Doctoral theses at NTNU, 2024:317

Martin Brattmyr

Patient-Reported Outcomes of Common Mental Health Disorders in Clinically Representative Therapy: Treatment, Measurement, and Clinical Subgroups

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Social and Educational Sciences
Department of Psychology

NTNU
Norwegian University of
Science and Technology

Martin Brattmyr

Patient-Reported Outcomes of Common Mental Health Disorders in Clinically Representative Therapy: Treatment, Measurement, and Clinical Subgroups

Thesis for the Degree of Philosophiae Doctor

Trondheim, September 2024

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Social and Educational Sciences
Department of Psychology

ISBN 978-82-326-8226-3 (printed ver.)
ISBN 978-82-326-8225-6 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2024:317

Printed by NTNU Grafisk senter

Table of Contents

Svenskt Sammandrag (Swedish Summary)

Psykiatrisk behandling har flera kunskapsluckor. Vi vet lite om innehållet och om utfallet av ordinär behandling, och den lilla kunskap som finns är ofta inhämtad från metaanalyser av kontrollbetingelsen *treatment as usual*. Men denna har i sin tur ofta brister i kvalitet och stringens, vilket påverkar generaliserbarheten av i stort sett all annan syntetiserad kunskap vi har om psykologisk behandling för vanliga psykiska åkommor.

Istället utvärderas ofta vården av psykiatriska åkommor utifrån kvalitetsindikatorer, systematiskt inhämtat kunskapsunderlag som täcker flera kvalitetsaspekter av vården. Dock kritiseras dagens användning av kvalitetsindikatorer av framförallt två grunder: relevant information inhämtas inte på majoriteten av patienterna och de indikatorer som inhämtas är ofta ämnade för slutenvårdsbehandling. I kontrast behandlas majoriteten av patienterna i öppenvård för depression eller ångestångesstörningar.

Det finns flera sätt att mäta depression och ångest, bland annat som icke direkt observerbara *latenta* variabler utifrån ett antal symptom. De instrument som vanligtvis används behöver dock vara valida för den målgrupp som ska mätas. Vidare är komorbiditeten ofta hög bland patienter med depression och ångest, men åsikterna går isär till hur stor grad, och hur komorbidtet ska mätas. Givet dessa kunskapsluckor så presenteras i denna avhandlingen tre publicerade artiklar med tre olika aspekter av depression och ångest: effekten av ordinär behandling, mätegenskaper av två vanligt förekommande instrument och latenta subgrupper.

I artikel I gjordes en systematisk litteraturstudie och meta-analys på nordisk öppenvårdsbehandling där vi fann 11 studier som ansågs vara kliniskt representativa, med en låg till medelstark inomstudie-effektstorlek (g = 0.49) av behandlingarna. Effektstyrkan verkade vara lägre än evidensbaserade interventioner, men i paritet med andra interventioner

som genomfördes i studierna (g = -0.21, justerat för publikationsbias g = -0.06). Slutsatsen var att nordisk öppenvårdsbehandling karaktäriserades av en stor spridning av patient-, terapeut- och utfallsvariabler.

I artikel II undersöktes faktorstruktur och mätinvarians av Patient Health Questionnaire (PHQ-9) och Generalized Anxiety Scale (GAD-7), samt deras samtidiga validitet med ett instrument som mätte grad av funktionsnedsättning. Då en bi-faktormodell hade goda mätegenskaper drogs slutsatsen att instrumenten framförallt laddade en faktor. Dessutom fann vi att jämförelse mellan flera olika grupper är berättigat, och att depressiva symptom var mer associerade med nedsatt funktion. Därmed kunde vi dra slutsatsen att kvinnliga patienter och de med komorbida tillstånd hade generellt högre symptomtryck och funktionsnedsättning.

I artikel III undersöktes symptomtryck mellan olika diagnosgrupper och latenta subgrupper baserat på symptom. Patienter med komorbida ångest- och depressionsdiagnoser hade högre symptomtryck, men också lägre prevalens än förväntat. En latent normalfördelad faktor-subgruppsanalys av PHQ-9 och GAD-7 visade tre grupper, där 33% av patienterna hamnade i gruppen med mest aggregerade problem. Denna grupp karaktäriserades av höga somatiska symptom av depression och ångest, samt högre sannolikhet att få fler konsultationer och psykiska diagnoser.

Sammanfattningsvis visar vi att flera aspekter av depression och ångest kan vara viktiga kvalitetsindikatorer. Först presenterar vi effekten och innehållet av ordinär skandinavisk öppenvårdsbehandling för vanliga psykiska problem, även om målet var att jämföra hela Norden. Trots att resultatet är heterogent så kan detta fungera som ett riktmärke för vad ordinär behandling innehåller och vad effekten kan tänkas vara på gruppnivå. Vi visar också att PHQ-9 och GAD-7 har goda mätegenskaper vid start av behandling, och kan

användas som kvalitets-mätinstrument för depression och ångest. Vidare visar vi att hög nivå av somatiska depressionssymptom kan vara viktiga indikatorer för patienter som kan vara svåra att behandla. Detta kan ha forskningsimplikationer för andra som vill studera ordinär behandling, men också för de som ska jämföra interventioner med ordinär behandling. Det har även kliniska implikationer: för att skapa en större förståelse för ordinär behandling, för att mäta symptomnivå hos patienter som startar behandling, och för att tidigt kunna identifiera patienter med eventuellt större vårdbehov. Avhandlingen ska dock inte ses som en komplett beskrivning av ordinär behandling, utan mer forskning behövs för att kunna komma vanliga patienter till gagns, som effektstudier med hög kvalitet, process-studier av kliniskt representativa behandlingar, studier som utvärderar PHQ-9 och GAD-7 som diagnostiska instrument, och undersöka om *de facto* tidig identifiering av potententiellt svårbehandlade patienter kan ha klinisk nytta. Målet med avhandlingen är att ge ordinär behandling den plats i forskningsfältet som den förtjänar, för de många med vanligt förekommande psykiska problem som behandlas inom psykiatrisk öppenvård.

English Summary

Psychiatric treatment has several knowledge gaps. We know little about the content and outcomes of ordinary treatment, and the little knowledge that exists is partly derived from studies where treatment as usual has served as the control condition. Yet this in turn often lacks quality and stringency, which affects the generalisability of virtually all other synthesised knowledge we have about psychological treatment for common mental health disorders.

An alternative source of knowledge is retrieved through quality indicators, a systematically obtained knowledge base that covers several quality aspects of care. However, today's use of quality indicators is criticised for mainly two reasons: relevant information is not obtained on the majority of patients, and the indicators that are collected are often intended for inpatient treatment. In contrast, the majority of outpatients are treated at outpatient facilities for depressive or anxiety problems, so called common mental health disorders (CMHD).

There are several ways to measure CMHD, including non-directly measured latent variables based on several symptoms. However, the instruments that are typically used must be valid for the target group. Furthermore, comorbidity is often high among patients with CMHD, but opinions differ to what extent, and how the comorbidity should be measured. This project includes three published articles concerning three different aspects of CMHD: the effect of ordinary treatment, measurement properties of two commonly used instruments, and latent subgroups.

In paper I, a systematic literature study and meta-analysis were conducted for Nordic outpatient treatment, where we found 11 studies that were considered to be clinically representative, with a low to medium effect size ($g = 0.49$), when measured before and after

treatment. The strength of the effect appeared to be lower than evidence-based interventions, but on a par with other interventions conducted in the studies (g = -0.21, adjusted for publication bias, g = -0.06). The conclusion was that Nordic outpatient treatment was characterised by widespread patient and therapist variables and outcomes.

In paper II, the factor structure and measurement invariance of the Patient Health Questionnaire (PHQ-9) and the Generalized Anxiety Scale (GAD-7) were investigated, together with their concurrent validity with an instrument measuring functional impairment. Since a bi-factor model had good measurement properties, these instruments were found to be mainly unidimensional. In addition, we found that comparisons between several different groups were justified and that depressive symptoms were more associated with impaired functioning. Female patients and those with comorbid conditions thus seemed to have greater deal of symptoms and functional impairment on a group level.

In paper III, diagnoses and subgroups based on symptoms were examined. There was a lower prevalence than expected of patients with comorbid depression and anxiety diagnoses. A latent normally distributed subgroup analysis of the PHQ-9 and GAD-7 showed three groups, where 33% of the patients ended up in the group with the most aggregated problems, characterised by high somatic symptoms of depression and anxiety, as well as a greater probability of having more consultations, and of being diagnosed with several CMHD.

In conclusion, we show that several aspects of depression and anxiety can be important quality indicators. We present the effect and content of ordinary Scandinavian outpatient treatment for common mental health problems, even though the goal was to compare the entire Nordic region. Although the result is heterogeneous, this can serve as a benchmark for what ordinary treatment contains and what the effect may be on a group level.

We also show that PHQ-9 and GAD-7 have good measurement properties at the start of treatment and can be used as quality measurement instruments for depression and anxiety. Furthermore, we show that a high level of somatic depression symptoms can be important indicators for patients who may be difficult to treat. This may have research implications for others who want to study usual treatment, but also for those who want to compare interventions with the usual treatment. It also has clinical implications; to create a better understanding of ordinary treatment, measuring symptom severity at the start of ordinary treatment, and be able to identify patients with potentially greater care needs at an early stage. However, the thesis should not be seen as a complete description of ordinary treatment, since more research is needed to be able to benefit ordinary patients, such as high-quality effect studies, process studies of clinically representative therapies, investigating the diagnostic utility of PHQ-9 and GAD-7, and investigating whether de facto early identification of difficult-to-treat patients has some clinical benefit. This thesis offers the argument that ordinary treatment should be given the place in the research field to which it really should be entitled, namely for the many people with common mental health problems who are treated in outpatient care.

Scientific Environment

This thesis was carried out in the PhD program at the Faculty of Social and Educational Sciences, Norwegian University of Science and Technology, NTNU. The financial support was provided by “Ramme Strategi- og Omstillingsmidler” from the Faculty of Social and Educational Sciences, NTNU. My four-year position included one year of teaching duties, mostly within the field of clinical psychology, and one year as president of the doctoral and postdoctoral interest organization at NTNU. In turn, this included one year of participation in the research council of NTNU, representing temporary scientific employees.

I was part of a research group that included Associate Professor Audun Havnen, Professor Stian Solem and Professor Odin Hjemdal, and two fellow Ph.D.-candidates: Martin Schevik Lindberg and Jakob Lundqvist, all affiliated with the Department of Psychology.

The overarching quality-insurance project was developed by representatives from NTNU (Audun, Stian, Odin), together with representatives from Nidaros District Psychiatric Center (DPS), and user representatives.

My main supervisor was Audun, and my co-supervisors were Stian and Odin. They also co-authored the papers in this thesis. Martin co-authored all papers and Jakob papers I and III. Professor emeritus Lars-Göran Öst, affiliated with Stockholm University, co-authored paper I, and associate professor Frederick Anyan, affiliated with NTNU, co-authored paper III. Data for paper I was retrieved from published papers across several databases. Data for papers II and III were collected from Nidaros DPS.

Acknowledgement

After my wife and I had our first child five years ago, we moved from Västerås in Sweden to Trondheim in Norway, intending to stay for just a year. However, during my parental leave, we had a change of plans since I got the offer to start my Ph.D. During my first day, I was told by several colleagues that I was very fortunate to be affiliated with a research group with such experienced and well-liked supervisors. They were right. Therefore, I would like to start by directing my gratitude to them.

Audun Havnen: Your kindness and your sharpness have been outstanding. You have always been available for support, no matter what, and your feedback has been exceptionally accurate. Thank you for all your encouragement. To my co-supervisor Stian Solem: Your wisdom in psychology is remarkable, and you always seem to have a joke up your sleeve. Thank you for always brightening up my workdays. Odin Hjemdal: You have an ability to see beyond mundane problems – such as not having converging models – and you have helped me to remember that there will come a day after this Ph.D. So, to all of you, I am sincerely grateful for having had the opportunity to work with you.

To my colleagues, office roommates and friends, Jakob Lundqvist and Martin Schevik Lindberg. Discussing relevant, and irrelevant topics with you has made my workdays rich and purposeful. I look forward to continuing our friendship after I am finished with my Ph.D. I would also like to direct a thank you to my co-authors Lars-Göran Öst and Frederick Anyan: I have learnt a lot from your great methodological expertise, and your co-authorship has elevated the quality of the papers in this thesis.

To my colleagues during my one-year practical duty for psychologists at child psychiatric care in Härnösand. Your kind-hearted and creative approaches to meeting distressed families have been a true inspiration for my clinical and academic work. A special

thanks to my former supervisor Karin Nordhage. You have taught me a lot, more than just to stay awake during hours of assessment. Thank you for being a great role model.

To my former colleagues in Västerås: Mårten Tyrberg, Gunilla Schiölde, Karin Carlsson. I am deeply thankful for having worked with you. Your empathic and professional skills have inspired me. To my other friends and colleagues: Thank you for keeping my spirit up. There is not enough room to thank all of you, so to name a few who have checked in with my progress of this thesis: thank you Erik, Johan, Evald, Milad, Heather, Dani, Franci, Håvard, Anne-Elise, Lars, Adrian, Steffen, Line, Espen, Tora, and many more.

To the participating patients at Nidaros DPS. I realise the distress one can experience before starting treatment, thus, a sincere thank you for giving your time to participate in the research project. Also, a thank you to all mental health workers at Nidaros DPS, who have administered, and reminded patients to answer the instruments, and to all the leaders who have made this project possible.

I would like to thank my mother, father, and sisters for your support during my upbringing. Since I was pathologically curious, you served me with fact books that gave me information on everything I needed and didn't need to know – making it possible to lecture you on disastrous train accidents during train rides. Thank you for forming me. I would also like to direct a thanks to my in-laws with family, who also have shown interest in my research, and well-being. For information Torbjörn, my *h*-index is currently 2.

A final thanks to my closest family. Hedvig, I am so happy having you by my side, twenty years after we first kissed at OK-hotellet. Thank you for being my greatest support, through thick and thin. And to my children Ingrid and Freja, you have been the greatest motivation for wrapping up my work, thank you for being you.

Acronyms and Abbreviations

CBT: Cognitive Behaviour Therapy

CFA: Confirmatory Factor Analysis

CI: Confidence Interval

CMHD: Common Mental Health Disorders

COVID-19: Coronavirus Disease 2019

DSM: Diagnostic and Statistical Manual of Mental Disorders

DPS: District psychiatric centers

EBT: Evidence-based treatments

ES: Effect-size

FMM: Factor Mixture Modelling

GAD-7: The seven-item Generalized Anxiety Disorder scale-7

HiTOP: Hierarchical Taxonomy of Psychopathology

ICD: International Statistical Classification of Diseases and Related Health Problems

LCA: Latent Class Analysis

MADD: Mixed Anxiety-Depression Disorder

MI: Measurement Invariance

OECD: The Organisation for Economic Co-operation and Development

PAS: Patient Administration System

PHQ-9: The nine-item Patient Health Questionnaire-9

PTSD: Post-Traumatic Stress Disorder

PROM: Patient-Reported Outcome Measures

RCT: Randomized Controlled Trials

RDoC: Research Domain Criteria

ROM: Routine Outcome Monitoring

SEM: Structural Equation Modeling

TAU: Treatment As Usual

WSAS: The Working Social Adjustment Scale

List of Papers

Paper I – Brattmyr, M., Lindberg, M. S., Lundqvist, J., Öst, L., Solem, S., Hjemdal, O., & Havnen, A. (2024). Clinically representative therapy for Nordic adult outpatients with common mental health problems: A systematic review and meta-analysis. *Scandinavian Journal of Psychology*, *65*(2), 311–320. https://doi.org/10.1111/sjop.12976

Paper II – Brattmyr, M., Lindberg, M. S., Solem, S., Hjemdal, O., & Havnen, A. (2022). Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder scale-7 in a Norwegian psychiatric outpatient sample. *BMC Psychiatry*, *22*(1), 461. https://doi.org/10.1186/s12888-022-04101-z

Paper III – Brattmyr, M., Lindberg, M. S., Lundqvist, J., Solem, S., Hjemdal, O., Anyan, F., & Havnen, A. (2023). Symptoms and prevalence of common mental disorders in a heterogenous outpatient sample: An investigation of clinical characteristics and latent subgroups. *BMC Psychiatry*, *23*(1), 804. https://doi.org/10.1186/s12888-023-05314-6

List of Tables

List of Figures

1. Introduction

Mental health disorders are globally burdensome, across age groups and genders (James et al., 2018). They are considered the costliest health conditions, and in Norway their cost exceeds NOK 80 billion per year, mainly due to lost production and healthcare expenditure (Kinge et al., 2017). Two of the groups of disorders most frequently encountered in psychiatric outpatient treatment, and the general population, are depressive and anxiety disorders, also called common mental health disorders (CMHD). These are among the leading causes of years lived with disability (James et al., 2018). Over their lifespan, women are overrepresented twice over with major depressive disorder (Hasin et al., 2005) and anxiety disorders (Baxter et al., 2013). Women, but also younger adults report higher levels of psychological distress than elderly (Drapeau et al., 2014), and correspondingly, women and younger adults are overrepresented in seeking Norwegian adult outpatient treatment (Pedersen & Lilleeng, 2019).

The estimates of CMHD prevalence vary across the globe: in the Norwegian general population, the estimated point prevalence of depression (3.2%) seems less frequent than the global mean (3.4%) and the European mean (3.5%), while anxiety disorders seem more frequent (6.7%) than the global mean (3.8%) and the European mean (4.6%; Dattani et al., 2023). In Norwegian District Psychiatric Centres (DPS), mood disorders (31%) and anxiety disorders (36%) are also most prevalent (Pedersen & Lilleeng, 2019), and these account for almost half of the outpatient visits in Norway (Schem et al., 2018).

Estimates could change over time: the estimated global prevalence increased with the coronavirus disease (Covid-19), with a change in depressive disorders from 2.5% to 3.2% and in anxiety disorders from 3.8% to 4.8% (Santomauro et al., 2021). However, the suggested increase in prevalence rates has been challenged by a study from Trondheim, Norway, which

showed a reduced self-reported prevalence of mental disorders in general, from 15.3% to 8.7%, in the initial phase of the pandemic (Knudsen et al., 2021). There is a great discrepancy in estimated prevalence between these studies, and it should be noted that self-reported prevalence and clinician reported prevalence could yield different results (Hyland & Shevlin., 2024).

There is general consensus that acknowledging comorbidity is important, since patients with several distinguished mental health disorders have lower remission rates in routine care (Fava et al., 2008). There are both organisational and methodological differences in estimating comorbid CMHD, however, and the estimates might vary. For example, the estimated prevalence rates for depression from medical registers are quite similar across the Scandinavian countries, yet, being diagnosed with an anxiety disorder after being diagnosed with a depressive disorder was 2.5 times more common in Sweden compared to Norway and Denmark (Pasman et al., 2023).

In recent decades, there have been several organisational changes in mental health services. They have shifted from institutionalised care to outpatient treatment, to better meet patients' needs (World Health Organization, 2021). While the number of inpatient stays has been reduced by 50% in the last 20 years in Norway, the growth in staff numbers has mainly been in outpatient facilities (Pedersen & Lilleeng, 2019). Furthermore, evidence-based treatment (EBT) for CMDH has increasingly been recommended as the first-line treatment (National Institute for Health and Care Excellence, 2022). However, there are reasons to believe that most people with CMHD do not receive EBT (Bergmark et al., 2022; Harvey & Gumport, 2015; Johnson et al., 2016).

Even though CMHD is severely burdensome for society and individuals, mental health is an underrepresented research field compared to other health-related fields (Hazo et

al., 2019). Even though research output has grown tremendously in recent decades, several fundamental questions regarding patients with CMHD in routine facilities remain.

There have been several calls for a better understanding of routine mental health care, both internationally and in Norway. One project that disseminated knowledge gaps in the field suggested, among other things, three areas in need of improvement in Europe (Haro et al., 2014). First, they identified a fundamental lack of knowledge of the deployment of psychological treatment in Europe, such as its contents and outcomes. Therefore, they concluded that there is a dire need for valid knowledge of the quality of public mental health care, such as research on regional differences in primary and specialist mental health treatment. Second, they identified a need to change the services' relationship with their patients. To enhance patient-centred services, novel initiatives are needed, to gain better knowledge of routine outcome monitoring merged with electronic patient records. Third, they identified limitations in current classification systems, such as conceptualising comorbidity. Since patients with comorbid disorders are associated with increased service use and poorer prognosis, novel methods to identify subtypes of patients were declared to be needed (Haro et al., 2014).

Similar key messages were formulated in the Lancet World Psychiatric Association Commission on Depression (Herrman et al., 2022): a need to identify which treatment works for whom, engaging people with lived experience, and acknowledging the complexity of depression. Similar conclusions were also acknowledged in the Norwegian Ministry of National Health and Hospital Plan 2020-2023 (Ministry of Health and Care Services, 2019): there is a need to understand what works for whom, and a need to develop systems for evaluation of the efficacy of treatment.

Consequently, policy documents from the World Health Organization (World Health Organization, 2021), the European Union (Directorate-General for Health and Food Safety, 2023), the Organisation for Economic Co-operation and Development (OECD; Hewlett & Moran, 2014), and Norway (Ministry of Health and Care Services, 2023) have stated that patients need to be better involved with their care, fostering patient-centred mental health treatment through the use of monitoring systems.

One important way to improve knowledge of what works for whom and the effectiveness of mental health services is by routinely collecting patient-reported outcome measures (PROM). By having service users report on a set of quality measures of the treatment, increased knowledge about routine treatment can be gained. Yet there is a need to connect PROM data with register data, to answer complex questions. This thesis aims to cover three central aspects of patients with depression and anxiety in routine care: examining treatment as usual in the Nordic countries, assessing two common patient-reported outcome measures on CMHD, and examining the prevalence and subgroups of CMHD. Two wider themes will follow in the introduction: what is treatment as usual, and what is the clinical utility of PROM as quality measures/indicators?

1.1. Treatment As Usual and Clinically Representative Therapy

Treatment as usual (TAU) is an umbrella term for ordinary practices deployed in routine care. Other, similar concepts are routine treatment, care as usual, existing practice, normal care, or for a more stringent definition, *clinically representative therapy* (Shadish et al., 1997). Since TAU and clinically representative therapy should reflect ordinary treatment, there is a need to understand their content, so as to examine the representativeness and effectiveness of ordinary treatment, for patients in ordinary settings.

TAU has traditionally mostly meant patients who receive usual treatment instead of being randomised to an intervention group in randomised controlled trials (RCTs), which in turn is the golden standard for *efficacy research* (Shean, 2014). In this research paradigm, high internal validity is key to establishing evidence of an intervention, such as cognitive behaviour therapy (CBT). Internal validity is optimised by using strict inclusion criteria, and simultaneously comparing results with valid control groups. Although double-blinded studies are the *modus operandi* for RCTs in medical research, it is virtually impossible to blind patients and therapists in mental health care research (Shean, 2014). Both patients and practitioners will know that a control-group treatment has been delivered, and thus they may expect inferior treatment, for both psychological (Freedland et al., 2011) and pharmacological mental health treatment (Cuijpers et al., 2015). Especially psychological treatment could be affected by patients' expectations, since outcomes are often evaluated using self-rated instruments, such as PROM (Enck & Zipfel, 2019).

Except for TAU, there are mainly two other categories of control groups: psychological placebo and waiting list (Faltinsen et al., 2022). Psychological placebos are often defined as treatments without the active ingredient from the intervention, for example, behavioural exercises which mimic exposure interventions in CBT (Enck & Zipfel, 2019). While a placebo could control for unwanted noise in clinical medical research, there are

theoretical and methodological problems associated with psychological placebos. Besides the blinding issue, the dividing lines between ordinary psychotherapy and psychological placebo are not easily disentangled, so that the concept is criticised for being flawed (Wampold et al., 2016). Since there are practical difficulties in disentangling specific and common factors in psychotherapy, it is problematic to draw conclusions about the effects of a specific factor on an outcome (Enck & Zipfel, 2019).

Regarding waiting lists, it has been suggested that patients with depression who are on waiting lists might see a reduction of symptoms due to treatment expectations (Høstmælingen et al., 2023), or that a waiting list could be even worse than no treatment (Furukawa et al., 2014). No further clarification came from a Cochrane study of five sub-studies for a range of mental disorders, which found non-significant differences between TAU and waiting list (Faltinsen et al., 2022). Although some continue to argue for waiting lists to be better than TAU, since the latter is often very heterogenous (Munder et al., 2019), waiting lists are of little use for patients and are often considered inadequate to estimate the efficacy of psychological treatments (Cristea, 2019).

Therefore, a TAU control group could be a better alternative, since it does not withhold treatment from patients, and can simultaneously be used to control for several common factors (Kazdin, 2015). Yet it is also associated with several problems. First and foremost, it is often unknown to what extent TAU is representative of usual treatment. Specifically, researchers rarely describe in detail what TAU consists of. A systematic review of the use of TAU for CMHD in primary healthcare research found that a majority, 56% (k = 18), of the studies included only gave a basic description of TAU, and only 9% (k = 3) had an advanced description of it (Petersson et al., 2023). This limitation is also stated in other systematic reviews examining TAU for CMHD: relevant information on TAU is often omitted, such as facility, patient, therapist and treatment characteristics (Wampold et al.,

2011; Watts et al., 2015). Furthermore, patients in TAU often receive less attention, and therapists receive less supervision and training compared to the intervention in focus (Wampold et al., 2011). The use of TAU has thus been criticised for being *laissez-faire* (Freedland et al., 2011), and even worse, not even intended to be therapeutic (Wampold et al., 2011). And since TAU is often of poor quality, the effect size (ES) of the intervention of interest could be artificially inflated (Cuijpers & Cristea, 2016). The poor quality of control groups, TAU included, could thus lead to severe internal validity problems for efficacy studies.

Additionally, there are often problems with the ecological validity and therefore the external validity of efficacy studies. Their strict inclusion criteria mean that several patient types are not eligible for treatment, such as patients with comorbid disorders (Humphreys et al., 2015). Since a high degree of comorbidity seems to be the rule, rather than the exception in real-world settings, efficacy studies might not be generalisable to clinical real-world populations.

