = 2.422; \beta = 7.578$          | Beta                |
| Activity intensity by HS            | $\alpha = 5; \beta = 5$                  | Beta                |
| Opportunity cost for HS             | $a = 0; b = 269; c = 236$                | Triangular          |
| Other healthcare costs              | Min = -8 500; Max = 8 500                | Uniform             |

Housing staff per unit                      Min = 1; Max = 12                      Uniform

*For Beta distributions,  $\alpha$  (alpha) and  $\beta$  (beta) represent shape parameters. For Triangular distributions,  $a$  = min value,  $b$  = max value, and  $c$  = mode (most likely value).*

### *Ad-hoc sensitivity analysis of the impact of other healthcare costs*

As an ad-hoc sensitivity analysis, the potential impact on other healthcare costs were explored using previously published data from the REFOCUS trial <sup>111</sup>. Including an effect on other healthcare costs in the base-case would be a way too optimistic assumption. Given the contextual differences between the UK and Sweden, two types of services were selected: inpatient care and physician visits. This represents six of the sixteen services included in the study.

Costs were calculated by multiplying the proportion of participants using each service (Table 9) by the mean number of contacts among users (Table 10) and the unit cost per contact. Unit costs for each services were based on rates from the Northern Health Care Region (2025) and set as 2 406 SEK per visit in primary care, and 7 877 SEK per night for adult psychiatric inpatient care <sup>234</sup>. The overall impact on other healthcare costs was then estimated by calculating the difference-in-differences based on the change in costs for each group.

*Table 9 Proportion (%) using each service in the REFOCUS trial*

| Service               | REFOCUS (n=139) |           | Control (n=127) |           |
|-----------------------|-----------------|-----------|-----------------|-----------|
|                       | Baseline        | Follow-up | Baseline        | Follow-up |
| General practitioner  | 84%             | 83%       | 77%             | 82%       |
| Psychiatrist          | 66%             | 55%       | 61%             | 65%       |
| Other doctors         | 21%             | 17%       | 21%             | 14%       |
| Psychologist          | 11%             | 9%        | 17%             | 13%       |
| Psychiatric inpatient | 9%              | 4%        | 8%              | 6%        |
| Physical inpatient    | 4%              | 5%        | 5%              | 10%       |

*Table 10 Mean number of contacts among users in the REFOUCS trial*

| Service               | REFOCUS  |           | Control  |           |
|-----------------------|----------|-----------|----------|-----------|
|                       | Baseline | Follow-up | Baseline | Follow-up |
| General practitioner  | 3.5      | 3.2       | 3.7      | 3.3       |
| Psychiatrist          | 2.9      | 2.3       | 2.6      | 2.4       |
| Other doctors         | 2.3      | 2.6       | 5.6      | 2.1       |
| Psychologist          | 8.1      | 6.0       | 8.6      | 10.4      |
| Psychiatric inpatient | 30.6     | 59.7      | 44.0     | 67.3      |
| Physical inpatient    | 3.5      | 6.0       | 3.4      | 7.7       |

### Budget impact analysis

By combining process data from the RCT, with demographic data from national public registries from Statistics Sweden, we estimated the budgetary impact of a full-scale implementation of ELR, in a hypothetical municipality with 100 000 inhabitants (see Table 11). By using data from participating municipalities, the average number of residents per capita and average budget spent on supported accommodation per capita were calculated. To protect participant confidentiality, particularly given the small size of some municipalities, the estimates are not presented.

*Table 11 Parameters in the budget impact analysis*

| Parameter               | Sample (Mean) | National      | Hypothetical |
|-------------------------|---------------|---------------|--------------|
| Population (n)          | 30 131        | 10 609 460    | 100 000      |
| Total Budget (SEK)      | 33 676 000    | 6 996 610 000 | 94 747 000   |
| Per-capita rates        |               |               |              |
| Budget per capita (SEK) | 947.47        | 659.5         | 947.47       |
| Participants per capita | 0.000568      | NA            | NA           |
| Expected participants*  | 17.1          | 6026          | 56,8         |

Process data allowed us to estimate the average number of housing staff and occupational therapists per resident. As spending time with residents is already part of the usual practices, most housing staff components were assumed to not incur any additional financial costs. Costs related to housing staff was therefore treated as fixed, remaining constant regardless of the number of participating residents. In contrast, the time spent by occupational therapists was considered a variable cost that increases with each additional resident. For this analysis, it was assumed that occupational therapists provided services at the same rate as in the trial, which was approximately 4.5 residents per occupational therapist.

## Process Evaluation (Study IV)

Study IV is an embedded process evaluation on the early implementation process in Study II. Qualitative data were generated through focus group interviews with housing staff (n = 21) from seven units who had participated in the trial. A purposive sampling strategy was used, which targeted maximum variation in implementation outcomes<sup>235</sup>. This was done by asking our municipal contact persons to nominate units that exemplified either a smooth or problematic implementation process. The overall aim was to examine how organisational discourse and contextual determinants shaped the implementation of ELR. As such, the material was analysed using critical discourse analysis (CDA). The focus group interviews resulted in roughly 268 minutes of recorded material. Each session involved two to five staff and lasted an average of 46 minutes.

### Data generation: Focus group interviews

To explore staff's discourses in response to the early implementation process seven focus groups were performed with units who had delivered ELR as part of the six-month long RCT. When exploring barriers and intended consequences, frontline practitioners are considered good interview subjects, as they often are left "*picking up the pieces*" after an implementation process<sup>236</sup>. While focus groups were chosen for practical reasons, they also allow for capturing expressions of power dynamics within the group. A semi-structured interview guide was developed using the CFIR web-application<sup>237</sup>, and further refined using Brönnimann's framework for developing realist-inspired interview questions<sup>238</sup>.

Table 12 Summary of focus group interviews

| Interview | Staff (n) | Wave | Residents (n) | Format   | Length    |
|-----------|-----------|------|---------------|----------|-----------|
| 0103      | 5         | 1    | 3             | Physical | 40:41 min |
| 0104      | 4         | 1    | 1             | Physical | 48:28 min |
| 0105      | 2         | 1    | 4             | Digital  | 25:17 min |
| 0201      | 3         | 3    | 3             | Physical | 49:27 min |
| 0301      | 5         | 4    | 2             | Digital  | 41:25 min |
| 0401      | 2         | 2    | 4             | Physical | 62:43 min |
| 0402      | 2         | 2    | 4             | Physical | 56:53 min |

### Data analysis: Critical Discourse Analysis

In study IV, we applied the form of CDA developed by Norman Fairclough<sup>239</sup>, whose work explicitly aligns with critical realism<sup>240</sup>. CDA enables exploration of how organisational change is mediated through discursive strategies and how representations are negotiated and contested in practice<sup>241</sup>. Discourse analysis tends to focus on naturally occurring language use, but includes a multitude of approaches, from linguistics to critical studies. We adopt a more critical approach, where our interest are not in language itself, but in a specific social phenomenon<sup>242</sup>.

Critical theory is believed to originate from the Frankfurt school, with Horkheimer being one of the founders<sup>243</sup>. This paradigm represents a practical-ethical agenda, where research is valued based on its ability to distinguish injustice and changing society. A common misunderstanding is that CDA is applicable only for very negative or serious political issues<sup>242</sup>. All forms of texts and language are open for critical analysis. What is essential is that language is understood as a medium of power, and that the aim is to challenge taken-for-granted assumptions and raise agent's awareness of their own interests.

In Fairclough's three-dimensional analytical framework discourse are conceptualised simultaneously as a text, a discursive practice, and a social practice<sup>244</sup>. Events, experiences and reflections expressed by housing staff were treated as discursive and was analysed using this framework. Discourses are considered to be relatively stable representations of reality, implying that multiple discourses are able to exist at the same time<sup>245</sup>. What is perceived as "*common sense*" is then a form of habit, which

emerges through engagement in particular social contexts <sup>246</sup>. Even if actors are ideologically positioned, they still retain agency through personal and social cognition, including memories and opinions <sup>247</sup>.

To be able to analyse the corpus, interviews were first transcribed verbatim using a form of denaturalised transcription, which means that we emphasised content over linguistic features <sup>248</sup>. Extracts were then selected to identify discursive patterns. An inductive coding approach where used, by assigning several non-exclusive colour codes to each extract <sup>249</sup>. During the whole analysis we focused on patterns, power dynamics, conflicts, and emotional content in order to highlight key tensions. Extracts were systematically compared with the corpus, paying special attention to deviant and contradictory cases. Codes and extracts were then abstracted into discursive representations through an interpretive analysis.

## Ad-hoc analysis of implementational outcomes

It is important to understand discourses from a contextual perspective. As Study IV focused on discursive aspects of the transcribed material, we had limited opportunities to describe the process data available from the trial. Therefore, I have attempted to frame some of the findings using Proctor’s taxonomy of implementational outcomes as an ad-hoc analysis <sup>92</sup>. These findings represent a combination of the discursive representations from Study IV and quantitative process data from Study I and II. This section will outline how each of the implementational outcomes have been operationalised. Note that these are short-term implementation outcomes observed during the 6-month long trial, other studies are on-going exploring the long-term implementation and sustainability of the intervention.

*Table 13 Proctor's taxonomy of implementation outcomes*

| Outcome         | Components  |
|-----------------|---|
| Acceptability   | Satisfaction with various aspects of the innovation       |
| Adoption        | Uptake; utilisation; intention to try                     |
| Appropriateness | Perceived fit; compatibility; suitability; practicability |
| Feasibility     | Actual fit or utility; practicability                     |
| Fidelity        | Adherence; integrity; quality of program delivery         |

|                     |   |
|---------------------|---|
| Implementation Cost | Marginal cost; cost-effectiveness; cost-benefit       |
| Penetration         | Level of institutionalisation; Spread; Service access |
| Sustainability      | Maintenance; continuation; institutionalisation       |

To assess acceptability among residents, contentment scores of the intervention, staff and occupational therapists reported in Study I were utilised. These scores are aggregated from wave one, two and the delayed intervention from wave one. A narrative analysis has been published that explores the stories of residents who partake in the trial on how they make sense of their engagements in everyday life activities and their recovery processes <sup>124</sup>. In addition, we have access to qualitative data of staff's satisfaction with the intervention and its components from Study IV.

As participating units have been recruited as part of a trial, it is difficult to assess adoption. However, we are able to use process data on the recruitment process from Study I to assess a crude measure of early adoption. There are no available data on the total number of eligible housing units in each municipality, making it impossible to distinguish adoption rates for these. In addition, to assess initial implementation and intention to try, we refer to some qualitative data found in written process diaries presented in Study I, as well as interview data with housing staff found in Study IV.

Appropriateness are operationalised using qualitative data from Study IV, drawing on housing staff's perception of the intervention and their own residents. Feasibility is similarly defined, by drawing primarily on housing staff's perceived feasibility of the intervention from Study IV. However, attrition rates from Study II are also presented as a proxy for feasibility among residents. Fidelity is estimated from process data on the total number of views on each module in the web-based education. From this it is possible to estimate the number of participating housing staff. In addition, qualitative data from Study IV is used to deduce fidelity to the educational package and the educational material.

Penetration was estimated by calculating reach based on process data and written information sent in by the municipalities before the trial. All municipalities were asked to report how many units were eligible, how many residents who lived there, and how many potential participants in the trial. As there was low adherence rate for this item, we only have usable data from 36% of the municipalities.

Implementation costs are addressed in the cost-effectiveness analysis and are therefore not reported here. There is also currently no available evidence on the sustainability of the intervention, so this outcome is not reported, although different scenarios are explored in the health economic sensitivity analyses.

## Ethical Considerations

The research project has been reviewed and approved by the Swedish Ethical Review Authority (2020-06220). An adjusted ethical application (2023-04148-02) was also approved for the changes following the internal pilot. All study participation was voluntary and based on informed consent.

This project is grounded in the principle of health equity, recognising residents right to high-quality services and respecting their autonomy to decide by themselves whether to take part in research. However, others have flagged for that persons with SMI have diminished capacity to make truly informed decisions <sup>250</sup>. In the light of ELR, it is essential to actually allow persons with SMI to participate and make autonomous decisions over their everyday life. Thorough considerations were made to ensure that residents fully grasp what their participation entails. Invitation letters and written brochures were sent out with clear information and instructions; residents could then mail their informed consent to the university administrators. Last chance to provide written consent was during the baseline measurement, where blinded assessors orally presented what the project and could answer any potential questions.

The intent of the intervention is to promote agency by fostering a person-centred and recover-oriented approach. However, the community-based setting in supported accommodation increases the risk of privacy breaches <sup>251</sup>, where information about activities and progress could leak to other residents. In addition, an unequal power relationship exists between housing staff/occupational therapists and residents, which may create external pressure on residents <sup>252</sup>. Because residents are continuously encouraged to participate in activities, it may be difficult for them to express a desire to withdraw or to avoid specific activities. This is particularly problematic as this would occur under the belief that they are done for beneficial reasons.

There is a risk that residents are unable to distinguish between usual care and research, a phenomenon known as therapeutic misconception, which

is common among persons with SMI <sup>253</sup>. The idea is that ELR should be integrated into usual practices even after the trial through the maintenance period. During which, occupational therapists outline how progress can be maintained. This process could potentially highlight that residents actually have a legal right to rehabilitative efforts and provide a foundation on how activities can be integrated into the usual practices. Lastly, it is important to acknowledge that staff are both providers and study participants in this context, and verbal consent was obtained from all housing staff before the interviews.

## Let's get philosophical

Before diving into the findings, I would like to take the opportunity to anchor and reframe the previous sections of EBM in the light of scientific philosophy. The direction and growth of this movement is in many ways rooted in epistemological and ontological assumptions of the world. These complicated terms refer to what we know about the world and what is real in the world <sup>254</sup>. Critique against EBM are often framed as an attack towards its perceived alignment with positivism. However, I see modern EBM as fundamentally post-positivist, grounded in fallibilism, hypothesis testing, and error control, which, at least in principle, is directly opposed to positivism. The idea of “*best-available evidence*” aligns much more closely with a realist ontology, rather than the more anti-realist position of traditional positivism.

*“For I saw now how closely realism was connected with facts, procedures, principles I valued and that it had helped to bring them about while positivism merely described the results in a rather complicated way after they had been found: realism had fruits, positivism had none” - Paul Feyerabend <sup>255</sup>“*

My position would likely be described as a form of experimentalism, but I refer to myself as a critical realist. This means that this thesis is written from a realist ontology and subjectivist epistemology, which together with judgmental relativism forms the holy trinity of critical realism <sup>256</sup>. This contemporary position is often associated with, but not exclusive, to the work of Roy Bhaskar (1944 – 2014). The EBM paradigm adheres well to a realist ontology by acknowledging a mind-independent reality and viewing theories are approximately true descriptions of reality.

Compared to empiricist approaches, critical realism adopts a subjective epistemology, which means that all knowledge is assumed to be fallible

and theory dependent. According to Bhaskar it is essential that we avoid the epistemic fallacy, which is when a phenomenon is reduced to what we are able to know about it <sup>257</sup>. Critical realism argues for an ontological stratification, which is sometimes called “*depth realism*” <sup>258</sup>. Reality is then constructed as three layers, and to put it simply, it is a stratification of experiences, events and causal mechanisms <sup>259</sup>. Bhaskar originally referred to his approach as “*transcendental realism*”, which was developed through abductive reasoning about what reality must be like for experimental knowledge to be possible <sup>260</sup>.

By rejecting a sharp divide between natural and social sciences, critical realism adheres to a form of naturalism. While experimental knowledge is considered valid in both fields, a principle of “*methodological unity-in-difference*” is adopted, arguing that methods must be tailored to the studied phenomenon <sup>261</sup>. Only observing event regularities does not support causal claims. Even when observations appear lawlike, which they rarely do, they do not explain a phenomenon <sup>262</sup>. Critical realists argue that events are generated through the interaction of multiple causal powers and generative mechanisms, and that it is these interactions that we should attempt to explain. This means that causal inference is based on narrative descriptions, rather than statistical associations alone <sup>263</sup>.

Experiments can be considered to have a life of their own, one that is independent of theories and theorising <sup>264</sup>. Deborah Mayo is part of a school of thinking which goes by the name “*new experimentalism*” <sup>265</sup>. It is characterised by severe frequentist testing and heavy use of error-control methods. Echoing the words of Donald Campbell, she insists on the use of piecemeal experiments and manipulation to study demi-regularities. The goal of an experiment is to create an experimental narrative based on an argument from error <sup>266</sup>. Error is considered absent after it have failed “*severe testing*”, meaning those that have a high probability of failing the hypothesis if it is false.

To summarise, I believe that the aim of an evaluation is not to provide evidence for large-scale theories, rather its purpose is to test hypothesis in such a way that we learn something about the generative mechanisms <sup>267</sup>. It is tempting to take an agnostic stance between realism and anti-realism, but it has an impact on how we approach evaluations. Effect estimates should be viewed as an input to causal inference, rather than its final step, serving to prompt explanations of generative mechanisms behind a phenomenon. In this thesis, I adopt a realist ontology, centred around narrative explanation, while still recognising that piecemeal experiments are essential for evaluating competing hypotheses.

# Findings

In this chapter, data and results from the four included studies will be presented, for a brief summary of the main findings, see Table 14. Even if the findings mostly follow the timeline of the published articles, they are not necessarily structured around the specific publications. In the first section of this chapter baseline demographics of residents is presented. This is followed by a section on the effectiveness of the intervention, including how outcomes are impacted when transformed to utility. The third section includes the health economic evaluation, sensitivity analyses, and the budget impact analysis. Finally, the chapter concludes with two sections on the process evaluation, including implementational outcomes and discourses among housing staff.

*Table 14 Summary of main findings for each study*

| Study | Main findings   |
|-------|---|
| I     | Several adjustments to the protocol, including an increased recruitment area, re-dimensioning the study to detect a difference of 10 points on ReQoL and a fourth wave.   |
| II    | Those who received ELR had significantly greater improvements on both ReQoL and RAS-DS at 6 months compared to the control.   |
| III   | The intervention was found to be cost-effective compared to usual care, with an estimated incremental cost-effectiveness ratio of 133 494 SEK per QALY.   |
| IV    | Discourses among staff reveal structural limitations, such as vague leadership and scarce resources, which constrained staff's agency, making them rely on discursive strategies to rationalise resistance and gatekeeping. This often manifested as a fear of residents own choices and led to a downgrading of their capacity for change. |

## Demographics

After the fourth wave a total of 73 housing units from 16 municipalities had been recruited and randomised (see Figure 4). However, before participant recruitment, 13 housing units choose to drop out from the

trial. As result the final sample consisted of 60 housing units, whereof 33 units were allocated to ELR and 27 were allocated to TAU. From these units a total of 161 participants were recruited. At six months, the overall attrition rate was 24% in both groups. At the cluster level, the intervention group lost three units (9%), while the control group lost two units (7%).

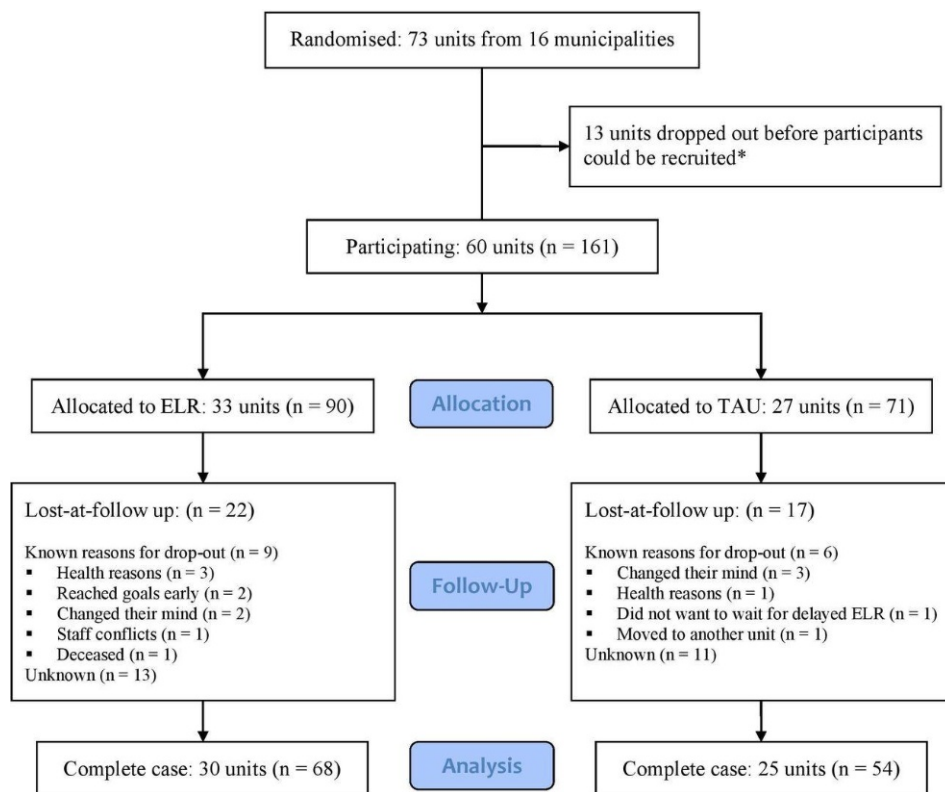


Figure 4 Flowchart

Table 15 presents demographic characteristics of residents who participated in the trial. The median age was 39 years, ranging from 21 to 77 years, and the sample was nearly evenly divided by gender. On average, residents had lived in supported accommodation for seven years. Most were born in Sweden and had attained upper secondary education. The most reported diagnosis was autism, followed by psychosis. In Study II, it was concluded that there were no major relevant differences between the two groups, suggesting that randomisation worked as intended.