Therefore, other research paradigms have been acknowledged, such as *effectiveness research*. By researching effectiveness – examining treatments given in real-world settings, using ordinary therapists and ordinary patients – the external validity could be said to be optimised. Although several studies have established the effectiveness of treatments for several specific disorders (for example Öst, Brattmyr et al., 2023), and guidelines recommend that professionals deliver treatments based on the best available evidence (the Norwegian Psychological Association, 2024), many therapists state explicitly that they do not follow these recommendations in their clinical practice (Bergmark et al., 2022; Harvey & Gumport, 2015; Johnson et al., 2016). There are several reasons that therapists do not use *evidence-based practices* in real-world treatments, ranging from personal beliefs and attitudes, to high caseloads and little supervision (Harvey & Gumport, 2015; Lilienfeld et al., 2013). As a

result, professionals use methods that have received little empirical examination, which could make evidence from both efficacy and effectiveness studies of little use to the majority of patients.

There have also been emerging research paradigms for practice-oriented research, covering patient-focused research, practice-based evidence, and practice research networks (Castonguay et al., 2021). Common factors, such as goal consensus, alliance and treatment expectations have been found more important than specific factors, such as treatment differences, and adherence to protocols (Wampold, 2015). For example, therapist factors, such as interpersonal skills, seem to have a central role in yielding good treatment results (Heinonen & Nissen-Lie, 2020). Critics argue, on the other hand, that this does not mean that all treatments work equally well, since it is still not well-known how psychotherapies work (Cuijpers, Reijnders, et al., 2019). Nevertheless, integrative psychotherapeutic approaches have been suggested as an alternative to specific modalities (Zarbo et al., 2016), namely for therapists to choose evidence-based methods that fit patient needs, while others argue that such approaches could hinder effective learning to become a skilful therapist (Byrne et al., 2018). As a result, therapists could adhere to eclectic approaches, choosing methods without empirical evidence, or theoretical underpinnings (Zarbo et al., 2016). Although there is limited research on the number of therapists who identify as being eclectic, empirical evidence points to how a majority of real-world therapists use several approaches in their everyday work (Thoma & Cecero, 2009). Research of ordinary practices in mental health research has thus been mostly overlooked, and little is known about what is done and its effects. Given that TAU should reflect ordinary practices, it should thus gain empirical knowledge of its own (Kazdin, 2015).

1.1.1. The Effects of TAU/Clinically Representative Therapy

Parallel to the evolving research paradigms of the efficacy and effectiveness of psychological treatments, the meta-analytical framework has emerged. In response to the conclusion that psychotherapy was probably not effective (Eysenck, 1952), the first modern meta-analysis was conducted by Smith and Glass (1977). They found a standardised mean difference of 0.68 in favour of psychotherapy over other control groups (Smith & Glass, 1977). Although Eysenck replied in his title that this was *An exercise of mega-silliness* (Eysenck, 1978), nowadays there is little dispute that he probably was wrong – psychological treatments can be assumed to have some effect on a range of disorders. However, the debate continues regarding how much effect psychological treatments have, because the poor quality of studies, mainly due to poor control groups, is an obfuscating factor (Cuijpers, Karyotaki et al., 2019). In other words, since the ES of an intervention is relative to the control group, meta-analysing the results of psychological treatments could bias the ES, if neither the quality nor the stringency of the control group is controlled. Studies of poor quality will therefore lead to *garbage in, garbage out* (Eysenck, 1978).

Furthermore, if the content of TAU is not clarified, the results can be difficult to interpret. For example, an umbrella review of the efficacy of mental health treatments (including 26 meta-analyses of psychological treatments) found that psychotherapies had a standardised mean difference of 0.36 over TAU (Leichsenring et al., 2022). However, if TAU was ordinary treatment, a great deal of treatment within the control group should be psychological treatment. This example shows the ambiguity TAU can be associated with. Since TAU is seldom described, controlled for, and is often associated with low quality, a criticism could be that such an approach uses bad apples to compare apples with.

There are thus limitations to the generalisability of meta-analyses to real-world contexts. In an umbrella review of four meta-analyses of children and adolescents, almost all

of the studies reviewed were *research therapy*, in contrast to *clinic therapy* (Weisz et al., 1992). Inspired by these findings, another team sent letters to authors of every meta-analysis they could find, inquiring about the representativeness of their included studies (Shadish et al., 1997). They used three cumulative criteria for representativeness. First, studies were conducted in non-university settings, with patients referred through ordinary routes, with ordinary therapists having ordinary caseloads. Second, there was no reliance on a treatment manual, and the treatment was not monitored. Third, patients had to be heterogeneous regarding their demographics, and mental health problems, with therapists not being trained for the study, and with therapists free to choose their procedures. They contacted 48 authors of 59 published meta-analyses, of whom 24 authors replied (Shadish et al., 1997). Out of 486 sub-studies, 56 passed the first criterion, 15 the second, and only 1 study passed the final one.

Later, they included studies that passed the first criterion, as 40 randomly sampled studies from the first meta-analysis, and nine already proposed clinical therapy studies for children and adolescents (Shadish et al., 2000; Weisz et al., 1995). Then they developed ten criteria for clinical representativeness: 1) participants had mental health problems; 2) in a setting that was primarily a service-delivery site; 3) referred through usual clinical routes; 4) practising professionals at the site; 5) not structured or representatively structured therapies; 6) not monitored in a way that could influence therapists' behaviour; 7) patients with heterogeneous problems; 8) therapists had no extra training before the study; 9) therapists were free to use multiple techniques; and 10) no fixed limits on the number of sessions.

Out of the 90 analysed studies, ten were suggested as clinically representative, with an estimated within-study ES of $d = 0.41$, which was the same for all 90 studies included (Shadish et al., 2000). However, their study already included then outdated studies, spanning from 1942 to 1993. Furthermore, it is not stated in their manuscript which studies were suggested as clinically representative. This gives reasons to believe that they were all

outdated studies of child or adolescent therapies. Therefore, it is potentially misleading to generalise their findings with treatment for adults delivered today.

A few meta-analyses have sought to establish the effect of TAU on CMHD, presenting sub-studies that use samples with a mix of these disorders (see Table 1). For example, a study of TAU compared to CBT for patients with depression and anxiety disorders found five studies that treated both depression and anxiety, with a between-group ES of $g = 0.44$ for symptoms of depression, and $g = 0.34$ for symptoms of anxiety in favour of CBT (Watts et al., 2015). However, TAU was delivered by general practitioners in three studies, through telephone assessment in one, while only one was mainly psychological treatments delivered by a range of different professionals (Den Boer et al., 2007).

In another meta-analysis of TAU versus EBT for CMHD, three studies had a TAU which was psychological treatments, with a between-group ES of $d = 0.33$ in favour of EBT (Wampold et al., 2011). However, therapists received additional supervision in the EBT-intervention group. Furthermore, these three studies were diagnosis-specific: one panic disorder, one depression and one post-traumatic stress disorder (PTSD). In an updated review, they found a between-group ES of $g = 0.40$ after 0–4 months of treatment, and $g = 0.20$ after 12–18 months, with smaller ES with more active TAU (Flückiger et al., 2014). However, their included studies were also diagnose-specific – depression was examined in 11 studies, social anxiety disorder in two, generalised anxiety disorder in one, and panic disorder in one, and many of these were not active treatments, and therapists often received additional training.

Another meta-analysis examined TAU for heterogenous CMHD amongst children and adolescents and found a within-study ES of $g = 0.52$ (Bear et al., 2020). However, their sub-sample of 11 studies with mixed disorders mainly used only one technique, such as

pharmacological treatment or psychoanalytic treatment. Furthermore, the transferability between child and adolescent treatment and adult treatment is disputable (Cuijpers et al., 2020). It should also be noted that within-group ES, that is pre-post scores, has been criticised for being influenced by natural processes, but could still be informative when examining routine treatment (Cuijpers et al., 2017).

Regarding their clinical representativeness, previous meta-analyses do not pass the strictest criteria from Shadish et al. (2000). This is mainly because patients in the sub-studies were not expressing a mix of problems, and because therapists only used a specific treatment modality (See Table 1). Still, an important finding among these meta-analyses is a substantive heterogeneity of TAU, both qualitatively and quantitively (Wampold et al., 2011; Watts et al., 2015). To reduce such heterogeneity, it has been suggested that meta-analyses should be conducted within one country and one setting, with an expected number of studies subject to consideration (Cuijpers et al., 2021). This motivates an updated study of clinically representative therapy, in a restricted area. In this case, the Nordic countries.

Table 1

Meta-Analyses of TAU as Psychological Interventions with Heterogenous Sub-Studies of both Depressive and Anxiety Disorders

	***k* sub-samples using both depression and anxiety (*K* all studies)**	**Intervention**	**Effect size in studies with both depression and anxiety, and pooled effect size**	**Limitations of TAU**
Studies with within-study effect size				
Bear et al., 2020	11 (38)	Treatment as usual for children and adolescents	$g = 0.52$. For all studies $g = 0.74$.	TAU was defined as all treatments unaltered by study design and therefore included such studies such "pharmacotherapy only", "psychoanalytic treatment only", "manualised CBT only", etc.
Shadish et al., 2000	Unknown (90, whereas 10 were considered fully representative)	Clinically representative therapy	Unknown. For 10 clinically representative therapies, the estimated $d = 0.41$, which was the same for pooled ES.	Although representativeness was coded through 10 criteria, treatment methods in TAU were not presented, and articles are outdated (from 1942 to 1993).
Studies with between-study effect size				
Flückiger et al., 2014	2 (11)	Evidence-based treatment	Unknown for these two studies, but all 11 showed $g = 0.40$ after 0-4 months, and $g = 0.20$ after 12–18 months.	An update to Wampold et al., 2011, and thus therapists also received additional training and supervision in intervention compared to TAU. Furthermore, studies examined diagnosis-specific treatments.
Leichsenring et al., 2022	0 (26 meta-analyses)	Psychotherapy	Not able to retrieve. For all studies, standardised mean difference = 0.34.	Does not describe TAU. Only two meta-analyses examine both depression and anxiety. None of these include studies of depression and anxiety in combination, together with psychological treatment.
Wampold et al., 2011	3 (14)	Evidence-based treatment	$d = 0.33$. For all studies, $d = 0.45$	Therapists in intervention groups received additional training and supervision. Diagnosis-specific treatments.
Watts et al., 2015	5 (48)	CBT for depression and anxiety	Symptoms of depression $g = 0.44$, anxiety $g = 0.34$. For all studies, $g = 0.55$	Only one study examined TAU as a psychological treatment.

Note. Between studies effect sizes favouring other interventions than TAU.

1.2. Measuring the Quality of Services for Patients with CMHD

Although often used interchangeably, a quality indicator indicates quality, while a quality measure measures it. Quality indicators are thus always connected with a secondary appraisal (Quentin, 2019). To measure quality, patient-reported outcome measures (PROM) are often used. These are standardised questionnaires that are mainly used to measure symptoms, health-related quality of life, or functional status, directly from patients (Churruca et al., 2021).

A common argument for implementing health quality indicators is to improve stakeholder involvement (Schang et al., 2021). In turn, such indicators could help people involved in treatment – patients, mental health workers, and decision-makers – to make informed choices based on the quality of care. A plethora of indicators exist; for example, a systematic review identified 53 different indicators for depression in primary mental health care (Petrosyan et al., 2017). To make such diversity more comprehensible, frameworks such as the Donabedian quality of care model are often applied – namely to categorise them into structural, process, and outcome indicators (Donabedian, 1966). Structural indicators are factors that affect the context of treatment, such as the number of hospital beds, or the availability of staff. Process indicators are the actions performed at the facility, such as diagnosis or treatment methods. Outcome indicators are often the results of treatment, such as changes in symptoms.

Although structural indicators could be more accessible, other indicators are needed to get a better picture of the quality of the treatment (Quentin, 2019). For example, outcome indicators are essential for examining the effectiveness of care, yet they are often in the minority (Petrosyan et al., 2017). Furthermore, indicators that are relevant for inpatient care are often well-covered, in contrast to outpatient care, at the expense of the many patients with mild-to-moderate disorders (Hewlett & Moran, 2014). As a consequence, stakeholders'

attention could be directed towards existing superfluous quality indicators, at the expense of more important aspects that could be more difficult to measure (Quentin, 2019).

This is the case in several countries, such as Norway. A recently developed quality register for mental disorders has been established (National service for medical quality registers, 2024), which has already been criticised for not reflecting the complex reality (Frahm Jensen et al., 2022). Since data is unpublished as yet, information regarding the quality of care for patients with CMHD has been covered within a national healthcare quality indicator system (Norwegian Directorate of Health, 2024). Although there are several indicators, only a few are of relevance for patients with CMHD in outpatient treatment. These indicators are mainly formal policies, such as whether patients are offered treatment within due time. Since more suitable indicators are lacking, this system has been criticised as insufficient to examine the most central aspects of mental health treatment (Office of the Auditor General of Norway, 2021). This was further supported by their survey of leaders of DPSs, of whom many perceived minimal attention from policymakers about the effectiveness of treatment conducted at their facilities.

1.2.1. Quality Measurement of CMHD using Patient-Reported Outcome Measures

In recent years, the focus in the quality assessment field has shifted from measuring the delivery of care, to clinical outcomes (Druss, 2018). As a consequence, policymakers across the world have promoted scaling up the amount of shared patient-centred data in connection with routine treatment. WHO has encouraged countries to collect data to map patient needs. A prior target for 2020 was for 80% of the countries to collect and report a core set of mental health indicators (World Health Organization, 2013). This was not achieved, since only 31% of the WHO member states comply with this target (Mental Health Atlas 2020, 2021). As many aspects were not fulfilled in 2020, the plan was revised with an extension to 2030, to amongst other things double the research output for mental health care

(World Health Organization, 2021). Similar upscaling initiatives for patient-centred data generation and research can be found within the OECD (Hewlett & Moran, 2014), EU (Directorate-General for Health and Food Safety, 2023), and Norway (Ministry of Health and Care Services, 2023).

A common suggestion is to monitor through PROM. This could facilitate service planning (Gelkopf et al., 2022), and assist in the diagnostic process for CMHD (Pinho et al., 2021). However, the instruments need to be appropriate (Quentin, 2019), and measurements could be improved according to five recommendations: CMHD measures need to be validated; they should be retrieved through accessible technologies; outcomes should be assessed routinely; it should be possible to link data sources across settings; and finally, it should be possible to identify subpopulations in need of quality improvement (Kilbourne et al., 2018).

Several initiatives have implemented and researched data from PROM. One example is Improving Access to Psychological Therapies in England (Clark, 2018). Since 2008, they have routinely collected data for patients treated with CBT for CMHD. Among other things, they have found that organisational factors affect the outcomes of treatment: shorter waiting times, high levels of attendance, higher numbers of sessions, and higher treatment focus are associated with better outcomes (Clark, 2018). Furthermore, a meta-analysis of 47 studies found a within-study ES of $d = 0.87$ for depression measured with PHQ-9, and $d = 0.88$ for anxiety measured with GAD-7 (Wakefield et al., 2021). Another project that has implemented PROM is the Leiden Routine Outcome Monitoring Study in the Netherlands (De Beurs et al., 2011). Their collection started in 2002 and has covered over 10,000 patients (Leiden University Medical Centre, 2016), with several papers being published.

An important feature of PROM is the possibility to link information with register data on service utilisation and disorders (Al Sayah et al., 2021). This can, for example, be used to examine comorbid CMHD and symptom severity. However, only a few published studies have applied this (Gelkopf et al., 2022). For example, a PROM study of patients treated for anxiety disorders found that 45% had comorbid depression, and comorbidity was associated with more severe symptoms before and after treatment (Klein Breteler et al., 2021). Nevertheless, PROM studies examining other aspects, such as service use, are scarce.

1.2.2. Measurement Considerations for CMHD

Using PROM, one common approach is to compare patients' mean symptom scores across different mental disorders. Disorders are mainly diagnoses from either one of the two most commonly applied classification systems: the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2013), and the International Statistical Classification of Diseases and Related Health Problems (ICD; World Health Organization, 2016). These are widely applied in research, clinical practice, administration and communication between healthcare providers, and are used to inform patients and caregivers (First et al., 2019). Although these taxonomies have been successful in creating a common language for mental disorders, they have given rise to several controversies. Two common criticisms that are of relevance for comparing symptom scores might seem contradictory: their criteria are seldom formally applied by clinicians, and clinicians have an over-reliance on these criteria rather than contextual factors (Stein et al., 2022). The first is a procedural issue of *diagnosis*; in a survey of almost 2,000 psychiatrists and psychotherapists, 50% never used a structured diagnostic interview, and only 15% of patients were ever interviewed (Bruchmüller et al., 2011). Even when they were applied, many CMHD diagnoses showed critically low inter-rater reliability (Freedman et al., 2013). The second issue is a conceptual issue of *diagnoses*. Although DSM/ICD were constructed by expert

consensus and are claimed to be atheoretical and purely descriptive, this view has been challenged. For example, there are reasons to believe that DSM makes an implicit neurobiological understanding of the underlying disorders (Castiglioni & Laudisa, 2015).

Although disorders from ICD/DSM are classified by using a set of criteria, it is not clear whether these criteria are to define the disorders or should only be used to measure them. In a correspondence, Kendler (2017) proposed an *indexical* relationship over a *constitutive* one, suggesting that these criteria should only be used to measure their corresponding disorders, and not define them. In a reply, Van Loo and Romeijn (2018) suggested a double role for the disorders, to both measure and define them, just like any other symptom disorder, such as migraine. However, a wide range of measures exist to measure CMHD, and they include different symptoms. For example, in an analysis of seven commonly applied depression scales, 52 disparate symptoms were reported (Fried, 2017).

Due to this disparity, the International Consortium for Health Outcomes Measurement (Obbarius et al., 2017), and some of the largest mental health research funders (Farber et al., 2023), have suggested a core set of outcome measures for depression and anxiety: the patient health questionnaire (PHQ-9) and the generalized anxiety disorder scale (GAD-7). However, these recommendations have been criticised, among other things for instruments not being researched enough in settings where they are used the most (Patalay & Fried, 2021). More specifically, they point to conflicting results of their unidimensionality and measurement invariance (MI), that is whether they are measuring the same construct across different groups. Therefore, there is a need to research their validity in real-world clinical settings (Patalay & Fried, 2021), which motivates us to examine their properties, among Norwegian psychiatric outpatients.

Another central aspect of these classification systems is for disorders to cause clinically significant functional impairment. However, the relationship might not be as

straightforward as it may seem. For example, the relationship between functional impairment and symptoms of depression (r^2= .25) and anxiety (r^2= .12) was found to be small in two systematic reviews (McKnight et al., 2016; McKnight & Kashdan, 2009). Thus, the concurrent relationship between CMHD and functional impairment seems weaker than how it is theoretically positioned. At the same time, these reviews found great variance in reported functional impairment between and within CMHD, and accounting for redundant noise could theoretically improve the model.

1.2.3. Measurement Considerations for Comorbid CMHD

Neither DSM nor ICD have included a comorbid depressive and anxiety disorder diagnosis, but a sub-threshold mixed anxiety-depression disorder (MADD) was incorporated in ICD-10 and later ICD-11. In contrast, DSM-IV placed this disorder in its research appendix and did not include it in DSM-5, due to validity and reliability issues. Nevertheless, DSM-5 has included an anxious distress specifier for depressive disorders. However, the legitimacy of MADD as a sub-clinical or clinical disorder is disputed. Some researchers argue that a sub-threshold depressive and anxiety disorder diagnosis should be included as a separate diagnostic entity, partly due to the condition's high prevalence (Möller et al., 2016). Others argue for MADD to be a clinical, rather than a sub-clinical disorder (Shevlin, Hyland, et al., 2022).

However, to treat disorders as categorical constructs could struggle to deal with the great covariances between depression and anxiety disorders. Both DSM and ICD have therefore been criticised for not being able to sufficiently separate them (e.g. Demyttenaere & Heirman, 2020). While some argue that findings of high prevalence of comorbid CMHD could be explained by a common distress factor (for example by negative affectivity in the tripartite model by Clark & Watson, 1991), others argue that comorbidity is purely artefactual, since disorders share many symptoms in terms of how they are defined (Maj,

2005). Correspondingly, two alternative frameworks have emerged with a dimensional approach to mental disorders: the Hierarchical Taxonomy of Psychopathology (HiTOP) and the Research Domain Criteria (RDoC).

The HiTOP shares several similarities with DSM/ICD, but uses symptoms to form a hierarchical system. Depression and anxiety form sub-factors that load an internalising spectrum, which together with other spectra loads on a superspectrum – the *p*-factor. RDoC, on the other hand, was developed to be a translational research framework with a biological assumption of mental disorders (Insel et al., 2010) and has among other things been applied in the research of depression and anxiety disorders (Böttger et al., 2023). However, both initiatives have been criticised for theoretical issues and have little clinical utility in everyday practice. For example, describing CMHD comorbidity through a common internalising factor is criticised for re-labelling disorders into one even more heterogeneous group (Haeffel et al., 2022). RDoC, on the other hand, was not designed to be a clinical tool, and its biological assumption of brain disorders could create a reified understanding of mental disorders (Haeffel et al., 2022).

1.2.4. CMHD as Categories, Dimensions or Both?

There are several approaches to model the heterogeneity of CMHD: from subtypes in DSM/ICD to dimensional conceptualisations in RDoC and HiTOP. While subtypes of categorical disorders could be criticised by the same reification logic that DSM/ICD is criticised for, dimensional approaches have been suggested to be a better alternative, since this approach moves away from the disease model (Ross & Margolis, 2019). That is, dimensional models of CMHD could be conceptualised as extreme tails of continuous distributions of depression and anxiety. To overcome the heterogeneity problem with great overlap between disorders, CMHD can be modelled across multiple dimensions. However, modelling for several dimensions simultaneously is not feasible with limited computer power,

due to the *curse of dimensionality* – an exponential decrease in performance to find outliers in relation to increased dimensions (Feczko et al., 2019). At the same time, a dimensional approach makes the assumption that all individuals belong to the same group. For example, intelligence is often assumed to be normally distributed in the general population, but a purely dimensional model might oversee another normally distributed subpopulation of people with chromosomal disorders at the lower end of the spectrum (Ross & Margolis, 2019).

An alternative approach is to model subgroups using computer-driven subtypes (Feczko et al., 2019). Since non-observable subgroups of patients might be present, an increasingly popular approach in mental health research is to examine finite mixtures of Gaussian distributions, through latent class analysis (LCA; Kongsted & Nielsen, 2017). Several papers exist on LCA using symptoms of both depression and anxiety simultaneously, in both clinical (Eaton et al., 1989; Podlogar et al., 2018), and non-clinical adult populations (Curran et al., 2022; Curran et al., 2020; Das-Munshi et al., 2008; Hettema et al., 2015; Lei et al., 2022; Liu et al., 2021; Rhebergen et al., 2014; Singham et al., 2022). For a further review of LCA solely on depression, see Ulbricht et al., 2018.

However, LCA has a severe limitation when applied to CMHD: namely the assumption of conditional or local independence (Clark et al., 2013). That is, the manifest variables (symptoms) should not be dependent on each other, only on their respective classes, which is unrealistic due to the large symptom overlap of CMHD (Van Loo et al., 2018). If this is violated, spurious classes may occur. Hybrid model variations of LCA and factor analysis, called factor mixture models (FMM) can overcome this problem (Clark et al., 2013). Using FMM, disorders can be conceptualised as having both continuous and categorical properties, thus making this a hybrid approach (Borsboom et al., 2016). However, it has been little applied to simultaneously model depression and anxiety (for example

Shevlin et al., 2022; Ten Have et al., 2016), and rarely with a clinical outpatient sample. This motivates an examination of subgroups of CMHD using FMM.

1.3. Aims and Research Questions

The thesis aims to answer questions about quality aspects of routine treatment for patients with CMHD, by the application of PROM. The knowledge gaps previously reported are a lack of understanding of the content and effects of ordinary treatment, the validity of PHQ-9 and GAD-7 in a clinically representative setting, and potentially unknown subgroups of CMHD.

1. The first aim of the thesis was therefore to examine the content and effects of clinically representative real-world adult mental health outpatient therapy of common mental health disorders in the Nordic countries, using a systematic review and meta-analytical design.

- We hypothesised a great variability of patient, therapist, and treatment characteristics of clinically representative therapy
- We expected a great variation in descriptions of treatment content
- We also expected effect sizes comparable to other TAU studies

2. The second aim of the thesis was to assess the validity of two commonly applied clinical self-reporting instruments: the patient health questionnaire-9 (PHQ-9), and the generalized anxiety disorder scale-7 (GAD-7), which measure symptoms of depression and anxiety, respectively. Specific aims were to:

- Evaluate the factor structure of PHQ-9 and GAD-7
- Investigate measurement invariance across gender, disorder and comorbidity
- Evaluate the concurrent validity with functional impairment

3. The third aim of the thesis was to estimate the rates of CMHD, and:

- Assess their associations with symptom severity
- Identify clinical latent subgroups
- Analyse the subgroups' associations with service use and CMHD

2. Methods

2.1. Context

2.1.1. Mental Health Outpatient Facilities in the Nordic Countries

There are several commonalities between the Nordic healthcare systems, such as being publicly available, mainly tax-financed, and having low patient fees. Generally, basic mental health treatment is conducted within primary care services, while more severe cases are referred to specialist psychiatric care facilities, such as outpatient treatment, often through a referral from a general practitioner (Pasman et al., 2023). While all healthcare treatment is free of charge in Denmark, both Sweden and Norway have a fee for visits, although with a yearly cap.

However, it is difficult to disentangle the content of ordinary treatment, since there is a general lack of research into what treatment is delivered (Kazdin, 2015). On the other hand, there is grey literature that could be informative. In a survey of Norwegian outpatient mental health workers, they reported using psychopharmacological treatment (28%), supportive psychotherapy (49%), counselling (14%), crisis interventions (8%), cognitive psychotherapy (22%), psychodynamic psychotherapy (13%), interpersonal psychotherapy (9%), and other individual treatment (13%; Gråwe et al., 2008). Since therapists could use several modalities, this adds up to over 100%. Together with the empirical finding that many mental health workers do not adhere strictly to treatment manuals (Johnson et al., 2016), we believe that many mental health workers adhere to integrative/eclectic approaches in clinically representative therapies.