The demographics indicate a heterogeneous population that is diverse in terms of age, diagnosis, experiences and needs. Given the high eligibility

requirements for supported accommodation in Sweden, a higher proportion of individuals with psychosis were expected. Instead, autism was the most common diagnosis, which is often not included in the SMI concept at all. One potential explanation is an underreporting of more severe conditions, although it may also reflect a general shift in the diagnostic composition of this population.

*Table 15 Demographics*

| Characteristics                          | TAU, N = 71* | ELR, N = 90* |
|--|--------------|--------------|
| Age                                      | 41 (32, 52)  | 37 (32, 52)  |
| Men                                      | 38 (56%)     | 48 (53%)     |
| Singel?                                  | 49 (72%)     | 80 (90%)     |
| Have children?                           | 9 (13%)      | 18 (20%)     |
| Years living in supported housing        | 8 (4, 16)    | 7 (4, 13)    |
| Born in Sweden                           | 63 (93%)     | 87 (97%)     |
| <i>Education level</i>                   |              |              |
| Low                                      | 10 (15%)     | 17 (19%)     |
| Middle                                   | 55 (81%)     | 65 (72%)     |
| High                                     | 3 (4.4%)     | 8 (8.9%)     |
| No organised activities in last 2 weeks  | 9 (13%)      | 29 (33%)     |
| No meetings with friends in last 2 weeks | 38 (56%)     | 43 (48%)     |
| <i>Mental Health Conditions</i>          |              |              |
| Addiction                                | 9 (13%)      | 14 (16%)     |
| ADHD/ADD                                 | 14 (21%)     | 21 (24%)     |
| Psychosis                                | 25 (37%)     | 28 (31%)     |
| Autism/Asperger's                        | 26 (38%)     | 37 (42%)     |
| Bipolar Disorder                         | 6 (8.8%)     | 4 (4.5%)     |
| Personality Disorder                     | 6 (8.8%)     | 5 (5.6%)     |
| Other                                    | 22 (32%)     | 32 (36%)     |

\* Median (Interquartile range); n (%)

## Did ELR promote personal recovery?

Yes, it did. The statistical analysis of the primary outcome shows that those who received the intervention reported statistically significantly larger change in scores compared to the control <sup>14</sup>. At baseline, those in the control had a mean ReQoL score of 26, while those receiving ELR had only a slightly higher mean score of 28. By visually inspecting Figure 5, it is clear that the overall distribution of scores was quite similar between the two groups. After six months, those in the control, on average, showed little to no change between the two time points, while those who received ELR reported substantially higher scores after 6 months (see Table 16).

*Table 16 Scores on the primary and secondary outcome*

| Characteristic         | TAU, n = 71* | ELR, n = 90* |
|------------------------|--------------|--------------|
| Baseline ReQoL scores  | 26 (18, 46)  | 28 (21, 45)  |
| Endpoint ReQoL scores  | 26 (18, 49)  | 49 (43, 60)  |
| Baseline RAS-DS scores | 71 (62, 108) | 78 (63, 101) |
| Endpoint RAS-DS scores | 72 (62, 114) | 99 (87, 111) |

\* Median (Interquartile range)

### *Primary outcome measurement*

In the mixed-effects model, units receiving ELR showed an estimated mean increase of 20.1 points on the ReQoL instrument (95% CI: 15.8 to 24.4) relative to TAU. While there are few studies who have used ReQoL, a change of 10 has been suggested to represent a minimum important difference, referring to the smallest change that is considered clinically important. As the confidence interval's lower bound remains well above the threshold for clinical significance, this provides robust evidence on the effectiveness of the intervention.

### *Secondary outcome measurement*

For the secondary outcome RAS-DS, those who received ELR reported slightly higher mean scores at baseline (see Table 16). From the visual inspection of Figure 6, we also observed larger deviations in the distribution of scores between the two groups. Nonetheless, consistent with the primary outcome, units receiving ELR demonstrated significantly greater improvements in scores at six months compared to the control, whose mean score remained essentially unchanged. In the

mixed-effect model, units receiving ELR showed an estimated mean increase of 19.0 points on RAS-DS (95% CI: 13.8 to 24.1) relative to TAU.

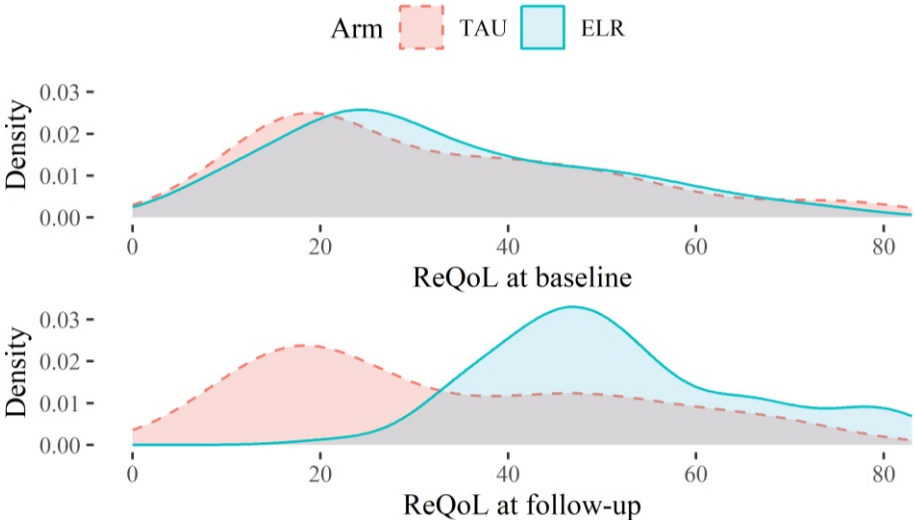


Figure 5 Distribution of ReQoL scores

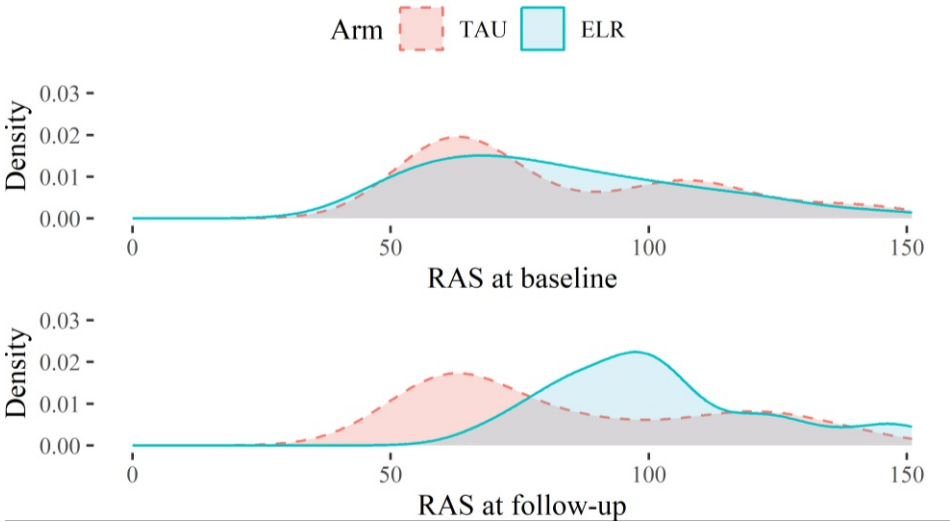


Figure 6 Distribution of RAS-DS scores

*Ancillary analysis*

Ancillary analyses were also conducted, based on gender and adherence to ELR, but none of these analyses suggested any meaningful differences (see Table 17). Our conclusion from Study II was that ELR demonstrated

robust effectiveness when it comes to promoting recovery-oriented quality of life among residents with SMI living in supported accommodation. These findings provide evidence supporting the hypothesis that focusing on meaningful everyday activities can lead to a clinically significant improvement in personal recovery and quality of life.

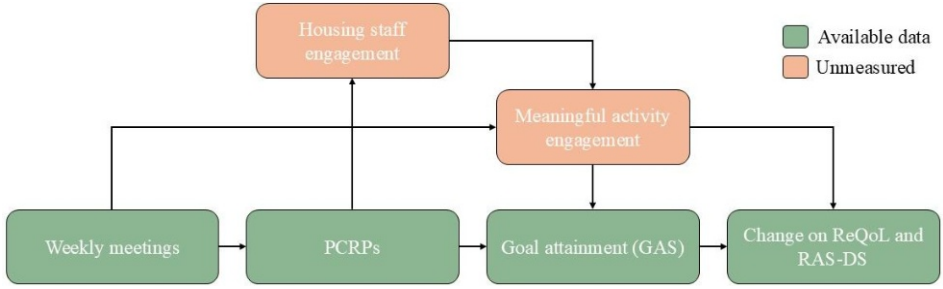
*Table 17 Ancillary analysis*

| Analysis                  | ReQoL (95% CI)      | RAS-DS (95% CI)     |
|---------------------------|---------------------|---------------------|
| ITT analysis              | 20.1 (15.8 to 24.4) | 19.0 (13.8 to 24.1) |
| Complete case-analysis    | 20.3 (15.9 to 24.7) | 22.3 (17.1 to 27.6) |
| <i>Sub-group analysis</i> |                     |                     |
| Men                       | 19.8 (14.7 to 24.8) | 18.0 (11.1 to 24.9) |
| Women                     | 20.9 (15.3 to 26.5) | 21.3 (14.8 to 27.8) |
| Per-protocol analysis*    | 20.5 (16.1 to 24.8) | 22.3 (17.1 to 27.6) |

\* Excluding those with less than 70% adherence to the weekly sessions

### Estimation of the level of exposure

All residents completing the trial submitted a process diary, indicating at least some contact with their occupational therapist. However, the completeness of these documents varies, limiting the available data on potential generative mechanisms and levels of exposure. Below, is a report of the process data currently available which could be considered relevant for assessment of exposure (see Figure 7).



*Figure 7 Illustration of available process data*

*Influence of exposure on effectiveness*

In Study II, a per-protocol analysis excluding those with less than 70% adherence to the weekly sessions (n = 5) were performed. Surprisingly, the effect estimate remained almost completely unchanged (MD: 20.5, 95% CI: 16.1 to 24.8). Despite small sample size and low data validity, this means we were unable to demonstrate a dose-response relationship. Furthermore, given the weak reporting and delegation of activities to housing staff, adherence to the weekly sessions should be considered a weak proxy for exposure.

*Goal attainment among residents*

In Study II, 89% of the residents reported that they had met or exceeded their expected goals (see Table 18), which resulted in a mean T-score on GAS of 53, with above 50 representing adequate goal attainment. Only eight individuals reported that they had progressed more poorly than expected. These findings would suggest that residents had an “adequate” exposure to activity engagement and the intervention.

*Table 18 Goal Attainment (GAS)*

| Scores on GAS              | Number of residents (%) |
|----------------------------|-------------------------|
| Worse than expected (<50)  | 8 (10.5%)               |
| As expected, (=50)         | 42 (55.3%)              |
| Better than expected (>50) | 26 (34.2%)              |

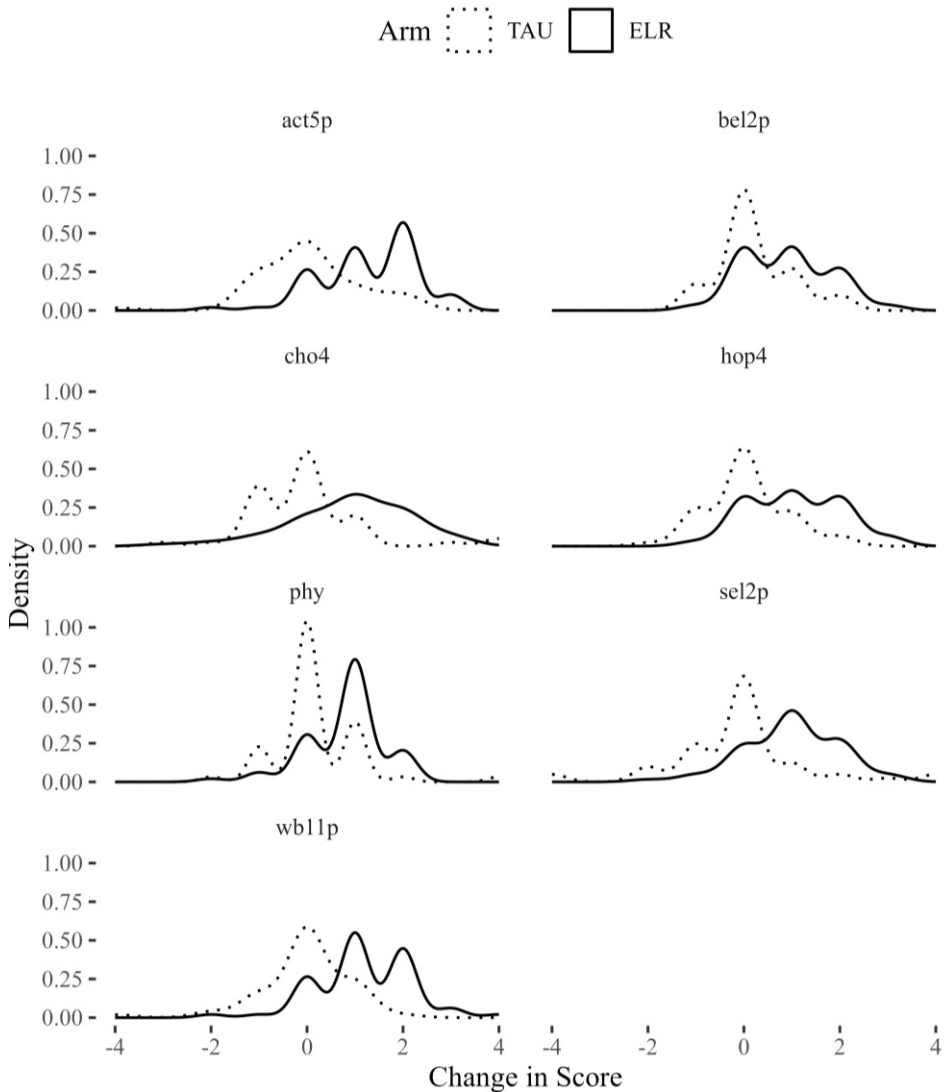
*Involvement of housing staff*

Overall, there is a lack of formal data on housing staff involvement in the intervention. In Study III, the total number of views on the web-based educational package were used to estimate that 147.6 housing staff participated in the intervention, resulting in around 4.5 housing staff per unit (see the methodological section “Identification and quantification of resource use” for more details). During Study IV, several housing staff expressed that their resident had not engaged with that many activities, or at least meaningful ones. Many of them also expressed that they only felt marginally involved in the overall process.

**Impact of transforming ReQoL to QALYs**

For the health economic evaluation in Study III, the results from the RCT were transformed to utility scores. This was done for all residents who had





*Figure 9 Change of score in each item between baseline and 6 months*

*Analysis of the interaction effect in ReQoL-UI*

As an ad hoc analysis, we examined how changes in individual item scores affected the final utility value (see Figure 10). Each item was varied across its full response range (0–4) while all other items were held constant. These results are aggregated across both study arms and time points. The most impactful item in the ReQoL-UI is, without a doubt, the physical item. Adjusting other items resulted in only a small change in utility, ranging from 0.052 to 0.088. In contrast, changing the score of the

physical item caused the average utility across the entire sample to vary dramatically, from 0.37 to 0.84.

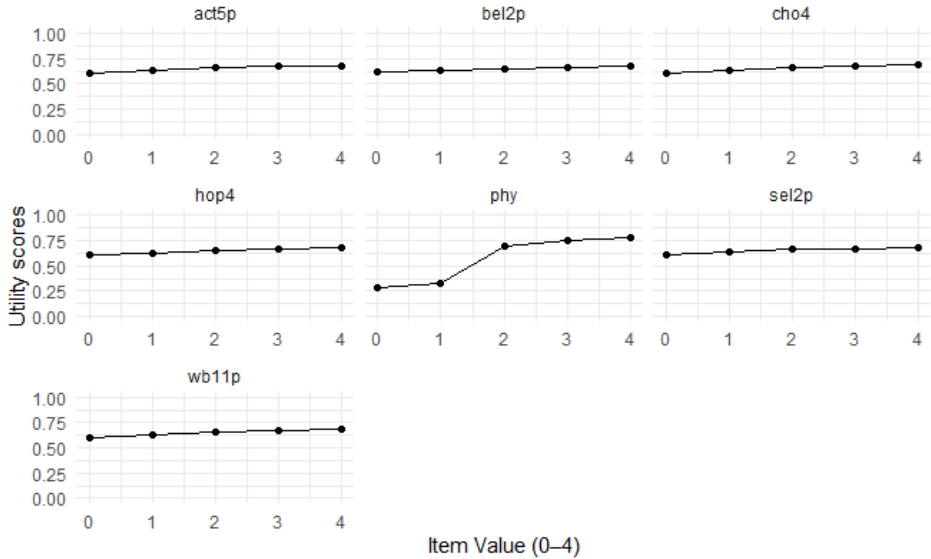


Figure 10 Sensitivity analysis of items in the ReQoL-UI

*Estimated effect on quality-adjusted life years*

After imputation of missing values and adjusting for baseline differences the estimated marginal mean QALYs was estimated to 1.43 for ELR (95% CI: 1.40-1.46) and 1.26 for TAU (95% CI: 1.23- 1.30). The estimates are very similar in the complete case dataset, where we did not impute missing values. Differences between the group were larger after adjusting for baseline differences.

Table 19 Estimated means QALY’s gain for 24 months

| Arm (n)                      | No adjustment for baseline | After adjustment           |
|------------------------------|----------------------------|----------------------------|
| <i>Imputed dataset</i>       |                            |                            |
| ELR (n = 90)                 | 1.39 (95% CI: 1.31 - 1.46) | 1.43 (95% CI: 1.4 - 1.46)  |
| TAU (n = 71)                 | 1.32 (95% CI: 1.23 - 1.40) | 1.26 (95% CI: 1.23 - 1.30) |
| <i>Complete case dataset</i> |                            |                            |

|              |                            |                            |
|--------------|----------------------------|----------------------------|
| ELR (n = 68) | 1.37 (95% CI: 1.29 - 1.46) | 1.43 (95% CI: 1.4 - 1.46)  |
| TAU (n = 54) | 1.32 (95% CI: 1.23 - 1.42) | 1.25 (95% CI: 1.21 - 1.29) |

## Is the intervention cost-effective?

Yes, it is. The base-case analysis in Study III estimated an incremental cost-effectiveness ratio (ICER) of 133 494 SEK per QALY (see Table 20). Results did not change in the complete-case analysis, which yielded a slightly lower change in QALYs for the control, resulting in an estimated ICER of 123 391 SEK. Both of these is below the established WTP threshold of 350 000 SEK, indicating that the intervention is cost-effective. The analysis yields a net health benefit of 7.1 QALYs and a net monetary benefit of 2 477 148 SEK, suggesting that gained benefits exceed the associated opportunity costs. Assuming that the WTP threshold reflects the marginal productivity of the system, it is estimated that 0.38 QALYs are foregone elsewhere for every QALY gained by the intervention.

*Table 20 Base case analysis (n = 90)*

| Δ Costs       | Δ Benefits  | ICER                 |
|---------------|-------------|----------------------|
| 1 526 852 SEK | 11.44 QALYs | 133 494 SEK per QALY |

## Estimated costs of the intervention

The first step of health economic evaluations is to quantify and value resource use. As we in this evaluation mainly talk about staff resources, the concept of opportunity cost and value of alternative uses become central. Table 21 presents the estimated unit costs of each interventional component included in the health economic model. Unit costs were calculated by multiplying salary data by the estimated time use of each profession, while the events and time columns reflects the extent of each component during the trial.

*Table 21 Unit costs for the intervention*

| Cost component        | Unit cost | Events | Time |
|-----------------------|-----------|--------|------|
| Introductory workshop | 6 106 SEK | 10     | 20   |

|  |             |      |      |
|--|-------------|------|------|
| Educational package                    | 4 873 SEK   | 37   | 131  |
| Development of material                | 4 212 SEK   | 1    | 10   |
| Collaborative start-up meeting         | 1 924 SEK   | 33   | 33   |
| Supervision meetings with OT           | 1 327 SEK   | 42   | 63   |
| Monthly reflection sessions            | 404 SEK     | 198  | 50   |
| Weekly rehabilitation sessions with OT | 309 SEK (h) | 2340 | 2340 |
| Activities together with HS            | 236 SEK (h) | 4113 | 1472 |

*OT: Occupational therapist; HS: Housing staff*

After summing up the unit costs in Table 21, the estimated cost of delivering ELR, and maintaining it for up to 24 months, was estimated to around 1.5 million SEK (see Table 22). This represents the cost of developing and delivering the intervention as part of the experimental trial. The most expensive component is, unsurprisingly, the weekly rehabilitation sessions with the occupational therapist (723 323 SEK), followed by the educational package (192 179 SEK). Most costs are incurred during the intervention phase (0 to 6 months), followed by substantially lower cost during maintenance (6 to 24 months).

*Table 22 Total costs for the intervention*

| Cost component                 | Intervention | Maintenance | Total   |
|--------------------------------|--------------|-------------|---------|
| Weekly rehabilitation sessions | 723 323      | 0           | 723 323 |
| Activities together with HS    | 137 908      | 209 004     | 346 912 |
| Educational package            | 181 726      | 10 453      | 192 179 |
| Monthly reflection sessions    | 79 925       | 0           | 79 925  |
| Collaborative meeting          | 63 484       | 0           | 63 484  |
| Introductory workshop          | 61 062       | 0           | 61 062  |

|                              |           |         |           |
|------------------------------|-----------|---------|-----------|
| Supervision meetings with OT | 55 754    | 0       | 55 754    |
| Development of material      | 4 212     | 0       | 4 212     |
| Total                        | 1 307 395 | 219 457 | 1 526 852 |

*OT: Occupational therapist; HS: Housing staff*

For occupational therapists, it is quite clear how much time they are expected to spend on each resident during the intervention. They are supposed to meet with the residents at least once a week, in addition to the collaborative start-up meetings and the educational package. The same cannot be said for the housing staff, who are involved in activities to a varying degree depending on what is decided during the weekly sessions. From interviews in Study IV, it seems like the amount of time staff has spent on activities varies a lot, whereas some perceived that they organised everything, some claims do have done very little.

Study III relies on a conservative assumption that occupational therapist spends one hour per week with each resident throughout the intervention period (0 to 6 months), and that the time staff spends, increases linearly from 0 to 30 minutes per week. During the maintenance phase (6 to 12 months) the occupational therapist is assumed to spend no time with residents, while staff continue engaging with residents for 30 minutes per week. In Study IV, housing staff expressed that the intervention did not have any major impact on their workload, which strengthens the reliability of this assumption.

### How sensitive are the conclusions?

The deterministic sensitivity analysis (Figure 11) identified the extrapolated long-term effectiveness of the intervention as the most critical assumption in Study III. While the inclusion of other healthcare costs had an even large impact, this is a theoretical assumption that is mostly beneficial for the intervention. However, both of these assumptions will be discussed in their own section. In summary, none of the individual parameters in the deterministic sensitivity analysis showed enough uncertainty (ICERs from 66 606 SEK to 254 033 SEK) to change the conclusion of the base-case analysis.

Varying the effectiveness within the lower and upper bounds of the confidence interval only moderately affected the results (110 992 SEK to 167 440 SEK). Similarly, altering time-use up and down with a generic

20% had only a moderate impact on the ICER (106 795 SEK to 160 193 SEK), which was similar for salary estimates (112 861 SEK to 154 127 SEK). Even when accounting for the opportunity cost of resident’s time, the intervention remained cost-effective (175 946 SEK). In a similar fashion, assuming a lower opportunity cost for housing staff would improve the ICER (103 163 SEK).

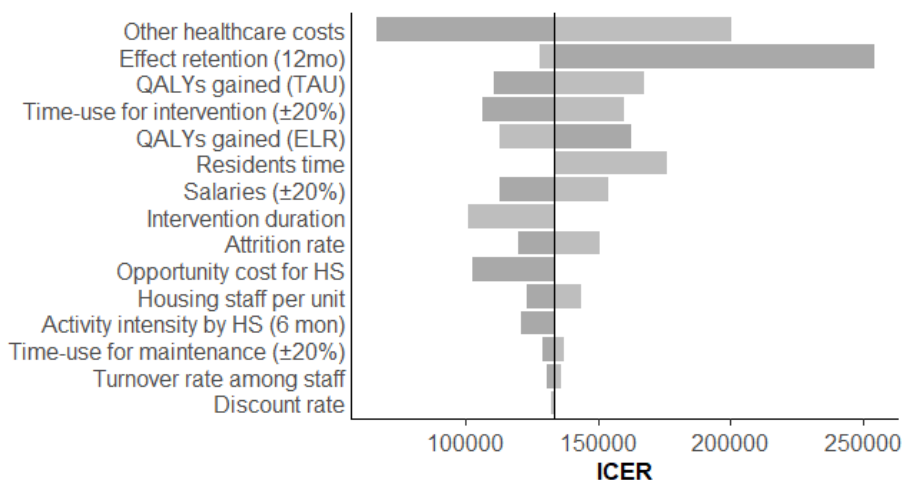


Figure 11 Deterministic sensitivity analysis

### Impact on healthcare consumption

As seen in the deterministic sensitivity analysis, the most sensitive assumption is related to the parameter “Other healthcare costs”. This parameter represents an ad-hoc analysis done on request from reviewers in Study III, who asked us to model the potential impact of the intervention on other healthcare costs. This is relevant due to the societal perspective, but also challenging, as no data on service utilisation were collected during the trial. To accommodate the request, data from the REFOCUS trial were incorporated into the analysis, to extrapolate change in service-use for some specific services (see Table 23).

Assuming that ELR is associated with a similar trend in service utilisation as observed in the REFOCUS trial, the intervention is estimated to reduce other healthcare costs to a value of 8500 SEK per person. When included in the model it results in a substantial lower ICER (66 605 SEK). For consistency, the potential impact on an increase in healthcare consumption were also modelled, which resulted in an ICER of 200 382 SEK. This assumption is subject to considerable uncertainty, and research

is needed to establish causality between ELR and healthcare consumption. These findings are mainly used to illustrate the potential resource saving capabilities of these types of interventions.

*Table 23 Sensitivity analysis of other healthcare costs*

| Service               | Intervention | Control      | Difference        |
|-----------------------|--------------|--------------|-------------------|
| General practitioner  | -658 SEK     | -368 SEK     | -290 SEK          |
| Psychiatrist          | -1592 SEK    | -64 SEK      | -1 528 SEK        |
| Other doctors         | -119 SEK     | -2148        | 2 029 SEK         |
| Psychologist          | -857 SEK     | -72          | -785 SEK          |
| Psychiatric inpatient | -2244 SEK    | 1929 SEK     | -4 173 SEK        |
| Physical inpatient    | 1190 SEK     | 4943 SEK     | -3 753 SEK        |
|                       |              | <b>Total</b> | <b>-8 500 SEK</b> |

*Sustainability of the interventions effect*

The most critical assumption in the model are the longevity of the interventions effect. While the observed mean change in scores reflects a form of demi-regularity and provides robust evidence on the benefits during that period, there are currently limited data on what happens when the trial has ended. The trial only lasted for six months, and we do not know how much of the gained benefits are retained over time. The model relies on data from a previous feasibility study, which found that participants who received ELR, on average retained 92% of the gained benefits six months after the intervention.

Assuming that all benefits are immediately lost once the intervention ends, with services continuing as planned during the maintenance phase, the ICER is estimated to 435 040 SEK per QALY (see Figure 12). In reality, gained benefits are not immediately lost, and therefore this represents an unrealistic “*worst-case*” scenario. Additionally, the intervention specifically includes a maintenance phase to avoid that this happens. It is more realistic to assume that benefits are lost linearly over time, which is also what is assume in the base-case analysis.

By adjusting the retainment rate so all gained benefits instead are lost linearly during the maintenance phase, the ICER is estimated at 254 033 SEK per QALY, with higher retention rates leading to even lower ICERs. This analysis suggests that, although retention is a critical assumption, the

conclusions are robust and not highly sensitive to variations in the level of effect retention.

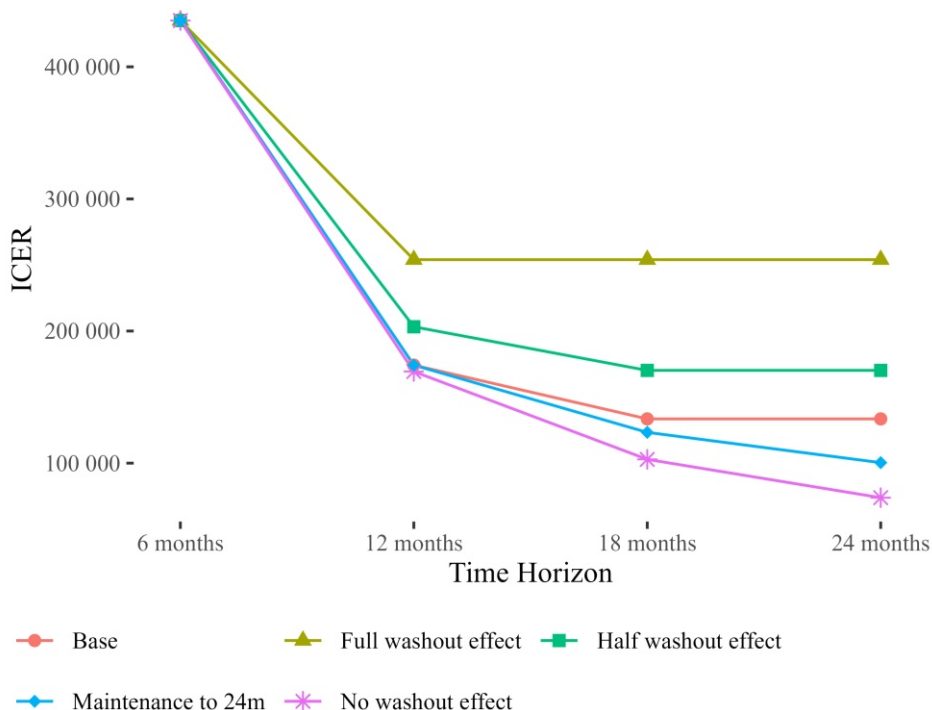


Figure 12 Sensitivity analysis of effect retainment

### Probabilistic sensitivity analysis

In addition to the deterministic sensitivity analysis, a probabilistic sensitivity analysis was also performed. Across 10 000 Monte Carlo simulations the mean ICER was estimated to 162 587 SEK per QALY. With the mean incremental cost of the intervention being 1 855 761 SEK (95% CI: 874 557 SEK - 2 866 270 SEK) and the mean QALYs gained was 125 (95% CI: 120 - 129) for ELR and 114 (95% CI: 111 - 117) for TAU. At a WTP threshold of 350 000 SEK per QALY, the probability of the intervention being cost-effective was 98% (see Figure 13).

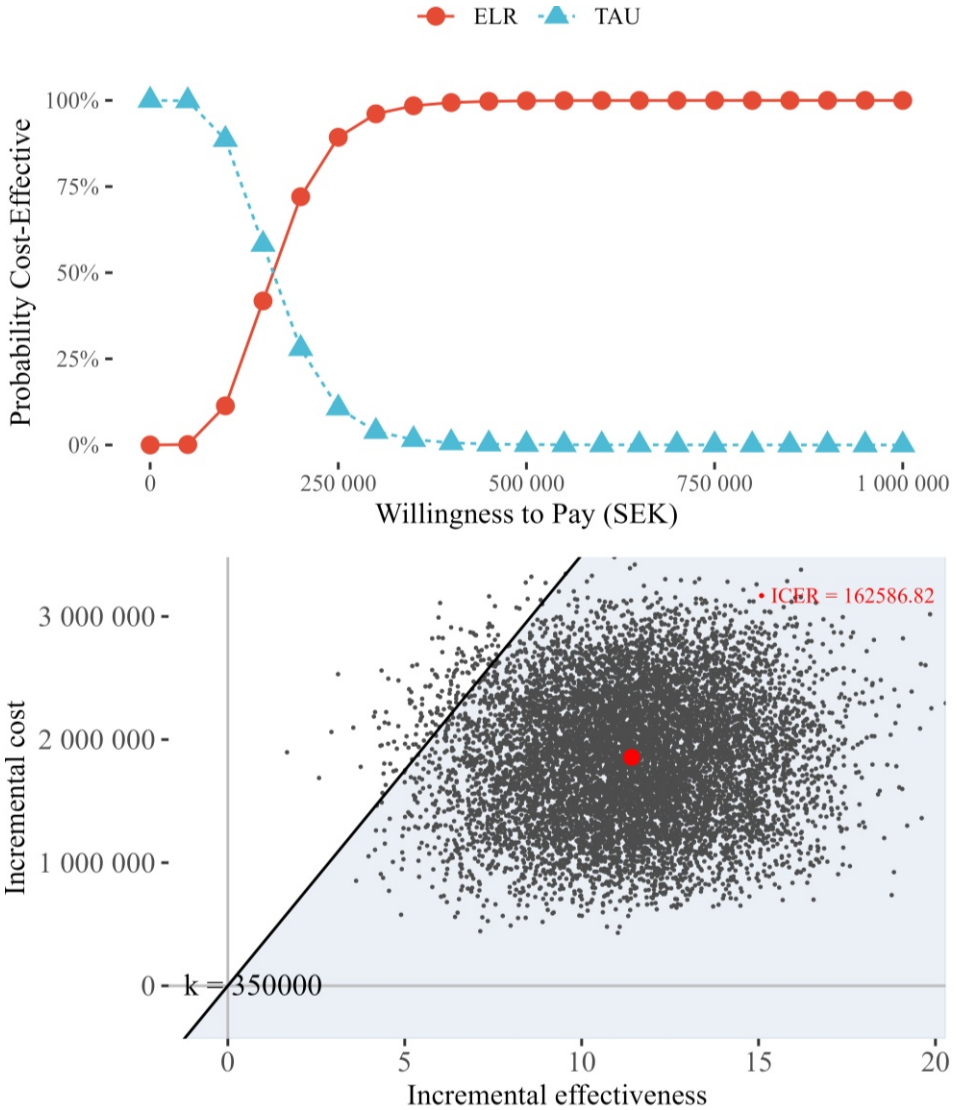
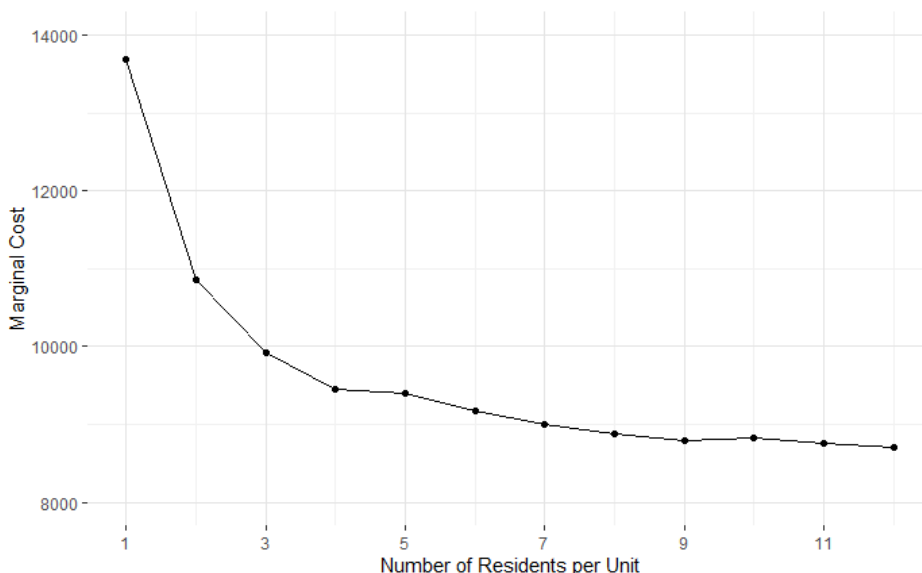


Figure 13 Cost-effectiveness acceptability curve and plane

### Budget impact analysis

As the service delivery is provided at a unit level, the marginal cost of providing the intervention is affected by the number of residents participating (see Figure 14). The marginal cost of offering ELR to one additional resident for one year was estimated to range between 8 699 to 13 681 SEK depending on the number of participating residents per unit.

To evaluate the budget impact of implementing ELR, in Study III, we created scenario using a hypothetical and generic Swedish municipality with 100 000 inhabitants. The participating municipalities had on average 56.8 residents receiving ELR per 100 000 inhabitants. Scaling up the intervention to this level, while maintaining the same staff-to-resident ratio, would require 147.6 housing staff and 19.8 occupational therapists. Depending on the number of housing units involved, providing ELR to these residents for one year is estimated to require labour resources valued between 503 224 to 886 364 SEK. This corresponds to an estimated average cost of 10 118 SEK per resident for one year of ELR.



*Figure 14 Marginal costs per unit*

## Discourses among housing staff

In Study IV the results from the process evaluations are presented as five discursive representations (see Table 24). These will not be presented in the same format in these thesis as they are well-described in the publication. Instead, in the following sections, I will attempt to highlight key parts based on how they influence and are influenced by the contextual challenges. This represents an ad-hoc unstructured attempt to contextualise the findings based on their position in the causal chain.

*Table 24 Discursive representations found in the process evaluation*

| Discursive representations  |
|---|
| Diminishing novelty of ELR and legitimising familiar practices      |
| Restricting resident potential in response to organisational norms  |
| Guarding the boundary between insiders and outsiders                |
| Intersectoral collaboration and issues of ownership and power       |
| Blurred accountability, low expectations and inevitable constraints |

### Contextual challenges in this setting

In Study IV, staff's discourses in response to the early implementation process of ELR was examined, and the corpus allows us to explore the social and organisational structures that influenced the implementation process. This gives insight into existing contextual challenges as they are perceived by staff, while this analysis is not comprehensive, it sheds new light on this specific setting.

#### *Resource scarcity and a strained organisation*

Resource scarcity was a frequent issue raised by housing staff and it is clear that they perceive that they are constrained in what services they are able to provide. Consequently, several respondents constructed the intervention as a potential threat to organisational stability. Staff often assumed their own residents had more severe disabilities than other units in the municipality, emphasising the additional strain they perceive that their specific housing unit have.

#### *The importance of the inner setting*

The inner setting of the organisation was often constructed around a shared experience of working "*on the floor*", fostering a collective identity and creating a sense of sameness. This was always framed as being beneficial, and the staffing group described themselves as closely knitted and having a good working environment. Even when managers were perceived as being hands-off, they were often well liked and clearly accepted as belonging to the inner setting of the organisation.

#### *Isolation and the potential of external actors*

Staff described being isolated and having little contact with other professions or organisations. Staff described a quite distant relationship

with their manager and even responded with laughter when asked about contact with upper management. While this isolation appears to foster a close, long-term relationship with both staff and residents, it also create a sense of stagnation, where staff felt that they had depleted most of their motivational strategies. External actors were therefore framed as a solution to this perceived organisational stagnation.

### How were discourse influenced by context?

The contextual challenges observed in this setting has on impact on how staff talk about the intervention, the organisation and their residents. This allows us to explore how these discourses have been influenced by existing contextual challenges.

#### *The inevitability of the organisational constraints*

Based on the perceived strain of the organisation a discourse of inevitable constraints tended to emerge among staff. The perceived scarcity of resources, over which they had limited influence, often manifested as an attempt to gatekeep available resources to ensure organisational stability. This can be exemplified by staff's use of hypothetical, exaggerated scenarios of resource-intensive activities to emphasise the perceived risks of residents choosing activities freely.

Paradoxically, these hyperbolic scenarios coexisted with the perception that the chosen activities were not disruptive enough. Which would imply that staff believe that they are capable of providing more support than they currently doing. Even if this discourse has developed as a response to real resource scarcity, it embeds a managerial discourse of resource management into their practices. It is unclear why these constraints are viewed as inevitable, but it could likely be due to previous failed experiences of changing practices and policies.

#### *Culture of cohesiveness*

Staff positioned themselves as being unique when compared to other units, often praising themselves for their good work environment. Many had worked together for a long time and explained that they had learned that for the organisation to function, consensus was necessary. Together with the emphasis of sameness this seems to foster a strong culture of cohesiveness, in which deviations from the norm is strongly discouraged. Staff constructs this culture as something beneficial, but it could also make the introduction of new actors and perspectives challenging.

In the corpus of Study IV, we observed that managers tended to be discursively protected by staff, which partly could be explained by their position as an insider of the inner setting. This assumption is strengthened, by observing how the occupational therapist, who in contrast were framed as an outsider, tended to be criticised. Despite being new in this setting, the occupational therapist often intervened and exercised authority over staff, sometimes more actively than managers typically did. One potential interpretation is that the resistance towards the occupational therapist is not necessarily founded on practical concerns but from a failure to establish strong interpersonal relationships and alignment with the local culture of the inner setting.

### *Vague leadership and low expectations*

As managers usually provided quite vague leadership, staff were used to taking responsibility for the delivery of services. This provides individual staff with a lot of informal power, which they are able to use to influence attitudes of others. During the intervention, lack of managerial involvement created a social vacuum, which forced staff and occupational therapists to negotiate responsibilities between themselves. Even if the occupational therapist is considered an outsider, the legal frameworks provide them with formal decision-making mandate in the organisation. This positions them organisationally above housing staff, making them a threat to established power dynamics.

In the corpus, occupational therapists was frequently constructed as being the one who responsible for the implementation of ELR. This is in stark contrast to how the accountability of managers was framed. Despite holding formal organisational responsibility, managers were portrayed as incapable to influence the situation, mirroring this discourse of inevitable constraints found among staff. This blurring of accountability further showcases the importance of aligning with the inner setting to be accepted by the staff.