2.1.2. Nidaros Psychiatric Outpatient Clinic

St. Olav Hospital consists of two DPSs. One of them, Nidaros DPS, covers mental health services for a population of around 115,000. In 2017, Nidaros DPS reported 5,244

consultations/10,000 inhabitants, compared to the Norwegian mean of 3,613, and the regional mean of 4,067 consultations (Pedersen & Lilleeng, 2019). Furthermore, Nidaros DPS had 949 discharges/10,000 inhabitants, compared to the national mean of 963. Upon admission, patients were invited to complete electronic questionnaires, and informed consent was collected. Some patient groups were treated at other facilities, such as patients with substance use, obsessive-compulsive disorder, schizophrenia, and the elderly, and were thus seldom included in the current studies.

2.2. Study Designs

For a comparison of study designs, see Table 2. Paper I was a systematic review and meta-analysis, with published papers on clinically representative therapies in the Nordic countries. In papers II and III, all patients who were to start treatment at Nidaros DPS were invited to participate. Those who gave informed consent and filled in at least one item on the surveys were included. Therefore, these studies were cross-sectional, but paper III had a prospective element, examining service use and CMHD up until one year after treatment started. Data collection began in February 2020 and ran until November 2020 for paper II, and February 2022 for paper III. Information about diagnoses and service utilisation was retrieved in November 2020 for paper II and February 2023 for paper III.

2.3. Recruitment and Sample

2.3.1. Paper I

The sample consisted of patients in studies examining clinically representative therapies of adults in the Nordic countries. In the final sample, 12 studies with a total of 1,604 patients were retrieved for a qualitative synthesis, and for the quantitative synthesis, 11 studies with 1,413 patients were examined. Females were in the majority, by 72%, and the estimated mean age was 31.7 years.

Table 2

Study Designs for Papers Included in the Thesis

	Paper I	**Paper II**	**Paper III**
Design	Systematic review and meta-analysis	Cross-sectional	Cross-sectional and prospective
Time period and setting	All published papers from Nordic countries until November 2022	Pre-data February-November 2020, from Nidaros DPS.	Pre-data February 2020-February 2022, Nidaros DPS Treatment characteristics extracted February 2023
Participants	Patients in clinically representative therapy	Patients starting routine outpatient treatment	Patients starting routine outpatient treatment
Data collection			
Measure	Generic symptom measures	Primary PHQ-9 & GAD-7, secondary WSAS	Primary PHQ-9 & GAD-7, secondary WSAS
Additional data	Patient and therapist characteristics	Administrative patient data	Administrative patient data
Statistical methods	Meta-analysis	Structural equation modelling	Mixture modelling
Specific methods	Sub-group analysis Meta-regression	Confirmatory factor analysis, measurement invariance testing, latent path analysis	Confirmatory factor analysis, latent class analysis, factor mixture modelling

2.3.2. Paper II & III

Since patients from paper II were also included in paper III, there is an overlap. In paper II, a total of 857 patients consented to participate, while 26 were removed due to being completely missing. In paper III, 2,519 consented, of whom 46 were removed due to being completely missing. Women were in the majority in both studies (61% for paper II and 63%

for paper III), with a mean age of 30 years in both papers (median = 27, lower quartile = 23, upper quartile = 34 years).

Since the dataset from paper III was more complete, further information on demographic and diagnostic statistics will be provided from this dataset. Men were more often single (64%) compared to women (52%; total 57%; $\chi^2(1) = 28.92$, $p < .001$). Thirty-two percent of the sample were on sick leave before treatment started, with no statistically significant gender differences. Regarding CMHD, depressive disorders were most common (36%), followed by anxiety disorders (25%), PTSD (10%), bipolar (5%), somatisation (4%), and MADD (2%). Women were at greater risk of having an anxiety disorder (27%), compared to men (23%; $\chi^2(1) = 5.78$, $p = .016$), but no other gender differences were found. Other diagnoses were attention deficit hyperactivity disorder (10%), personality disorders (4%), and R45, symptoms and signs involving emotional state (21%).

One year after treatment started, patients attended a median of 11 sessions (lower quartile = 5, upper quartile = 18). There were no gender differences in the number of assessment sessions (median = 3, lower quartile = 1, upper quartile 5), but women attended more psychotherapy sessions (median = 7, lower quartile = 1, upper quartile = 12), than men (median = 5, lower quartile = 1, upper quartile = 10; $z = 3.79$, $p < .001$). Using a newer dataset (N = 3,676 patients) with patients answering post-treatment questionnaires (n = 934) the median treatment length was 177 days (lower quartile = 100, upper quartile = 315 days), with no gender differences in treatment length.

2.4. Data Collection and Instruments

2.4.1. Patient-Register Data

Register data was retrieved from a patient administration system (PAS). Gender and age were extracted from patients' national identity numbers. Patients answered questions regarding relationship and employment status in free text, which was manually converted to

dummy variables. Diagnoses using ICD-10 were retrieved from PAS. In paper II, some patients were still in treatment during the extraction process, so that diagnostic information from ICD-10-chapter V: Mental and behavioural disorders, was not retrieved from 23% of the patients. In paper III, diagnoses up to 1 year after treatment started were retrieved, and thus only 3% had no diagnostic data. In paper II, depression was retrieved from disorders found under ICD section V block F30-F39 (Mood [affective] disorders), and anxiety as F40-F48 (Anxiety, dissociative, stress-related, somatoform and other nonpsychotic mental disorders). In paper III, we also retrieved clusters from the most common disorders in each subsection, namely: bipolar, depressive, PTSD, other anxiety disorders and somatisation disorder, together with MADD. Comorbid CMHD was defined as having a concurrent F30-F39 and F40-F48 disorder during the treatment period. Furthermore, the number of direct consultations was retrieved from PAS and categorised into assessments and psychotherapy.

2.4.2. Electronic Self-Report Measures

Generic PROM, measuring a broad concept of mental ill health, was used in paper I, while specific PROM, measuring disorder-specific problems, was used in papers II and III. In the meta-analysis, three studies used the "Symptom check list-90-revised global severity subscale", and three studies used the "Behaviour and symptoms identification scale". The five other studies used: "36-item short form health survey mental component summary score", "Hospital anxiety and depression scale", "Clinical outcomes in routine evaluation outcome measure problems/symptoms-subscale", "Outcome questionnaire symptoms subscale", and "Brief symptom inventory". In papers II and III, the following measures were used:

The nine-item Patient Health Questionnaire (PHQ-9) is used to screen and measure the severity of major depressive disorder, using the symptoms from DSM-IV (Kroenke et al., 2001). It evaluates symptoms experienced during the last two weeks, using a 4-point Likert

scale from 0 (not at all) to 3 (almost every day), thus higher scores indicate higher levels of symptoms. Its psychometric properties have been evaluated in several studies (Obbarius et al., 2017), among others in a meta-analysis (k = 100; Negeri et al., 2021), and in Norway for adolescents and females with and without eating disorders (Burdzovic Andreas & Brunborg, 2017; Wisting et al., 2021). Different variations of its factor structure have been suggested. A systematic review found 19 studies supporting unidimensionality, and 12 two-factor, one three-factor and one study with a bi-factor solution (Lamela et al., 2020). Since then, several papers have found support for bi-factor models, for example in a study covering 58,272 participants (Bianchi et al., 2022), and another suggesting a bifactor (S − 1) model (De Man et al., 2021). There is also empirical support for MI across gender, ethnicity, students, levels of care, marital status, education levels and clinical conditions (Lamela et al., 2020), and across several countries (Shevlin, Butter, et al., 2022).

The seven-item Generalized Anxiety Disorder scale (GAD-7) was designed to assess generalised anxiety disorder (Spitzer et al., 2006) but has become a widely used measure for all sorts of anxiety-related problems (Obbarius et al., 2017). It uses an identical Likert scale to PHQ-9, and its psychometric properties have also been evaluated extensively, for example among outpatients with CMHD (Rutter & Brown, 2017), and amongst psychiatric out- and inpatients in Norway (Johnson et al., 2019). Unidimensionality (Shevlin, Butter, et al., 2022), two factors (Beard & Björgvinsson, 2014), and bifactor-(S − 1) have been suggested (De Man et al., 2021). Several aspects of MI have been found, among others across gender (Rutter & Brown, 2017), age (Bolgeo et al., 2023), and countries (Shevlin, Butter, et al., 2022).

The five-item Work and Social Adjustment Scale (WSAS) assesses functional impairment across five domains (work, home chores, social leisure, private leisure and relationships; Mundt et al., 2002). It uses a 9-point Likert scale from 0 (not at all) to 8 (very

severely impaired). Its psychometric properties have among other things been examined for people with and without personality disorders in Norway (Pedersen et al., 2017), and is suggested as a good supplement to PHQ-9 and GAD-7 (Zahra et al., 2014). The correlation between WSAS and a cognitive and somatic factor of GAD-7 and PHQ-9 has shown higher correlations between functional impairment and depression, than anxiety (Boothroyd et al., 2018).

2.5. Analyses

2.5.1. Paper I: Systematic Review and Meta-Analysis

A search string was created with a university librarian, using variations of mental health problems, routine treatment, outcome measures and the Nordic countries. A systematic search was conducted across several databases with no time limit. Abstracts were screened by two researchers, blinded from each other, and eligible studies were read in full text. Reference lists and citing papers were read, to find further potential studies. Qualitative descriptions of TAU and potential moderators were extracted, together with pre- and post-data from the studies included. To assess the quality of the studies, we used the Downs and Black checklist, since it could be used across randomised and non-randomised studies (Downs & Black, 1998). This was modified, in a similar procedure to another meta-analysis on TAU (Bear et al., 2020).

Comprehensive Meta-Analysis (version 3.3.070; Borenstein et al., 2014) was used as the statistical program for meta-analysing the results. Between- and within-group random effect sizes were estimated using Hedges *g*, based on pre-post scores, together with follow-up data. Pre-post-treatment correlations were imputed to 0.5. Heterogeneity was estimated with Q and I^2. Publication bias was analysed using Egger's regression intercept and Duval and Tweedie's trim-and-fill method. Subgroup analysis and meta-regression were conducted when information was available from ten or more studies.

2.5.2. Paper II and III: Descriptive Statistics

Stata (version 17; StataCorp, 2021) was used for descriptive statistics in papers II and III. Student's *t*-test and Pearson's x^2 were used to examine gender differences between demographic, diagnostic and symptom scores in paper II. In paper III, Student's *t*-test was used to examine symptom differences between non-comorbid and comorbid CMHD, while one-way analysis of variance with a Bonferroni post hoc test was used to examine sum-scores of PHQ-9 and GAD-7 across non-comorbid diagnostic categories.

2.5.3. Paper II and III: Structural Equation Modelling

A common way to measure levels of depression and anxiety is by counting the number of symptoms, or symptom scores, to an aggregated value. However, this will also include redundant noise – or *measurement error* – to the aggregated variable. To overcome this problem, structural equation modelling (SEM) could be used. This statistical framework uses factor analysis as its core and can be used to model depression and anxiety as *latent variables*, since they cannot be measured directly. It uses the shared variance of *observable variables,* or in our case, symptoms reported by the patients themselves. It also has the advantage of estimating the measurement and structural parts of the model simultaneously. Mplus (version 8.8; (Muthén & Muthén, 1998-2017) was used as the statistical software, to conduct the following applications of SEM:

Papers II and III Confirmatory factor analysis (CFA): To find an acceptable *measurement model,* which is a prerequisite for almost all other SEM approaches, CFA is conducted. In contrast to exploratory factor analysis, CFA should be guided through already established theory and empirical evidence, by the use of a series of fit indices. In paper II, we used an estimator for categorical data: mean and variance-adjusted weighted least square to examine PHQ-9 and GAD-7 independently. However, in paper III we used maximum

likelihood since other estimators are prone to not having models to converge for the more computer-demanding analysis that was conducted.

In paper II, we examined unidimensional factor structures, two-factor structures, and bi-factor models. The latter has been increasingly popular in psychology (Eid et al., 2018). They are composed of a general factor and specific factors. However, due to overfitting, they can yield a good global fit, even though the bifactor does not describe data well (Bornovalova et al., 2020). Thus, global fit is not enough to select a bifactor model over other models. Related to this, bifactors are also prone to anomalous results when traditional sampling methods are conducted (Eid et al., 2018). Except for changing the sampling design, bifactors can use a reference indicator: a bifactor-(S − 1) model. This also increases the interpretability of the model, since the specific factor will be the result in relation to the reference factor, while the general factor is the result of the specific factor, corrected for measurement error (Eid et al., 2018). Both PHQ-9 and GAD-7 have been suggested to be well-suitable with this model (De Man et al., 2021). Furthermore, we also examined internal consistency by using *composite reliability*, and the unidimensionality of the bifactor structure was examined using the *omega hierarchical*.

In paper III, we modified PHQ-9. We removed item 8 assessing movement symptoms, due to its ambiguity, since high scores can mean both too little, or too much movement. Furthermore, GAD-7 also assesses restlessness, with its item 5. Previous studies found that PHQ-9 motor symptoms had a strong connection to the restlessness symptom in GAD-7, but also had stronger connections to other GAD-7 symptoms, than PHQ-9 (Beard et al., 2016). Both theory and empirical findings thus led us to this modification. The modelling strategy was then to create a measurement model of depression and anxiety separately.

Paper II Measurement invariance: Multigroup CFA can be used to examine whether there is empirical support for groups of patients interpreting the latent construct similarly, namely by testing MI across these groups. We examined MI for PHQ-9 and GAD-7 across gender, diagnosis, and comorbidity. Since Mplus did not have an automatic procedure for examining MI for categorical data, a manual approach was pursued as a series of steps: to assess configural, metric and scalar invariance. If factor loadings and item intercepts were invariant, we concluded that there was support for scalar invariance, and that group comparisons could be justified.

Paper II Latent path analysis: When the measurement models were established, the structural part of SEM was used to test the relationship between the latent variables. Since we assumed that functional impairment was affected by depression and anxiety, we examined different parameters of the models and the results across different groups of patients.

Paper III Mixture models: One limitation of the variable-centred approach we took in paper II is the assumption that all individuals belong to the same homogeneous population (Clark et al., 2013). To examine alternative strategies to overcome the heterogeneity problem with CMHD, mixture models could be used to examine a finite number of latent subpopulations (Wang & Wang, 2012). Thus, in contrast to the variable-centred approach, we took a person-centred approach in paper III, to examine non-directly observed subgroups, by using Gaussian mixture models, or mixture models for short. Following other researchers' recommendations (Clark et al., 2013), we used a series of steps to examine latent subgroups: first, by establishing the measurement model. Second, we used latent class analysis (LCA) to gain an estimation of the upper number of classes to retrieve. Third, we examined factor-mixture models procedurally by starting with more restricted models. This took place with a set of information criteria and levels of entropy, together with theoretical implications.

There are several variants of FMM: FMM-1 is a non-parametric model, assuming no within-class heterogeneity (Clark et al., 2013). The parametric corresponding model, FMM-2, has freely estimated factor variances and loadings, making the assumption of within-class heterogeneity. In FMM-3, classes are determined by item thresholds, in comparison to FMM-1 and FMM-2, which are determined by factor means and variances. In FMM-4, all parameters are invariant, making the assumption that factors could be interpreted differently between classes (Clark et al., 2013). When the best model was retrieved, the predicted class membership associations with demographic variables were explored using the Mplus R3STEP option, and outcomes with the DCAT and BCH option.

2.6. Ethical Considerations

The research in this thesis complies with the 7th revision of Declaration of Helsinki, and laws and regulations in Norway. In paper I, we published a pre-registered protocol, in the International Prospective Register of Systematic Reviews (PROSPERO), to enhance the transparency of the research process (register number: CRD42020213988). Further, disagreements during the extraction process were discussed in the research group, until full consensus was reached.

For papers II and III, research was approved by the Regional Committee for Medical and Health Ethics (REK Midt-Norge, registration number: 2019/31836). Participation was voluntary, and there were no consequences for patients choosing not to participate. Informed consent was collected electronically, and participants were free to withdraw their consent at any time. The Norwegian Centre for Research Data approved the privacy impact assessment of the project for protecting sensitive personal information (registration number: 2020/605327). Data were collected using an electronic system provided by CheckWare. Both patients and therapists authorized themselves with a high level of data-safety verification.

Data was stored on a high security server, provided by Services for Sensitive Data, only accessible to project members at NTNU.

Several other ethical considerations were made for the overarching quality insurance project. The PROMs were to be valid, reliable, appropriate, and not too extensive. Thus, considerations were made regarding patients' burden and mental health workers' workload. Eventual distress from answering questions was handled by mental health workers since patients were already in a treatment context. Thus, the potential for harm was considered mitigated.

The research from this project has been published in international peer-reviewed journals and presented at international conferences. Further, we have provided workers and leaders at Nidaros DPS with data and statistics, to ensure that the findings are being utilized.

3. Summary of results

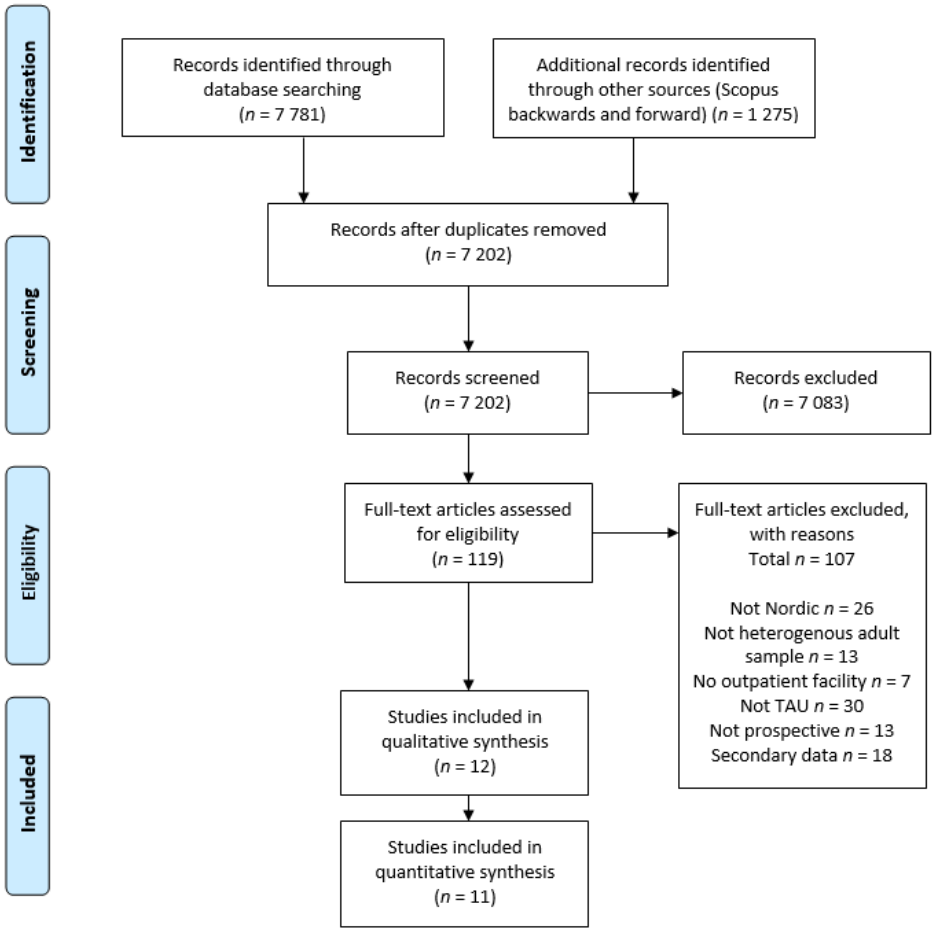
3.1. Paper I – Clinically Representative Therapy for Nordic Adult Outpatients with Common Mental Health Problems: A Systematic Review and Meta-Analysis

Twelve studies were included in a qualitative synthesis, out of which 11 were included in a quantitative analysis (see Figure 1). Their quality varied: none showed excellent quality, two good, six fair, and four studies were of poor quality. Six studies were conducted in Sweden, five in Norway, one in Denmark, and none were conducted in Finland or Iceland. One study examined both specialised and primary healthcare, seven specialised, and four primary mental health care. Eight of the studies reported CMHD, with an estimated proportion of 43% depressive disorders and 34% anxiety disorders. The most common treatment interventions were CBT, but also meta-cognitive, psychodynamic, support, systemic, humanistic, existential therapies, and psychoeducation were reported. Six studies reported a mean number of interventions, with an estimated mean of 7.13 sessions/patient. Four reported mean duration, totalling 100 days.

For a forest plot of the ESs, see Figure 2. The within-studies random effect model of 11 studies (12 ESs) showed a small to medium ES (g = 0.49), although with a substantial heterogeneity (I^2 = 90%). Four studies reported follow-up results, and the pooled ES was not significantly different from the post-scores. Between-studies random ES for eight studies (nine ESs) were small and favoured other interventions than TAU (g = -0.21) with moderate heterogeneity (I^2 = 50%). Follow-up for between-study ES was also not significantly different. Publication bias did not seem to bias the within-study ES, but between studies could be biased, having an adjusted ES of g = -0.06.

Figure 1

PRISMA Flow Chart of Included Studies

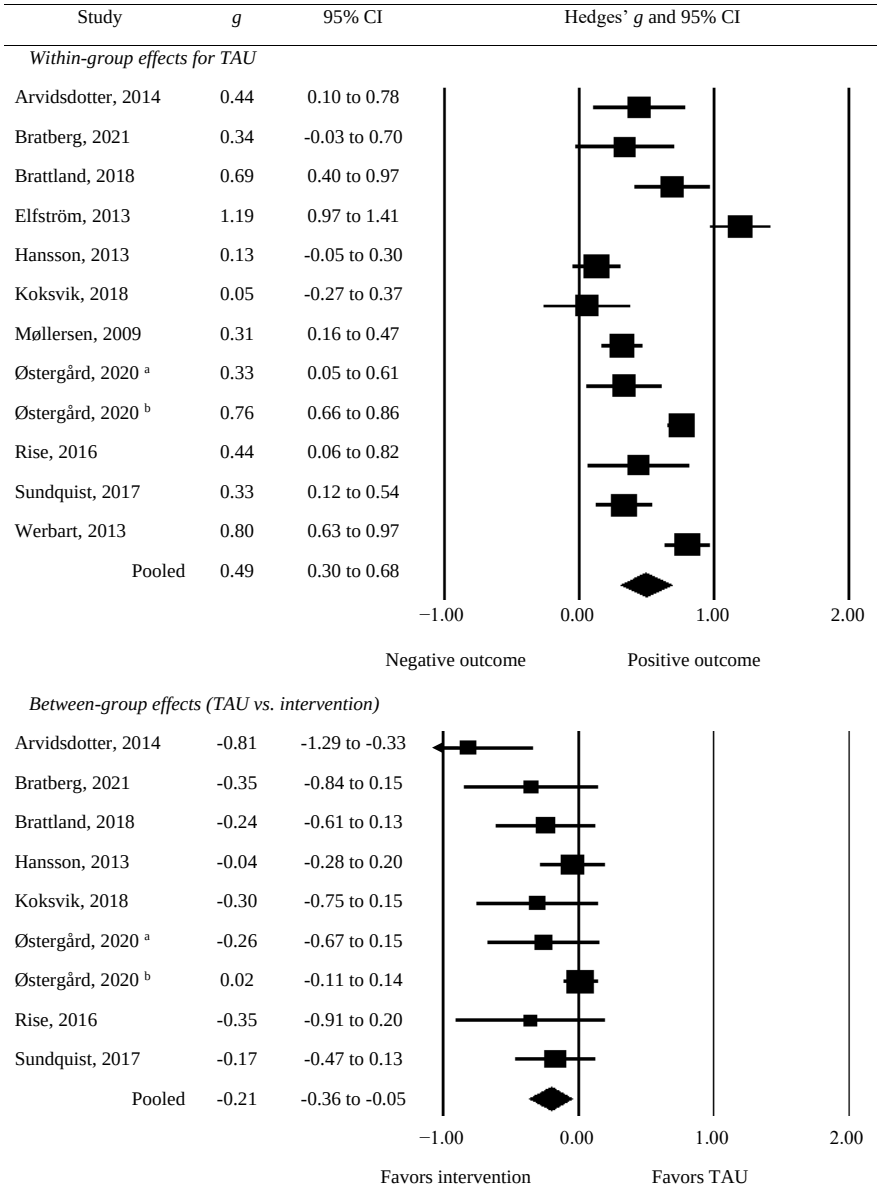


Note. From "Clinically representative therapy for Nordic adult outpatients with common mental health problems: A systematic review and meta-analysis" by M. Brattmyr., M. S. Lindberg., J. Lundqvist., L. Öst., S. Solem., O. Hjemdal., & A. Havnen, 2023. *Scandinavian Journal of Psychology*, 65(2), p. 314 (https://doi.org/10.1111/sjop.12976). CC BY 4.0.

Therefore, the publication bias-adjusted difference between TAU and other non-evidence interventions could be close to zero. Moderator analysis showed significantly lower ES and less heterogeneity in studies conducted in RCTs ($g = 0.33$, $I^2 = 59\%$) compared to open trials ($g = 0.68$, $I^2 = 92\%$). No other statistically significant differences were found.

Figure 2

Forest Plots Depicting Random Effect Sizes at Post-Treatment for TAU and Comparisons with other Interventions



Note. [a] Group treatment. [b] Individual treatment. From "Clinically representative therapy for Nordic adult outpatients with common mental health problems: A systematic review and meta-analysis" by M. Brattmyr., M. S. Lindberg., J. Lundqvist., L. Öst., S. Solem., O. Hjemdal., & A. Havnen, 2023. *Scandinavian Journal of Psychology*, *65*(2), p. 317 (https://doi.org/10.1111/sjop.12976). CC BY 4.0.

3.2. Paper II – Factor Structure, Measurement Invariance, and Concurrent Validity of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder scale-7 in a Norwegian Psychiatric Outpatient Sample

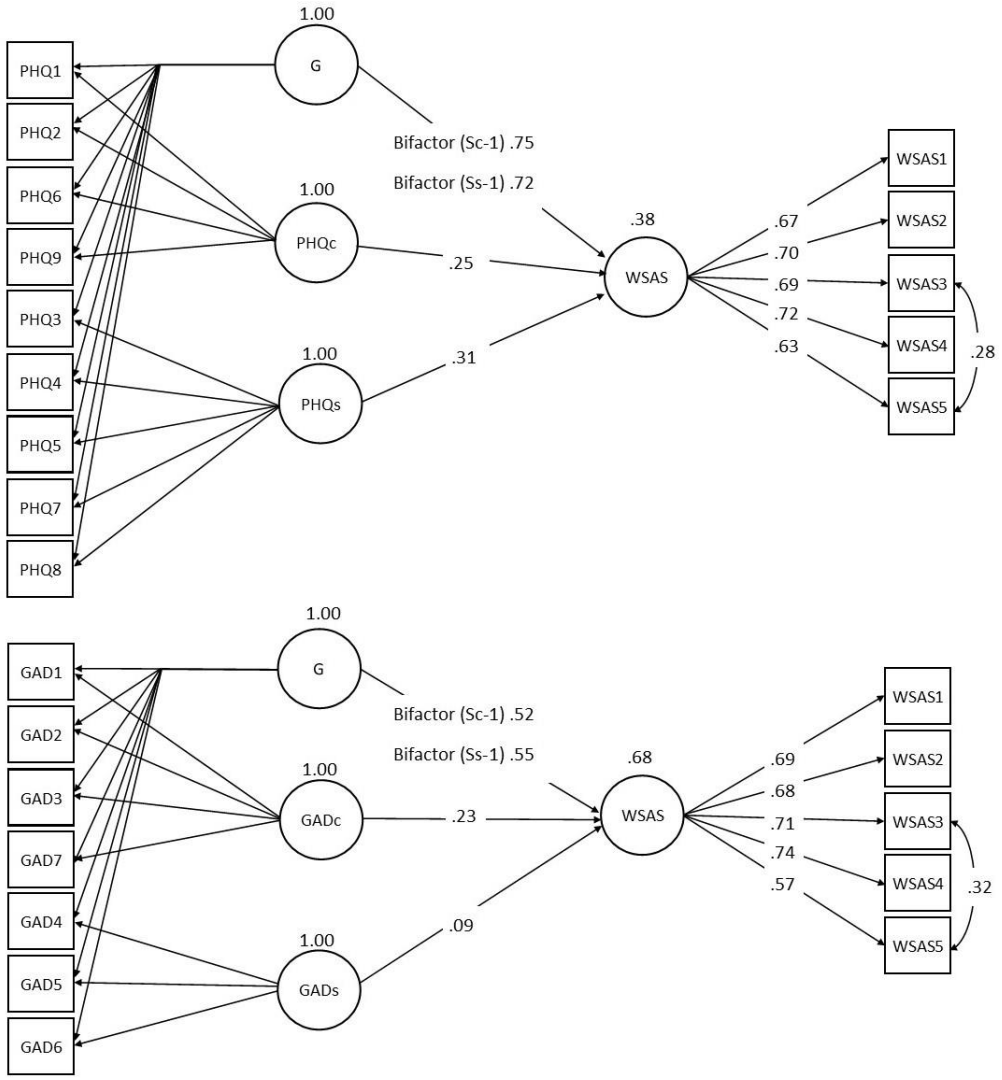
Unitary factor structures of PHQ-9 and GAD-7 showed poor model fit, while two-factor solutions specifying a somatic and cognitive factor of each instrument were acceptable. The symptoms that loaded a cognitive factor of PHQ-9 were: anhedonia, sadness, worthlessness, and suicidal ideation, while problems of sleep, fatigue, appetite, concentration, and being slow/restless loaded a somatic factor. For GAD-7, being nervous, not able to stop worrying, worrying too much, and being afraid loaded a cognitive factor, while trouble with relaxing, being unable to sit still, and being easily annoyed loaded a somatic factor.

Since there was a high correlation between the factors, we also examined a bifactor-(S − 1) solution of both instruments. In turn, these showed adequate fit, were theoretically justified, and were therefore accepted as the model for further examinations. The composite reliability showed good internal consistency for both instruments. Omega-hierarchical showed minor, but negligible issues with PHQ-9, while GAD-7 showed overall adequate properties. Thus, both instruments mainly showed unidimensional results.

We also found support for scalar invariance across gender, CMHD and comorbidity for both instruments. Therefore, group comparisons were justified. We found higher symptom severity for women and comorbidity, but no significant differences between patients diagnosed with anxiety versus depression. In the latent path analysis, higher PHQ-9 scores showed a greater degree of functional impairment, measured by WSAS, than GAD-7 (see Figure 3). The model where WSAS was regressed on the general factor of PHQ-9 also explained a higher degree of variance (r^2 = .62), than GAD-7 (r^2 = .32).

Figure 3

Latent Path Model of PHQ-9 and GAD-7 as Bifactor Models



Note. Over WSAS is the unexplained variance of the endogenous variable (ζ).
From "Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder scale-7 in a Norwegian psychiatric outpatient sample," by M. Brattmyr., M. S. Lindberg., S. Solem., O. Hjemdal., & A. Havnen, 2022. *BMC Psychiatry*, *22*(1), 461, p. 8 (https://doi.org/10.1186/s12888-022-04101-z). CC BY 4.0.

3.3. Paper III – Symptoms and Prevalence of Common Mental Disorders in a Heterogenous Outpatient Sample: an Investigation of Clinical Characteristics and Latent Subgroups

Out of all the patients in this study, 63% were diagnosed with a CMHD, while 14% of the sample had a comorbid CMHD. Correspondingly, 22% of the patients with a CMHD had a comorbid depressive and anxiety disorder. Patients with a comorbid CMHD reported more severe symptoms than non-comorbid counterparts. Patients with a comorbid CMHD had higher GAD-7 scores, compared to patients with non-comorbid depression, and higher PHQ-9 scores compared to non-comorbid anxiety disorders. There were no statistically significant gender differences in having a comorbid CMHD, compared to non-comorbid CMHD.

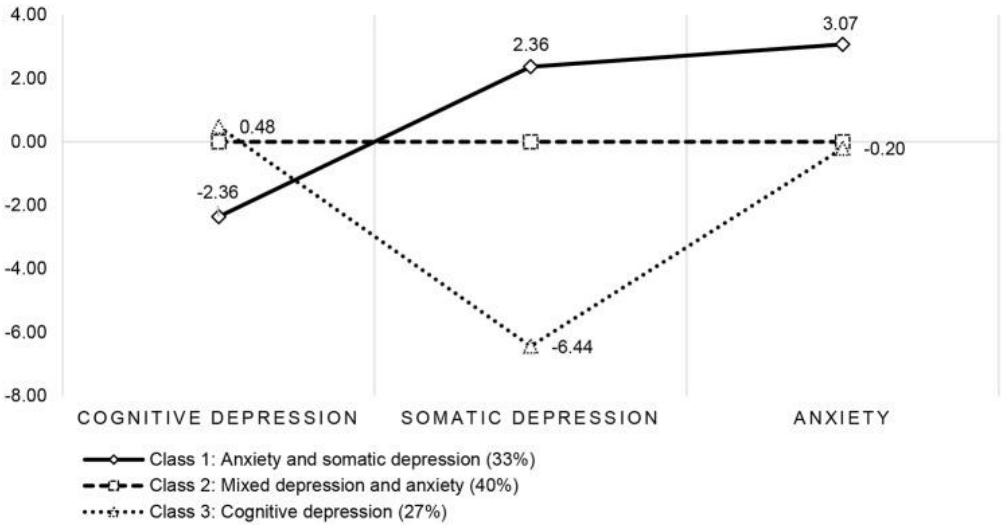
In a stepwise procedure, a three-factor solution of depression and anxiety was established, with adequate model fitness, specifying two PHQ-9 factors (somatic and cognitive factors), and a unitary structure of GAD-7 with two error covariances. We also conducted LCA, in the process of estimating the upper limit of classes, where three classes were found to be most appropriate. Applying the three-factor structure, we continued with FMM-1, which was not allowed for within-class variance and was not chosen due to theoretical implications, and non-suitable criterion fit indices. The FMM-2 three-factor four-class showed best criterion fit indices, but resulted in a spurious class, and were not picked. We therefore considered the three-factor, three-class FMM-2 as the best model, due to its second-best criterion fit indices, and best theoretical implication: a hybrid conceptualisation of CMHD as having both categorical and dimensional properties.

The classes were named Class 1: *Anxiety and somatic depression* (33%) with higher degrees of somatic depression and anxiety compared to the other classes; Class 2: *Mixed depression and anxiety* (40%) with a higher degree of cognitive depression than class 2, and a higher degree of somatic depression than class 3; Class 3: *Cognitive depression* (27%) with

equal degrees of cognitive depression and anxiety compared to class 2, but a lower degree of somatic depression compared to class 2 (see Figure 4).

Figure 4

Three-Factor Three-Classes Factor Mixture Model Latent Variable Factor Means



Note. From "Symptoms and prevalence of common mental disorders in a heterogenous outpatient sample: An investigation of clinical characteristics and latent subgroups," by M. Brattmyr., M. S. Lindberg., J. Lundqvist., S. Solem., O. Hjemdal., F. Anyan., & A. Havnen, 2023. *BMC Psychiatry*, *23*(1), 804, p. 6 (https://doi.org/10.1186/s12888-023-05314-6). CC BY 4.0.

For a description of the class probabilities associated with clinical covariates, see Figure 5. The anxiety and somatic depression class was characterised by older patients, with a higher probability of being single and being diagnosed with depression, anxiety disorders, and comorbid CMHD compared to the other classes. They also had the highest degree of functional impairment, a greater risk of being on sick leave, and the highest number of consultations.

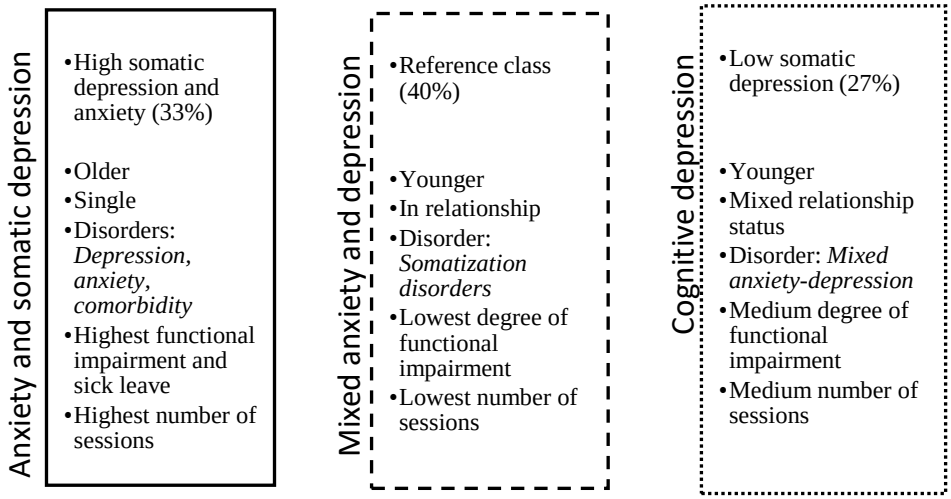
The mixed depression and anxiety class was characterised by patients in relationships, who had the lowest risk of being on sick leave, and a higher risk of somatisation disorders

compared to the cognitive depression class, but also the lowest risk of a range of CMHD and comorbidity. This class was associated with the fewest number of consultations and the lowest degree of functional impairment.

The cognitive depression class was characterised by patients with a medium degree of functional impairment, and number of sessions, who were at greater risk of a range of CMHD, compared to the mixed depression and anxiety class, but also had a higher association with MADD, compared to the anxiety and somatic depression class.

Figure 5

Description of Class Probabilities Association with Clinical Covariates



4. Discussion

This thesis examined several knowledge gaps concerning adults in routine outpatient facilities. We investigated the content and the effects of ordinary treatment, factor-related measurement properties of PHQ-9 and GAD-7, and whether latent subgroups could yield supporting information for patients starting up treatment.

In paper I, we found great variability in descriptions of clinically representative therapy, regarding therapists, patients and treatment contents. The within-group effect size was medium to low (g = 0.49), and the between-group effect size was close to minimal (g = -0.21; adjusted for publication bias g = -0.06). Although several sub-study comparisons and moderators were tested, the only statistically significant finding was that clinically representative therapies in RCTs yielded lower effect sizes than non-RCTs.

In paper II we used a bi-factor structure to examine factor-related properties of PHQ-9 and GAD-7 and found that there are psychometric justifications to assume unidimensionality with clinically representative samples. We also found support for MI across gender, CMHD, and comorbidity. In other words, there was no indication that different subgroups understood the scales' underlying constructs differently. We also found depressive symptoms to be associated with greater functional impairment, and that women and comorbid patients had common mental health problems to a higher degree.

In paper III, we found a lower-than-expected estimated prevalence of comorbid disorders. A hybrid latent subgroup model showed three classes. One third of the patients were associated with the most severe symptom class, which in turn was associated with more sessions, and were at greater risk of depressive and anxiety disorders. These findings will be further elaborated in the following section.

4.1.1. Paper I: Clinically representative therapy

- *We hypothesised a great variability of patient, therapist, and treatment characteristics of clinically representative therapy*

We found a great diversity of patient, therapist and treatment characteristics, which supports previous systematic reviews and meta-analyses, (Wampold et al., 2011; Watts et al., 2015). But also in line with previous studies, we found a lack of detailed descriptions of clinically representative therapy (Petersson et al., 2023). One major reason for this heterogeneity and lack of information is assumedly the complexity associated with researching mental health treatment. To get to the heart of the matter, all patients and mental health workers are unique. Therefore, Kazdin (2015) suggests three levels of sources of the variability of TAU: First, therapists often tailor their treatment to each patient, creating potential for great individual inter-therapist variability. Second, therapists at the same facility will differ in terms of preferred methods and experience, creating great intra-therapist variability. Third, there is no clear definition of usual treatment across facilities, giving even more variability (Kazdin, 2015).

This reasoning is in line with the finding that many therapists do not adhere to treatment guidelines (Johnson et al., 2016), and many mental health workers use integrative/eclectic approaches (Thoma & Cecero, 2009). However, it is only an assumption that most patients are given integrative/eclectic treatment, since there is a lack of research on what treatment modalities ordinary therapists use (Kazdin, 2015). There are, however, empirical findings on what effective therapists do in treatment. For example, interpersonal skills seem to be an important therapist factor, yet results seem embedded in the context where they are studied (Heinonen & Nissen-Lie, 2020), making studies in clinically representative settings called for.

This gives a need for studies to examine the content of clinically representative therapy, such as therapists' use of methods, but also therapists' and patients' general experience and satisfaction with routine care. Also, research extracting highly detailed information about clinically representative therapies would be informative, such as qualitative studies, or research using videotaped sessions of ordinary treatment to assess the content of the given treatment. There are thus several designs and approaches that could increase the knowledge of clinically representative therapy.

- *We also expected effect sizes comparable to other TAU studies*

We found similar effects of clinically representative therapies, compared to other meta-analyses of active TAU on mixed populations (Shadish et al., 2000; Wampold et al., 2011; Watts et al., 2015). An important finding is the lower effects of clinically representative therapy in RCTs compared to open trials. This finding supports the expectancy/attention argument by amongst others Freedland et al., (2011) and Kazdin (2015); patients who are not randomised into a specific intervention might be disappointed, and/or the therapists might give less attention to these patients. Therefore, study design could be an important factor when comparing clinically representative therapies, especially when results are evaluated by self-reported measures (Enck & Zipfel, 2019).

Since our meta-analysis could be used to compare, or benchmark against other studies, the logical question is: how well does Scandinavian clinically representative therapy perform in relation to other interventions, such as EBT? This is a complex question, since even if patient, therapist and treatment characteristics are equal, the specificity of instruments should be considered when benchmarking studies (Minami et al., 2008). There is namely a difference between disorder-specific PROM and generic PROM (Churruca et al., 2021). Specific PROM could be more responsive to treatment changes compared to generic PROM

when examining specific groups (De Beurs et al., 2019), and in these situations could yield higher effects (Shadish et al., 2000). Therefore, one could expect better accuracy in instruments measuring depression, for patients with depression, treated for depression, by therapists who are experts on depression. This would be less of a problem in efficacy studies, where there is a strict inclusion criterion and comorbid patients, for example, are excluded. However, patients in clinically representative therapies express a greater range of problems, and comorbidity is often highly prevalent. Therefore, both types of PROM complement each other, but generic PROMs could be more appropriate for benchmarking at the organisational or system level (Churruca et al., 2021).

Given a situation where all characteristics and measures are equal, there is still no given benchmarking method for mental health treatment (Delgadillo et al., 2014). A very conservative approach would be to compare the upper and lower 95% confidence interval (CI). In paper I, the pooled within-study ES 95% CI was 0.30–0.68, and the 95% CI in open trials was 0.41–0.95. Many effectiveness studies of CBT often yield smaller 95% CI over 1.00 (L.-G. Öst et al., 2022, 2023a, 2023b). However, CBT manuals are often designed to be used for specific disorders and are often measured by specific PROMs, making this comparison possibly deceptive. Psychodynamic therapies may sometimes be considered for treating patients with CMHD (National Institute for Health and Care Excellence, 2022). However, there are fewer effectiveness meta-analyses on this subject. One meta-analysis with sub-studies measuring general psychiatric symptoms, and also using no manual (k = 7), found a within-study effect size of d = 1.13, with a lower CI of 0.86 (Town et al., 2012). There is thus a difference in ES for non-manualised psychodynamic therapies since it had a higher ES compared to the pooled ES in our study. However, comparing the lower 95% CI from this study with the higher 95% CI from non-RCTs in paper I did not show any statistically significant difference.

Given that we found a neglectable difference between clinically representative therapy and other interventions in paper III, there is thus no substantial evidence of greater effects with non-manualized psychotherapies, compared to Scandinavian clinically representative therapy. Conversely, clinically representative therapy seems less effective than therapies guided by manuals since the collected empirical findings point in this direction. However, since we found great heterogeneity, and since we cannot account for every patient, therapist, and facility variable, we cannot conclude with any great certainty that this is a fact. This gives a need for larger meta-analyses of clinically representative therapies, to have a better understanding of ordinary treatment and its effects compared to other interventions.

Following the knowledge gaps identified in Paper 1, we have some recommendations for future research. If we were to recommend an up-to-date meta-analysis of clinical representative therapy, a comprehensive multilevel approach would be preferable, accounting for the levels of variation described by Kazdin (2015), as intra-therapist, inter-therapist, and facility variations. Studies should be of high quality, preferably using methods of high methodological quality, such as intention-to-treat instead of completer analysis. They should also provide a large number of potential moderators, for example: information regarding patient variables (age, gender, diagnosis, exclusion, attrition, expectations), therapist variables (experience, methods, supervision, caseload), treatment variables (number of sessions, duration, frequency), for both TAU and the specific intervention. If studies were to be found across several parts of the world, it would also be possible to link studies with regional data (continent, human development index, access to care, and so on). Ideally, all studies should have used the same set of instruments, or at least used a broad set of generic PROMs to measure a wide range of mental health problems, for direct comparison. If possible, other measures, such as specific PROMs, functional impairment, or measures of

client satisfaction, could yield further important information on clinically representative therapy.

Since validated measures are a prerequisite for high-quality, clinically representative studies, there should be extra focus on well-validated and commonly used instruments. However, many PROMs are not extensively validated in the settings in which they are used the most, creating uncertainties about their transferability to clinical settings (Patalay & Fried, 2021). This led us to our next research aim, investigated in paper II.

4.1.2. Paper II: Factor Properties of PHQ-9 and GAD-7

One concern expressed by Patalay and Fried (2021) was the transferability of PHQ-9 and GAD-7 to clinical settings, since most studies have used non-clinical samples. General populations could be more heterogeneous than clinical samples, creating greater variance, and items would be more likely to load one factor (Petersen et al., 2015). For example, other studies using a general population have found unidimensionality of these instruments without the need to specify a bifactor model (e.g. Shevlin, Butter, et al., 2022). At the same time, we also found support for unidimensionality of PHQ-9 and GAD-7 in a representative clinical sample, using a bifactor model. Thus, findings of unidimensionality of PHQ-9 and GAD-7 are consistent across non-clinical and clinical populations, although there are still exceptions (for example, Beard & Björgvinsson, 2014).

However, the unidimensional findings are not proof that these are the "right" structures of depression and anxiety. One problem concerns statistical equivalence: bi-factor models are prone to give a better model fit than other models, while higher order and network models could also have given an equally good fit (Bornovalova et al., 2020). This does not mean these models are merely alternative representations of the same data (Guyon et al., 2017). Yet if one aims to examine depression and anxiety as latent constructs, our findings

from paper II indicate that PHQ-9 and GAD-7 seem to capture these attributes well. Therefore, these instruments seem to measure what they were designed to measure, also for clinically representative patients at the start of treatment. This might answer the first concern by Patalay and Fried (2021).

At the same time, if they are used simultaneously, some considerations should be made. For example, both instruments have an item that covers restlessness. But the movement/restlessness item in PHQ-9 is ambiguously phrased; high scores could mean both moving slowly or being restless. Thus, item-specific similarities should be managed when they occur in models using both depression and anxiety. Furthermore, the somatic factors were highly correlated, and there are factor-specific considerations that need to be made. If these instruments or their corresponding taxonomies were to be revised, a recommendation would be to thoroughly reconsider the somatic symptoms of CMHD. Namely, should somatic symptoms of depression and anxiety be conceptualised as a common negative affectivity factor, as suggested by Clark & Watson (1991), or do we assume different somatic symptoms for each disorder? Since item and factor considerations need to be made, we also need to explore that these are understood in the same way across different groups of patients. Thus, the second aim was to:

- *Investigate measurement invariance across gender, disorder, and comorbidity*

Another concern raised by Patalay and Fried (2021) was the lack of studies investigating MI with clinical samples. In paper II, we found empirical support for scalar invariance across gender, CMHD, and comorbidity/non-comorbidity. In other words, we did not find evidence that patients from different groups interpret depression and anxiety differently, meaning that there are psychometric justifications to compare different groups of

patients. Women and comorbid patients had higher latent means on both PHQ-9 and GAD-7, but we found no significant difference between patients with anxiety versus depression.

Although the mechanisms behind the symptom differences across gender, CMHD and comorbid/non-comorbid disorders were outside the scope of the current thesis, it is important to note the complexity of the underlying factors which could create these differences. First of all, the aetiology of CMHD is unknown, and thus empirical findings and theoretical assumptions are needed to disentangle some of its complexity. Regarding gender, findings of women reporting higher psychological distress corroborate with other research (for example, Drapeau et al., 2014). Specifically, women tend to be more prone to use internalising coping strategies (Altemus et al., 2014). One theory to explain this is hormonal differences. However, after controlling for age, women still seem to be at greater risk of internalising disorders (Van Loo et al., 2023). In turn, this makes the hormonal explanation unlikely, since the gender differences have been found to persist even after menopausal age (Van Loo et al., 2023). Another theory is that cultural factors may lead to higher depression rates for women. For example, social inequalities have been shown to be associated with higher gender disparity regarding depression, and also men generally suffer more in highly unequal societies (Yu, 2018). Another empirical finding is that men to a wider extent use externalising strategies to tackle depressive and anxiety problems, and could therefore be more prone to substance use disorders (Altemus et al., 2014). In Norway, patients with these problems are treated at facilities specialised in substance use disorders. As a result, men with higher levels of distress could be underrepresented in clinical mental health studies.

Although the finding of differences regarding comorbid/non-comorbid disorders might seem trivial, namely patients with comorbid disorders having higher distress, comorbidity is complex. Anxious mood has been shown to be concurrent with, but also to precede, depressive mood to a higher degree than the other way around (Starr & Davila,

2012). At the same time, the fluctuations seem best explained by short intervals (Starr & Davila, 2012). There could thus be a temporal aspect of comorbid CMHD, which we were not able to examine since we used a cross-sectional design.

The non-finding of differences between PHQ-9 and GAD-7 for depressive and anxiety disorders seems counterintuitive. Patients with anxiety disorders should intuitively score higher on GAD-7, while patients with depressive disorders should score higher on PHQ-9. Since we were not able to extract diagnostic data from all patients, some were patients still in treatment during the extraction process. Thus, this non-finding could be due to noise in the diagnostic dataset. Therefore, a more stringent approach to sampling diagnostic information could have been beneficial, and paper III provides an example of this. The final aim for paper II was to:

- *Evaluate the concurrent validity of PHQ-9 and GAD-7 with functional impairment*

Since clinically significant levels of functional impairment are a prerequisite for a diagnosis of CMHD to be made, a high overlap between symptom severity and impairment could be assumed. Consistent with previous studies (McKnight et al., 2016; McKnight & Kashdan, 2009), we also found that depression had a stronger association with functional impairment than anxiety. However, previous studies have shown weaker associations than might be expected. One reason could be that measurement error has not been accounted for. When we used a latent path model that accounted for measurement error, this model explained a higher degree of variance in functional impairment (depression r^2 = .62, anxiety r^2 = .32) compared to pooled associations (depression r^2=.25, anxiety r^2=.12) in the systematic reviews by McKnight et al., (2016) and McKnight and Kashdan (2009). Measurement error seem thus to account for a great deal of unexplained variance.

Yet there was still much unexplained variance, potentially coming from several sources. One reason might be unaccounted feedback loops: symptoms of CMHD could lead to functional impairment, but functional impairment can lead to symptoms of CMHD. This is an important subject for further studies. Another potential source of unexplained variance could be derived from sample properties (McKnight & Kashdan, 2009). This hypothesis, together with the criticism of CMHD as separate disorders, led us to examine symptoms of CMHD from a population-based approach in paper III, instead of a variable-based approach.

4.1.3. Paper III: Symptom Severity and Latent Subgroups of CMHD

- *Estimated prevalence rates of CMHD and their associations with symptom severity*

Since clinically representative therapies do not have fixed time limits, we examined the prevalence rates one year after treatment started. With this approach, the prevalence rates for CMHD were similar to Norwegian DPS estimates (Pedersen & Lilleeng, 2019). However, we only found comorbid CMHD amongst 22% of patients with a CMHD, or 14% of the whole sample. This was lower than expected, also when compared to a PROM study from the Netherlands (Klein Breteler et al., 2021).

In contrast to Paper II, we found that patients with comorbid CMHD had more severe symptoms of anxiety, compared to patients with non-comorbid depression, and more severe symptoms of depression compared to patients with non-comorbid anxiety disorders. This could be due to power issues, since the results in paper II were close to being significant, and paper III had a larger sample size. In other words, patients with a major depressive disorder and no concurrent anxiety disorder had a higher degree of depression than patients with non-comorbid PTSD, anxiety and somatisation disorders. At the same time, we found no significant differences in anxiety scores between non-comorbid disorders.