### How did discourse influence the uptake of ELR?

Lastly, discourses found among staff influence how the intervention was implemented and manifested in practice. This allows us to identify specific implementational challenges but also attempt to explain how these have been shaped by the discourses found among staff.

### *Risk aversion and organisational defensiveness*

This idea of inevitable constraints fosters risk-averse tendencies within the organisation, which limits the organisational tension for change. As a result, staff tend to stress the importance of avoiding potential failures or establishing unrealistic expectations among their residents. Lack of resources, such as vehicles, staffing and economical compensation, were often used to rationalise this organisational defensiveness. The organisational gap between occupational therapists and housing staff create a sense of conflicting priorities, where tasks are perceived as being delegated without consideration for their existing workload.

### *Insiders versus outsiders*

Constructing the occupational therapist as an outsider, who are not part of the inner setting, leads to several implementational challenges. By discursively positioning occupational therapists as accountable for the intervention, the inner setting effectively avoids taking ownership over the implementation process. While this could function as defensive measure, it also reinforces the sense of inevitability and undermines staff's agency to influence the outcome of the intervention. When occupational therapists has taken an advocative role, attempting to enforce the will of the residents, this has been perceived as confrontational. In these cases, the boundaries of insiders and outsiders amplify the conflicts, making intersectoral collaboration challenging. The abdication of ownership combined with vague leadership imposes a disproportionate burden for implementing the intervention on the occupational therapists.

### *Conflation of practices*

The conflation of ELR with usual practices diminishes staff's willingness to change, as it creates a feeling of adherence, even while adapting or ignoring specific interventional components. This could be interpreted as a discursive strategy to rationalise a lack of effort but also functions as a way to establish the value of current practices. One alternative explanation is that staff have not fully internalised the content of the intervention, making the conflation of practices simply an expression of unawareness. More successful units highlighted how the intervention legitimise specific beneficial activities in the organisation, but this seems to require that housing staff do not perceive it as a threat to the available resources.

### *Absence of accountability*

The vagueness of the leadership provided by managers has normalised the absence of support to such an extent that staff no longer anticipate receiving any at all. Rather than holding the organisation accountable for implementation failures, blame was directed at the occupational therapist, who was constructed as responsible for the intervention. The absence of managerial involvement creates a binary situation, where it is either the occupational therapist or housing staff who are responsible for any failures in delivering the intervention. This further reinforces the organisational defensiveness, making the staff more prone to avoid taking risks. Given staff's limited capacity to influence either resident activity engagement or organisational resources, blaming the occupational therapist could be interpreted as a deflection of frustration.

## Early implementation outcomes

During the publication of the individual studies there has been limited opportunity to present a consolidated summary of the process data collected during trial. While some of these have been used in different aspects of the evaluation process, they also represent proxy measures, which can be used to operationalise implementational outcomes. This section represents an ad-hoc analysis of qualitative data from Study IV in combination with process data collected during the RCT.

### Acceptability: Conflation of practices

In Study I, the mean contentment score of ELR in the internal pilot was 7.67 of 10. Residents reported higher contentment with the collaboration with occupational therapists (mean score of 8.44) than with housing staff (mean score of 6.94). As these scores are generic and descriptive, they are quite difficult to interpret, but we believe that it suggests a relatively high satisfaction with the intervention. This conclusion are supported by an independent qualitative study of residents' recovery pathways <sup>124</sup>. In this study, residents emphasised the importance of the intervention, but especially the collaboration with the occupational therapist.

In Study IV, staff explicitly stated that they would like more support from external actors, and that there exists a need for ELR, or similar interventions, in the organisation. However, many believed that the language used in the material was too complex. Some claimed that their residents had chosen the "*wrong goals*", which they believed would be solved if they were allowed to have more influence in the development of

the PCR. We also observed a discursive tendency to conflate the intervention with their usual practices. While this could ease organisational buy-in, it could also make it difficult for staff to assess the relative advantage of the intervention. However, since staff appear to be comfortable equating the intervention with usual practices, it must align sufficiently well with the organisation norms and objectives.

### Adoption: High willingness, but lack of effort

Of all municipalities invited to participate in the early waves in Study I, 16 agreed to attend separate information meetings, of which 7 ultimately agreed to join the study, corresponding to an estimated acceptance rate of 44%. We do not know how many units in each municipality was eligible to participate in the study. The fact that municipalities themselves selected eligible housing units increases the risk of introducing selection bias.

In Study I, several units had problems starting up the intervention in time, especially during the first wave. This is also apparent in Study IV, where housing staff from wave one, explicitly told us about the stressful situation of coming back from vacation and feeling unprepared to start the intervention, as housing managers had forget to tell them they were part of the RCT. Following the first wave, introductory workshops was introduced to better prepare the organisation for the work needed to implement the intervention. After this we received significantly less indications of delays. Our interpretation was that both municipalities and the units were willing to adopt the intervention, but that the lack of organisational structures often was a barrier.

### Appropriateness: Important, just not for their residents

All units in Study IV expressed explicit support for the theoretical rationale of the intervention, suggesting that housing staff believe that ELR is an appropriate tool for their residents. Housing staff also claimed that the intervention did not have a negative effect on their workload, strengthening the assumption that the content is well suited to be used in practice. Some of the more optimistic units even highlighted that the intervention was able to legitimise beneficial practices in the organisation.

Despite acknowledging that the intervention was well suited for the current context, many units paradoxically believed that it would be more effective in other units. Even if the intervention is developed specifically for those with severe manifestations of SMI, staff often expressed that their own residents had to complex disabilities to be able to engage with the intervention effectively. This is even more interesting considering the

large effectiveness observed in Study II. As these interviews occur after the intervention is finished, we would expect housing staff to have noticed the large positive benefits incurred by their residents.

### Feasibility: A hypothetical threat to organisational stability

In the internal pilot, some municipalities experienced challenges in organising and implementing the intervention, which led to delays in starting up the intervention in some units. Even if staff in Study IV often claimed that the intervention did not impede on their everyday practice, they were often worried about what activities residents could potentially choose. They feared that residents would engage in complex activities, like trips and events, which was rarely the case in practice. This hypothetical threat could be caused by a feeling of lack of control among staff. We interpret this as a discursive strategy developed as response to contextual constraints, rather than an expression of the unfeasibility of ELR.

In addition, we observed relatively high attrition rates during the intervention (24%). However, these rates were similar in both of the study arms, suggesting that there was nothing in the intervention that led to the high attrition rates. Staff in Study IV frequently construct their residents as being difficult to motivate and that their mood shifts heavily, which strengthens our picture of this population, where relatively high attrition rates are to be expected.

### Fidelity: Low adherence to the educational components

The educational material was delivered digitally, allowing us to track the total number of views ( $n = 1676$  views). However, as the web-education did not require identification we are unable to determine the exact number of staff exposed to the material. In Study III, it was estimated that approximately 148 housing staff participated in the intervention. This assumes all occupational therapists viewed all modules, only eligible participants accessed the material, and that no modules were re-watched.

In Study IV, housing staff were almost entirely unaware that the monthly reflection sessions should have taken place or what their purpose was, from which we infer that managers have not organised these as planned. Housing staff also seemed to conflate the intervention with their usual practices, suggesting that the content had not been fully internalised. In the previous section “*Estimation of the level of exposure*”, we presented process data showcasing that most residents had regular contact with their occupational therapist, and achieved their set goals, suggesting adequate fidelity to those interventional components.

## Penetration: Unclear reach among participants

In Study III, it was estimated, based on available process data, that approximately eight residents lived at each housing unit. The number of participating residents per unit ranged from 1 to 8 (mean = 2.68, median = 2). This corresponds to an estimated reach of 33%, which resembles the estimate reported in the internal pilot (36%). While this could imply that the whole target population has not been reached, the number is consistent the estimated average cluster size assumed in the original sample size calculation (2 per unit).

# Discussion

In conclusion, ELR has been found to not only be an effective intervention, but also a cost-effective one. This represents robust empirical evidence demonstrated through applied experimental research. In the trial, we observed a statistically, and clinically, significant effect on resident's quality of life. But stopping here means we do not really understand why and how these effects occur, or the nature of the causal statements we are making. Trust me, I am not saying this to be pessimistic. From a decision-making perspective, the choice to implement the intervention is quite straightforward. If the goal is to provide an evidence-based intervention to this, otherwise often forgotten population, ELR is a great choice.

The intervention is highly effective in promoting quality of life, but is also quite affordable, with an estimated opportunity cost of around 10 000 SEK per resident and year. In addition, when considering commonly used WTP thresholds, there is a high probability that ELR is cost-effective compared to usual practices. Both the analysis and conclusions are based on conservative assumptions, which means that the expected cost of providing the intervention in practice is likely to be even lower. If ELR is sustained over time, with resources shifted from low-value activities, and at least some benefits are retained, the intervention is associated with extremely low ICERs. Nonetheless, even in the base-case, the intervention appears cost-effective, with an ICER below the marginal productivity of the healthcare sector <sup>268</sup>, implying good value for taxpayer's money.

Units who received the intervention showed a significant improvement in ReQoL scores, rising from 28 to 49, corresponding to an average gain of 20 points relative to the control group. Given that scores above 50 are typically found among the general population <sup>204</sup>, residents has been able to rise to the upper threshold of the clinical range (0 to 49) after just six months of ELR. However, from a realist perspective, we still have not explained how this large effect has emerged. There remain significant unknowns regarding the intervention's mechanisms and the contextual challenges involved. Let us try to dive even deeper into the causal powers at play and explore how more translational approaches help us disentangle some of the assumptions on which these conclusions rests.

## Why did ELR succeed when others failed?

Evidence on the value of recovery-oriented practices remains inconclusive, and several interventions have failed to demonstrate significant effects in experimental trials <sup>113</sup>. This raises the question: why did we observe such a strong effect as we did? Is there something special about this particular intervention? The effect emerged despite the pragmatic trial design offering limited control over exposure, and several implementational challenges. One potential explanation for this could be that some of the intervention's components or underlying mechanisms are particularly potent.

### The role of the occupational therapist

ELR is a multimodal intervention that, compared with REFOCUS and CARE, adopts a relatively modest approach to staff education. Having difficulties introducing recovery-oriented practices among staff are not unique for supported accommodation, and challenges have been observed across various settings <sup>269-271</sup>. Changing staff practices seems to be difficult, and individual-level strategies, such as educational packages, tend to be associated with minor improvements at best <sup>272</sup>.

Given the strength of the effect, the limited evidence on staff education, and the low engagement among housing staff, targeting individual “*champions*” may be more fruitful than targeting entire staff groups. The integration of occupational therapists in ELR could therefore potentially explain some of the findings. In the process evaluation, staff expressed a desire for support from external actors, and occupational therapists often provide competency, perspectives, and resources that are not currently available within the units. Facilitating activity engagement might also require skilled labour, as untrained staff tend to select activities poorly matched to the needs of those with severe disabilities <sup>273</sup>. In addition, the occupational therapist represents a novel, non-stagnated, relationship for the resident, which potentially make it easier for them to motivate residents in engaging meaningful activities.

The main concern with forcing occupational therapists to be champions due to a lack of organisational implementation structures, is the potential opportunity cost of their involvement <sup>274</sup>. While they may be effective, drawing them from a legitimated profession risks removing expertise from other areas of the organisation. Study III demonstrates that implementing ELR, in a similar scale as in the RCT, would require almost 20 occupational therapists for a mid-sized generic municipality. This

estimate is quite theoretical, as it assumes that the same staff-per-resident ratio as during the trial. In practice, municipalities would be able to adjust the staff-per-resident-ratio, but also the scope of the implementation by limiting the number of units involved at a time, which would significantly reduce the organisational burden. Alternatively, it is likely that other responsibilities need to be reduced, to allow occupational therapist to dedicate more time specifically to this population.

## The importance of meaningful activities

In Sweden, those living in supported accommodation have the right to receive opportunities to engage in everyday community life and live like everyone else. Which means that engagement in meaningful activities should already be a natural part of usual practice, still their use remains inconsistent and often lacks structured frameworks<sup>167, 168, 171, 175, 178</sup>. Given the expensiveness of these services, it is essential to consider how resources are best allocated<sup>159, 275</sup>. The updated Social Care Act (2025:400) stresses the importance of both preventive and evidence-based services, while also emphasising that all individuals should be able to live a dignified life. It is important to prioritise tasks that yield the greatest benefit to residents, in order to improve the quality and efficiency of community-based mental health services.

The explicit hypothesis, that engagement in self-chosen, meaningful everyday activities, out in the real world, promotes personal recovery, is one of the unique components of ELR. Residents in supported accommodation is a vulnerable population whose self-determination often has been compromised by long-term symptoms, marginalisation and service dependency<sup>164–166, 276</sup>. Leisure activities is key for active living, but the form of the activities are often secondary to the meaning derived from the experience<sup>277</sup>. Almost any activity can be considered leisure, as it is defined not by the activity itself, but by one's attitude toward it and the freedom to choose it<sup>278</sup>. Due to the severity of resident's disabilities, we must accept that not everyone will be able to do everything and instead focus on what activities they would like to do.

Participation in activities is associated with higher levels of quality of life and affects health and wellbeing through a multitude of mechanisms<sup>279, 280</sup>. Leisure activities is especially important for those unable to achieve satisfactory levels of meaning through work or family<sup>278</sup>. For this population, engagement in everyday life activities, out in society, could function as an "*antidote to potential lifelong dependency on formal mental health services*"<sup>281</sup>. By applying a life course perspective, activity

engagement also address sedentary behaviour, which could be beneficial for not only quality of life, but also physical health and longevity.

## Explaining deviant cases and uncertainties

While the program theory and the interventional components could partly explain why ELR achieved such good outcomes, there still exists uncertainties and deviations which are not accounted for. Overall, it is important to note that all included studies evaluate short-term effectiveness, with observable data only being available for a 6-month period, and lot of work remains when it comes to evaluating the long-term impact and sustainability of the intervention.

### Threats towards causal inference

In Study II we observed a large association (strength of the effect), consistent findings over four waves (consistency<sup>2</sup>), controlled exposure at two timepoints (temporality), a well-developed program theory (plausibility), and an experimental design (experiment) <sup>282</sup>. All together this provides robust evidence for inferring causality between ELR and the observed effect. However, two issues remain: why did we not observe a dose-response relationship (biological gradient) and what alternative causes are we not able to rule out (specificity)? To invoke these criteria's, we would have needed better control over the measurement of exposure.

Our method of assessing adherence through free-form text in clinical documents is inherently crude and imprecise <sup>283</sup>. Although the experiment shows that ELR, as a package, were able to promote personal recovery, it provides little evidence on the generative mechanisms involved. Personal recovery may have been promoted by activity engagement, it could also have emerged from non-specific treatment effects (attention from an external actor) or observation bias, where participants changed their behaviour due to being studied (Hawthorne effect) <sup>284</sup>. The lack of dose-response and the potential influence of observation bias confound our ability to attribute recovery specifically to activity engagement. Future studies that explore the causal role of individual components should also attempt to account for varying levels of exposure.

In ELR, the appropriate strategies and level of engagement, is determined collaboratively between residents and occupational therapists. It is reasonable to assume that occupational therapist adjusts the level of

---

<sup>2</sup> Not a completely accurate interpretation of consistency which refers to replicated experiments.

engagement based on needs of the individual resident. This means that even those with low session adherence could have received “adequate exposure” to the intervention, which could explain the strong effect observed in the per-protocol analysis. While the optimal “dose” of different components remains unknown, exploring the ways occupational therapists worked in practice could be an important direction for future research. One interpretation is that receiving opportunities to exercise agency and self-determination is just as important as engaging in the activities themselves. However, it is important to acknowledge that this is a pragmatic, rather than explanatory trial, which is designed to assess whether the intervention is effective under real-world conditions, rather than to explain its underlying mechanisms<sup>201</sup>. Additional studies focusing on mechanisms of ELR are planned or on-going.

## Resistance and gatekeeping among staff

How come that even though housing staff support the theoretical rationale of the intervention, they still appeared to be quite indifferent of the progress of the residents? There appears to be a gap between the observed scores on ReQoL, and the discourses observed among housing staff. In addition, almost all units explicitly expressed that they want support from external actors and that they looked forward to the collaboration with the occupational therapist. However, when faced with the practical implementation, several units were disappointed with both the level of collaboration and activity engagement.

The process evaluation suggested that staff rarely resisted the actual intervention, but rather the way it was implemented or how outcomes were measured. As the primary outcome is a patient-reported outcomes measure, the scores are inherently subjective. While research has shown that people can estimate other’s utilities independently of their own preferences<sup>285</sup>, but these proxies show remarkable variability in the direction of bias, typically overreporting impairment while underestimating quality of life<sup>286</sup>. The fact that many housing staff thought they were not that involved in the process, also suggest that they are not the most reliable assessors of resident’s progress to begin with.

In Study IV, we describe staff’s tendency to use hyperbole language to amplify the perceived risk of residents selecting activities freely, but when criticising activities actually performed, they instead trivialised them, arguing that they were not disruptive enough. One might ask, why did such contradictor rhetorical strategies emerge? Risk management is one of the predominant discourses in mental health, despite it in most cases

being countertherapeutic<sup>287</sup>. Recovery-oriented practices require creative risk-taking, and staff must feel comfortable supporting residents in exercising their right to necessary risks<sup>288</sup>. One possible explanation is that staff felt a responsibility for gatekeeping scarce resources, which effectively blocked their receptiveness for change. Increased involvement of the manager could potentially reduce the perceived need for gatekeeping among housing staff. Manager buy-in has been suggested as an important mechanism to promote risk taking, and to produce lasting impact following recovery-oriented staff training<sup>289</sup>.

On the other hand, these contradictory rhetorical strategies could also be interpreted as a direct response to a perceived loss of control and decision-making power. Since the activities were mainly negotiated between the occupational therapist and residents, there is a de facto loss of power for housing staff, even though they still are expected to carry out the plans. Greater housing staff involvement in the decision-making process, could reduce resistance towards the choice of activities, but could also risk diluting residents' autonomy. Mental health services are not characterised by ideological consensus, and staff advocates for different perspectives depending on their educational and professional background<sup>290</sup>.

The organisational and legal separation between housing staff and occupational therapists may give rise to a hierarchical dynamic, potentially amplifying problems arising from conflicting priorities, which have been found to make the introduction of recovery-oriented practices challenging<sup>291</sup>. As occupational therapists sometimes are required to advocate for residents, this may create additional strain in the relationship with housing staff. There is also a potential risk that residents might idealise occupational therapists who advocate for them, while antagonising housing staff who attempt to uphold organisational constraints<sup>292</sup>. To reduce the risk of rivalry, competition, and professional stereotyping, managers may need to adopt a more of "*hands-on-approach*" during the implementation process<sup>293</sup>.

## Sustainability, vague leadership and accountability

Supported accommodation are delivered in a domestic context with high staff availability, which should in theory be an excellent setting for implementing recovery-oriented and person-centred practices. In Study IV we see how housing staff express a physical and emotional closeness, not only to their residents, but also to each other. However, this familiarity also seems to foster a stagnated organisational culture. The services were consistently portrayed as being bound by resource scarcity beyond their

control, producing a discourse of inevitable constraints. There might be a need to cultivate a stronger sense of ownership among housing staff to empower them to take initiative in shaping their own practices.

It has been suggested that educational interventions promoting recovery-oriented practices often are too brief, and underestimate the amount of supervision and guiding that will be necessary <sup>294</sup>. While occupational therapists hold formal responsibility for the rehabilitation process, they were also often inadvertently given informal ownership of the whole implementation process. As housing staff are unlikely to sustain the intervention independently, and it is important that housing managers actively work to ensure compliance to both the PCRCP and the self-care plan following the intervention, in order to maintain achieved progress.

Findings on long-term effectiveness are limited by the short time horizon, making it difficult to hypothesise about the sustainability of the intervention. However, a previous feasibility study suggests good retainment of effect up to six months after the intervention <sup>9</sup>. The process evaluation suggests that for the implementation to be successful, continuous engagement from the housing manager is likely required during the intervention's maintenance phase.

This redistribution of responsibility also raises another fundamental question: What are the managers' responsibilities when it comes to service delivery? In the elderly care it has been noted that when person-centred practices are implemented, staff are often neglected by management, which can make them feel out of control and pushed into more instrumental forms of care <sup>295</sup>. Managers have also reported a discrepancy in the communication used by upper management and the local level <sup>296</sup>, which may create a feeling of "voicelessness", making it challenging for them to translate directives between the different levels.

In Study IV, managers were framed as leaving it up to housing staff and occupational therapists to resolve conflicting demands. When managers fail to take adequate accountability, the responsibility for distributing resources is inherited by staff, a phenomenon which tends to lead to defensive behaviour within organisations <sup>297</sup>. Among mental health staff there is also an expectation that change should be initiated from a higher organisational level <sup>298</sup>. Therefore, it is important that managers are involved in the process, and that they are ready to alter schedules, remove tasks, and reallocate resources, to accommodate the needs of both staff and residents <sup>299</sup>.

## Prioritising welfare services

Determining whether an intervention is cost-effective requires more than accounting for the financial resources consumed in its delivery, it also requires considering the benefits foregone by not using those resources elsewhere. In addition, what society and organisations are willing to pay for benefits might also vary, especially when alternative methods have been used to derive preference-based utility scores.

### Validity of ReQoL: Is a QALY always a QALY?

A key limitation in Study III is the use of ReQoL-UI, as well as the absence of country-specific preference weights for Sweden, which limits the validity of direct comparisons between ELR and other interventions. While EQ-5D have been found to have poor sensitivity for persons with SMI <sup>82-84</sup>, this does not imply that ReQoL is the optimal alternative. It is important to consider the validity and relevance of the ReQoL instrument in the Swedish context.

The ad-hoc sensitivity analysis in Study III showcases the large influence of the physical activity item when calculating utility scores using ReQoL-UI. By incorporating an interaction effect between physical activity and other items, the scoring approach differs conceptually from that used in EQ-5D <sup>300</sup>. This favours interventions that have a large impact on physical activity and have a direct impact on policy decisions. There might be an argument for capturing the synergistic effects of physical and mental health <sup>301, 302</sup>, but this is nonetheless a procedure of which evaluators and decision-makers must be aware.

Despite the extra-welfarist approach having been developed to broaden the evaluative space, applied research has almost exclusively focused on health <sup>77</sup>. Several attempts have been made to widen the evaluative space, most notably through Amartya Sen's capability approach, as exemplified by instruments such as ICECAP <sup>303</sup> and capability-adjusted life years <sup>304</sup>. While alternative preference measures provide an opportunity to consider other aspects than health, it also raises concerns about the comparability of QALYs derived from different instruments. One of the main strengths of QALYs is its ability to compare interventions between different populations and settings, and this development reinvigorates the classical debate: is a QALY always a QALY? <sup>305</sup>

## What are we willing to pay for societal benefits?

Willingness-to-pay (WTP) represents an essential metric in health economic evaluations but often receives quite limited attention in practice. WTP can be operationalised using two distinct approaches: the consumer price of a QALY, denoted as  $v$ , or the marginal productivity of the health care sector, denoted as  $k$ . Empirical research typically demonstrates that  $v$  exceeds  $k$ , suggesting that we value health benefits more highly, than the marginal cost of producing them <sup>268, 306, 307</sup>.

In an optimal healthcare system with unconstrained budgets, relying on  $v$  would be the appropriate choice. However, most tax-funded healthcare systems operate under quasi-fixed budget constraints. In this context, the relevant consideration is not what we are willing to pay for a QALY, but rather the opportunity cost of reallocating resources. In situations where we are not able to predict which services will be displaced, applying estimates of the average marginal productivity of the whole sector is a reasonable approach <sup>268</sup>. Siverskog estimated the marginal cost of a life year in Sweden's healthcare sector at approximately 370 000 SEK, which they translate to an expected cost per QALY of 180 000–430 000 SEK <sup>268</sup>.

It is important to acknowledge that traditional threshold values have been developed in the healthcare sector, which means that they are not necessarily transferable to the rest of the welfare sector. As these services does not have health maximisation as their sole objective, it is not advisable to make prioritisation decisions solely on the basis of QALYs, it also makes estimating marginal productivity more challenging. In Sweden, it is common with references towards a WTP of 500 000 SEK per QALY. While this is not a formally established threshold, it is likely an direct translation of the arbitrary threshold of \$50 000 that appeared in the literature during the 1990s <sup>308</sup>. This threshold lacks both theoretical and empirical justification. In Study III, we relied on revealed WTP thresholds based on government decisions in the Swedish pharmaceutical market, where the mean cost per QALY was estimated at approximately 350 000 SEK. This resembles Siverskog's results and, being much lower than the generic 500 000 SEK threshold, represents a conservative assumption, even if it is an uncertain one.

Residents in supported accommodation may not be in acute need, but they do have a severe and complex, often persistent, disability, resulting in a life expectancy that is 15–20 years shorter than average <sup>309, 310</sup>. Considering the severity of these conditions, and their substantial impact on both autonomy and quality of life, they should have a high priority in

resource allocation according to the Swedish model for open priorities in health care <sup>311</sup>. This is already reflected in national guidelines, where conditions such as schizophrenia and psychosis are classified as severe <sup>227</sup>. Accordingly, the WTP for services targeting this population should be higher, and for pharmaceuticals it is not uncommon with ICERs up to up to 1 000 000 SEK and higher <sup>226</sup>.

## Future of EBM: The case for realism

Throughout this thesis I argue from a realist position, which, as described in the introduction, closely aligns with EBM. Critic against EBM is often centred around issues of ranking different types of evidence. The evidence hierarchy, often illustrated as a pyramid, is one of the most attacked manifestations of EBM, and is frequently framed as objectivist, contextless, and as enforcing dogmatism <sup>312</sup>. Arranging designs hierarchically does not imply that only the highest levels should be used, it simply establishes which designs are best at controlling for confounders, a foundational principle of most research.

While the idea of “*best available evidence*” is inherent to EBM, and requires some sort of ranking, it does not mean that other methods are incapable of producing evidence. There is an inherent contextualism in attempting to explain under which conditions interventions are effective <sup>313</sup>. Pragmatism is sometimes raised as a competing philosophical framework to critical realism. While pragmatism also advocates for a fallibilist understanding of knowledge, it takes an explicit stance against foundationalism. This tends to manifest as opposition towards evidence rankings and the golden status of the RCT. Despite this, experimental designs in themselves are very much in line with pragmatism, as they temporarily suspends prior knowledge and promotes open inquiry <sup>314</sup>.

While the authoritarian tendencies that have accompanied the EBM paradigm warrant attention, they fall a little bit outside the scope of this thesis; I will instead redirect the debate toward the ontological implications of these two positions. According to me, one of the more important contributions of EBM are not these hierarchies, but an increased transparency on what grounds causal statements are made. Pragmatism represents a form of anti-theory, and, as such, tends to resist strong causal claims, a tendency which has been described dismissively as “*...pragmatists are phobic about truth*” <sup>315</sup>, referring to their unwillingness to make causal claims.

## Why ontology matters

Traditional positivism is long since dead in applied science, this form of objectivist anti-realism has since been replaced with different forms of instrumentalism. When critiquing contemporary realist approaches, many mistakenly conflate them with positivism, despite it rather representing a form of anti-realism, where knowledge is reduced to what can be positively verified through experience <sup>316</sup>. This means that concepts are adopted not because they are true, but because we have observed them, or in the case of pragmatism, we found them useful.

This is one of the major flaws of pragmatisms, but the important demarcation here is not from anti-realism per se, but from approaches that leave their ontological commitments implicit and unexamined <sup>317</sup>. By adopting an agnostic stance toward ontology, we often act as our theories are real but avoid epistemologically committing to them. Pragmatists tend to criticise realists for having unnecessary complicated ontological frameworks <sup>256</sup> and instead preferring action-based approaches. They are not that concerned with the details of the “*black box*”, what matters is output and predictive power. Theories are tools to use when needed and discard when not <sup>318</sup>.

Contemporary debates in medical science are typically framed as quantitative versus qualitative, or positivism versus relativism. This epistemological distinction is an intellectual hay-man which often rests on exaggerated versions of the epistemology of science <sup>319</sup>. Framing it only as an epistemological issue, also transform it into a debate of naturalism. Namely, should social science rely on the same methods as natural science or not? Qualitative research often reproduces exactly the same form of empiricism found in positivism <sup>320</sup>. “*I heard it and therefore it exists*”, it is just applied to a subjective phenomenon.

## Risky predictions and experiments

Designing experiments requires making risky predictions but also force us to be explicit of ontological assumptions. Of course, not everything is suitable for testing in a RCT, but any planned inquiry including deliberate and reliable procedures designed to detect error, could be considered an experiment <sup>321</sup>. This also relates to the critical realist notion of the epistemic fallacy, stressing the importance of not conflating what we know about the world, with what is true in it. Accepting the relativistic notion of all types of inquiry and observations are equally valid, leaves us without a way to distinguish between contradictory theories. While making

ontological commitments does not eliminate disagreement, it directs the debate to the right place.

Some argue that experimental trials are not compatible with a realist approach <sup>322</sup>, a sentiment which I reject. Although RCTs are not necessary for causal inference, they represent one of the most effective methods available for testing hypotheses. There are also practical examples on how realism and RCTs are able to complement each other <sup>323</sup>, which I believe yield stronger and more informative causal statements. One valid critique is the reliance on randomisation to control for unknown variables, where pseudo-experiments have been found to achieve comparable results <sup>26, 27</sup>. The problem with matching is that it is impossible to know when bias have been introduced and when it has not been. In addition, matching is not exclusive for pseudo-experiments, differences in baseline covariates even after randomisation are not only acceptable but expected. It is important that we balance what we can and randomise what we can not <sup>324</sup>.

*“Randomization is merely a practical tool to reduce confounding. It does not fundamentally change the nature of the way we view or research the social world, or affect how we will use comparative empirical data to test hypotheses about mechanisms” Chris Bonell <sup>325</sup>”*

## The role of qualitative research in evaluations

One of the most classical debates in research is the use of quantitative versus qualitative data <sup>326</sup>. Karl Popper, for example, supposedly described qualitative research as *“hopelessly unfocused”* <sup>98</sup>. It is important to acknowledge potential limitations of qualitative research, such as the lack of accuracy measures, difficulties to generalise, and the potential influence of the researcher <sup>327–329</sup>. Nevertheless, some aspects of life are simply not captured by quantitative data. Claiming that a survey checkbox would contain more valuable information than an interview response would be absurd, and there is no inherent value in stripping away information by transforming qualitative into quantitative data.

The term qualitative relates to the nature of a phenomenon, often with a focus on interpreting attributed meaning, with the core question being: what is this really about? <sup>243</sup> As described in the philosophical statement, I reject relativistic ontological positions, which holds true even when working with qualitative data. There is an abundance of theoretical schools adhering to relativistic principles, such as constructivism, interpretivism, hermeneutics, many of which hold prominent positions in the qualitative research. One of the main issues with relativism is its lack

of clear criteria for deciding between competing alternatives, whereas critical realism argues that it is possible to rationally determine which interpretations have greater explanatory power <sup>330</sup>.

According to a critical realist, the choice of methods should be based on the nature of the research question, a position which neutralises some of the issues of “*paradigm switching*” <sup>331</sup>. Meanings held by individuals, and other mental processes, causally influence events in the physical world, and visa versa, events in the physical world impact individual’s beliefs <sup>332</sup>. This implies that qualitative and quantitative data simply have different logical forms for describing properties of objects <sup>333</sup>. An implication of this is that mental entities, such as meaning, agency and reason, are crucial when developing causal statements.

It has been suggested that health economics should apply qualitative methodology to a larger extent, as it is needed to be able to explain the mechanisms producing outcomes <sup>334, 335</sup>. One could argue that health economists indirectly are already relying on qualitative data to populate models, but in the form of expert opinions and assumptions. While variables have to be mathematically operationalised, qualitative data can be used to explain, support, interpret, and challenge the assumptions applied. This can be done for both behavioural assumptions, how individuals choose to engage with an intervention, but also structural assumptions, as the transition probabilities between health states.

Even when it is difficult to generalise results from qualitative studies, they should at least be considered more reliable than informed guesses by armchair economists. For example, in Study III, we discuss the assumed opportunity cost of the time use of housing staff. From this pragmatic trial alone, it is impossible to distinguish exactly what services was forsaken in order to deliver the intervention. Current praxis in health economics is to assume that market salaries is an adequate proxy for the value of the opportunity cost. However, in Study IV we observed how staff themselves claim that they were not that involved, and that they did not experience any negative impact on their workload. Suggesting that the opportunity cost of ELR is likely much lower than what is assumed in the base-case.

## Disentangling the evaluation, and its limitations

The premise of the causal statement in this thesis, is that ELR has been found to improve personal recovery in an experimental trial in Swedish supported accommodation. This type of evidence do not really tell us that

an intervention works, rather that it has worked previously, in the specific population and setting in which the trial was conducted<sup>336</sup>. However, the logic of the statement is that efforts to promote meaningful everyday activities, by integrating outreach-based rehabilitation and staff training into usual practice, could potentially generate comparable outcomes in similar settings.

Provided that resource use resembles what was observed in the trial, the intervention is estimated to yield benefits at a level considered cost-effective within Swedish healthcare. There seems to exist a high demand for internal and external support within the organisation, and the intervention requires substantial involvement from the occupational therapist. Relational legitimacy also appears to play a significant role within the inner setting, which can create challenges when integrating new actors from outside of it. To mitigate the impact of implementational barriers, greater involvement of housing managers might be required. Holding managers accountable and involving them more actively could likely substantially reduce implementational friction.

### For whom, under what circumstances, and how?

The title of this section is based on Ray Pawon's well-known slogan for realist evaluations: "*for whom, under what circumstances?*"<sup>102</sup>. While this thesis does not include a realist evaluation and therefore does not construct any context-mechanism-outcome (CMO) configurations, this slogan prompts important questions relevant for all evaluations. Which is why it is borrowed here to reflect on the variability and external validity of the findings.

As the estimated effectiveness is aggregated on a housing level, it is difficult to determine for whom the intervention is most beneficial. Although the intervention was developed for those with severe manifestations of mental illness, Study IV suggest that staff perceive it as better suited to younger, more active residents. This is particularly important given that the mean age among participating residents were 39 years, which signals a conflicting view between the potential of the intervention and the underlying values held by the organisation. Recovery through activity is not limited to those who are "*well-functioning*", and ELR is designed to support even those with severe symptoms. This is also an example on how mixed methods can highlight potential issues which would not be found using only quantitative research.

The pragmatic design offers limited opportunity to control for disease severity, and the available data do not allow testing of which subgroups benefited most from the intervention. While there might be a potential risk that “*well-functioning*” residents are overrepresented in the sample, eligibility for supported accommodation requires extensive needs, ensuring that all participating residents have severe disabilities. An implication and strength of the pragmatic design is that the recruited sample most likely resembles what would be expected in practice. However, future research should still consider that the effectiveness, and adoption, of ELR could potentially differ across subgroups differing in diagnosis and severity of psychiatric conditions.

Despite observing a statistical and clinically significant effect on recovering quality of life among residents, the process evaluation revealed large implemental issues in several units. These were mainly related to housing staff and managers, including delays in the start-up of the intervention, limited internalised of the material, and a general sense of detachment from the process. This is partly expected due to the purposive sampling strategy, which targeted maximum variation, but surprisingly we noticed similar tendencies in more successful units. Based on the findings, we hypothesise that it is likely the occupational therapist that are one of the primary contributors to the intervention’s success. Further research is on-going on the experiences among occupational therapist, which could potentially shed further light on this hypothesis.

Many housing staff appears not to have fully internalised the material, implying that additional efforts may be needed to ensure full adoption. This is in line with previous research, which has suggested that educational efforts may be a weak component in promoting recovery-oriented practices <sup>113</sup>. Whether or not more intense educational efforts could provide value remains debatable, but this thesis suggests that alternative approaches may be more worthwhile pursuing. There is also a need to ensure the involvement of housing managers and adequate organisational readiness for the implementation process, as practices are unlikely to change without addressing structural limitations.

There exist several contextual challenges that hindered the uptake of the intervention. In Study IV, staff predominantly describes a distant relationship with their manager, alongside a strong desire for external support. Despite the occupational therapist being considered an outsider, they were often given ownership of the implementation process. The manager on the other hand, who already has relational legitimacy, was rarely referred to as having any mandate or responsibility at all.

Redefining the manager's role to promote more hands-on involvement in day-to-day practice has the potential to address many of the observed implementation challenges.

This thesis provides evidence supporting the conclusion that ELR, as an interventional package, has a large short-term effect on residents' quality of life. This informs us that further research on long-term effectiveness, the mechanism of change, and implementation sustainability, is warranted. Given the absence of a dose-response and the uncertainties regarding the respective contributions of housing staff and occupational therapists, studies of mechanisms are particularly important. Quasi-experimental designs or case studies could be used to test different implementation strategies, while also ensuring consistent adherence. Another viable approach may be longitudinal follow-up of organisations implementing the intervention in practice, allowing interprofessional relationships to be observed as they develop over time.

## Limitations of the studies

To be able to interpret findings from the studies, it is important to acknowledge the potential limitations which threaten their validity. Given the critical perspective adopted in Study IV, it seems only right to start with acknowledging that most findings could be influenced by the project teams theoretical commitments and assumptions. The principal investigator of this project is also the developer of ELR. To reduce the potential risk of researcher influence, she has not been directly involved in neither the statistical nor economic analysis. In addition, external assessors and university administrators collected and stored data, keeping it blinded from the researcher's conducting the analyses.

### *Limitations related to the measures*

For the whole thesis, it is important to recognise the limited time horizon of this experiment. The findings are inherently bound by the six-month study period, which significantly increases uncertainty for all assumptions related to aspects like, long-term effectiveness, sustainability, and societal impact. When it comes to the long-term implementation of the intervention, and its translation into practice, substantial knowledge gaps remain. To avoid drawing faulty or to optimistic conclusions, conservative assumptions were used heavily in the health economic analysis. Given these conservative assumptions, and the fact that is a low-resource intervention designed to integrate well with usual practice, there is a high probability that additional long-term data would favour the intervention.

The use of subjective patient-reported outcomes lower the certainty of evidence <sup>337</sup>, but it is important to highlight that this aligns with the intervention's emphasis on personal recovery, which is simply not captured by objective measures. Another key limitation is the use of ReQoL-UI in Study III, and its absence of country-specific preference weights for Sweden. It has been suggested that national value-sets could have limited impact on the findings, and that focus instead should be on the validity of the instrument in itself <sup>338</sup>. However, as seen in the results, ReQoL-UI uses a distinct scoring approach which could limit the validity of direct comparisons between ELR, and other interventions evaluated using alternative instruments.

Overall, the pragmatic design makes it challenging to make causal statements on the impact of different generative mechanisms. Avoiding invasive design features strengthens external validity, but also forces us to rely on non-invasive proxies when estimating levels of exposure <sup>339</sup>. The personal free-text reporting used in the process diaries lowers the validity of data collected on adherence and make it difficult to accurately analyse the impact of exposure. As the occupational therapists were formally employed by the municipalities, they operated under a typical caseload and retained other duties during the trial. While this strengthens the external validity, there is a potential risk of contamination, as occupational therapists could be responsible for serving residents in both arms. However, as services provided in the control are brief and sporadic, the overall risk for contamination is still considered low.

### *Statistical limitations*

Although Study II include a large number of units, the small cluster sizes increase sensitivity to within-cluster variation. Nevertheless, the cluster sizes remained consistent with those estimated in the original power calculation <sup>12</sup>. Some imbalances in demographics were also observed, but as there were no major deviations in the randomisation process, they are assumed to be caused by random variation. A relatively high attrition rate was observed, which could be a potential threat for the external validity, but we found no systematic differences in attrition between the groups. In addition, the attrition rates were consistent with those observed in other trials conducted in similar contexts <sup>110, 111, 340</sup>.

Planned subgroup analysis by diagnosis was not performed due to small subgroups and low data validity, representing a deviation from the study protocol. The number of housing units and residents in each study arm varied, as thirteen randomised units dropped out before resident

recruitment could begin. As a result, baseline data for these potential residents were not available for imputation. In addition, as residents themselves decided whether to participate or not, some units enrolled more or fewer participants than anticipated, affecting the distribution of participants. While these imbalances are not considered to have introduced any major systematic bias, they represent a potential source of unrepresentativeness.

### *Trustworthiness in the qualitative analysis*

Interviews create a specific, co-constructed discursive setting, or genre, which is shaped by both the interviewer and participants. This is important, as the focus in Study IV were not on the conversation in itself, or discourses that only emerge during interviews. Consequently, the analysis of discursive practices does not necessarily represent naturally occurring interactions. It is quite likely that some discourses have not been fully captured, where a corpus based on more natural forms of communication would be required.

While the procedures in Study IV were documented and communicated as explicit as possible, practical constraints limit how much raw data can be published. When identifying, selecting, and presenting appropriate excerpts, there is always a risk of selective reporting. To ensure the trustworthiness of claims, and that arguments are substantiated by data, excerpts were presented in their full context where possible. When analysing discourses, it is important to properly balance the presentation of data against its interpretation, as data are not assumed to “*speak for themselves*”<sup>341</sup>. In the manuscript, the distinction between text, discursive practice, and social practice effectively narrates the interpretation process, tracing the movement from raw data to discursive representations<sup>341</sup>.

### What about the evaluation monster?

Having reached the end of this evaluation of ELR, what have we learned about the evaluation monster? I hope that this work have showcased how complex interventions can be approached using RCTs in combination with translational approaches to enrich experimental findings. The process evaluation, for example, describes how housing staff employed discursive strategies to justify gatekeeping and risk-averse practices. Building on this, it is hypothesised that strengthening relational legitimacy and managerial accountability may be critical for addressing challenges rooted in organisational power dynamics and resource scarcity in this context.

Similarly, the health economic evaluation highlights key knowledge gaps, such as the opportunity cost of housing staff and the retention of effect, while still providing a structured framework for transparent decision-making, despite uncertainties. As demand for evidence grows, it is important to acknowledge that not every phenomenon can, or should, be studied with large-scale trials. While I teasingly refer to this impulse as an *evaluation monster*, it is by no means a rebuttal of experiments, quite the opposite. However, most decisions involves uncertainty, and both the societal cost of conducting research, and the opportunity cost of delaying decisions, must be considered <sup>342</sup>.