Since there are several methods to estimate the prevalence rates of CMHD and comorbid CMHD in clinical samples, we cannot claim that these findings are the "true" prevalence rates. For example, one Norwegian study examined diagnostic discrepancies between an experienced psychologist/researcher with access to structural interviews and patients' journal data, and diagnoses made by ordinary clinicians (Øiesvold et al., 2013). They found that the experienced psychologist gave more diagnoses, and the agreement between her, and other clinicians regarding anxiety disorders, and therefore also comorbidity, was critically low. Since clinicians gave fewer diagnoses, the authors criticised the validity of diagnostic registers. However, another Norwegian study compared diagnoses from structural interviews with register data and concluded that comorbidity rates could be elevated in health registers (Torvik et al., 2018). This could be a particular concern when comparing clinical rates with the general population.

Given that current taxonomies could be conceptualised as pragmatic diagnosis tools (Van Loo & Romeijn, 2018), the prevalence rates of CMHD in clinically representative samples should be estimated by ordinary procedures. In our case, CMHD was common, but comorbidity was not as common as it hypothetically should be. Since registers of clinician administered disorders could be prone to excessive measurement error (Hyland & Shevlin., 2024), we also aimed to use patient's self-reported distress to:

- *Identify clinical latent subgroups*

After using CFA to identify an adequate three-factor structure of PHQ-9 and GAD-7, and using LCA to estimate three classes as the upper limit, we identified three classes based on patients' reported PHQ-9 and GAD-7 symptoms. This FMM-2 three-factor three classes could also be conceptualised as a hybrid model of a categorical and dimensional CMHD. In other words, these classes were permitted for within-class heterogeneity and were labelled

anxiety and somatic depression, characterised by a high degree of anxiety and somatic depression; *mixed depression and anxiety*, used as a reference class; and *cognitive depression,* characterised by a lower degree of somatic depression compared to the reference.

- *Analyse the subgroups' associations with service use and CMHD*

No gender differences were found between these classes, but patients of a higher age, and being single, were more likely to be associated with the most severe anxiety and somatic depression class. This class was also associated with higher service utilisation, risk of being on sick leave, and being diagnosed with depressive and anxiety disorders. Although the degree of somatic depression separated the other two classes, the group with the lowest somatic depression, the cognitive depression class, had more severe functional impairment, was more often on sick leave, had higher service utilisation, and a greater probability of being associated with a range of CMHD than the mixed depression and anxiety class.

Although our latent path model in paper II described more variance than reviews have found (McKnight et al., 2016; McKnight & Kashdan, 2009), since it accounted for measurement error, there is another potential source of the variance, namely sample properties. Thus, when accounting for latent subgroups, we found that one group had lower levels of somatic symptoms of depression, and at the same time higher levels of functional impairment in comparison to another group. This might have been due to different life circumstances: patients in the mixed depression and anxiety class experience greater somatic symptoms of depression, but might still be able to work/study, and therefore face more daily stress factors. On the other hand, patients associated with the cognitive depression class have greater functional impairment and higher service needs, but might also be less exposed to daily stress factors, since they are not working. This gives reason to believe that some functional impairment heterogeneity could be explained by groups of patients having

different life circumstances. For example, another study found that higher somatic, rather than cognitive, symptoms of depression were more associated with biomarkers for stress, which in turn was associated with longer chronicity of the disorder (Iob et al., 2020).

Why could somatic symptoms of depression be a good indicator for different subgroups of patients? One reason might be that people with chronic disorders have more somatic consequences, due to lifestyle factors. Studies taking lifestyle factors into account have still found an increased risk of other somatic problems, such as cardiovascular problems, stroke, diabetes and obesity (Penninx et al., 2013). Somatic symptoms of depression might thus be an indicator of other problems than depression, also affecting functional impairment, for example, musculoskeletal disorders. It is therefore recommended that other PROM studies also examine somatic aspects of the patient's health, to examine potential colliders.

4.1.4. Why are not More Clinically Representative Studies Using PROM Reported?

An important question is why more studies using PROM data for CMHD are not conducted. In paper I, we found 12 representative studies in the systematic review, which is a clear minority of all studies conducted in the field. This lack of studies also motivated paper II, where few studies have examined the factor properties of PHQ-9 and GAD-7 in clinically representative samples (Patalay & Fried, 2021). In paper III, we found that only a few studies have linked PROM with other registers, creating difficulties in comparing our findings, and very few have used factor mixture statistics to model CMHD, especially using clinically representative samples.

There are probably several reasons for this scarcity of clinically representative research. One aspect is the lack of clarity gained from researching these samples. Since RCTs are often regarded as the best design to gain trustworthy results, it is argued that the sum of

high-quality efficacy studies is best to yield correct evidence of treatments (Skarpsno, 2019). However, it is unrealistic to conduct an infinite amount of RCTs to cover all patient groups, and therefore efficacy studies are suggested to be of little use in clinical practice (Juul, 2019). Since many therapists do not adhere to specific treatment modalities (Johnson et al., 2016), an infinite amount of RCTs would still not give a realistic perspective on real-world treatments.

A second aspect is the limitations of the current reward systems: statistically significant findings and novel interventions are rewarded in the form of funding, publication in prestigious journals, and academic promotion, at the cost of the quality of the studies (Ioannidis et al., 2014). This could make clinically representative therapies of little interest to researchers, since they often lack the novelty and large effect sizes that are often funded. At the same time, since mental health research is underfinanced, and current research strategies could be non-beneficial to real-world patients, there have been calls for changing the research agenda (Hazo et al., 2019). Consequently, several policy documents, from WHO (World Health Organization, 2021), the OECD (Hewlett & Moran, 2014), the EU (Directorate-General for Health and Food Safety, 2023) and Norway (Ministry of Health and Care Services, 2023) have advocated more real-world research, but it takes time to implement these initiatives.

The third aspect relates to practical research. Many initiatives have implemented PROM, but do not collect informed consent for research purposes (Al Saya et al., 2021). This means that results from many PROM initiatives are not published, at least not in peer-reviewed papers. To answer complex research questions, however, PROM data needs to be integrated with other register data, which could be inaccessible due to jurisdictional limitations (Al Saya et al., 2021). For example, since real-world data seldom has fixed time points, and PROM is often collected before and after treatment, there is a need to collect

supplementary data, to control treatment dosage. Since PROM data lacks diagnostic data, it can be difficult to compare clinically relevant patient groups. PROM studies with integrated register data are scarce (Gelkopf et al., 2022), which has led to a call for integrating approaches, to fully enhance PROM's potential (Al Saya et al., 2021).

We have hereby presented three approaches to studying clinically representative samples using PROM. In Paper I, we extracted all studies of clinically representative therapies in the Nordic countries we could find that reported PROM, together with potential moderators. In paper II, we showed how diagnostic data could inform several aspects of two commonly used PROMs. In paper III, we further analysed latent subgroups of PROM, together with service use and other clinically relevant variables. However, the low number of similar published studies reduced the opportunity for relevant comparisons.

There is, however, an increasing number of other real-world informing initiatives, such as patient-focused studies using routine outcome monitoring (ROM). Although ROM and PROM are often used interchangeably, ROM are mainly feedback tools, deployed at every treatment session. On the other hand, PROM is mainly used pre- and post-treatment. ROM can thus yield valuable process details for routine treatments, but the effects compared to no-feedback seem inconclusive (Kendrick et al., 2016). Furthermore, on monitoring the patients, the behaviour of therapists might arguably change, and thus it is not considered a clinically representative therapy according to the criteria of Shadish et al (1997). Thus, PROM, but not ROM, can be informative for clinically representative therapy, if patients and therapists are not blinded to the results. At the same time, ROM and PROM, together with other approaches, can complement each other to gain real-world evidence.

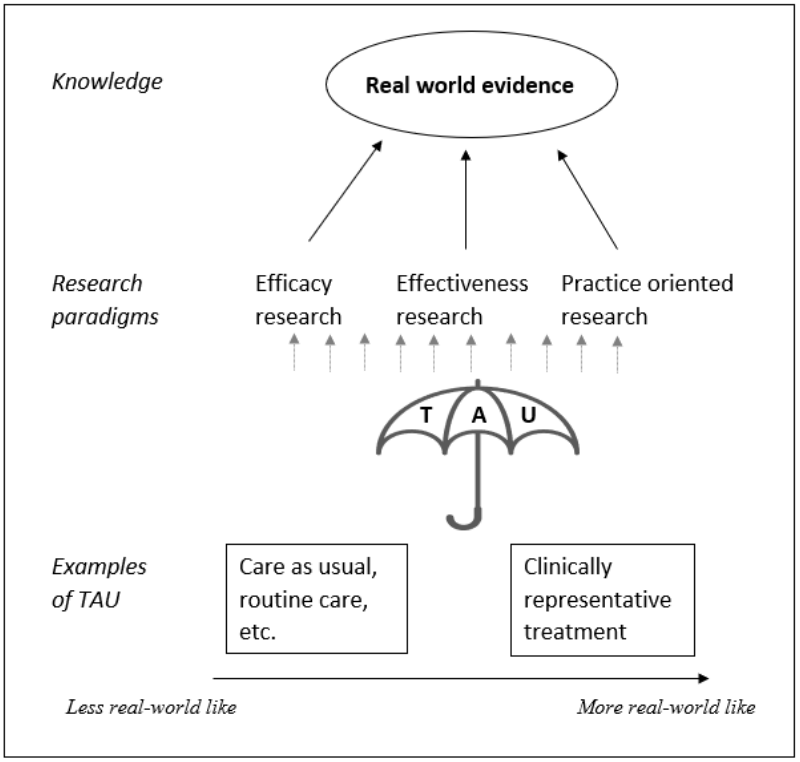
4.1.5. Contextualization of Clinically Representative Therapy

The definition of clinically representative therapy seems to be inconsistent in the literature. For example, clinically representative research has been set as an overarching research paradigm for effectiveness research and practice-oriented research, at the same time distinct from efficacy research (Cahill, 2013). Although there are reasons to agree that studies of clinically representative research should be a central research agenda, their conceptualisation misses the point of the strictness defined by, among others, Shadish et al., (1997). As a consequence, there seems to be a lack of studies of clinically representative therapy in mental health care, at least as published in the Nordic countries. Therefore, setting clinically representative research as an overarching research paradigm is unjustified. The results from paper I showed that in an efficacy study, a TAU condition can sometimes be defined as a clinically representative therapy. This makes it arguably more suitable for clinically representative therapy to be one of many TAUs, which in turn can be used to gain real-world evidence through several research paradigms (see figure 6).

In this conceptualisation, other control categories are not represented, since waiting lists, psychological placebo, and sham treatments should not be considered to be treatment as usual. Furthermore, it should not be interpreted that all efficacy, effectiveness and practice-oriented research are suggested to yield real-world evidence, although they might, depending on the sampling procedure and research design. Accordingly, TAU is more suitable as an umbrella term for ordinary practice, while clinically representative therapy could be the most restrictive, and possibly the most real-world-like category of TAU.

Figure 6

TAU is an Umbrella Term to Gain Real-World Evidence through Different Research Paradigms



4.1.6. The Epistemology of PROM

Measurement is at the core of psychology and is a central aspect of this thesis. However, it is disputed how well measures capture the reality. Traditionally, psychological measures have been suggested to work like other quantitative measures, such as in physics. As a consequence, it has been justified to add up symptoms of CMHD to an aggregated value. This was the approach in paper I. However, except for adding redundant noise to the variables, simply quantifying psychological attributes has axioms that cannot be fulfilled (Sherry, 2011). This is related to the ontology of psychological constructs. Mental attributes, such as depression, differ ontologically from physical attributes. Since mental disorders are embedded within a context, they cannot be reduced to a disorder of the brain (Olbert & Gala,

2015). While some initiatives have explicitly stated that CMHDs are brain disorders (e.g. RDoC; Insel et al., 2010), and others make this connection implicitly (DSM; Castiglioni & Laudisa, 2015), there is reason to believe that this logic is flawed (Olbert & Gala, 2015).

At the same time, depression and anxiety should not be conceptualised merely as social constructs. Even though diagnostic categories from classification systems were created by expert consensus, conceptualisations of depression and anxiety have existed for thousands of years and seem to be a universal human experience. The concepts thus touch upon something that can be assumed to exist in the real world, but in the form of a subjective experience for the patients (Guyon et al., 2017). These experiences can be collected through PROM, but sensible statistical considerations are necessary for conceptualising them in epistemologically justified ways. Therefore, we took a latent variable approach in papers II and III.

From a latent variable perspective, symptoms of depression and anxiety could be conceptualised as manifestations of their underlying attributes (Guyon et al., 2017), in our case, CMHD. However, this conceptualisation has been criticised. Underlying attributes are suggested to imply a natural categorical understanding of mental disorders, which leads to the assumption of brain disorders (Fried & Nesse, 2015). As an alternative and assumedly *more real* approach, mental disorders have been suggested to be modelled better as complex dynamic systems, as in the case with network models – where, for example, symptoms of CMHD mutually influence each other (Beard et al., 2016). In turn, traits may emerge from these systems, such as neuroticism (Cramer et al., 2012). However, the epistemological problem is to explain these new psychological attributes, without implying circular arguments (Guyon et al., 2017). Instead, latent variable advocates, such as Guyon et al., argue that these attributes do not exist as entities, but they do in fact correspond to reality. This reality, for mental disorders, does not have to be entities like brain disorders, but they are rather a

simplified model of the complex reality – making it possible to take a *pragmatistic-realistic* epistemological approach (Guyon et al., 2017). From this perspective, the ontology of mental disorders can be framed as psychological attributes, within a certain social context, and not as valid truths in every setting (Guyon et al., 2017). This legitimises the current studies of CMHD using PROM from a latent variable perspective, in a Norwegian real-world outpatient context. At the same time, the generalisability of the results could be limited to the context in which it was studied. Some of these contextual limitations will be discussed in the next section.

4.2. Method Discussion

4.2.1. Paper I

Selection Bias in Paper I.

Since we used a strict inclusion criterion, which paradoxically was quite wide: only studies with a wide variety of patients, and therapists with a great deal of freedom, this could have stopped us from finding studies that were also delivered in real-world settings. For example, there are reasons to believe that some facilities only provide CBT for a specific disorder, or that facilities provide a fixed number of sessions. Since we do not have empirical data for how many of these services exist in the Nordic countries, we do not know how representable clinically representative therapy actually is. Furthermore, we did not find any moderating effects of attrition or the number of excluded patients. Yet there could be systematic selection biases for which we were not able to control, and more high quality studies of clinical representative therapy are recommended, to yield a better estimate of the content and effect of clinically representative therapy.

Is it Possible to Synthesise and Compare Heterogeneous Results?

Meta-analyses have been criticised for comparing apples and oranges. This should first and foremost be a problem when using fixed effect sizes, making assumptions of a true

effect size across all studies. However, this is an unrealistic assumption in psychological research, making random effect models more suitable, since there should be several effects across the studies (Dettori et al., 2022). Assuming different true effect sizes, studies of apples and oranges could thus give important information on fruits (Deeks et al., 2019). At the same time, the results come with a serious caveat: our findings of a small to moderate within-study effect size are not to say that all patients have a small to medium effect of ordinary treatments. Instead, it is an estimate of the general effects for the many people treated within routine facilities. Although we did not have the power to conduct one, a multilevel approach would be even more suitable, accounting for higher clusters, such as regional-specific variance.

The great heterogeneity of real-world treatments does not, however, necessarily have to be a problem. First, there are statistical considerations: The I^2 is the estimated proportion of variance due to heterogeneity. However, I^2 is not only imprecise, but it can also be biased, due to small sample sizes (Von Hippel, 2015). Furthermore, when using subsamples of TAU in RCTs, we found similar I^2 as other meta-analyses (Wampold et al., 2011; Watts et al., 2015), revealing that this concern is not unique to paper I. Secondly, there are clinical considerations: the heterogeneity itself gives an opportunity to examine subgroups (Segal et al., 2023), thereby emphasising the importance of alternative strategies to examine subgroups, which took place in paper III.

4.2.2. Paper II & III

Selection Bias in Papers II & III.

A major limitation is that we could not control for several aspects of attrition. First, the number of patients who did not receive invitations to participate is unknown. Second, we were not able to analyse patterns of patients who did not consent to participate, due to ethical considerations. This makes it difficult to examine selection bias in papers II and III.

Previous studies have found that patients with comorbid CMHD could be overrepresented in clinical samples (Torvik et al., 2018). This suggests a possibility of the Berkson's paradox, namely that a patient with comorbid disorders may be overrepresented in clinical studies, since they have a higher probability of searching for treatment (Berkson, 1946). At the same time, the estimates of comorbidity in our studies were lower than expected, which corresponds to another study (Øiesvold et al., 2013). Mental health workers could be focused on finding a main diagnosis, but also have a working hypothesis of a comorbid disorder. At the same time, if comorbid CMHD is not formally addressed using a formal diagnosis, the estimated prevalence in diagnostic registers could be underestimated, and therefore under-communicated (Øiesvold et al., 2013). Thus, from the approach of estimating prevalence from diagnostic data, we still know little of the "true" estimates of comorbid CMHD, but we gained information on what is formally diagnosed during clinically representative therapy. In turn, these estimates correspond to a low degree with the latent subgroups we extracted.

Furthermore, since patients with substance abuse were treated at other facilities, they are not represented in the samples included. In turn, this could have affected the findings of higher symptoms of depression and anxiety in women. Men tend to have a higher risk of developing substance use syndromes (Altemus et al., 2014). If men with higher levels of distress use alcohol or other substances to tackle their problems, they could be disqualified from ordinary treatment in mental health services, since patients with these problems are treated by other specialised facilities for alcohol or substance abuse. We therefore cannot conclude that women suffer more than men in general, but only that this was a finding for the sample we used. It is therefore recommended that clinically representative research more thoroughly examine Berkson's paradox and other selection biases in the sampling process.

Measurement Aspects that were Not Examined.

Given the common argument that quality measures should be effective regarding administration time and only measure the most important aspects of patients' problems (Schang et al., 2021), it is important not to overload patients and mental health workers with unnecessary measures. Yet as a consequence, other important aspects might have been overlooked. The two instruments we assessed and used, PHQ-9 and GAD-7, are closely related to internalising disorders. Concerning the previously mentioned risk of sampling bias, where men tend to be prone to substance use disorders, they also use externalising coping mechanisms to a greater degree than women (Altemus et al., 2014). Measures examining externalised problems could thus potentially capture another important dimension of problems found in clinically representative samples.

The sampling period, partly coinciding with the Covid-19 lockdown, may limit the generalisability of papers II and III. Since data collection was conducted during the Covid-19 pandemic, there is a risk of increased symptom burden due to social restriction measures, as studies have suggested an increased symptom burden during the pandemic (Santomauro et al., 2021). However, other studies have shown a decrease in the prevalence of CMHD in Trondheim, Norway, during the initial phase of the pandemic (Knudsen et al., 2021). These conflicting results make it difficult to conclude to what degree Covid-19 affected our results.

Except for being suitable measures of symptom severity, PHQ-9 and GAD-7 are sometimes proposed for being suitable diagnostic tools. Due to low specificity when compared to clinical interviews, however, their diagnostic utility has been questioned (Pranckeviciene et al., 2022). They have therefore been suggested as more suitable as screening tools. However, such an approach assumes that clinical interviews reveal better estimates for CMHD, which could be disputed (Hyland & Shevlin, 2024). As an alternative design to assess their diagnostic properties, PHQ-9 and GAD-7 latent scores could be

compared with the patient's expressed focal mental distress, or possibly by using other mixed-method designs.

Is a Hybrid Approach a More Real Understanding of Mental Disorders?

The ontological assumptions of mental disorders as brain disorders by DSM/ICD/RDoC, do not seem to be supported since theory does not correspond to empirical findings. Further, researchers who seek to find the "true" form of mental disorders use methods with epistemological flaws, since all models are mere simplifications of the real world. Given the flexibility of modelling mental disorders, the structure of for example HiTOP cannot represent mental disorder in a real sense, since it is like […] *pouring water into an ice tray, freezing it, and then claiming the ice cubes are the empirical structure of ice* (page 286; Haeffel et al., 2022). Mental disorders can sometimes be better modelled as categorical constructs, and sometimes dimensional, thus, a pluralistic approach has therefore been suggested as a better alternative (Stein et al., 2022). In our case, we used a hybrid approach in paper III, since it could theoretically overcome this flexibility problem (Borsboom et al., 2016). We make no claim, however, that this will carve nature by its joints, since it makes most theoretically sense to place latent constructs within a pragmatistic-realistic epistemological theoretical framework. Our results will therefore be embedded in the context it was studied, while it does not ignore the real distress patients experience.

Many papers using latent variable approaches suffer from weak theoretical aims (Fried, 2020), for which especially paper III could be criticised. Although Fried (2020) has criticised authors of latent-variable papers for not explicitly stating that their goal has been data exploration, we did state this as the aim of the paper. On the other hand, one suggested strength of mixture models has been the hypothesis-driven nosology (Feczko et al., 2019). This was not the primary goal of the paper, and the model results were not unequivocal, nor was the information criterion easily interpretable, and the class results were ambiguous, since

mainly one class predicted depressive and anxiety disorders. Either we modelled CMHD wrong, or the complexity of CMHD is reflected in our results. We did not find evidence of a separate anxiety, depression and comorbid class like previous studies did (Shevlin, Hyland, et al., 2022), but rather three classes characterised by different degrees of symptom severity. As a result, our data-driven subgroups of PROM did not correspond well to the diagnostic categories, since their intercorrelations were ambiguous (Kendler, 2017).

If DSM should be used to both measure and define CMHD, as suggested by van Loo and Romeijn (2018), there is a need for more bottom-up research, with real-world data, to see how well symptoms of CMHD correspond to their respective clinician administered diagnoses. If there is consistent empirical evidence that there are no corresponding symptom clusters in clinically representative samples, there will either be a need to re-conceptualise CMHD, as they are formulated in current taxonomies, or accept that clinician and patient assessments do not match each other well. Either way, we should not exclude that hybrid approaches could potentially mitigate some of the theoretical, and methodological problems associated with DSM, ICD, HiTOP and RDoC. Given their nature, however, taxonomies of mental health disorders are only models, and models will never become true representations of the complex reality.

4.3. Research Implications

Since there is limited evidence about the effect of routine treatment for patients in public mental healthcare, there is a need for high-quality research on clinically representative therapy. For studies to be of high quality, they should at least provide detailed information concerning patient-, therapist-, and treatment variables. If clinically representative therapy is a control group in RCTs, there is a need to control for patients' and therapists' expectations of this treatment arm. If researchers believe the intervention could disappoint patients, or make therapists less efficacious, observational designs would be preferable to estimate the effects

of clinically representative therapy. Since evaluations using PROM could be affected by patient expectations (Enck & Zipfel, 2019), it could also be of scientific value to examine if the conflicting results of deterioration/improvement of patients being on the waiting list could be explained by study design (Faltinsen et al., 2022; Furukawa et al., 2014; Høstmælingen et al., 2023). In other words, is there an anticipation effect for patients on a waiting list and in treatment in observational studies, while there is a disappointment effect for patients not randomised into a novel treatment?

Further, PHQ-9 and GAD-7 seem well suitable for examining patients starting treatment and for comparing their symptom scores across several groups. However, researchers should be aware of the epistemological limitations of aggregating their sums without accounting for measurement error. Using latent models of PHQ-9 and GAD-7 could therefore reduce a great deal of the unexplained variance, for example when examining the association between patients' symptom scores with functional impairment.

The estimated prevalence of comorbidity in register data was lower than expected. Accounting for latent subgroups, one-third of the patients were associated with the more severe class. Latent subgroups could potentially mitigate the heterogeneity problem of clinically representative samples. Therefore, further research on factor mixture models of CMHD for clinically representative samples is needed.

We have shown applications of PROM connected with diagnostic and service usage data. A next step could be to assess process indicators with clinically representative therapy, such as examining the effects of treatment intensity, modelled by good-enough-levels. These could be measured by a range of outcomes, such as symptom reduction, gained functionality, improved health-related quality of life, and client satisfaction. Other valuable scientific insights could come from examining therapist factors (Heinonen & Nissen-Lie, 2020). Such

studies could examine possible mediators and moderators of treatment outcomes in clinically representative therapy. Such factors could relate to structural factors (e.g., waiting time, therapists' caseloads), therapist factors (e.g., self-relatedness), patient factors (e.g., treatment expectations), common factors (Wampold, 2015) such as the therapeutic working alliance, and other suggested mechanisms of change (e.g., change in dysfunctional cognitions).

Studies on clinically representative therapy could be important for bridging the gap between research and clinical work. Since studies on clinically representative treatment are scarce, and often poorly described, there is enormous potential in increasing research on real world clinics, for the better to the common patients and the common health workers.

4.4. Clinical Implications

Mental health workers should be aware of the knowledge gap between researched therapies and what is done in everyday treatment. This gap creates an uncertainty of the effects of therapeutic methods. With the current surge of PROM in mental health treatment, the research of clinically representative therapy could potentially increase, and could therefore be used to yield better estimates of the effects of a range of mental health treatments.

The increased availability of PROM could also be directly beneficial for people involved in mental health treatment. Clinicians could gain information on patients' self-reported status, making it a potentially useful clinical tool for identifying patients' degree of distress and for making a prognosis. PROM is also a tool for fostering patient-centred treatment, which could improve the patients' agency of treatment. There is, however, a need to not overload patients and mental health workers with excessive assessment. Two short instruments, PHQ-9 and GAD-7, have in this thesis been found to be suitable and may be used to measure symptom severity and to compare different groups at the start of treatment.

Together with other short instruments, measuring clinically relevant aspects of patients' distress, such as functional impairment, health-related quality of life, and client satisfaction, PROM could increase the knowledge of several dimensions of patients' distress. In turn, this may be important information about the patient for the therapist to better tailor the treatment provided.

If quality measures are to become quality indicators, however, there is a need for more studies with representative samples. For instance, this thesis has shown that somatic symptoms of depression could be an indicator of patients with higher treatment needs. A possible implication that should be investigated in future research is whether interventions aimed at reducing somatic symptoms may be beneficial for difficult-to-treat patients.

Since empirical evidence and theoretical underpinnings point toward mental health disorders as psychological attributes, not as entities, there are justifications for conceptualising patients as experts on their distress, and clinicians' experts on administrating treatment. Therefore, by fostering patient-centred treatment and research using PROM, the voices of the patients might become clearer, and they could be positioned as central actors in mental health treatment and research, together with mental health workers.