Both health economics and implementation science are a natural part of contemporary medical research, and I believe they could provide considerable value in the welfare sector as well, if applied thoughtfully. Evaluations are based on an desire to understand a previously unknown condition <sup>107</sup>, which means that the numeric estimates they produce are not necessarily the most important finding. Estimates of effect shows what changed, in a specific time and place, but provides little information about how interventions will manifest in new contexts. The same intervention, implemented with the same fidelity, can produce different outcomes when context differ, meaning that effects are not a single fixed thing, but are inherently relative and situational.

There is an allure in treating observations as stable facts, rather than contingent outcomes shaped by specific contexts and generative mechanisms. The challenge for evaluators is therefore to move beyond identifying empirical regularities, toward identifying underlying causal mechanisms that generate observed outcomes. Everyone occasionally falls victim to the evaluation monster, believing that more or better data will inevitably lead to easier decisions. This monster feeds on our desire for simple answers and definitive results. Rather than attempting to eliminate uncertainty, the focus should be on understanding, explaining, and communicating it transparently and honestly.

## Concluding remarks

The overall aim of this thesis was to provide an evaluation of the effectiveness, cost-effectiveness and early implementation of ELR, while using the project as a canvas to contribute to the broader debate on the role of evaluations within welfare settings. Throughout this thesis, I have attempted to merge and contrast different analytical approaches in order to construct a comprehensive understanding of the trial, the intervention,

and the setting. Whether I succeeded in this objective or not, must be left to the reader to decide.

The first objective of this thesis was to test the hypothesis that ELR are able to promote personal recovery and quality of life by engaging individuals in meaningful everyday activities. In Study II, a statically and clinically significant difference was observed between those who received ELR and those in the control, which supports the conclusion that ELR promote personal recovery. Frequentist statistics suggest that it is unlikely that we would have observed such a large difference without a causal influence of the intervention. Although it remains unclear to what extent the effect was caused by activity engagement, the findings support a causal relationship between the intervention and recovering quality of life. Findings from Study IV suggest that the relationship with the occupational therapist is a potential key component. However, the trial was not designed to formally test the influence of different interventional components.

The second objective was to quantify the comparative effectiveness of ELR and conduct an economic evaluation, assessing its incremental benefits, costs, and overall impact relative to usual practice. The findings suggest that ELR are likely to be cost-effective in promoting quality-adjusted life years for persons with SMI living in supported accommodation. Despite uncertainties regarding long-term effects and resource consumption, the robust effect, combined with conservative assumptions, makes it likely that the intervention would be cost-effective if implemented in practice.

The third objective was to explore housing staff's discourses and reasoning around the implementation of ELR, with particular attention to how social dynamics and contextual conditions influenced the implementation process. The process evaluation demonstrates how the staff employed several discursive strategies to rationalise resistance and gatekeeping. This often manifested as a fear of residents own choices and led to a downgrading of their capacity for change. In addition, attempts were made to explain how these discourses emerged from contextual challenges, such as organisational isolation, vague leadership, and resource scarcity.

Supported accommodation is one of society's most intensive welfare services, which also targets one of our most vulnerable and marginalised populations. Tax-based systems have a duty to be transparent when implementing and prioritising among services. Such justification should not rely on financial arguments alone, and the impact on utility must be

considered when prioritising shared resources. I argue that we have a responsibility to ensure that community-based services do not reproduce the patterns of paternalism, neglect and institutionalisation that persons with psychiatric conditions have historically faced <sup>343-345</sup>.

Significant work remains in examining the normative and ethical foundations underlying the adaptation of evaluative approaches to the welfare sector. Recovery-oriented interventions, such as ELR, requires a fundamental shift in organisational and societal values: from a system focused on control and protection to a system centred on rights and flourishing. To make reliable decisions, we must dare to critically and systematically evaluate not only novel interventions, but also current practices. Regardless of methodological disagreements, making uncertainties and assumptions explicit, and asking critical questions, are sound first steps towards disentangling complex interventions.

# Acknowledgement

First and foremost, I want to thank my supervisor Maria Lindström for her invaluable guidance, support, and encouragement throughout these four years. ELR is her creation, and being allowed to tear it apart has been both a privilege, and at times challenging, though I suspect she enjoyed watching me struggle. It is hard to imagine a better supervisor than Maria, and I will always be grateful for her unwavering support.

I would also like to thank my co-supervisors Lars Lindholm and Linda Richter Sundberg, as well as my informal supervisor Per Liv, for their constructive advice and critical feedback, which made this work considerably better than it would otherwise have been.

Academia was never for me. Growing up working-class in rural northern Sweden, graduating with below-average grades from the second-worst high school in the country, higher education was not really in the cards. When I, despite this, applied to university, my family asked why the hell I would do that. Why pursue something so impractical and pointless?

When I first started, I felt out of place, never quite sure I belonged. That changed, largely thanks to my teachers, colleagues, and friends at the Department of Epidemiology and Global Health. They showed me that good research does not have to be dull, but can remain warm, playful, and down to earth. At some point, the feeling of being an outsider disappeared.

While my family would probably still call my thesis pretentious, and in some ways, it probably is, I know they appreciate the work that went into it. My mother will be in the audience at my defence, so I may take the opportunity to ask her. For my father: I miss you. And to Miriam, my Vilhelmina boys, and my friends, thank you for being part of my life, you will remain an important part of it for many years to come.

/André Sjöberg

# References

1. Nutley S, Boaz A, Davies H, Fraser A (2019) New development: What works now? Continuity and change in the use of evidence to improve public policy and service delivery. *Public Money & Management* 39:310–316
2. Fleetwood S (2017) The critical realist conception of open and closed systems. *Journal of Economic Methodology* 24:41–68
3. Mullen EJ, Shlonsky A, Bledsoe SE, Bellamy JL (2005) From concept to implementation: challenges facing evidence-based social work. *Evidence & Policy* 1:61–84
4. SFS 2025:400 (2025) Socialtjänstlag.
5. Deaton A, Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* 210:2–21
6. Lindström M, Lindberg M, Sjöström S (2011) Responsiveness or resistance : views of community care workers encountering a new rehabilitation model.
7. Lindström M (2011) Promoting agency among people with severe psychiatric disability : occupation-oriented interventions in home and community settings.
8. Lindström M, Isaksson G (2017) Personalized Occupational Transformations: Narratives from Two Occupational Therapists' Experiences with Complex Therapeutic Processes. *Occupational Therapy in Mental Health* 33:15–30
9. Lindström M, Hariz G-M, Bernspång B (2012) Dealing with Real-Life Challenges: Outcome of a Home-Based Occupational Therapy Intervention for People with Severe Psychiatric Disability. *OTJR: Occupation, Participation and Health* 32:5–14
10. Lindström M (2022) Development of the Everyday Life Rehabilitation model for persons with long-term and complex mental health needs: Preliminary process findings on usefulness and implementation aspects in sheltered and supported housing facilities. *Front Psychiatry* 13:954068

11. Lindström M, Sjöström S, Lindberg M (2013) Stories of Rediscovering Agency: Home-Based Occupational Therapy for People With Severe Psychiatric Disability. *Qual Health Res* 23:728–740
12. Lindström M, Lindholm L, Liv P (2022) Study protocol for a pragmatic cluster RCT on the effect and cost-effectiveness of Everyday Life Rehabilitation versus treatment as usual for persons with severe psychiatric disability living in sheltered or supported housing facilities. *Trials* 23:657
13. Lindström M (2007) *Vardagslivets Rehabilitering – Del A: Modellverksamhet för rehabilitering inom socialpsykiatri: Pilotprojekt i svensk kommun med modellverksamhet baserad på OTIPM (Occupational Therapy Intervention Process Model)*. Umeå Universitet
14. Sjöberg A, Liv P, Lindström M (2025) Effect of Everyday Life Rehabilitation on recovering quality of life in individuals with serious mental illness in supported accommodation: a pragmatic cluster randomised controlled trial. *BMJ Ment Health*. <https://doi.org/10.1136/bmjment-2025-301757>
15. Sjöberg A, Liv P, Lindholm L, Lindström M (2026) Evaluating the cost-effectiveness of Everyday Life Rehabilitation (ELR) for individuals with serious mental illness in supported accommodation. *Eur J Health Econ*. <https://doi.org/10.1007/s10198-026-01905-8>
16. Howick J (2015) *The Philosophy of Evidence-Based Medicine*.
17. Guyatt G, Cairns J, Churchill D, et al (1992) Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA* 268:2420–2425
18. Ratnani I, Fatima S, Abid MM, Surani Z, Surani S (2023) Evidence-Based Medicine: History, Review, Criticisms, and Pitfalls. *Cureus* 15:e35266
19. Godlee F (2007) Milestones on the long road to knowledge. *BMJ* 334 Suppl 1:s2-3
20. Davies H, Nutley S, Tilley N (2000) What works? Evidence-based policy and practice in public services.

21. Jacobsson K, Meeuwisse A (2020) 'State governing of knowledge' – constraining social work research and practice\*. *European Journal of Social Work* 23:277–289
22. Berg H, Fjelland R (2026) Numbers and measurement: a critique of evidence-based practice in psychology. *Humanit Soc Sci Commun* 13:463
23. Berg H (2024) Is evidence-based practice justified?—A philosophical critique. *Journal of Evaluation in Clinical Practice* 30:855–859
24. Worrall J (2002) What Evidence in Evidence-Based Medicine? *Philosophy of Science* 69:S316–S330
25. Ogden T (2020) RCTs in Development Economics, Their Critics and Their Evolution. pp 126–151
26. Shadish WR, Clark MH, Steiner PM (2008) Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments. *Journal of the American Statistical Association* 103:1334–1344
27. Waddington HS, Villar PF, Valentine JC (2023) Can Non-Randomised Studies of Interventions Provide Unbiased Effect Estimates? A Systematic Review of Internal Replication Studies. *Eval Rev* 47:563–593
28. Borgerson K (2009) Valuing evidence: bias and the evidence hierarchy of evidence-based medicine. *Perspectives in biology and medicine*. <https://doi.org/10.1353/pbm.0.0086>
29. Dahler-Larsen P (2011) *The Evaluation Society*. Stanford University Press
30. Sundell K, Olsson T (2021) Svenska effektutvärderingar av beteendemässiga, psykologiska och sociala insatser 1990-2019. Forte – Forskningsrådet för hälsa, arbetsliv och välfärd, Stockholm
31. Socialstyrelsen (2020) *Evidensbaserad praktik i socialtjänsten 2019*.
32. Statistics Sweden (2024) *Costs per area of activity for regions in 2020–2024*.

33. Statistics Sweden (2024) Costs per area of activity for municipalities in 2020–2024.
34. Vårdanalys (2021) Nära vård i sikte?
35. Ludvigsson JF, Bergman D, Lundgren CI, Sundquist K, Geijerstam J-L af, Glenngård AH, Lindh M, Sundström J, Kaarme J, Yao J (2025) The healthcare system in Sweden. *Eur J Epidemiol* 40:563–579
36. Fire M, Guestrin C (2019) Over-optimization of academic publishing metrics: observing Goodhart’s Law in action. *Gigascience* 8:giz053
37. Landhuis E (2016) Scientific literature: Information overload. *Nature* 535:457–458
38. Sogi G (2023) Research Waste. *Contemporary Clinical Dentistry* 14:179–179
39. Chalmers I, Glasziou P (2009) Avoidable waste in the production and reporting of research evidence. *The Lancet* 374:86–89
40. Altman DG (1994) The scandal of poor medical research. *BMJ* 308:283–284
41. Open Science Collaboration (2015) Estimating the reproducibility of psychological science. *Science* 349:aac4716
42. Hoffmann F, Allers K, Rombey T, Helbach J, Hoffmann A, Mathes T, Pieper D (2021) Nearly 80 systematic reviews were published each day: Observational study on trends in epidemiology and reporting over the years 2000-2019. *Journal of Clinical Epidemiology* 138:1–11
43. Eysenck HJ (1978) An exercise in mega-silliness. *American Psychologist* 33:517–517
44. Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S, Forbes A (2014) Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. *Cochrane Database Syst Rev* 2014:MR000035

45. Mayo-Wilson E, Li T, Fusco N, et al (2017) Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. *Journal of Clinical Epidemiology* 91:95–110
46. Bohlin I (2011) Evidensbaserat beslutsfattande i ett vetenskapsbaserat samhälle: Om evidensrörelsens ursprung, utbredning och gränser. In: *Evidensens många ansikten*. p 31
47. Puljak L (2022) The difference between evidence-based medicine, evidence-based (clinical) practice, and evidence-based health care. *Journal of Clinical Epidemiology* 142:311–312
48. Parrish DE (2018) Evidence-Based Practice: A Common Definition Matters. *Journal of Social Work Education* 54:407–411
49. Citrome L (2021) Evidence-based medicine: It's not a cookbook! *Current Psychiatry*. <https://doi.org/10.12788/cp.0124>
50. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312:71–72
51. Worrall J (2010) Evidence: philosophy of science meets medicine. *Journal of Evaluation in Clinical Practice* 16:356–362
52. Mellis C (2015) Evidence-based medicine: what has happened in the past 50 years? *J Paediatr Child Health* 51:65–68
53. Tugwell P, Knottnerus JA (2015) Is the 'Evidence-Pyramid' now dead? *Journal of Clinical Epidemiology* 68:1247–1250
54. Wieten S (2018) Expertise in evidence-based medicine: a tale of three models. *Philos Ethics Humanit Med* 13:2
55. Oliver K, Pearce W (2017) Three lessons from evidence-based medicine and policy: increase transparency, balance inputs and understand power. *Palgrave Commun* 3:43
56. Vedung E (2009) *Utvärdering i politik och förvaltning*. Studentlitteratur, Lund
57. Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV (2021) Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. *Eur J Epidemiol* 36:873–887

58. Sjöberg A (2025) Vad är effektutvärdering? In: Utvärdering, uppföljning och granskning i socialt arbete : inledning och centrala aspekter. Natur och kultur, pp 240–256
59. Shadish W, Cook T, Campbell D (2001) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, 2nd edition. Cengage Learning, Boston
60. Moore GF, Evans RE, Hawkins J, Littlecott H, Melendez-Torres GJ, Bonell C, Murphy S (2019) From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation (Lond)* 25:23–45
61. Diener E, Northcott R, Zyphur MJ, West SG (2022) Beyond Experiments. *Perspect Psychol Sci* 17:1101–1119
62. Wichman C, Smith LM, Yu F (2020) A framework for clinical and translational research in the era of rigor and reproducibility. *J Clin Transl Sci* 5:e31
63. Leppin AL, Mahoney JE, Stevens KR, et al (2019) Situating dissemination and implementation sciences within and across the translational research spectrum. *J Clin Transl Sci* 4:152–158
64. Westerlund A, Nilsen P, Sundberg L (2019) Implementation of Implementation Science Knowledge: The Research-Practice Gap Paradox. *Worldviews Evid Based Nurs* 16:332–334
65. Djulbegovic B, Paul A (2011) From Efficacy to Effectiveness in the Face of Uncertainty: Indication Creep and Prevention Creep. *JAMA* 305:2005–2006
66. Cartwright N, Munro E (2010) The limitations of randomized controlled trials in predicting effectiveness. *J Eval Clin Pract* 16:260–266
67. Cartwright N (2010) What are randomised controlled trials good for? *Philos Stud* 147:59–70
68. Cartwright N (2022) A Philosopher Looks at Science. <https://doi.org/10.1017/9781009201896>

69. Srikanthan A, Amir E (2015) Efficacy-effectiveness gap as an obstacle to translating clinical trials to clinical practice. *European Journal of Cancer* 51:905–906
70. Hawe P, Shiell A, Riley T (2004) Complex interventions: how “out of control” can a randomised controlled trial be? *BMJ* 328:1561–1563
71. Patsopoulos NA (2011) A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 13:217–224
72. Broqvist M, Branting Elgstrand M, Carlsson P, Eklund K, Jakobsson A (2011) National Model for Transparent Prioritisation in Swedish Health Care : Reviderad version. Linköping University Electronic Press
73. Krevers B, Bäckman K (2023) Modell för prioriteringsarbete i kommuner processer och verktyg vid resursfördelning i en övergripande ledningsprocess. Linköping University Electronic Press, Linköping
74. Brouwer WBF, Culyer AJ, van Exel NJA, Rutten FFH (2008) Welfarism vs. extra-welfarism. *Journal of Health Economics* 27:325–338
75. Blaug M (1997) *Economic Theory in Retrospect*, 5th ed. <https://doi.org/10.1017/CBO9780511805639>
76. Mwachofi A, Al-Assaf AF (2011) Health Care Market Deviations from the Ideal Market. *Sultan Qaboos Univ Med J* 11:328–337
77. Coast J (2009) Maximisation in extra-welfarism: A critique of the current position in health economics. *Social Science & Medicine* 69:786–792
78. Seixas BV (2017) Welfarism and extra-welfarism: a critical overview. *Cad Saúde Pública* 33:e00014317
79. The World Health Organization (2012) *The World Health Organization Quality of Life (WHOQOL)*.
80. Whitehead SJ, Ali S (2010) Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin* 96:5–21

81. The EuroQoL group (1990) EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 16:199–208
82. Connell J, O’Cathain A, Brazier J (2014) Measuring quality of life in mental health: are we asking the right questions? *Soc Sci Med* 120:12–20
83. Brazier J, Connell J, Papaioannou D, et al (2014) A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess* 18:vii–viii, xiii–xxv, 1–188
84. Benedetto V, Filipe L, Harris C, Tahir N, Doherty A, Clegg A (2023) Outcome measures for economic evaluations and cost-effectiveness analyses of interventions for people with intellectual disabilities: A methodological systematic review. *J Appl Res Intellect Disabil* 36:230–240
85. Brazier JE, Tsuchiya A, Roberts J, Busschbach J (2004) A comparison of the EQ-5D and the SF-6D across seven patient groups. *Health Economics* 13:873–884
86. Bauer MS, Kirchner J (2020) Implementation science: What is it and why should I care? *Psychiatry Res* 283:112376
87. Hulscher M, Wensing M (2020) Process Evaluation of Implementation Strategies. In: *Improving Patient Care*. John Wiley & Sons, Ltd, pp 369–387
88. Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N (2005) Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of Clinical Epidemiology* 58:107–112
89. Nilsen P (2015) Making sense of implementation theories, models and frameworks. *Implementation Sci* 10:53
90. Birken SA, Powell BJ, Shea CM, Haines ER, Alexis Kirk M, Leeman J, Rohweder C, Damschroder L, Pesseau J (2017) Criteria for selecting implementation science theories and frameworks: results from an international survey. *Implementation Sci* 12:124

91. Damschroder L, Aron D, Keith R, Kirsh S, Alexander J, Lowery J (2009) Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Sci* 4:50
92. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M (2011) Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda. *Adm Policy Ment Health* 38:65–76
93. Proctor EK, Bunger AC, Lengnick-Hall R, Gerke DR, Martin JK, Phillips RJ, Swanson JC (2023) Ten years of implementation outcomes research: a scoping review. *Implementation Sci* 18:31
94. Wensing M, Grol R (2019) Knowledge translation in health: how implementation science could contribute more. *BMC Med* 17:88
95. Moore J, Khan S (2024) A critique of implementation science. *Implementation Science*
96. Morris ZS, Wooding S, Grant J (2011) The answer is 17 years, what is the question: understanding time lags in translational research. *J R Soc Med* 104:510–520
97. Vignero L, Wenmackers S (2021) Degrees of riskiness, falsifiability, and truthlikeness. *Synthese* 199:11729–11764
98. Bonell C, Moore G, Warren E, Moore L (2018) Are randomised controlled trials positivist? Reviewing the social science and philosophy literature to assess positivist tendencies of trials of social interventions in public health and health services. *Trials* 19:238
99. Trocmé-Nadal L (2022) Epistemological Considerations about Big Data and Prediction in Ecology. *Proceedings* 81:86
100. Dehue T (2001) Establishing the Experimenting Society: The Historical Origin of Social Experimentation According to the Randomized Controlled Design. *The American Journal of Psychology* 114:283–302
101. Campbell DT (1991) Methods for the experimenting society. *Evaluation Practice* 12:223–260

102. Pawson R (2006) *Evidence-Based Policy: A Realist Perspective*, 1st edition. SAGE Publications Ltd, London ; Thousand Oaks, Calif
103. Coryn CLS, Noakes LA, Westine CD, Schröter DC (2011) A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. *American Journal of Evaluation* 32:199–226
104. Bryan CJ, Tipton E, Yeager DS (2021) Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nat Hum Behav* 5:980–989
105. Joyce K, Cartwright N (2019) Bridging the Gap Between Research and Practice: Predicting What Will Work Locally. *American Educational Research Journal* 57:000283121986668
106. Johnston R, Smith S (2010) How critical realism clarifies validity issues in theory-testing research: analysis and case. *Information Systems Foundations: The Role of Design Science*
107. Eriksson BG, Karlsson P-Å (2022) *Att utvärdera välfärdsarbete*. Gleerups, Malmö
108. Hawe P, Shiell A, Riley T (2009) Theorising Interventions as Events in Systems. *American Journal of Community Psychology* 43:267–276
109. Porter S, McConnell T, Reid J (2017) The possibility of critical realist randomised controlled trials. *Trials* 18:133
110. Bitter N, Roeg D, van Assen M, van Nieuwenhuizen C, van Weeghel J (2017) How effective is the comprehensive approach to rehabilitation (CARE) methodology? A cluster randomized controlled trial. *BMC Psychiatry* 17:396
111. Slade M, Bird V, Clarke E, Boutillier CL, McCrone P, Macpherson R, Pesola F, Wallace G, Williams J, Leamy M (2015) Supporting recovery in patients with psychosis through care by community-based adult mental health teams (REFOCUS): a multisite, cluster, randomised, controlled trial. *The Lancet Psychiatry* 2:503–514
112. Meadows G, Brophy L, Shawyer F, Enticott JC, Fossey E, Thornton CD, Weller PJ, Wilson-Evered E, Edan V, Slade M (2019) REFOCUS-PULSAR recovery-oriented practice training in specialist mental health care: a stepped-wedge cluster randomised controlled trial. *Lancet Psychiatry* 6:103–114

113. McPherson P, Lloyd-Evans B, Dalton-Locke C, Killaspy H (2021) A Systematic Review of the Characteristics and Efficacy of Recovery Training for Mental Health Staff: Implications for Supported Accommodation Services. *Frontiers in Psychiatry*. <https://doi.org/10.3389/fpsy.2021.624081>
114. Deacon B, McKay D (2015) The biomedical model of psychological problems: A call for critical dialogue. *Behavior Therapy*
115. Van Eck RM, Burger TJ, Vellinga A, Schirmbeck F, de Haan L (2018) The Relationship Between Clinical and Personal Recovery in Patients With Schizophrenia Spectrum Disorders: A Systematic Review and Meta-analysis. *Schizophrenia Bulletin* 44:631–642
116. Recovery. *Oxford English Dictionary*
117. Thornton T, Lucas P (2011) On the very idea of a recovery model for mental health. *J Med Ethics* 37:24–28
118. Skar-Froding R, Clausen HK, Saltyte Benth J, Ruud T, Slade M, Sverdvik Heiervang K (2021) The Importance of Personal Recovery and Perceived Recovery Support Among Service Users With Psychosis. *Psychiatr Serv* 72:661–668
119. Anthony WA (1993) Recovery from mental illness: The guiding vision of the mental health service system in the 1990s. *Psychosocial Rehabilitation Journal* 16:11–23
120. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M (2011) Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *Br J Psychiatry* 199:445–452
121. Speyer H, Roe D, Slade M (2025) Recovery-oriented psychiatry: oxymoron or catalyst for change? *The Lancet Psychiatry* 12:795–802
122. Leonhardt BL, Huling K, Hamm JA, Roe D, Hasson-Ohayon I, McLeod HJ, Lysaker PH (2017) Recovery and serious mental illness: a review of current clinical and research paradigms and future directions. *Expert Rev Neurother* 17:1117–1130
123. Bejerholm U, Eklund M (2004) Time Use and Occupational Performance Among Persons with Schizophrenia. *Occupational Therapy in Mental Health* 20:27–47

124. Aguiar R, Lindström M (2024) Stories of taking part in Everyday Life Rehabilitation - A narrative inquiry of residents with serious mental illness and their recovery pathway. *AIMS Public Health* 11:1198–1222
125. Nordaunet OM, Sælør KT (2018) How meaningful activities influence the recovery process. *Advances in Dual Diagnosis* 11:114–125
126. Fakhoury WKH, Priebe S, Quraishi M (2005) Goals of New Long-Stay Patients in Supported Housing: A UK Study. *Int J Soc Psychiatry* 51:45–54
127. Borg M, Davidson L (2008) The nature of recovery as lived in everyday experience. *Journal of Mental Health* 17:129–140
128. Topor A, Bøe TD, Larsen IB (2018) Small Things, Micro-Affirmations and Helpful Professionals Everyday Recovery-Orientated Practices According to Persons with Mental Health Problems. *Community Ment Health J* 54:1212–1220
129. Mezzich JE (2023) The Place of Person-Centeredness in Public Mental Health. *World Social Psychiatry* 5:116
130. Socialstyrelsen (2025) Uppföljning av omställningen till en god och nära vård.
131. Petrusson U, Hermansson C, Borkmann A (2025) Hälso- och sjukvård som norm: implementering av ”patientens rätt” och möjliggörande av personcentrerad hälso- och sjukvård. *Acta Universitatis Gothoburgensis*
132. Balint E (1969) The possibilities of patient-centered medicine. *The Journal of the Royal College of General Practitioners* 17:269
133. Ekman I, Swedberg K, Taft C, et al (2011) Person-Centered Care — Ready for Prime Time. *European Journal of Cardiovascular Nursing* 10:248–251
134. Santana MJ, Manalili K, Jolley RJ, Zelinsky S, Quan H, Lu M (2017) How to practice person-centred care: A conceptual framework. *Health Expectations: An International Journal of Public Participation in Health Care and Health Policy* 21:429

135. Beadle-Brown J, Mansell J, Ashman B, Ockenden J, Iles R, Whelton B (2014) Practice leadership and active support in residential services for people with intellectual disabilities: an exploratory study. *J Intellect Disabil Res* 58:838–850
136. Dustin D (2007) *The McDonaldization of social work*. Ashgate, Aldershot ; Burlington, VT
137. Viron MJ, Stern TA (2010) The Impact of Serious Mental Illness on Health and Healthcare. *Psychosomatics* 51:458–465
138. Lund C, Brooke-Sumner C, Baingana F, et al (2018) Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. *The Lancet Psychiatry* 5:357–369
139. Lauanders N, Dotsikas K, Marston L, Price G, Osborn DPJ, Hayes JF (2022) The impact of comorbid severe mental illness and common chronic physical health conditions on hospitalisation: A systematic review and meta-analysis. *PLOS ONE* 17:e0272498
140. Arias D, Saxena S, Verguet S (2022) Quantifying the global burden of mental disorders and their economic value. *eClinicalMedicine*. <https://doi.org/10.1016/j.eclinm.2022.101675>
141. GBD 2019 Mental Disorders Collaborators (2022) Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry* 9:137–150
142. Vigo D, Thornicroft G, Atun R (2016) Estimating the true global burden of mental illness. *Lancet Psychiatry* 3:171–178
143. Johnson DL (1997) Overview of severe mental illness. *Clinical Psychology Review* 17:247–257
144. Gonzales L, Kois LE, Chen C, López-Aybar L, McCullough B, McLaughlin KJ (2022) Reliability of the Term “Serious Mental Illness”: A Systematic Review. *Psychiatr Serv* 73:1255–1262
145. Newson JJ, Pastukh V, Thiagarajan TC (2021) Poor Separation of Clinical Symptom Profiles by DSM-5 Disorder Criteria. *Front Psychiatry*. <https://doi.org/10.3389/fpsy.2021.775762>

146. Martínez-Martínez C, Richart-Martínez M, Ramos-Pichardo JD (2020) Operational Definition of Serious Mental Illness: Heterogeneity in a Review of the Research on Quality-of-Life Interventions. *J Am Psychiatr Nurses Assoc* 26:229–244
147. Fekih-Romdhane F, Hajje R, Haddad C, Hallit S, Azar J (2023) Exploring negative symptoms heterogeneity in patients diagnosed with schizophrenia and schizoaffective disorder using cluster analysis. *BMC Psychiatry* 23:595
148. Solmi M, Radua J, Olivola M, et al (2022) Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol Psychiatry* 27:281–295
149. Whitley R, Palmer V, Gunn J (2015) Recovery from severe mental illness. *CMAJ* 187:951–952
150. Molstrom I-M, Nordgaard J, Urfer-Parnas A, Handest R, Berge J, Henriksen MG (2022) The prognosis of schizophrenia: A systematic review and meta-analysis with meta-regression of 20-year follow-up studies. *Schizophrenia Research* 250:152–163
151. Socialstyrelsen (2022) Nationell utvärdering av vård och stöd vid schizofreni och schizofreniliknande tillstånd – 2022 – Huvudrapport med förbättringsområden.
152. Sells D, Borg M, Marin I, Mezzina R, Topor A, Davidson L (2006) Arenas of Recovery for Persons with Severe Mental Illness. *American Journal of Psychiatric Rehabilitation* 9:3–16
153. Knutsson G, Pettersson S (1992) Psykiskt störda - välfärd och valfrihet : en översikt av Psykiatriutredningen. Publica :, Stockholm
154. Wetterberg L (2012) History of psychiatry in Sweden during a millennium. *Nordic Journal of Psychiatry* 66:42–53
155. Börjesson U, Skillmark M, Bülow PH, Bülow P, Vejkliint M, Wilińska M (2021) “It’s about Living Like Everyone Else”: Dichotomies of Housing Support in Swedish Mental Health Care. *Social Inclusion* 9:276–285
156. The Public Health Agency of Sweden (2023) Psykisk hälsa och suicid i Sverige – Statistik om nuläge och utveckling fram till 2022.

157. Socialstyrelsen (2023) Insatser och stöd till personer med funktionsnedsättning: lägesrapport 2023. Socialstyrelsen, Stockholm
158. Lilliehorn S, Fjellfeldt M, Högström E, Markström U (2023) Contemporary Accommodation Services for People with Psychiatric Disabilities – the Simple Taxonomy for Supported Accommodation (STAX-SA) Applied and Discussed in a Swedish Context. *Scandinavian Journal of Disability Research* 25:92–105
159. Killaspy H, Marston L, Green N, et al (2016) Clinical outcomes and costs for people with complex psychosis; a naturalistic prospective cohort study of mental health rehabilitation service users in England. *BMC Psychiatry* 16:95
160. Tiainen A, Rehnberg C (2010) The economic burden of psychiatric disorders in Sweden. *Int J Soc Psychiatry* 56:515–526
161. Statistikmyndigheten SCB (2023) Kommunernas kostnader för vård och omsorg för äldre och för personer med funktionsnedsättning år 2018–2022. In: Statistikmyndigheten SCB. <https://www.scb.se/hitta-statistik/statistik-efter-amne/offentlig-ekonomi/finanser-for-den-kommunala-sektorn/rakenskapssammandrag-for-kommuner-och-regioner/pong/tabell-och-diagram/kommun--och-regionsektorn-2022/kommunernas-kostnader-for-var-d-och-omsorg-for-aldre-och-for-personer-med-funktionsnedsattning-ar-20182022/>. Accessed 19 Mar 2025
162. Brolin R, Rask M, Syrén S, Baigi A, Brunt DA (2015) Satisfaction with housing and housing support for people with psychiatric disabilities. *Issues Ment Health Nurs* 36:21–28
163. Eklund M, Argentzell E, Bejerholm U, Tjörnstrand C, Brunt D (2017) Wellbeing, activity and housing satisfaction – comparing residents with psychiatric disabilities in supported housing and ordinary housing with support. *BMC Psychiatry* 17:315
164. Brolin R, Brunt D, Rask M, Syrén S, Sandgren A (2016) Striving for meaning—Life in supported housing for people with psychiatric disabilities. *Int J Qual Stud Health Well-being* 11:10.3402/qhw.v11.31249
165. Bengtsson-Tops A, Ericsson U, Ehliasson K (2014) Living in supportive housing for people with serious mental illness: A

paradoxical everyday life. *International Journal of Mental Health Nursing* 23:409–418

166. Petersen K, Hounsgaard L, Borg T, Nielsen CV (2012) User involvement in mental health rehabilitation: a struggle for self-determination and recognition. *Scand J Occup Ther* 19:59–67
167. Tjörnstrand C, Eklund M, Bejerholm U, Argentzell E, Brunt D (2020) A day in the life of people with severe mental illness living in supported housing. *BMC Psychiatry* 20:508
168. Kearns Murphy C, Shiel A (2019) Institutional injustices? Exploring engagement in occupations in a residential mental health facility. *Journal of Occupational Science* 26:115–127
169. Caple V, Maude P, Walter R, Ross A (2023) An exploration of loneliness experienced by people living with mental illness and the impact on their recovery journey: An integrative review. *Journal of Psychiatric and Mental Health Nursing* 30:1170–1191
170. Vancampfort D, Firth J, Schuch FB, et al (2017) Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta-analysis. *World Psychiatry* 16:308–315
171. Argentzell E, Tjörnstrand C, Brunt D, Eklund M, Bejerholm U (2023) Opportunities and barriers for occupational engagement among residents in supported housing. *Scand J Occup Ther* 30:125–135
172. Markström U, Högström E, Fjellfeldt M (2023) Mental health supported accommodation services in a post-deinstitutionalised era. *Alter European Journal of Disability Research* 39–56
173. Socialstyrelsen (2021) *Kompetens i LSS-boenden*.
174. Nordentoft M, Pedersen MG, Pedersen CB, Blinkenberg S, Mortensen PB (2012) The new asylums in the community: severely ill psychiatric patients living in psychiatric supported housing facilities. A Danish register-based study of prognostic factors, use of psychiatric services, and mortality. *Soc Psychiatry Psychiatr Epidemiol* 47:1251–1261

175. Ericsson U, Bengtsson-Tops A (2014) Housing support workers as equilibrists between instrumentality and situation. *Vulnerable Groups & Inclusion* 5:23755
176. Carlsson KS, Brommels M (2022) Integrated Health and Social Services for People With Chronic Mental Health Problems: People Are More Important Than Processes. Insights From a Multiple Case Study in Swedish Psychiatry. *Front Public Health* 10:845201
177. Bångsbo A, Dunér A, Dahlin-Ivanoff S, Lidén E (2022) Barriers for Inter-Organisational Collaboration: What Matters for an Integrated Care Programme? *Int J Integr Care* 22:22
178. Brodin R, Tjörnstrand C, Friis M, Argentzell E, Bejerholm U, Eklund M, Brunt D (2024) “It’s not just the residents who need to be motivated for activity”: a qualitative study of the perspectives of staff on providing activity support for people with psychiatric disabilities in supported housing in Sweden. *Frontiers in Psychiatry* 14:
179. Socialstyrelsen (2025) Statistikdatabas för kommunal hälso- och sjukvård (KVÅ).
180. Socialstyrelsen (2023) Hälso- och sjukvård i hemmet – Kunskapsstöd för personcentrerad vård och rehabilitering.
181. Malm U, Jacobsson L, Larsson NO (2002) Psychiatry in Sweden: Values, System, and Evidence for a Reform in Progress. *International Journal of Mental Health* 31:50–65
182. European Commission (2013) EuroPoPP-MH report: European profile of prevention and promotion of mental health - European Commission.
183. Ahgren B, Axelsson R (2011) A decade of integration and collaboration: the development of integrated health care in Sweden 2000–2010. *Int J Integr Care* 11:e007
184. Sacchetti L, Lindahl G (2024) Understanding Healthcare Design Transformations. Insights from the Swedish Experience. In: *Studies in health technology and informatics*. pp 237–249
185. SNS – Center for Business and Policy Studies (2021) SNS Research Brief 76. Hospital bed capacity and ED crowding – The impact on patient safety - SNS English.

<https://www.sns.se/en/publications/sns-research-brief-76-hospital-bed-capacity-and-ed-crowding-the-impact-on-patient-safety>. Accessed 21 Jan 2026

186. Elmer D, Endrei D, Németh N, Horváth L, Pónusz R, Kívés Z, Danku N, Csákvári T, Ágoston I, Boncz I (2022) Changes in the Number of Physicians and Hospital Bed Capacity in Europe. *Value in Health Regional Issues* 32:102–108
187. Myndigheten för Vårdanalys (2025) Omtag för omställning. Utvärdering av omställningen till en god och nära vård: slutrapport.
188. SFS 2001:453 Socialtjänstlag.
189. SFS 1993:387 Lag (1993:387) om stöd och service till vissa funktionshindrade.
190. Ehliasson K, Markström U (2020) Revealing the Ideas in the Swedish Social Services Act Regarding Support to Individuals with Disabilities. *Scandinavian Journal of Disability Research* 22:393–402
191. Janlöv N, Blume S, Glenngård AH, Hanspers K, Anell A, Merkur S (2023) Sweden: Health System Review. *Health Syst Transit* 25:1–236
192. SFS (2017:30) Health and Medical Services Act.
193. European Commission (2017) Towards Community-based and socially inclusive mental health care - European Commission.
194. Wilson NJ, Pracilio A, Morphet J, Kersten M, Buckley T, Trollor JN, Cashin A (2023) A scoping review of registered nurses' delegating care and support to unlicensed care and support workers. *Journal of Clinical Nursing* 32:6000–6011
195. Axelsson J, Elmståhl S (2004) Home care aides in the administration of medication. *Int J Qual Health Care* 16:237–243
196. Gransjön Craftman A, Hammar LM, von Strauss E, Hillerås P, Westerbotn M (2015) Unlicensed personnel administering medications to older persons living at home: a challenge for social and care services. *International Journal of Older People Nursing* 10:201–210

197. Socialstyrelsen (2025) Kunskapsstöd för egenvård – enligt egenvårdslagen – Stödet riktar sig till hälso- och sjukvård, socialtjänst och LSS-verksamheter.
198. Skivington K, Matthews L, Simpson SA, et al (2021) A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ* 374:n2061
199. Torgerson DJ (2001) Contamination in trials: is cluster randomisation the answer? *BMJ* 322:355–357
200. Williams HC, Burden-Teh E, Nunn AJ (2015) What Is a Pragmatic Clinical Trial? *Journal of Investigative Dermatology* 135:1–3
201. Thorpe KE, Zwarenstein M, Oxman AD, et al (2009) A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Journal of Clinical Epidemiology* 62:464–475
202. Bond C, Lancaster GA, Campbell M, Chan C, Eddy S, Hopewell S, Mellor K, Thabane L, Eldridge S (2023) Pilot and feasibility studies: extending the conceptual framework. *Pilot Feasibility Stud* 9:24
203. Keetharuth A (2017) How to interpret ReQoL-scores v1.0. <https://innovation.ox.ac.uk/wp-content/uploads/2016/08/How-to-interpret-ReQoL-scores-Clinicians-v1.0-130917.pdf>. Accessed 25 May 2023
204. Keetharuth AD, Brazier J, Connell J, et al (2018) Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties†. *The British Journal of Psychiatry* 212:42–49
205. Keetharuth A (2016) ReQoL-20 Scoring Guide. [https://www.sheffield.ac.uk/sites/default/files/2022-04/ReQoL20\\_Scoring\\_Guide\\_v0.3.pdf](https://www.sheffield.ac.uk/sites/default/files/2022-04/ReQoL20_Scoring_Guide_v0.3.pdf). Accessed 25 May 2023
206. Granholm Valmari E, Melander M, Hariz G-M, Naesström M, Lindström M (2023) Translation and linguistic validation of the Swedish recovering quality of life (ReQoL) - A brief research report. *Front Psychiatry* 14:1059406
207. Hancock N, Scanlan JN, Bundy AC, Honey A (2016) Recovery Assessment Scale - Domains and Stages (RAS-DS) Manual - Version 2.