5. Conclusions

There is a knowledge gap between treatment content and outcomes of routine treatment. Clinically representative therapy seems less effective than EBT, but there is not enough empirical evidence to conclude by how much. Since research is lacking on several aspects of clinically representative therapies, several initiatives have encouraged the implementation of PROM for patients with mental disorders. Linking PROM with already available data registers could be a powerful approach to gaining access to clinically important aspects of ordinary treatment. However, there is a need for research on these initiatives, to make information available to stakeholders and the research community.

This thesis has used three different approaches to using PROM to gain insights into some identified knowledge gaps. In paper I, we have shown that PROM can give knowledge of treatment content and effects; in paper II we have shown the validity of common measures; and in paper III we have shown potential latent subgroups. However, since there is a general lack of PROM research linked to register data, it is challenging to set guidelines for how to accommodate these results. Therefore, more research to explore how PROM may be used as quality indicators is needed.

Nevertheless, the results from this thesis can give some insights into the utility of PROM as quality indicators. The effects of clinically representative therapies can be used as a baseline when comparing implemented interventions with clinically representative populations, if patient, therapist, and treatment variables are similar. If the implemented intervention does not show statistical differences from clinical representative therapy, the treatment should not be regarded as better than ordinary treatment. Furthermore, PHQ-9 and GAD-7 can be used as quality measures in clinically representative samples. If patient groups have a pre-score that deviates significantly from the results presented in our papers, there is a

need to examine the reasons for this. Furthermore, somatic symptoms of depression could be an important indicator of the service needs of a patient. Patients with high scores for somatic depression, together with high anxiety levels, should be followed up closely.

6. References

Altemus, M., Sarvaiya, N., & Neill Epperson, C. (2014). Sex differences in anxiety and depression clinical perspectives. *Frontiers in Neuroendocrinology*, *35*(3), 320–330. https://doi.org/10.1016/j.yfrne.2014.05.004

American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). American Psychiatric Association.

Baxter, A. J., Scott, K. M., Vos, T., & Whiteford, H. A. (2013). Global prevalence of anxiety disorders: A systematic review and meta-regression. *Psychological Medicine*, *43*(5), 897–910. https://doi.org/10.1017/S003329171200147X

Bear, H. A., Edbrooke-Childs, J., Norton, S., Krause, K. R., & Wolpert, M. (2020). Systematic review and meta-analysis: Outcomes of routine specialist mental health care for young people with depression and/or anxiety. *Journal of the American Academy of Child & Adolescent Psychiatry*, *59*(7), 810–841. https://doi.org/10.1016/j.jaac.2019.12.002

Beard, C., & Björgvinsson, T. (2014). Beyond generalized anxiety disorder: Psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *Journal of Anxiety Disorders*, *28*(6), 547–552. https://doi.org/10.1016/j.janxdis.2014.06.002

Beard, C., Millner, A. J., Forgeard, M. J. C., Fried, E. I., Hsu, K. J., Treadway, M. T., Leonard, C. V., Kertz, S. J., & Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine*, *46*(16), 3359–3369. https://doi.org/10.1017/S0033291716002300

Bergmark, M., Sundberg, L. R., Markström, U., & Rosenberg, D. (2022). Evidence-based methods in rural areas – Knowledge and national guideline utilization in mental health service development. *Journal of Evidence-Based Social Work*, *19*(2), 161–184. https://doi.org/10.1080/26408066.2021.2001401

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, *2*(3), 47. https://doi.org/10.2307/3002000

Bianchi, R., Verkuilen, J., Toker, S., Schonfeld, I. S., Gerber, M., Brähler, E., & Kroenke, K. (2022). Is the PHQ-9 a unidimensional measure of depression? A 58,272-participant study. *Psychological Assessment*, *34*(6), 595–603. https://doi.org/10.1037/pas0001124

Bolgeo, T., Di Matteo, R., Simonelli, N., Dal Molin, A., Lusignani, M., Bassola, B., Vellone, E., Maconi, A., & Iovino, P. (2023). Psychometric properties and measurement invariance of the 7-item General Anxiety Disorder scale (GAD-7) in an Italian coronary heart disease population. *Journal of Affective Disorders*, *334*, 213–219. https://doi.org/10.1016/j.jad.2023.04.140

Boothroyd, L., Dagnan, D., & Muncer, S. (2018). Psychometric analysis of the Generalized Anxiety Disorder scale and the Patient Health Questionnaire using mokken scaling and confirmatory factor analysis. *Health and Primary Care*, *2*(4). https://doi.org/10.15761/HPC.1000145

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive Meta-Analysis* (version 3.3.070) [Computer software]. Biostat. https://meta-analysis.com/

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry*, *88*(1), 18–27. https://doi.org/10.1016/j.biopsych.2020.01.013

Borsboom, D., Rhemtulla, M., Cramer, A. O. J., Van Der Maas, H. L. J., Scheffer, M., & Dolan, C. V. (2016). Kinds *versus* continua: A review of psychometric approaches to

uncover the structure of psychiatric constructs. *Psychological Medicine*, *46*(8), 1567–1579. https://doi.org/10.1017/S0033291715001944

Böttger, S. J., Förstner, B. R., Szalek, L., Koller-Schlaud, K., Rapp, M. A., & Tschorn, M. (2023). Mood and anxiety disorders within the Research Domain Criteria framework of positive and negative valence systems: A scoping review. *Frontiers in Human Neuroscience*, *17*, 1184978. https://doi.org/10.3389/fnhum.2023.1184978

Brattmyr, M., Lindberg, M. S., Lundqvist, J., Öst, L., Solem, S., Hjemdal, O., & Havnen, A. (2024). Clinically representative therapy for Nordic adult outpatients with common mental health problems: A systematic review and meta-analysis. *Scandinavian Journal of Psychology*, *65*(2), 311–320. https://doi.org/10.1111/sjop.12976

Brattmyr, M., Lindberg, M. S., Lundqvist, J., Solem, S., Hjemdal, O., Anyan, F., & Havnen, A. (2023). Symptoms and prevalence of common mental disorders in a heterogenous outpatient sample: An investigation of clinical characteristics and latent subgroups. *BMC Psychiatry*, *23*(1), 804. https://doi.org/10.1186/s12888-023-05314-6

Brattmyr, M., Lindberg, M. S., Solem, S., Hjemdal, O., & Havnen, A. (2022). Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder scale-7 in a Norwegian psychiatric outpatient sample. *BMC Psychiatry*, *22*(1), 461. https://doi.org/10.1186/s12888-022-04101-z

Bruchmüller, K., Margraf, J., Suppiger, A., & Schneider, S. (2011). Popular or unpopular? Therapists' use of structured interviews and their estimation of patient acceptance. *Behavior Therapy*, *42*(4), 634–643. https://doi.org/10.1016/j.beth.2011.02.003

Burdzovic Andreas, J., & Brunborg, G. S. (2017). Depressive symptomatology among Norwegian adolescent boys and girls: The Patient Health Questionnaire-9 (PHQ-9) psychometric properties and correlates. *Frontiers in Psychology*, *8*, 887. https://doi.org/10.3389/fpsyg.2017.00887

Byrne, A., Salmon, P., & Fisher, P. (2018). A case study of the challenges for an integrative practitioner learning a new psychological therapy. *Counselling and Psychotherapy Research*, *18*(4), 369–376. https://doi.org/10.1002/capr.12185

Cahill, J. (2013). Clinically representative research in the psychological therapies: An emerging paradigm. *Journal of Research in Nursing*, *18*(2), 97–111. https://doi.org/10.1177/1744987112465079

Castiglioni, M., & Laudisa, F. (2015). Toward psychiatry as a "human" science of mind. The case of depressive disorders in DSM-5. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01517

Castonguay, L. G., Barkham, M., Youn, S. J., & Page, A. C. (2021). Chapter 6: Practice-based evidence- findings from routine clinical settings. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (7th edition, 50th anniversary edition). Wiley.

Churruca, K., Pomare, C., Ellis, L. A., Long, J. C., Henderson, S. B., Murphy, L. E. D., Leahy, C. J., & Braithwaite, J. (2021). Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. *Health Expectations*, *24*(4), 1015–1024. https://doi.org/10.1111/hex.13254

Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology*, *14*(1), 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*(3), 316–336. https://doi.org/10.1037/0021-843X.100.3.316

Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and Strategies for Factor Mixture Analysis: An Example Concerning the Structure Underlying Psychological Disorders. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(4), 681–703. https://doi.org/10.1080/10705511.2013.824786

Cramer, A. O. J., Van Der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414–431. https://doi.org/10.1002/per.1866

Cristea, I. A. (2019). The waiting list is an inadequate benchmark for estimating the effectiveness of psychotherapy for depression. *Epidemiology and Psychiatric Sciences*, *28*(03), 278–279. https://doi.org/10.1017/S2045796018000665

Cuijpers, P., & Cristea, I. A. (2016). How to prove that your therapy is effective, even when it is not: A guideline. *Epidemiology and Psychiatric Sciences*, *25*(5), 428–435. https://doi.org/10.1017/S2045796015000864

Cuijpers, P., Karyotaki, E., Andersson, G., Li, J., Mergl, R., & Hegerl, U. (2015). The effects of blinding on the outcomes of psychotherapy and pharmacotherapy for adult depression: A meta-analysis. *European Psychiatry*, *30*(6), 685–693. https://doi.org/10.1016/j.eurpsy.2015.06.005

Cuijpers, P., Karyotaki, E., Eckshtain, D., Ng, M. Y., Corteselli, K. A., Noma, H., Quero, S., & Weisz, J. R. (2020). Psychotherapy for depression across different age groups: A systematic review and meta-analysis. *JAMA Psychiatry*, *77*(7), 694. https://doi.org/10.1001/jamapsychiatry.2020.0164

Cuijpers, P., Karyotaki, E., Reijnders, M., & Ebert, D. D. (2019). Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. *Epidemiology and Psychiatric Sciences*, *28*(1), 21–30. https://doi.org/10.1017/S2045796018000057

Cuijpers, P., Quero, S., Papola, D., Cristea, I. A., & Karyotaki, E. (2021). Care-as-usual control groups across different settings in randomized trials on psychotherapy for adult depression: A meta-analysis. *Psychological Medicine*, *51*(4), 634–644. https://doi.org/10.1017/S0033291719003581

Cuijpers, P., Reijnders, M., & Huibers, M. J. H. (2019). The role of common factors in psychotherapy outcomes. *Annual Review of Clinical Psychology*, *15*(1), 207–231. https://doi.org/10.1146/annurev-clinpsy-050718-095424

Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, *26*(4), 364–368. https://doi.org/10.1017/S2045796016000809

Curran, E., Rosato, M., Ferry, F., & Leavey, G. (2020). Prevalence and factors associated with anxiety and depression in older adults: Gender differences in psychosocial indicators. *Journal of Affective Disorders*, *267*, 114–122. https://doi.org/10.1016/j.jad.2020.02.018

Curran, E., Rosato, M., Ferry, F., & Leavey, G. (2022). Prevalence and Risk Factors of Psychiatric Symptoms Among Older People in England During the COVID-19 Pandemic: A Latent Class Analysis. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-022-00820-2

Das-Munshi, J., Goldberg, D., Bebbington, P. E., Bhugra, D. K., Brugha, T. S., Dewey, M. E., Jenkins, R., Stewart, R., & Prince, M. (2008). Public health significance of mixed anxiety and depression: Beyond current classification. *British Journal of Psychiatry*, *192*(3), 171–177. https://doi.org/10.1192/bjp.bp.107.036707

Dattani, S., Rodés-Guirao, L., Ritchie, H., & Roser, M. (2023). *Mental health*. https://ourworldindata.org/mental-health

De Beurs, E., Den Hollander-Gijsman, M. E., Van Rood, Y. R., Van Der Wee, N. J. A., Giltay, E. J., Van Noorden, M. S., Van Der Lem, R., Van Fenema, E., & Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology & Psychotherapy*, *18*(1), 1–12. https://doi.org/10.1002/cpp.696

De Beurs, E., Vissers, E., Schoevers, R., Carlier, I. V. E., Van Hemert, A. M., & Meesters, Y. (2019). Comparative responsiveness of generic versus disorder-specific instruments for depression: An assessment in three longitudinal datasets. *Depression and Anxiety*, *36*(1), 93–102. https://doi.org/10.1002/da.22809

De Man, J., Absetz, P., Sathish, T., Desloge, A., Haregu, T., Oldenburg, B., Johnson, L. C. M., Thankappan, K. R., & Williams, E. D. (2021). Are the PHQ-9 and GAD-7 suitable for use in India? A psychometric analysis. *Frontiers in Psychology*, *12*, 676398. https://doi.org/10.3389/fpsyg.2021.676398

Deeks, J. J., Higgins, J. P., & Altman, D. G. (2019). Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (1st ed., pp. 241–284). Wiley. https://doi.org/10.1002/9781119536604.ch10

Delgadillo, J., McMillan, D., Leach, C., Lucock, M., Gilbody, S., & Wood, N. (2014). Benchmarking routine psychological services: A discussion of challenges and methods. *Behavioural and Cognitive Psychotherapy*, *42*(1), 16–30. https://doi.org/10.1017/S135246581200080X

Demyttenaere, K., & Heirman, E. (2020). The blurred line between anxiety and depression: Hesitations on comorbidity, thresholds and hierarchy. *International Review of Psychiatry*, *32*(5–6), 455–465. https://doi.org/10.1080/09540261.2020.1764509

Den Boer, P. C. A. M., Wiersma, D., Vaarwerk, I. T., Span, M. M., Dennis Stant, A., & Van Den Bosch, R. J. (2007). Cognitive self-therapy for chronic depression and anxiety: A multi-centre randomized controlled study. *Psychological Medicine*, *37*(03), 329. https://doi.org/10.1017/S0033291706009214

Dettori, J. R., Norvell, D. C., & Chapman, J. R. (2022). Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. *Global Spine Journal*, *12*(7), 1624–1626. https://doi.org/10.1177/21925682221110527

Directorate-General for Health and Food Safety. (2023). *Communication from the commission to the European Parliament, the council, the European economic and social committee and the committee of the regions on a comprehensive approach to mental health*. https://health.ec.europa.eu/document/download/cef45b6d-a871-44d5-9d62-3cecc47eda89_en?filename=com_2023_298_1_act_en.pdf

Donabedian, A. (1966). Evaluating the quality of medical care. *The Milbank Memorial Fund Quarterly*, *44*(3), 166. https://doi.org/10.2307/3348969

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health*, *52*(6), 377–384. https://doi.org/10.1136/jech.52.6.377

Drapeau, A., Marchand, A., & Forest, C. (2014). Gender differences in the age-cohort distribution of psychological distress in Canadian adults: Findings from a national longitudinal survey. *BMC Psychology*, *2*(1), 25. https://doi.org/10.1186/s40359-014-0025-4

Druss, B. G. (2018). Mental health quality improvement goes global. *World Psychiatry*, *17*(1), 44–45. https://doi.org/10.1002/wps.20487

Eaton, W. W., McCutcheon, A., Dryman, A., & Sorenson, A. (1989). Latent Class Analysis of Anxiety and Depression. *Sociological Methods & Research*, *18*(1), 104–125. https://doi.org/10.1177/0049124189018001004

Eid, M., Krumm, S., Koch, T., & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal of Intelligence*, *6*(3), 42. https://doi.org/10.3390/jintelligence6030042

Enck, P., & Zipfel, S. (2019). Placebo effects in psychotherapy: A framework. *Frontiers in Psychiatry*, *10*, 456. https://doi.org/10.3389/fpsyt.2019.00456

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, *16*(5), 319–324. https://doi.org/10.1037/h0063633

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, *33*(5), 517–517. https://doi.org/10.1037/0003-066X.33.5.517.a

Faltinsen, E., Todorovac, A., Staxen Bruun, L., Hróbjartsson, A., Gluud, C., Kongerslev, M. T., Simonsen, E., & Storebø, O. J. (2022). Control interventions in randomised trials among people with mental health disorders. *Cochrane Database of Systematic Reviews*, *2022*(4). https://doi.org/10.1002/14651858.MR000050.pub2

Farber, G. K., Gage, S., Kemmer, D., & White, R. (2023). Common measures in mental health: A joint initiative by funders and journals. *The Lancet Psychiatry*, *10*(6), 465–470. https://doi.org/10.1016/S2215-0366(23)00139-6

Fava, M., Rush, A. J., Alpert, J. E., Balasubramani, G. K., Wisniewski, S. R., Carmin, C. N., Biggs, M. M., Zisook, S., Leuchter, A., Howland, R., Warden, D., & Trivedi, M. H. (2008). Difference in treatment outcome in outpatients with anxious versus nonanxious depression: A STAR*D report. *American Journal of Psychiatry*, *165*(3), 342–351. https://doi.org/10.1176/appi.ajp.2007.06111868

Feczko, E., Miranda-Dominguez, O., Marr, M., Graham, A. M., Nigg, J. T., & Fair, D. A. (2019). The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. *Trends in Cognitive Sciences*, *23*(7), 584–601. https://doi.org/10.1016/j.tics.2019.03.009

First, M. B., Erlich, M. D., Adler, D. A., Leong, S., Dixon, L. B., Oslin, D. W., Goldman, B., Koh, S., Levine, B., Berlant, J. L., & Siris, S. G. (2019). How the DSM is used in clinical practice. *Journal of Nervous & Mental Disease*, *207*(3), 157–161. https://doi.org/10.1097/NMD.0000000000000953

Flückiger, C., Del Re, A. C., Munder, T., Heer, S., & Wampold, B. E. (2014). Enduring effects of evidence-based psychotherapies in acute depression and anxiety disorders versus treatment as usual at follow-up—A longitudinal meta-analysis. *Clinical Psychology Review*, *34*(5), 367–375. https://doi.org/10.1016/j.cpr.2014.05.001

Frahm Jensen, M. J., Høiby, Ø., & Garang Mofossbakke, R. (2022, November 18). *Nytt kvalitetsregister – uten å høyne kvaliteten?* [New quality register – without increasing quality?]. https://www.dagensmedisin.no/debatt-og-kronikk/nytt-kvalitetsregister-uten-a-hoyne-kvaliteten/107800

Freedland, K. E., Mohr, D. C., Davidson, K. W., & Schwartz, J. E. (2011). Usual and unusual care: Existing practice control groups in randomized controlled trials of behavioral interventions. *Psychosomatic Medicine*, *73*(4), 323–335. https://doi.org/10.1097/PSY.0b013e318218e1fb

Freedman, R., Lewis, D. A., Michels, R., Pine, D. S., Schultz, S. K., Tamminga, C. A., Gabbard, G. O., Gau, S. S.-F., Javitt, D. C., Oquendo, M. A., Shrout, P. E., Vieta, E., & Yager, J. (2013). The initial field trials of DSM-5: New blooms and old thorns. *American Journal of Psychiatry*, *170*(1), 1–5. https://doi.org/10.1176/appi.ajp.2012.12091189

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders*, *172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., Chen, P., Hunot, V., & Churchill, R. (2014). Waiting list may be a nocebo condition in psychotherapy trials: A contribution from network meta-analysis. *Acta Psychiatrica Scandinavica*, *130*(3), 181–192. https://doi.org/10.1111/acps.12275

Gelkopf, M., Mazor, Y., & Roe, D. (2022). A systematic review of patient-reported outcome measurement (PROM) and provider assessment in mental health: Goals, implementation, setting, measurement characteristics and barriers. *International Journal for Quality in Health Care*, *34*(Supplement_1), ii13–ii27. https://doi.org/10.1093/intqhc/mzz133

Gråwe, R. W., Ruud, T., & Hatling, T. (2008). *Bidrar utbyggingen av distriktspsykiatriske sentre til bedre tjenestetilbud og høyere brukertilfredshet? Resultater fra undersøkelser gjennomført i 2002, 2005 og 2007. Revidert september 2008* [Does the development of district psychiatric centers contribute to better service provision and higher user satisfaction? Results from investigations carried out in 2002, 2005 and 2007. Revised September 2008]. https://www.sintef.no/publikasjoner/publikasjon/1269070/

Guyon, H., Falissard, B., & Kop, J.-L. (2017). Modeling psychological attributes in psychology – An epistemological discussion: Network analysis vs. latent variables. *Frontiers in Psychology*, *8*, 798. https://doi.org/10.3389/fpsyg.2017.00798

Haeffel, G. J., Jeronimus, B. F., Kaiser, B. N., Weaver, L. J., Soyster, P. D., Fisher, A. J., Vargas, I., Goodson, J. T., & Lu, W. (2022). Folk classification and factor rotations: Whales, sharks, and the problems with the Hierarchical Taxonomy of Psychopathology (HiTOP). *Clinical Psychological Science*, *10*(2), 259–278. https://doi.org/10.1177/21677026211002500

Haro, J. M., Ayuso-Mateos, J. L., Bitter, I., Demotes-Mainard, J., Leboyer, M., Lewis, S. W., Linszen, D., Maj, M., Mcdaid, D., Meyer-Lindenberg, A., Robbins, T. W., Schumann, G., Thornicroft, G., Van Der Feltz-Cornelis, C., Van Os, J., Wahlbeck, K., Wittchen, H., Wykes, T., Arango, C., … Walker-Tilley, T. (2014). ROAMER: Roadmap for mental health research in Europe. *International Journal of Methods in Psychiatric Research*, *23*(S1), 1–14. https://doi.org/10.1002/mpr.1406

Harvey, A. G., & Gumport, N. B. (2015). Evidence-based psychological treatments for mental disorders: Modifiable barriers to access and possible solutions. *Behaviour Research and Therapy*, *68*, 1–12. https://doi.org/10.1016/j.brat.2015.02.004

Hasin, D. S., Goodwin, R. D., Stinson, F. S., & Grant, B. F. (2005). Epidemiology of major depressive disorder: Results from the national epidemiologic survey on alcoholism and related conditions. *Archives of General Psychiatry*, *62*(10), 1097. https://doi.org/10.1001/archpsyc.62.10.1097

Hazo, J.-B., Brunn, M., Wykes, T., McDaid, D., Dorsey, M., Demotes-Mainard, J., Van Der Feltz-Cornelis, C. M., Wahlbeck, K., Knappe, S., Meyer-Lindenberg, A., Obradors-Tarragó, C., Haro, J. M., Leboyer, M., & Chevreul, K. (2019). European mental health research resources: Picture and recommendations of the ROAMER project.

European Neuropsychopharmacology, *29*(2), 179–194. https://doi.org/10.1016/j.euroneuro.2018.11.1111

Norwegian Directorate of Health. (2024). *Kvalitetsindikatorer [Quality indicators]*. Psykisk helse for voksne. https://www.helsedirektoratet.no/statistikk/kvalitetsindikatorer/psykisk-helse-for-voksne

Heinonen, E., & Nissen-Lie, H. A. (2020). The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy research : journal of the Society for Psychotherapy Research*, *30*(4), 417–432. https://doi.org/10.1080/10503307.2019.1620366

Herrman, H., Patel, V., Kieling, C., Berk, M., Buchweitz, C., Cuijpers, P., Furukawa, T. A., Kessler, R. C., Kohrt, B. A., Maj, M., McGorry, P., Reynolds, C. F., Weissman, M. M., Chibanda, D., Dowrick, C., Howard, L. M., Hoven, C. W., Knapp, M., Mayberg, H. S., … Wolpert, M. (2022). Time for united action on depression: A Lancet–World Psychiatric Association Commission. *The Lancet*, *399*(10328), 957–1022. https://doi.org/10.1016/S0140-6736(21)02141-3

Hettema, J. M., Aggen, S. H., Kubarych, T. S., Neale, M. C., & Kendler, K. S. (2015). Identification and validation of mixed anxiety–depression. *Psychological Medicine*, *45*(14), 3075–3084. https://doi.org/10.1017/S0033291715001038

Hewlett, E., & Moran, V. (2014). *Making Mental Health Count: The Social and Economic Costs of Neglecting Mental Health Care*. OECD. https://doi.org/10.1787/9789264208445-en

Høstmælingen, A., Nissen-Lie, H. A., Wampold, B., & Ulvenes, P. G. (2023). Symptom change during waitlist for medicated and nonmedicated patients with chronic depression. *Psychotherapy and Psychosomatics*, *92*(5), 340–345. https://doi.org/10.1159/000533661

Humphreys, K., Blodgett, J. C., & Roberts, L. W. (2015). The exclusion of people with psychiatric disorders from medical research. *Journal of Psychiatric Research*, *70*, 28–32. https://doi.org/10.1016/j.jpsychires.2015.08.005

Hyland, P., & Shevlin, M. (2024). Clinician-administered interviews should not be considered the 'gold standard' method of assessing psychological distress. *New Ideas in Psychology, 73*, 101072. https://doi.org/10.1016/j.newideapsych.2023.101072

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Ioannidis, J. P. A., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., Schulz, K. F., & Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet*, *383*(9912), 166–175. https://doi.org/10.1016/S0140-6736(13)62227-8

Iob, E., Kirschbaum, C., & Steptoe, A. (2020). Persistent depressive symptoms, HPA-axis hyperactivity, and inflammation: The role of cognitive-affective and somatic symptoms. *Molecular Psychiatry*, *25*(5), 1130–1140. https://doi.org/10.1038/s41380-019-0501-6

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., … Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden

of Disease Study 2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Johnson, S. U., Hoffart, A., Havik, O. E., & Nordgreen, T. (2016). A survey of clinical psychologists' attitudes toward treatment manuals. *Professional Psychology: Research and Practice*, *47*(5), 340–346. https://doi.org/10.1037/pro0000108

Johnson, S. U., Ulvenes, P. G., Øktedalen, T., & Hoffart, A. (2019). Psychometric Properties of the General Anxiety Disorder 7-Item (GAD-7) Scale in a Heterogeneous Psychiatric Sample. *Frontiers in Psychology*, *10*, 1713. https://doi.org/10.3389/fpsyg.2019.01713

Juul, F. E. (2019, April 3). *Pasientutvalg representative for virkelig, klinisk praksis [Patient samples representative of real clinical practice]*. https://tidsskriftet.no/2019/04/kommentar/pasientutvalg-representative-virkelig-klinisk-praksis#comment-1821

Kazdin, A. E. (2015). Treatment as usual and routine care in research and clinical practice. *Clinical Psychology Review*, *42*, 168–178. https://doi.org/10.1016/j.cpr.2015.08.006

Kendler, K. S. (2017). DSM disorders and their criteria: How should they inter-relate? *Psychological Medicine*, *47*(12), 2054–2060. https://doi.org/10.1017/S0033291717000678

Kilbourne, A. M., Beck, K., Spaeth-Rublee, B., Ramanuj, P., O'Brien, R. W., Tomoyasu, N., & Pincus, H. A. (2018). Measuring and improving the quality of mental health care: A global perspective. *World Psychiatry*, *17*(1), 30–38. https://doi.org/10.1002/wps.20482

Kinge, J. M., Sælensminde, K., Dieleman, J., Vollset, S. E., & Norheim, O. F. (2017). Economic losses and burden of disease by medical conditions in Norway. *Health Policy*, *121*(6), 691–698. https://doi.org/10.1016/j.healthpol.2017.03.020

Klein Breteler, J., Ikani, N., Becker, E. S., Spijker, J., & Hendriks, G. (2021). Comorbid depression and treatment of anxiety disorders, OCD, and PTSD: Diagnosis versus severity. *Journal of Affective Disorders*, *295*, 1005–1011. https://doi.org/10.1016/j.jad.2021.08.146

Knudsen, A. K. S., Stene-Larsen, K., Gustavson, K., Hotopf, M., Kessler, R. C., Krokstad, S., Skogen, J. C., Øverland, S., & Reneflot, A. (2021). Prevalence of mental disorders, suicidal ideation and suicides in the general population before and during the COVID-19 pandemic in Norway: A population-based repeated cross-sectional analysis. *The Lancet Regional Health - Europe*, *4*, 100071. https://doi.org/10.1016/j.lanepe.2021.100071

Kongsted, A., & Nielsen, A. M. (2017). Latent class analysis in health research. *Journal of Physiotherapy*, *63*(1), 55–58. https://doi.org/10.1016/j.jphys.2016.05.018

Lamela, D., Soreira, C., Matos, P., & Morais, A. (2020). Systematic review of the factor structure and measurement invariance of the patient health questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings. *Journal of Affective Disorders*, *276*, 220–233. https://doi.org/10.1016/j.jad.2020.06.066

Lei, H., Chen, C., Hagiwara, K., Kusumi, I., Tanabe, H., Inoue, T., & Nakagawa, S. (2022). Symptom Patterns of the Occurrence of Depression and Anxiety in a Japanese General Adult Population Sample: A Latent Class Analysis. *Frontiers in Psychiatry*, *13*, 808918. https://doi.org/10.3389/fpsyt.2022.808918

Leichsenring, F., Steinert, C., Rabung, S., & Ioannidis, J. P. A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry*, *21*(1), 133–145. https://doi.org/10.1002/wps.20941

Leiden University Medical Center. (2016). *Institute presentation*. https://prism-project.eu/en/partner/leiden-university-medical-center/

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review*, *33*(7), 883–900. https://doi.org/10.1016/j.cpr.2012.09.008

Liu, Z., Liu, R., Zhang, Y., Zhang, R., Liang, L., Wang, Y., Wei, Y., Zhu, R., & Wang, F. (2021). Latent class analysis of depression and anxiety among medical students during COVID-19 epidemic. *BMC Psychiatry*, *21*(1), 498. https://doi.org/10.1186/s12888-021-03459-w

Maj, M. (2005). 'Psychiatric comorbidity': An artefact of current diagnostic systems? *British Journal of Psychiatry*, *186*(3), 182–184. https://doi.org/10.1192/bjp.186.3.182

McKnight, P. E., & Kashdan, T. B. (2009). The importance of functional impairment to mental health outcomes: A case for reassessing our goals in depression treatment research. *Clinical Psychology Review*, *29*(3), 243–259. https://doi.org/10.1016/j.cpr.2009.01.005

McKnight, P. E., Monfort, S. S., Kashdan, T. B., Blalock, D. V., & Calton, J. M. (2016). Anxiety symptoms and functional impairment: A systematic review of the correlation between the two measures. *Clinical Psychology Review*, *45*, 115–130. https://doi.org/10.1016/j.cpr.2015.10.005

Mental health atlas 2020. (2021). World Health Organization. https://iris.who.int/bitstream/handle/10665/345946/9789240036703-eng.pdf?sequence=1

Ministry of Health and Care Services. (2019). *Nasjonal helse- og sykehusplan 2020–2023* (Meld. St. 7). [National health and hospital plan 2020–2023]. https://www.regjeringen.no/contentassets/95eec808f0434acf942fca449ca35386/no/pdfs/stm201920200007000dddpdfs.pdf

Ministry of Health and Care Services. (2023). *Opptrappingsplan for psykisk helse (2023–2033)* (Meld. St. 23). [Action plan for mental health (2023–2033)]. https://www.regjeringen.no/contentassets/0fb8e2f8f1ff4d40a522e3775a8b22bc/no/pdfs/stm202220230023000dddpdfs.pdf

Möller, H.-J., Bandelow, B., Volz, H.-P., Barnikol, U. B., Seifritz, E., & Kasper, S. (2016). The relevance of 'mixed anxiety and depression' as a diagnostic category in clinical practice. *European Archives of Psychiatry and Clinical Neuroscience*, *266*(8), 725–736. https://doi.org/10.1007/s00406-016-0684-7

Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P., Rabung, S., Steinert, C., & Wampold, B. E. (2019). Is psychotherapy effective? A re-analysis of treatments for depression. *Epidemiology and Psychiatric Sciences*, *28*(03), 268–274. https://doi.org/10.1017/S2045796018000355

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, *180*(5), 461–464. https://doi.org/10.1192/bjp.180.5.461

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide. Eighth edition.* [Computer software]. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

National service for medical quality registers. (2024). *Nasjonale medisinske kvalitetsregistre status 2023* [National medical quality registers status 2023] (1/2024). https://www.kvalitetsregistre.no/sites/default/files/2024-02/Statusrapport%202023%20for%20medisinske%20kvalitetsregistre.pdf

National institute for health and care excellence. (2022). *Depression in adults: Treatment and management* (NICE Guideline NO. 222). https://www.nice.org.uk/guidance/ng222

Negeri, Z. F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., Bhandari, P. M., Neupane, D., Brehaut, E., Benedetti, A., & Thombs, B. D. (2021). Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: Updated systematic review and individual participant data meta-analysis. *BMJ*, *375*. https://doi.org/10.1136/bmj.n2183

Obbarius, A., Van Maasakkers, L., Baer, L., Clark, D. M., Crocker, A. G., De Beurs, E., Emmelkamp, P. M. G., Furukawa, T. A., Hedman-Lagerlöf, E., Kangas, M., Langford, L., Lesage, A., Mwesigire, D. M., Nolte, S., Patel, V., Pilkonis, P. A., Pincus, H. A., Reis, R. A., Rojas, G., … Rose, M. (2017). Standardization of health outcomes assessment for depression and anxiety: Recommendations from the ICHOM Depression and Anxiety Working Group. *Quality of Life Research*, *26*(12), 3211–3225. https://doi.org/10.1007/s11136-017-1659-5

Office of the Auditor General of Norway. (2021). *Riksrevisjonens undersøkelse av psykiske helsetjenester*. https://www.riksrevisjonen.no/globalassets/rapporter/no-2020-2021/psykiske-helsetjenester.pdf

Øiesvold, T., Nivison, M., Hansen, V., Skre, I., Østensen, L., & Sørgaard, K. W. (2013). Diagnosing comorbidity in psychiatric hospital: Challenging the validity of administrative registers. *BMC Psychiatry*, *13*(1), 13. https://doi.org/10.1186/1471-244X-13-13

Olbert, C. M., & Gala, G. J. (2015). Supervenience and psychiatry: Are mental disorders brain disorders? *Journal of Theoretical and Philosophical Psychology*, *35*(4), 203–219. https://doi.org/10.1037/teo0000023

Öst, L., Brattmyr, M., Finnes, A., Ghaderi, A., Havnen, A., Hedman-Lagerlöf, M., Parling, T., Welch, E., & Wergeland, G. J. (2023). Cognitive behavior therapy for adult eating disorders in routine clinical care: A systematic review and meta-analysis. *International Journal of Eating Disorders*, eat.24104. https://doi.org/10.1002/eat.24104

Öst, L.-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., Salomonsson, S., & Wergeland, G. J. (2022). Cognitive behavior therapy for obsessive-compulsive disorder in routine clinical care: A systematic review and meta-analysis. *Behaviour Research and Therapy*, *159*, 104170. https://doi.org/10.1016/j.brat.2022.104170

Öst, L.-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., Salomonsson, S., & Wergeland, G. J. (2023a). Cognitive behavior therapy for adult depressive disorders in routine clinical care: A systematic review and meta-analysis. *Journal of Affective Disorders*, *331*, 322–333. https://doi.org/10.1016/j.jad.2023.03.002

Öst, L.-G., Enebrink, P., Finnes, A., Ghaderi, A., Havnen, A., Kvale, G., Salomonsson, S., & Wergeland, G. J. (2023b). Cognitive behavior therapy for adult post-traumatic stress disorder in routine clinical care: A systematic review and meta-analysis. *Behaviour Research and Therapy*, *166*, 104323. https://doi.org/10.1016/j.brat.2023.104323

Pasman, J. A., Meijsen, J. J., Haram, M., Kowalec, K., Harder, A., Xiong, Y., Nguyen, T.-D., Jangmo, A., Shorter, J. R., Bergstedt, J., Das, U., Zetterberg, R., Tate, A., Lichtenstein, P., Larsson, H., Odsbu, I., Werge, T., Reichborn-Kjennerud, T., Andreassen, O. A., … Lu, Y. (2023). Epidemiological overview of major depressive disorder in Scandinavia using nationwide registers. *The Lancet Regional Health - Europe*, *29*, 100621. https://doi.org/10.1016/j.lanepe.2023.100621

Patalay, P., & Fried, E. I. (2021). Editorial Perspective: Prescribing measures: unintended negative consequences of mandating standardized mental health measurement.

Journal of Child Psychology and Psychiatry, *62*(8), 1032–1036. https://doi.org/10.1111/jcpp.13333

Pedersen, G., Kvarstein, E. H., & Wilberg, T. (2017). The Work and Social Adjustment Scale: Psychometric properties and validity among males and females, and outpatients with and without personality disorders: The work and social adjustment scale: psychometric properties and validity among males and females. *Personality and Mental Health*, *11*(4), 215–228. https://doi.org/10.1002/pmh.1382

Pedersen, P. B., & Lilleeng, S. E. (2019). *Distriktspsykiatriske tjenester 2017* [Psychiatric services in the communities 2017]. *Norwegian Directorate of Health*. https://www.helsedirektoratet.no/rapporter/distriktspsykiatriske-tjenester-driftsindikatorer-for-distriktspsykiatriske-sentre/Distriktspsykiatriske%20tjenester%202017.pdf

Penninx, B. W., Milaneschi, Y., Lamers, F., & Vogelzangs, N. (2013). Understanding the somatic consequences of depression: Biological mechanisms and the role of depression symptom profile. *BMC Medicine*, *11*(1), 129. https://doi.org/10.1186/1741-7015-11-129

Petersson, E.-L., Forsén, E., Björkelund, C., Hammarbäck, L., Hessman, E., Weineland, S., & Svenningsson, I. (2023). Examining the description of the concept "treatment as usual" for patients with depression, anxiety and stress-related mental disorders in primary health care research – A systematic review. *Journal of Affective Disorders*, *326*, 1–10. https://doi.org/10.1016/j.jad.2023.01.076

Petrosyan, Y., Sahakyan, Y., Barnsley, J. M., Kuluski, K., Liu, B., & Wodchis, W. P. (2017). Quality indicators for care of depression in primary care settings: A systematic review. *Systematic Reviews*, *6*(1), 126. https://doi.org/10.1186/s13643-017-0530-7

Pinho, L. G. D., Lopes, M. J., Correia, T., Sampaio, F., Arco, H. R. D., Mendes, A., Marques, M. D. C., & Fonseca, C. (2021). Patient-Centered Care for Patients with Depression or Anxiety Disorder: An Integrative Review. *Journal of Personalized Medicine*, *11*(8), 776. https://doi.org/10.3390/jpm11080776

Podlogar, M. C., Rogers, M. L., Stanley, I. H., Hom, M. A., Chiurliza, B., & Joiner, T. E. (2018). Anxiety, depression, and the suicidal spectrum: A latent class analysis of overlapping and distinctive features. *Cognition and Emotion*, *32*(7), 1464–1477. https://doi.org/10.1080/02699931.2017.1303452

Pranckeviciene, A., Saudargiene, A., Gecaite-Stonciene, J., Liaugaudaite, V., Griskova-Bulanova, I., Simkute, D., Naginiene, R., Dainauskas, L. L., Ceidaite, G., & Burkauskas, J. (2022). Validation of the patient health questionnaire-9 and the generalized anxiety disorder-7 in Lithuanian student sample. *PloS one*, *17*(1), e0263027. https://doi.org/10.1371/journal.pone.0263027

Rhebergen, D., Van Der Steenstraten, I. M., Sunderland, M., De Graaf, R., Ten Have, M., Lamers, F., Penninx, B. W. J. H., & Andrews, G. (2014). An examination of generalized anxiety disorder and dysthymic disorder by latent class analysis. *Psychological Medicine*, *44*(8), 1701–1712. https://doi.org/10.1017/S0033291713002225

Ross, C. A., & Margolis, R. L. (2019). Research Domain Criteria: Strengths, weaknesses, and potential alternatives for future psychiatric research. *Complex Psychiatry*, *5*(4), 218–236. https://doi.org/10.1159/000501797

Rutter, L. A., & Brown, T. A. (2017). Psychometric properties of the Generalized Anxiety Disorder Scale-7 (GAD-7) in outpatients with anxiety and mood disorders. *Journal of Psychopathology and Behavioral Assessment*, *39*(1), 140–146. https://doi.org/10.1007/s10862-016-9571-9

Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., … Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, *398*(10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7

Schang, L., Blotenberg, I., & Boywitt, D. (2021). What makes a good quality indicator set? A systematic review of criteria. *International Journal for Quality in Health Care*, *33*(3), mzab107. https://doi.org/10.1093/intqhc/mzab107

Schem, B.-C., Frich, J., Sandbu, H., & Hansen, F. H. (2018). *Plan for prioriterte fagområder for utvikling av nye medisinske kvalitetsregistre* [Plan for prioritized subject areas for the development of new medical quality registers]. https://www.kvalitetsregistre.no/sites/default/files/plan_nye_kvalitetsregistre_-_final_.pdf

Segal, J. B., Varadhan, R., Groenwold, R. H. H., Henderson, N. C., Li, X., Nomura, K., Kaplan, S., Ardeshirrouhanifard, S., Heyward, J., Nyberg, F., & Burcu, M. (2023). Assessing heterogeneity of treatment effect in real-world data. *Annals of Internal Medicine*, *176*(4), 536–544. https://doi.org/10.7326/M22-1510

Shadish, W. R., Matt, G. E., Navarro, A. M., Siegle, G., Crits-Christoph, P., Hazelrigg, M. D., Jorm, A. F., Lyons, L. C., Nietzel, M. T., Prout, H. T., Robinson, L., Smith, M. L., Svartberg, M., & Weiss, B. (1997). Evidence that therapy works in clinically representative conditions. *Journal of Consulting and Clinical Psychology*, *65*(3), 355–365.

Shadish, W. R., Navarro, A. M., Matt, G. E., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin*, *126*(4), 512–529. https://doi.org/10.1037/0033-2909.126.4.512

Shean, G. (2014). Limitations of randomized control designs in psychotherapy research. *Advances in Psychiatry*, *2014*, 1–5. https://doi.org/10.1155/2014/561452

Sherry, D. (2011). Thermoscopes, thermometers, and the foundations of measurement. *Studies in History and Philosophy of Science Part A*, *42*(4), 509–524. https://doi.org/10.1016/j.shpsa.2011.07.001

Shevlin, M., Butter, S., McBride, O., Murphy, J., Gibson-Miller, J., Hartman, T. K., Levita, L., Mason, L., Martinez, A. P., McKay, R., Stocks, T. V., Bennett, K. M., Hyland, P., Vallieres, F., Valiente, C., Vazquez, C., Contreras, A., Peinado, V., Trucharte, A., … Bentall, R. P. (2022). Measurement invariance of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) across four European countries during the COVID-19 pandemic. *BMC Psychiatry*, *22*(1), 154. https://doi.org/10.1186/s12888-022-03787-5

Shevlin, M., Hyland, P., Nolan, E., Owczarek, M., Ben-Ezra, M., & Karatzias, T. (2022). ICD-11 'mixed depressive and anxiety disorder' is clinical rather than sub-clinical and more common than anxiety and depression in the general population. *British Journal of Clinical Psychology*, *61*(1), 18–36. https://doi.org/10.1111/bjc.12321

Singham, T., Saunders, R., Brooker, H., Creese, B., Aarsland, D., Hampshire, A., Ballard, C., Corbett, A., Desai, R., & Stott, J. (2022). Are subtypes of affective symptoms differentially associated with change in cognition over time: A latent class analysis. *Journal of Affective Disorders*, *309*, 437–445. https://doi.org/10.1016/j.jad.2022.04.139

Skarpsno, E. S. (2019). Representative pasientutvalg er unødvendig i klinisk forskning [Representative patient samples are unnecessary in clinical research]. *Tidsskrift for Den norske legeforening*. https://doi.org/10.4045/tidsskr.19.0047

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*(9), 752–760. https://doi.org/10.1037/0003-066X.32.9.752

Starr, L. R., & Davila, J. (2012). Temporal patterns of anxious and depressed mood in generalized anxiety disorder: A daily diary study. *Behaviour Research and Therapy*, *50*(2), 131–141. https://doi.org/10.1016/j.brat.2011.11.005

StataCorp. (2021). *Stata statistical software* (Version 17) [Computer software].

Stein, D. J., Shoptaw, S. J., Vigo, D. V., Lund, C., Cuijpers, P., Bantjes, J., Sartorius, N., & Maj, M. (2022). Psychiatric diagnosis and treatment in the 21st century: Paradigm shifts versus incremental integration. *World Psychiatry*, *21*(3), 393–414. https://doi.org/10.1002/wps.20998

Ten Have, M., Lamers, F., Wardenaar, K., Beekman, A., De Jonge, P., Van Dorsselaer, S., Tuithof, M., Kleinjan, M., & De Graaf, R. (2016). The identification of symptom-based subtypes of depression: A nationally representative cohort study. *Journal of Affective Disorders*, *190*, 395–406. https://doi.org/10.1016/j.jad.2015.10.040

The Norwegian Psychological Association. (2024). *Prinsipperklæring om evidensbasert psykologisk praksis* [Statement of principles on evidence-based psychological practice]. *The Norwegian Psychological Association.* https://www.psykologforeningen.no/fag-og-politikk/psykologprofesjonen/prinsipperklaering-om-evidensbasert-psykologisk-praksis

Thoma, N. C., & Cecero, J. J. (2009). Is integrative use of techniques in psychotherapy the exception or the rule? Results of a national survey of doctoral-level practitioners. *Psychotherapy: Theory, Research, Practice, Training*, *46*(4), 405–417. https://doi.org/10.1037/a0017900

Torvik, F. A., Ystrom, E., Gustavson, K., Rosenström, T. H., Bramness, J. G., Gillespie, N., Aggen, S. H., Kendler, K. S., & Reichborn-Kjennerud, T. (2018). Diagnostic and genetic overlap of three common mental disorders in structured interviews and health registries. *Acta Psychiatrica Scandinavica*, *137*(1), 54–64. https://doi.org/10.1111/acps.12829

Town, J. M., Diener, M. J., Abbass, A., Leichsenring, F., Driessen, E., & Rabung, S. (2012). A meta-analysis of psychodynamic psychotherapy outcomes: evaluating the effects of research-specific procedures. *Psychotherapy (Chicago, Ill.)*, *49*(3), 276–290. https://doi.org/10.1037/a0029564

Ulbricht, C. M., Chrysanthopoulou, S. A., Levin, L., & Lapane, K. L. (2018). The use of latent class analysis for identifying subtypes of depression: A systematic review. *Psychiatry Research*, *266*, 228–246. https://doi.org/10.1016/j.psychres.2018.03.003

Van Loo, H. M., Beijers, L., Wieling, M., De Jong, T. R., Schoevers, R. A., & Kendler, K. S. (2023). Prevalence of internalizing disorders, symptoms, and traits across age using advanced nonlinear models. *Psychological Medicine*, *53*(1), 78–87. https://doi.org/10.1017/S0033291721001148

Van Loo, H. M., & Romeijn, J.-W. (2018). Letter to the Editor: Measuring and defining: the double role of the DSM criteria for psychiatric disorders. *Psychological Medicine*, *48*(5), 872–873. https://doi.org/10.1017/S0033291717001799

Van Loo, H. M., Wanders, R. B. K., Wardenaar, K. J., & Fried, E. I. (2018). Problems with latent class analysis to detect data-driven subtypes of depression. *Molecular Psychiatry*, *23*(3), 495–496. https://doi.org/10.1038/mp.2016.202

Von Hippel, P. T. (2015). The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Medical Research Methodology*, *15*(1), 35. https://doi.org/10.1186/s12874-015-0024-z

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., & Delgadillo, J. (2021). Improving Access to Psychological Therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence. *British Journal of Clinical Psychology*, *60*(1), 1–37. https://doi.org/10.1111/bjc.12259

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270–277. https://doi.org/10.1002/wps.20238

Wampold, B. E., Budge, S. L., Laska, K. M., Del Re, A. C., Baardseth, T. P., Flűckiger, C., Minami, T., Kivlighan, D. M., & Gunn, W. (2011). Evidence-based treatments for depression and anxiety versus treatment-as-usual: A meta-analysis of direct comparisons. *Clinical Psychology Review*, *31*(8), 1304–1312. https://doi.org/10.1016/j.cpr.2011.07.012

Wampold, B. E., Frost, N. D., & Yulish, N. E. (2016). Placebo effects in psychotherapy: A flawed concept and a contorted history. *Psychology of Consciousness: Theory, Research, and Practice*, *3*(2), 108–120. https://doi.org/10.1037/cns0000045

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus.* John Wiley & Sons.

Watts, S. E., Turnell, A., Kladnitski, N., Newby, J. M., & Andrews, G. (2015). Treatment-as-usual (TAU) is anything but usual: A meta-analysis of CBT versus TAU for anxiety and depression. *Journal of Affective Disorders*, *175*, 152–167. https://doi.org/10.1016/j.jad.2014.12.025

Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist*, *47*(12), 1578–1585. https://doi.org/10.1037/0003-066X.47.12.1578

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, *117*(3), 450–468. https://doi.org/10.1037/0033-2909.117.3.450

Wisting, L., Johnson, S. U., Bulik, C. M., Andreassen, O. A., Rø, Ø., & Bang, L. (2021). Psychometric properties of the Norwegian version of the Patient Health Questionnaire-9 (PHQ-9) in a large female sample of adults with and without eating disorders. *BMC Psychiatry*, *21*(1), 6. https://doi.org/10.1186/s12888-020-03013-0

World Health Organization. (2013). *Mental health action plan 2013-2020.* World Health Organization. https://iris.who.int/handle/10665/89966

World Health Organization. (2016). *International statistical classification of diseases and related health problems (10th ed.)*. https://icd.who.int/browse10/2016/en

World Health Organization. (2021). *Comprehensive mental health action plan 2013–2030*. https://iris.who.int/bitstream/handle/10665/345301/9789240031029-eng.pdf?sequence=1

Yu, S. (2018). Uncovering the hidden impacts of inequality on mental health: A global study. *Translational Psychiatry*, *8*(1), 98. https://doi.org/10.1038/s41398-018-0148-0

Zahra, D., Qureshi, A., Henley, W., Taylor, R., Quinn, C., Pooler, J., Hardy, G., Newbold, A., & Byng, R. (2014). The work and social adjustment scale: Reliability, sensitivity and value. *International Journal of Psychiatry in Clinical Practice*, *18*(2), 131–138. https://doi.org/10.3109/13651501.2014.894072

Zarbo, C., Tasca, G. A., Cattafi, F., & Compare, A. (2016). Integrative psychotherapy works. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.02021

Papers I–III

Paper I – Brattmyr, M., Lindberg, M. S., Lundqvist, J., Öst, L., Solem, S., Hjemdal, O., & Havnen, A. (2024). Clinically representative therapy for Nordic adult outpatients with common mental health problems: A systematic review and meta-analysis. *Scandinavian Journal of Psychology*, *65*(2), 311–320. https://doi.org/10.1111/sjop.12976

Paper II – Brattmyr, M., Lindberg, M. S., Solem, S., Hjemdal, O., & Havnen, A. (2022). Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder scale-7 in a Norwegian psychiatric outpatient sample. *BMC Psychiatry*, *22*(1), 461. https://doi.org/10.1186/s12888-022-04101-z

Paper III – Brattmyr, M., Lindberg, M. S., Lundqvist, J., Solem, S., Hjemdal, O., Anyan, F., & Havnen, A. (2023). Symptoms and prevalence of common mental disorders in a heterogenous outpatient sample: An investigation of clinical characteristics and latent subgroups. *BMC Psychiatry*, *23*(1), 804. https://doi.org/10.1186/s12888-023-05314-6

Scandinavian Journal of Psychology, 2024, 65, 311–320 DOI: 10.1111/sjop.12976

Review Article

Clinically representative therapy for Nordic adult outpatients with common mental health problems: A systematic review and meta-analysis

MARTIN BRATTMYR,[1] MARTIN SCHEVIK LINDBERG,[1,2] JAKOB LUNDQVIST,[1] LARS-GÖRAN ÖST,[3] STIAN SOLEM,[1] ODIN HJEMDAL[1] and AUDUN HAVNEN[1,4]

[1]*Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway*
[2]*Mental Health Care Services, Trondheim Municipality, Trondheim, Norway*
[3]*Department of Psychology, Stockholm University, Stockholm, Sweden*
[4]*Nidaros Community Mental Health Centre, Division of Psychiatry, St. Olavs University Hospital, Trondheim, Norway*

Brattmyr, M., Lindberg, M.S., Lundqvist, J., Öst, L.-G., Solem, S., Hjemdal, O. & Havnen, A. (2024). Clinically representative therapy for Nordic adult outpatients with common mental health problems: A systematic review and meta-analysis. *Scandinavian Journal of Psychology*, *65*, 311–320.