208. Honey A, Hancock N, Scanlan JN (2023) Staff perceptions of factors affecting the use of RAS-DS to support collaborative mental health practice. *BMC Psychiatry* 23:500
209. Hancock N, Scanlan JN, Honey A, Bundy AC, O'Shea K (2015) Recovery Assessment Scale – Domains and Stages (RAS-DS): Its feasibility and outcome measurement capacity. *Aust N Z J Psychiatry* 49:624–633
210. Kiresuk TJ, Sherman RE (1968) Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Ment Health J* 4:443–453
211. Turner-Stokes L (2009) Goal attainment scaling (GAS) in rehabilitation: a practical guide. *Clin Rehabil* 23:362–370
212. Carswell A, McColl MA, Baptiste S, Law M, Polatajko H, Pollock N (2004) The Canadian Occupational Performance Measure: A Research and Clinical Literature Review. *Can J Occup Ther* 71:210–222
213. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67:1–48
214. Buuren S van, Groothuis-Oudshoorn K, Vink G, et al (2025) mice: Multivariate Imputation by Chained Equations.
215. Kolada (2022) Nya tillsvidareanställda månadsavlönade, kommun, andel (%).
216. European Central Bank (2025) Euro exchange rates charts.
217. Statistikmyndigheten (SCB) (2025) Konsumentprisindex (1980=100), fastställda tal.
218. Statistikmyndigheten (SCB) (2023) Lönestatistik.
219. Verbooy K, Hoefman R, Exel J van, Brouwer W (2018) Time Is Money: Investigating the Value of Leisure Time and Unpaid Work. *Value in Health* 21:1428–1436
220. Krol M, Brouwer W (2015) Unpaid work in health economic evaluations. *Soc Sci Med* 144:127–137

221. Hagberg L, Lundqvist S, Lindholm L (2020) What is the time cost of exercise? Cost of time spent on exercise in a primary health care intervention to increase physical activity. *Cost Eff Resour Alloc* 18:14
222. Keetharuth AD, Rowen D, Bjorner JB, Brazier J (2021) Estimating a Preference-Based Index for Mental Health From the Recovering Quality of Life Measure: Valuation of Recovering Quality of Life Utility Index. *Value in Health* 24:281–290
223. Hunter RM, Baio G, Butt T, Morris S, Round J, Freemantle N (2015) An educational review of the statistical issues in analysing utility data for cost-utility analysis. *Pharmacoeconomics* 33:355–366
224. Manca A, Hawkins N, Sculpher MJ (2005) Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 14:487–496
225. Lenth RV, Banfai B, Bolker B, et al (2025) emmeans: Estimated Marginal Means, aka Least-Squares Means.
226. Svensson M, Nilsson FOL, Arnberg K (2015) Reimbursement Decisions for Pharmaceuticals in Sweden: The Impact of Disease Severity and Cost Effectiveness. *PharmacoEconomics* 33:1229–1236
227. Socialstyrelsen (2018) Nationella riktlinjer för vård och stöd vid schizofreni och schizofreniliknande tillstånd – Stöd för styrning och ledning.
228. Paulden M (2020) Calculating and Interpreting ICERs and Net Benefit. *PharmacoEconomics* 38:785–807
229. Siverskog J, Henriksson M (2021) On the role of cost-effectiveness thresholds in healthcare priority setting. *Int J Technol Assess Health Care* 37:e23
230. Rodrik D (2015) *Economics Rules: The Rights and Wrongs of the Dismal Science*, First Edition. W. W. Norton & Company, New York
231. Sjöberg A (2025) Att navigera i okända vatten med hjälp av hälsoekonomiska modeller. *SMT* 102:149–156

232. Morey RD, Hoekstra R, Rouder JN, Lee MD, Wagenmakers E-J (2016) The fallacy of placing confidence in confidence intervals. *Psychon Bull Rev* 23:103–123
233. Baio G, Dawid AP (2015) Probabilistic sensitivity analysis in health economics. *Stat Methods Med Res* 24:615–634
234. Norra sjukvårdsregionsförbundet (2025) Regional prislista för Norra sjukvårdsregionens länssjukvård och primärvård.
235. Emmel N (2013) Sampling and Choosing Cases in Qualitative Research: A Realist Approach. <https://doi.org/10.4135/9781473913882>
236. Manzano A (2016) The craft of interviewing in realist evaluation. *Evaluation* 22:342–360
237. Breimaier HE, Heckemann B, Halfens RJG, Lohrmann C (2015) The Consolidated Framework for Implementation Research (CFIR): a useful theoretical framework for guiding and evaluating a guideline implementation process in a hospital-based nursing practice. *BMC Nurs* 14:43
238. Brönnimann A (2022) How to phrase critical realist interview questions in applied social science research. *Journal of Critical Realism* 21:1–24
239. Fairclough N (2013) A dialectical–relational approach to critical discourse analysis in social research. *Critical Discourse Analysis: The Critical Study of Language*. <https://doi.org/10.4324/9781315834368>
240. Fairclough N (2005) Peripheral Vision: Discourse Analysis in Organization Studies: The Case for Critical Realism. *Organization Studies* 26:915–939
241. Fairclough N (2013) Critical discourse analysis, organisational discourse and organisational change. *Critical Discourse Analysis: The Critical Study of Language*. <https://doi.org/10.4324/9781315834368>
242. Wodak R, Meyer M (2016) Critical discourse studies: history, agenda, theory and methodology. *Methods of Critical Discourse Studies*

243. Bosch R (2025) Philosophy of Science for Qualitative Research</span>. <https://doi.org/10.2139/ssrn.5532118>
244. Fairclough N (2013) *Critical Discourse Analysis: The Critical Study of Language*, 2nd ed. <https://doi.org/10.4324/9781315834368>
245. Newman J (2020) Critical realism, critical discourse analysis, and the morphogenetic approach. *Journal of Critical Realism* 19:433–455
246. Regmi KD (2017) Critical Discourse Analysis: Exploring its Philosophical Underpinnings. *Méthod(e)s: African Review of Social Sciences Methodology* 2:93–107
247. Van Dijk TA (2015) Critical Discourse Analysis. In: *The Handbook of Discourse Analysis*. John Wiley & Sons, Ltd, pp 466–485
248. Oliver DG, Serovich JM, Mason TL (2005) Constraints and Opportunities with Interview Transcription: Towards Reflection in Qualitative Research. *Soc Forces* 84:1273–1289
249. Crinson I (2001) A Realist Approach to the analysis of Focus Group Data.
250. Gupta UC, Kharawala S (2012) Informed consent in psychiatry clinical research: A conceptual review of issues, challenges, and recommendations. *Perspect Clin Res* 3:8–15
251. Jain S, Kuppili PP, Pattanayak RD, Sagar R (2017) Ethics in Psychiatric Research: Issues and Recommendations. *Indian J Psychol Med* 39:558–565
252. Amer AB (2013) Informed Consent in Adult Psychiatry. *Oman Medical Journal* 28:228
253. Lidz CW, Appelbaum PS (2002) The therapeutic misconception: problems and solutions. *Med Care* 40:V55-63
254. Fryer T (2024) A short guide to ontology and epistemology: Why everyone should be a realist.
255. Feyerabend P (2010) *Against Method: Outline of an Anarchistic Theory of Knowledge*, 4th edition. Verso Books, London

256. Elder-Vass D (2022) Pragmatism, critical realism and the study of value. *Journal of Critical Realism* 21:261–287
257. Farasoo A (2024) The question of epistemic fallacy in practical research: the case of IR. *Journal of Critical Realism* 23:121–143
258. Collier A (1994) *Critical Realism: An Introduction to Roy Bhaskar's Philosophy*. Verso Books, London New York
259. Elder-Vass D, Fryer T, Groff RP, Navarrete C, Nellhaus T (2023) Does critical realism need the concept of three domains of reality? A roundtable. *Journal of Critical Realism* 22:222–239
260. Gorski P (2013) "What is Critical Realism? And Why Should You Care?" *Contemp Sociol* 42:658–670
261. van Ingen M (2021) Sublating the Naturalism/Anti-Naturalism Problematic: Critical Realism, Critical Naturalism, and the Question of Methodology. *Int Stud Rev* 23:835–861
262. Porpora D (2023) Do Realists Predict? *Journal for the Theory of Social Behaviour*. <https://doi.org/10.1111/jtsb.12404>
263. Porpora DV (ed) (2015) Do realists run regressions? In: *Reconstructing Sociology: The Critical Realist Approach*. Cambridge University Press, Cambridge, pp 31–64
264. Hacking I (1983) Representing and Intervening: Introductory Topics in the Philosophy of Natural Science. <https://doi.org/10.1017/CBO9780511814563>
265. Mayo DG (2018) Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars. <https://doi.org/10.1017/9781107286184>
266. Mayo DG (1994) The New Experimentalism, Topical Hypotheses, and Learning From Error. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1994:270–279
267. (2011) Toward Progressive Critical Rationalism - Exchanges with Alan Musgrave. *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*

268. Siverskog J, Henriksson M (2019) Estimating the marginal cost of a life year in Sweden's public healthcare sector. *European Journal of Health Economics* 20:751–762
269. Chisholm J, Hope J, Fossey E, Petrakis M (2026) Recovery-Oriented Practice in Public Mental Health Services: A Systematic Scoping Review of Multidisciplinary Clinicians' Views on Its Implementation. *International Journal of Mental Health Nursing* 35:e70188
270. Lorien L, Blunden S, Madsen W (2020) Implementation of recovery-oriented practice in hospital-based mental health services: A systematic review. *International Journal of Mental Health Nursing* 29:1035–1048
271. Piat M, Wainwright M, Sofouli E, Vachon B, Deslauriers T, Prefontaine C, Frati F (2021) Factors influencing the implementation of mental health recovery into services: a systematic mixed studies review. *Syst rev* 10:134
272. Lau R, Stevenson F, Ong BN, et al (2015) Achieving change in primary care—effectiveness of strategies for improving implementation of complex interventions: systematic review of reviews. *BMJ Open* 5:e009993
273. Heister N, Zentel P, Köb S (2023) Participation in Everyday Leisure and Its Influencing Factors for People with Intellectual Disabilities: A Scoping Review of the Empirical Findings. *Disabilities* 3:269–294
274. Santos WJ, Graham ID, Lalonde M, Demery Varin M, Squires JE (2022) The effectiveness of champions in implementing innovations in health care: a systematic review. *Implement Sci Commun* 3:80
275. Bitter N, Roeg D, van Nieuwenhuizen C, van Weeghel J (2020) Recovery in Supported Accommodations: A Scoping Review and Synthesis of Interventions for People with Severe Mental Illness. *Community Ment Health J* 56:1053–1076
276. Slettebø Å, Hellem E, Bruusgaard K, Madsen V, Alve G, Langhammer B (2012) Between power and powerlessness – discourses in the individual plan processes, a Norwegian dilemma. *Scandinavian Journal of Disability Research* 14:270–283

277. Iwasaki Y, Coyle CP, Shank JW (2010) Leisure as a context for active living, recovery, health and life quality for persons with mental illness in a global context. *Health Promot Int* 25:483–494
278. Iso-Ahola SE, Baumeister RF (2023) Leisure and meaning in life. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2023.1074649>
279. Fancourt D, Aughterson H, Finn S, Walker E, Steptoe A (2021) How leisure activities affect health: a narrative review and multi-level theoretical framework of mechanisms of action. *Lancet Psychiatry* 8:329–339
280. Ngamaba KH, Webber M, Xanthopoulou P, Chevalier A, Giacco D (2023) Participation in leisure activities and quality of life of people with psychosis in England: a multi-site cross-sectional study. *Ann Gen Psychiatry* 22:8
281. Tondora J, Miller R, Slade M, Davidson L (2014) *Partnering for Recovery in Mental Health: A Practical Guide to Person-Centered Planning*. John Wiley & Sons, Incorporated, Newark, UNITED KINGDOM
282. Fedak KM, Bernal A, Capshaw ZA, Gross S (2015) Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerg Themes Epidemiol* 12:14
283. Giovanazzi A, Jones K, Carr RM, Fairhurst CM, Backhouse MR, Adamson JA (2022) Current practice in the measurement and interpretation of intervention adherence in randomised controlled trials: A systematic review. *Contemporary Clinical Trials* 118:106788
284. Berkhout C, Berbra O, Favre J, Collins C, Calafiore M, Peremans L, Van Royen P (2022) Defining and evaluating the Hawthorne effect in primary care, a systematic review and meta-analysis. *Front Med*. <https://doi.org/10.3389/fmed.2022.1033486>
285. Tunney RJ, Ziegler FV (2015) Surrogate utility estimation by long-term partners and unfamiliar dyads. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2015.00315>
286. Andresen EM, Vahle VJ, Lollar D (2001) Proxy reliability: Health-related quality of life (HRQoL) measures for people with disability. *Qual Life Res* 10:609–619

287. Manuel J, Crowe M (2014) Clinical responsibility, accountability, and risk aversion in mental health nursing: a descriptive, qualitative study. *Int J Ment Health Nurs* 23:336–343
288. Martinelli A, Ruggeri M (2020) An overview of mental health recovery-oriented practices: potentiality, challenges, prejudices, and misunderstandings. *Official Journal of the Italian Society of Psychopathology* 26:
289. Gee M, Bhanbhro S, Cook S, Killaspy H (2017) Rapid realist review of the evidence: achieving lasting change when mental health rehabilitation staff undertake recovery-oriented training. *Journal of Advanced Nursing* 73:1775–1791
290. Maddock A (2015) Consensus or contention: an exploration of multidisciplinary team functioning in an Irish mental health context. *European Journal of Social Work* 18:246–261
291. Le Boutillier C, Slade M, Lawrence V, et al (2015) Competing priorities: staff perspectives on supporting recovery. *Adm Policy Ment Health* 42:429–438
292. Iversen HW, Riley H, Råbu M, Lorem GF (2025) Building and sustaining therapeutic relationships across treatment settings: a qualitative study of how patients navigate the group dynamics of mental healthcare. *BMC Psychiatry* 25:424
293. Littlechild B, Smith R (2013) *A Handbook for Interprofessional Practice in the Human Services: Learning to Work Together*. Routledge
294. Beckett P (2025) What works in the implementation of a strengths model training program: a realist evaluation. <https://doi.org/10.25959/27905460.v1>
295. Kadri A, Rapaport P, Livingston G, Cooper C, Robertson S, Higgs P (2018) Care workers, the unacknowledged persons in person-centred care: A secondary qualitative analysis of UK care home staff interviews. *PLOS ONE* 13:e0200031
296. Strömberg A, Engström M, Hagerman H, Skytt B (2019) First-line managers dealing with different management approaches. *Leadersh Health Serv (Bradf Engl)* 32:543–557

297. Rio JHM, Fuller J, Taylor K, Muir-Cochrane E (2021) The impact of workplace culture on the accountability of mental health nurses to involve consumers in care planning: A focused ethnography. *Nursing & Health Sciences* 23:255–262
298. Sandström B, Willman A, Svensson B, Borglin G (2014) ‘How do we know if this is the best?’ Mental health-care professionals’ views on national guidelines for psychosocial interventions. *International Journal of Mental Health Nursing* 23:221–231
299. Jönsson M, Edberg A-K, Wallin AO, Sundström M, Backman A (2025) Person-centred leadership in residential care for older people as described by first-line managers. *BMC Nurs* 24:666
300. Franklin M, Hunter RM, Enrique A, Palacios J, Richards D (2022) Estimating Cost-Effectiveness Using Alternative Preference-Based Scores and Within-Trial Methods: Exploring the Dynamics of the Quality-Adjusted Life-Year Using the EQ-5D 5-Level Version and Recovering Quality of Life Utility Index. *Value in Health* 25:1018–1029
301. Ishida M, Hulse ES, Mahar RK, et al (2020) The Joint Effect of Physical Multimorbidity and Mental Health Conditions Among Adults in Australia. *Prev Chronic Dis* 17:E157
302. O’Loughlin R, Hiscock H, Pan T, Devlin N, Dalziel K (2022) The relationship between physical and mental health multimorbidity and children’s health-related quality of life. *Qual Life Res* 31:2119–2131
303. Proud L, McLoughlin C, Kinghorn P (2019) ICECAP-O, the current state of play: a systematic review of studies reporting the psychometric properties and use of the instrument over the decade since its publication. *Qual Life Res* 28:1429–1439
304. Meili KW, Mulhern B, Ssegonja R, Norström F, Feldman I, Månsdotter A, Hjelte J, Lindholm L (2024) Eliciting a value set for the Swedish Capability-Adjusted Life Years instrument (CALY-SWE). *Qual Life Res* 33:59–72
305. Weinstein MC (1988) A QALY is a QALY--or is it? *J Health Econ* 7:289–290
306. Cameron D, Ubels J, Norström F (2018) On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an

absence of data: a systematic review. *Global Health Action* 11:1447828

307. Henriksson M, Siverskog J, Johannesen K, Eriksson T (2018) Tröskelvården och kostnadseffektivitet: innebörd och implikationer för ekonomiska utvärderingar och beslutsfattande i hälso- och sjukvården. Linköping University Electronic Press
308. Grosse SD (2008) Assessing cost-effectiveness in healthcare: history of the \$50,000 per QALY threshold. *Expert Rev Pharmacoecon Outcomes Res* 8:165–178
309. Hjorthøj C, Stürup AE, McGrath JJ, Nordentoft M (2017) Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. *The Lancet Psychiatry* 4:295–301
310. Laursen TM, Nordentoft M, Mortensen PB (2014) Excess early mortality in schizophrenia. *Annu Rev Clin Psychol* 10:425–448
311. Prioriteringscentrum (2017) Nationell modell för öppna prioriteringar inom hälso- och sjukvård: ett verktyg för rangordning. Linköping University Electronic Press
312. Rosner AL (2012) Evidence-based medicine: Revisiting the pyramid of priorities. *Journal of Bodywork and Movement Therapies* 16:42–49
313. Thomas SJ (2023) Critically Appraising Pragmatist Critiques of Evidence-Based Medicine: Is EBM Defensible on Pragmatist Grounds? *J Med Philos* 48:73–83
314. Goldenberg MJ (2009) Iconoclast or Creed?: Objectivism, Pragmatism, and the Hierarchy of Evidence. *Perspectives in Biology and Medicine* 52:168–187
315. Zotzmann K, Barman E, Porpora DV, Carrigan M, Elder-Vass D (2022) Round table: is the common ground between pragmatism and critical realism more important than the differences? *Journal of Critical Realism* 21:352–364
316. Brinkmann S (2017) British Philosophies of Qualitative Research: Positivism and Realism. In: Brinkmann S (ed) *Philosophies of Qualitative Research*. Oxford University Press, p 0

317. Lawson T (2011) Anti-Realism or Pro-Something Else? *Erasmus Journal for Philosophy and Economics* 4:53–66
318. Chalmers A (2013) *What Is This Thing Called Science?* University of Queensland Press, St Lucia, UNITED STATES
319. Avis M (2005) Is there an epistemology for qualitative research. *Qualitative Research In Health Care*
320. St. Pierre EA (2016) The Empirical and the New Empiricisms. *Cultural Studies ↔ Critical Methodologies* 16:111–124
321. Mayo DG (1996) *Error and the Growth of Experimental Knowledge.* University of Chicago Press, Chicago, IL
322. Van Belle S, Wong G, Westhorp G, Pearson M, Emmel N, Manzano A, Marchal B (2016) Can “realist” randomised controlled trials be genuinely realist? *Trials* 17:313
323. Warren EA, Melendez-Torres GJ, Bonell C (2022) Are realist randomised controlled trials possible? A reflection on the INCLUSIVE evaluation of a whole-school, bullying-prevention intervention. *Trials* 23:82
324. Senn S (2013) Seven myths of randomisation in clinical trials. *Statistics in medicine.* <https://doi.org/10.1002/sim.5713>
325. Bonell C, Warren E, Fletcher A, Viner R (2016) Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al. *Trials* 17:478
326. Griffiths P, Norman I (2013) Qualitative or quantitative? Developing and evaluating complex interventions: time to end the paradigm war. *International Journal of Nursing Studies* 50:583–584
327. Popper K, Popper K (2005) *The Logic of Scientific Discovery*, 2nd ed. <https://doi.org/10.4324/9780203994627>
328. Ochieng PA (2009) AN ANALYSIS OF THE STRENGTHS AND LIMITATION OF QUALITATIVE AND QUANTITATIVE RESEARCH PARADIGMS. *Problems of Education in the 21st Century* 13:

329. Mwita K (2022) Strengths and weaknesses of qualitative research in social science studies. *International Journal of Research in Business and Social Science* (2147-4478) 11:618–625
330. Robert Isaksen K (2016) Reclaiming Rational Theory Choice as Central: A Critique of Methodological Applications of Critical Realism. *Journal of Critical Realism* 15:245–262
331. McEvoy P, Richards D (2006) A critical realist rationale for using a combination of quantitative and qualitative methods. *Journal of Research in Nursing* 11:66–78
332. Maxwell J (2012) *A Realist Approach for Qualitative Research*, 1st edition. SAGE Publications Inc, Thousand Oaks, CA London Neu Delhi Singapur
333. Scott D (2007) Resolving the quantitative–qualitative dilemma: a critical realist approach. *International Journal of Research & Method in Education* 30:3–17
334. Obermann K, Scheppe J, Glazinski B (2013) More Than Figures? Qualitative Research in Health Economics. *Health Economics* 22:253–257
335. Coast J, McDonald R, Baker R (2004) Issues arising from the use of qualitative methods in health economics. *J Health Serv Res Policy* 9:171–176
336. Cartwright N (2012) *Evidence-based policy: a practical guide to doing it better*, Second edition. Oxford University Press, New York
337. Kluzek S, Dean B, Wartolowska KA (2022) Patient-reported outcome measures (PROMs) as proof of treatment efficacy. *BMJ Evidence-Based Medicine* 27:153–155
338. Khan MA, Richardson J (2019) Is the Validity of Cost Utility Analysis Improved When Utility is Measured by an Instrument with ‘Home-Country’ Weights? Evidence from Six Western Countries. *Soc Indic Res* 145:1–15
339. Janiaud P, Hemkens LG (2024) Modern trials are most useful when they are pragmatic and explanatory – there is no continuum. *Journal of Clinical Epidemiology*. <https://doi.org/10.1016/j.jclinepi.2024.111566>

340. Cardoso G, Papoila A, Tomé G, Killaspy H, King M, Caldas-de-Almeida JM (2017) A cluster randomised controlled trial of a staff-training intervention in residential units for people with long-term mental illness in Portugal: the PromQual trial. *Soc Psychiatry Psychiatr Epidemiol* 52:1435–1445
341. Greckhamer T, Cilesiz S (2014) Rigor, Transparency, Evidence, and Representation in Discourse Analysis: Challenges and Recommendations. *International Journal of Qualitative Methods* 13:422–443
342. Tuffaha HW, Gordon LG, Scuffham PA (2014) Value of information analysis in healthcare: a review of principles and applications. *Journal of Medical Economics* 17:377–383
343. Pelto-Piri V, Wetterling J (2024) From Paternalism to Social Inclusion? User Organisations' Narratives of Psychiatric Services in Sweden. Brill Academic Publishers, pp 79–97
344. Engwall K, Jørgensen AS, Katsui H (2026) Transformed Institutionalization. A Literature-Based Study of 'Independent Living' Among Persons with Intellectual Disabilities in Three Nordic Countries. *Scandinavian Journal of Disability Research*. <https://doi.org/10.16993/sjdr.1329>
345. Topor A, Andersson G, Bülow P, Stefansson C-G, Denhov A (2016) After the Asylum? The New Institutional Landscape. *Community Ment Health J* 52:731–737