There is a knowledge gap regarding clinically representative therapy given in routine settings, that is treatment as usual (TAU), for patients with common mental health problems (CMHP). This review and meta-analysis aimed to investigate what characterizes clinically representative therapy in Nordic routine clinics and meta-analyze the outcome of such treatment. Databases (PubMed, EMBASE, PsychINFO, and SveMed+) were searched for TAU, CMHP, and Nordic countries, together with backward and forward search in Scopus (7 November 2022). Studies were either randomized controlled trials (RCT) or open trials, using prospective study designs, examining heterogeneous outpatient groups in routine treatment. Within- and between-group effect sizes (ES), using random effects model, and moderator analyses were calculated. Eleven studies (n = 1,413), demonstrated a small to moderate within-group ES with high heterogeneity (g = 0.49, I^2 = 90%). ESs in RCTs were significantly smaller than in open trials. TAU had a marginally smaller ES (g = −0.21; adjusted for publication bias g = −0.06) compared to a broad set of clinical interventions. Clinically representative therapy in the Nordic countries demonstrated a wide variety of characteristics and also a marginally lower ES compared to other interventions. The ESs were smaller than other meta-analyses examining evidence-based treatments in routine treatment.

Key words: Clinically representative treatment, treatment as usual (TAU), care as usual, common mental health problems, meta-analysis.

Martin Brattmyr, Norwegian University of Science and Technology NTNU, Department of Psychology, NO-7491, Trondheim, Norway. Email: martin.brattmyr@ntnu.no

INTRODUCTION

Research on the treatment outcome of routine care for patients with common mental health problems (CMHP), such as depressive or anxiety disorders (National Institute for Health and Care Excellence [NICE], 2011), has been neglected (Hewlett & Moran, 2014). Routine mental health treatment may be equated to treatment as usual (TAU). Rather than viewing TAU as a generic control condition to evidence-based treatments (EBT), the effect of TAU may also be used as an indicator of what outcome to expect for patients who undergo treatment in mental health care facilities (Kazdin, 2015). Meanwhile, research on TAU is challenging, mainly due to its ambiguity (Freedland, Mohr, Davidson & Schwartz, 2011; Kazdin, 2015; Wampold, Budge, Laska *et al*., 2011; Watts, Turnell, Kladnitski, Newby & Andrews, 2015). One of the biggest limitations of TAU research is that many researchers do not describe the contents of it (Wampold *et al*., 2011; Watts *et al*., 2015). Therefore, TAU is often a *laissez faire* (Freedland *et al*., 2011) and in many randomized control trials (RCTs) not even intended to be therapeutic (Wampold *et al*., 2011). As a result, weaker TAU leads to greater effect sizes (ES) for the intervention of interest in direct comparisons (Cuijpers, Karyotaki, Reijnders & Ebert, 2019; Cuijpers, Quero, Papola, Cristea & Karyotaki, 2021; Wampold *et al*., 2011; Watts *et al*., 2015).

A second challenge is the use of TAU across different research contexts. Patients randomized to TAU can expect inferior treatment or receive less attention than the intervention group (Freedland *et al*., 2011). A third challenge comes with the variation in the availability and utilization of outpatient services across the world (World Health Organization [WHO], 2021). This in turn may affect the ES of TAU, for example between the US compared to Scandinavia (Löfholm, Brännström, Olsson & Hansson, 2013). Thus, Cuijpers *et al*. (2021) recommended meta-analyses to recruit studies within one country and one setting, although with the number of expected studies and resources in consideration.

Overall, there is substantial evidence of the effectiveness of psychological and psychopharmacological treatment for patients over a wide range of disorders (Leichsenring, Steinert, Rabung & Ioannidis, 2022), and EBT in routine care (Wakefield, Kellett, Simmonds-Buckley, Stockton, Bradbury & Delgadillo, 2021), but few studies have examined TAU restricted to so-called clinically representative therapy. In a review of meta-analyses, three increasingly stringent and cumulative criteria defined clinically representative treatment (Shadish, Matt, Navarro *et al*., 1997). First, patients had to be referred in a conventional manner into routine clinics, with regular therapists having regular caseloads. Second, treatments had to be unaltered by the researchers. Finally, patients had to have a spread of mental health problems and background characteristics, while therapists were free to use a variety of techniques, and not trained for the specific study. With over 500 studies examined, only one fulfilled all these criteria

(Shadish *et al*., 1997). An updated review found more studies (Shadish, Matt, Navarro & Phillips, 2000). The random effect size was $d = 0.41$ and effects increased with larger treatment dose and use of specific outcome measures. However, no statistically significant association between the degree of clinical representativeness and ES was found.

Others have reviewed TAU for depression and anxiety but restricted to be a control condition to cognitive behavior therapy (CBT; Watts *et al*., 2015), EBT (Wampold *et al*., 2011) or guideline-adherent interventions (Setkowski, Boogert, Hoogendoorn, Gilissen & van Balkom, 2021). In studies where TAU was a psychotherapy intervention for depression and anxiety disorders, five studies had a between-group ES for depression of $g = 0.44$ and for anxiety of $g = 0.34$ in favor of CBT (Watts *et al*., 2015). A more stringent criterion of TAU as an active treatment revealed three studies with a between-group ES of $d = 0.33$ in favor of EBT (Wampold *et al*., 2011). However, the heterogeneity for TAU in general was substantial in both these meta-analyses. Guideline-adherent therapies have also shown larger ES than TAU across nine different diagnosis-specific studies, with a between-group ES of $d = 0.29$ (Setkowski *et al*., 2021). Further, an umbrella review of 102 meta-analyses compared psychotherapy with active TAU, with an ES of $d = 0.36$ in favor of psychotherapy (Leichsenring *et al*., 2022). Within-group analysis of routine treatment for patients with depression and anxiety, treated with EBT in the Improving Access to Psychological Therapies (IAPT) program, had an ES of $d = 0.87$ and 0.88 for symptoms of depression and anxiety (Wakefield *et al*., 2021).

To increase knowledge about the treatment that most people with CMHP receive, an updated systematic review of clinically representative therapies is necessary. In this review the Nordic countries were selected, due to their many similarities in population characteristics and health care utilization. The Nordic region is often considered a distinct region in international comparisons of health care systems, characterized by high-trust, high-taxation, open economies (Lyttkens, Christiansen, Häkkinen, Kaarboe, Sutton & Welander, 2016). Despite a policy for evidence-based treatments across the Nordic countries, there is reason to believe that there is a great deal of non-adherence to these by therapists (Bergmark, Sundberg, Markström & Rosenberg, 2022). Although there is a lot of evidence regarding the effectiveness of evidence-based treatment, there appears to be a knowledge gap regarding the effect of the treatment provided in routine mental health care. A systematic review and meta-analysis on TAU in the Nordic region may serve as a benchmark of the effect of treatment that most patients receive in routine mental health care facilities, while at the same time unveiling the characteristics of TAU in these countries.

In this review we aimed to examine: (1) the contents of clinically representative mental health outpatient treatment for adults with CMHP in the Nordic countries; and (2) provide a meta-analysis regarding its treatment effects.

METHOD

In the present review, TAU was defined in accordance with the strictest definition of clinically representative therapy (Shadish *et al*., 1997). Treatment had to be active, unaltered by the researchers, and conducted in publicly available outpatient facilities where patients are referred through usual clinical routes. Thus, research done in private care and university clinics were excluded. Common mental health problems (CMHP) were defined according to, but not limited to, the clinical guidelines by NICE (2011), which use the term for depressive and anxiety disorders. Unlike common mental disorders (Hewlett & Moran, 2014), CMHP was conceptualized to include subclinical populations. Most patients with CMHP are treated within primary health care (NICE, 2011), but CMHP are also the most common disorders in heterogeneous outpatient treatment in secondary care. Thus, heterogeneous outpatient facilities were defined as publicly available primary or secondary care facilities aimed to treat CMHP, including depressive and anxiety disorders as well as related mental health problems.

Patients had to display heterogeneity regarding demography and mental health problems. Records examining only specific disorder groups, or for example, severe mental health disorders (defined by the authors themselves), or only suicidal patients were excluded. Therapists could also not rely on a specific technique, thus records where for example, CBT was termed TAU and was the only treatment intervention, were excluded. Both RCTs and open trials using prospective study designs were included.

Search strategy

A systematic search in the electronic databases PubMed, EMBASE, PsycINFO, and SveMed+ was conducted using no time-limit, which is further described in a preregistered protocol on PROSPERO (CRD42020213988). The search string is presented in Appendix S1. First, a search strategy for PubMed was designed. It was then adapted to the other databases regarding syntax and search field tags. The search strategy included variations and synonyms of mental health problems, "routine outpatient treatment" and outcome measurements, together with the Nordic countries: Denmark, Finland, Iceland, Norway, and Sweden.

Terms were combined with Boolean operators (OR/AND) along with truncation. Nordic countries were searched in all fields, other terms in title/abstract and corresponding index terms. Duplicates were discarded first in Endnote and then in the web tool Rayyan (Ouzzani, Hammady, Fedorowicz & Elmagarmid, 2016). Reference lists and citing reports of the records read for full text were identified using scopus.com, thus backwards and forward citation searching was conducted. Three researchers (MB, ML & JL), independently and blinded from each other, screened titles and abstracts for eligibility according to inclusion and exclusion criteria. Conflicting results were discussed to reach consensus. Records eligible for full text reading, but later excluded were documented, and the reason for the first discovered exclusion criterion was plotted. The following criteria for exclusion were used: not Nordic, no heterogeneous adult sample, not outpatient facility, not TAU, and not prospective study. Records using the same dataset as one already excluded or included were marked as secondary data if no new information of interest were presented. A data extraction table was designed, piloted, and used to extract data. For this review, PRISMA guidelines were followed (see Appendix S2).

Quality assessment

The Downs and Black (1998) checklist for assessment of the methodological quality of both randomized and non-randomized studies of health care interventions was applied. The checklist uses subscales regarding reporting, external validity, bias of the measurement/outcome, confounding selection of study subjects and power. The instrument was modified (equal to Bear, Edbrooke-Childs, Norton, Krause & Wolpert, 2020; questions regarding the intervention group were not used, that is, items 14 and 21–24), thus yielding a maximum score of 23, where higher scores indicated better quality. Correspondingly, cut-offs (as suggested by Hooper, Jutai, Strong & Russell-Minda, 2008) were adjusted, where less than 10 meant the study demonstrated poor, 10–14 fair, 15–20 good, and 21–23 excellent methodological quality. For the power analysis item, the research group determined a maximum score of 1 for sufficient power, and power for all studies was manually calculated. Two researchers

conducted the assessment independently and blinded, and intraclass correlation was calculated. Conflicting results were discussed in the research team until full consensus was reached. The scores were used to analyze the quality regarding external and internal validity, but also using an overall score in meta-regression analysis. The interrater reliability for the quality scale resulted in a moderate level of agreement (intraclass correlation coefficient = 0.74, 95% confidence interval [CI] = 0.66–0.79, $p < 0.001$).

Statistical analyses

IBM SPSS Statistics (version 27) and Comprehensive Meta-Analysis (version 3.3.070; Borenstein, Hedges, Higgins & Rothstein, 2014) were used for statistical analysis. Both between-group and within-group meta-analyses were conducted with random effects model using Hedges' *g*, presented by forest plot. Generic patient reported outcome measures or subscales measuring broader change in symptoms were used as primary outcome measures. Well established global psychotherapeutic outcome measures (e.g., as reviewed by Tarescavage & Ben-Porath, 2014) were preferred. If more than one measure of interest was reported, the most suitable measure was decided in the research group.

For studies reporting several points of assessment after the post-treatment assessment, the final point of assessment was used as follow-up. The intervention was assumed to affect the post-measure and standard deviation, thus as recommended by Lakens (2013), the following formula for ES was used: (*M*pre − *M*post)/*SD*pre. Due to unknown pre-post-treatment correlations, this was imputed at 0.5 (Follmann, Elliott, Suh & Cutler, 1992). For studies presenting results for completers only analysis, the post sample size was used. Heterogeneity was estimated with the *Q*-value and I^2 (Higgins, Thomas, Chandler *et al*., 2021). Risk of publication bias was analyzed by inspection of Egger's regression intercept (Egger, Davey Smith, Schneider & Minder, 1997), and by Duval and Tweedie's (2000) trim-and-fill method.

The following variables were extracted for qualitative synthesis and to examine sources of heterogeneity: main intervention, study design (RCT or open trial), generic outcome measure, other outcome measures, country, level of care (primary or secondary care), therapist profession, format of therapy, type of psychological intervention, percent declined, percent excluded before study started, non-starters, attrition to follow-up, quality of the studies, data-collection years, number of patients at start of treatment, percent female, mean age, working status, civil status, non-nativity, education, diagnosis, session mean, mean duration, percent pharmacotherapy, and weeks between pre and post measurement. To examine sources of heterogeneity, subgroup analysis was used for categorical variables, and meta-regression for continuous moderators. The Cochrane handbook recommends at least 10 studies as the lowest number for conducting subgroup analysis or meta-regression (Higgins *et al*., 2021). Thus, subgroup-analysis and meta-regression were applied if variables from at least 10 studies provided data on the variable in question.

RESULTS

Study selection

Four database searches (October 14, 2020; November 30, 2021; March 8, 2022, and November 7, 2022) and four backwards and forward searches were conducted (December 7, 2020; January 3, 2022; March 23, 2022, and November 7, 2022; see Fig. 1). A total of 7202 records were screened. Out of these, 119 reports were read in full text, and 12 studies were accepted (see Appendix S3 for full texts excluded with reason). Out of these, four articles provided follow-up data (Arvidsdotter, Marklund & Taft, 2014; Bratberg, Leira, Granan *et al*., 2021; Koksvik, Linaker, Gråwe, Bjørngaard & Lara-Cabrera, 2018; Rise, Eriksen, Grimstad & Steinsbekk, 2016). Since most articles had results by completers only (except Arvidsdotter *et al*., 2014, who provided intention to treat results), per protocol was used.

Eleven studies provided a generic outcome measure (see Appendix S4). For two studies, the 32-item behavior and symptom identification scale (BASIS-32) was chosen as it was more comprehensive than the four-item outcome rating scale (ORS). Three studies provided disorder-specific outcome measures, four studies provided measures of self-reported health, and three studies provided outcomes measuring social functioning. For meta-analytic data-synthesis, only generic outcome measures were chosen, due to too few studies providing other outcomes of interest.

Quality assessment

The included studies demonstrated a great variation in methodological quality measured by a modified version of Downs and Black (1998) (see Appendix S5). The mean quality score was 11.5 ($SD = 3.3$) out of 23 (range 8–17). Four studies demonstrated poor quality, six fair, and two good quality, while none demonstrated excellent quality. None of the included studies described adverse effects of the intervention, two described sufficiently the patient characteristics of attrition, and two controlled for it in the analyses.

Study characteristics

Design and attrition. Seven studies were RCTs and five were open trials. As a generic outcome measure, three studies used the 90-item Symptom Checklist (SCL-90), three BASIS-32, and five used others (see Appendix S4). One study did not present outcome using a generic symptom measure, and was excluded for quantitative synthesis (Ramirez, Ekselius & Ramklint, 2008). Three studies did not present clear inclusion or exclusion criteria except reasons why patients declined (Østergård, O'Toole, Svendsen & Hougaard, 2020;, Rise *et al*., 2016; Werbart, Levin, Andersson & Sandell, 2013). Number of declined, excluded, non-starting patients, and drop-outs (attrition) spanned considerably (see Appendix S4). However, many studies did not specify non-starters, thus attrition statistics could have been negatively affected.

Facility and patient characteristics. Six studies were conducted in Sweden, five in Norway, and one in Denmark (see Table 1). No studies conducted in Finland or Iceland were included. Four studies examined treatment in primary care, seven examined specialist mental health services or secondary care, and one study examined both primary care and specialist mental health services (correspondingly 30% and 70% of the patients; Werbart *et al*., 2013). All studies had a majority of female patients (average 72%, range 55%–90%), with mean age of 31.7 years (range 22–42 mean years).

Diagnoses were reported in eight studies, three using DSM-IV criteria (Hansson, Rundberg, Österling, Öjehagen & Berglund, 2013; Ramirez *et al*., 2008; Werbart *et al*., 2013), three ICD-10 criteria (Bratberg *et al*., 2021; Brattland, Koksvik, Burkeland *et al*., 2018; Møllersen, Sexton & Holte, 2009) and

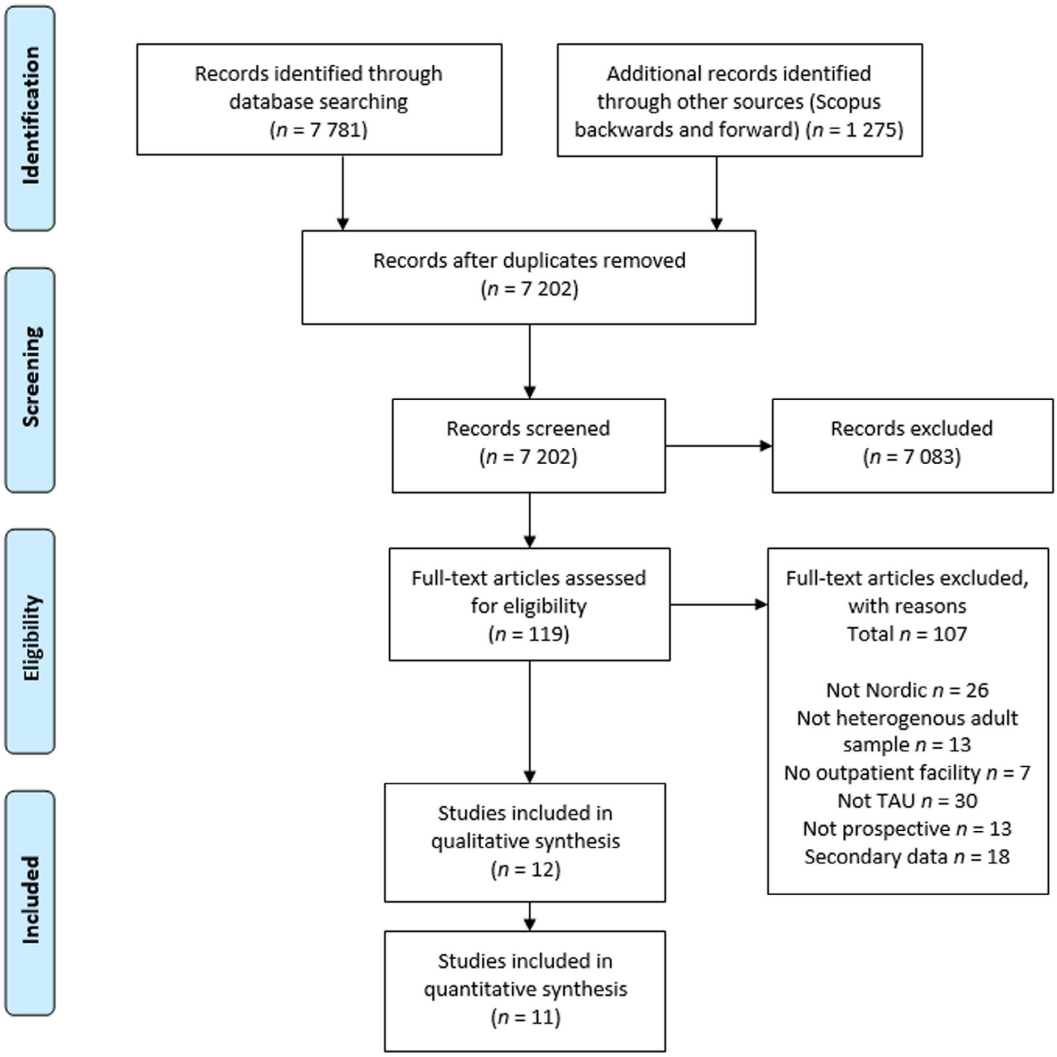


Fig. 1. PRISMA flow chart of included studies.

two without a specified diagnostic system (Arvidsdotter *et al*., 2014; Østergård *et al*., 2020). Patients with depressive, or anxiety disorders/mental health problems were reported in eight studies (depression range 28%–73%, anxiety range 20%–67%), with an estimated total proportion of diagnoses of 43% depression and 34% anxiety disorders.

Therapists and treatment characteristics. The most commonly reported professions at the primary health care level were psychologists and social workers (see Table 2). At the secondary level, psychiatrists and psychologists were most frequently reported, but also social workers, psychiatric nurses, nursing assistants, physiotherapists, occupational therapists, and milieu therapists were reported. Four studies did not report health care professionals' background.

Type of treatment also varied: two studies reported psychotherapy as the only treatment given, seven studies reported psychotherapy and psychopharmacotherapies used in combination. Three studies did not explicitly report the format of therapy (Koksvik *et al*., 2018; Ramirez *et al*., 2008; Rise *et al*., 2016). All treatments were given face-to-face individually or in group. The most common treatment interventions were CBT, but also metacognitive-, psychodynamic-, support-, psychoeducational-, systemic-, humanistic-, and existential-therapies were reported. Seven studies did not report if specific treatment interventions were given.

Six studies presented mean number of sessions per patient (M = 7.13, range 4–27 sessions), and one study presented range, where 52% had 1–10 visits (Hansson *et al*., 2013). Four articles presented mean duration (M = 100 days, range 56 days–

Table 1. *Facility and patient characteristics*

Study	Data collection	Country	Level of care	TAU *n*	Female %	Age *M* (*SD*)	Working status	Civil status	Disorders/problems
Arvidsdotter, Marklund, and Taft (2014)	2010–2011	Sweden	Primary	40	88	40 (9.1)	Unknown	Unknown	Dep 30% Anx 20%
Bratberg *et al.* (2021)	2014–2016	Norway	Secondary	35	64	31.7 (10.7)	56% employed	Unknown	Dep 28% Anx 59%
Brattland *et al.* (2018)	2012–2016	Norway	Secondary	85	60	34.6 (12)	48% not working	51% single	Dep 28% Anx 28%
Elfström *et al.* (2013)	Unknown	Sweden	Primary	133	71	42.4 (12.9)	Unknown	Unknown	Unknown
Hansson *et al.* (2013)	2007–2008	Sweden	Secondary	186	69	39 (14.1)	39% unemployed	66% single	Dep 33% Anx 24%
Koksvik *et al.* (2018)	2009–2013	Norway	Secondary	40	76	37.09 (12.8)	23.9% employed	40% married	Unknown
Møllersen *et al.* (2009)	1999–2001	Norway	Secondary	335	61	36.01 (11.9)	60% employed	Unknown	Dep 43% Anx 31%
Østergård *et al.* (2020)	2014–2016	Denmark	Primary	740	75	25.19 (4.6)	Unknown	Unknown	Dep 43% Anx 30%
Ramírez *et al.* (2008)	2002–2004	Sweden	Secondary	191	80	22.4 (1.9)	Unknown	Unknown	Dep 73%[b] Anx 67%
Rise *et al.* (2016)	2010– Unknown	Norway	Secondary	38	55	29.2 (unknown)	15.8% working	37% living alone	Unknown
Sundquist *et al.* (2017)	2012	Sweden	Primary	105	90	41 (11)	Unknown	65% married[a]	Unknown
Werbart *et al.* (2013)	2007–2010	Sweden	Primary & Secondary	180	74	36 (10.9)	Unknown	Unknown	Dep 42% Anx 24%

[a]Information provided in Sundquist *et al.*, 2019,
[b]Depressive disorders, dysthymia and bipolar disorder.

7 months), and three presented frequencies, between weekly to monthly sessions. Three studies provided mean number of sessions for both the primary intervention and TAU: one study reported more sessions provided to the intervention (5%; Bratberg *et al.*, 2021), and two studies reported more sessions provided to TAU (7% in Brattland *et al.*, 2018; 21% for group treatment, 5% for individual treatment in Østergård *et al.*, 2020). One study stated no difference in number of sessions (Hansson *et al.*, 2013), and one provided the mean of the intervention together with TAU (Rise *et al.*, 2016).

Five studies reported use of medication, of which two explicitly reported pharmacological drug of relevance for mental health condition (33% in Bratberg *et al.*, 2021; 47.4% in Rise *et al.*, 2016). One study reported drug categories (antidepressant 35%, anxiolytics 16% in Sundquist, Palmér, Johansson & Sundquist, 2017). Two studies presented percentage of patients with pharmacological drug use, without explicitly reporting if it was of relevance for their mental health condition (34.9% in Møllersen *et al.*, 2009; 41% in Werbart *et al.*, 2013).

Results of synthesis

Power analysis. For the within-group meta-analysis, 12 ESs were calculated with a mean number of treated participants of 117.8 (n = 1,413, SD = 126.6, range 28–480), and for the between-group analysis there were nine ESs with a mean number of treated participants of 103.8 (intervention n = 937, SD = 149.6, range 22–492, TAU n = 932, SD = 144.8, range 28–480). According to the formulas for power analysis in meta-analyses by Valentine, Pigott, and Rothstein (2010), there would be a 75.6% power for within-group and 57.8% power for between-group to detect a small ES (0.20), when assuming that the heterogeneity of ESs was high.

Meta analysis. Random effects model for overall within-group resulted in a significant small to moderate ES (g = 0.49, 95% CI = 0.30–0.68, $p < 0.001$; forest plot is displayed in Fig. 2). However, the heterogeneity between the studies was substantial, with a Q-value of 107.4, $df(11)$, $p < 0.001$, and I^2 = 90%. For the studies that reported follow-up data (n = 4), the follow-up ES (g = 0.64, 95% CI = 0.44–0.84, $p < 0.001$) was non-significantly different from the post ES (Q = 1.16, p = 0.281). For these, a non-significant test of heterogeneity was found, Q = 0.49, $df(3)$, p = 0.920, $I^2 < 1\%$ (see Appendix S6). Nine ESs were extracted for TAU compared to an intervention (see Fig. 2)

The random effects model for between-group resulted in a significant small ES in favor of the interventions (g = −0.21, 95% CI = −0.36 to −0.05, p = 0.010). The Q-value, Q = 15.90, $df(8)$, p = 0.044, was significant and the I^2 showed moderate heterogeneity (50%). For the few studies that reported follow-up data (n = 4), a non-significantly different ES was obtained (Q = 0.08, p = 0.782; g = −0.25, 95% CI = −0.52 to −0.02, p = 0.066), with smaller heterogeneity (Q = 3.21, $df(3)$, p = 0.361, I^2 = 6%; see Appendix S6).

Publication bias. Publication bias did not seem to be a critical issue regarding within-group analysis, with a non-significant Egger's regression intercept (t = −1.21, p = 0.253), and Duval

Table 2. *Therapists and treatment characteristics*

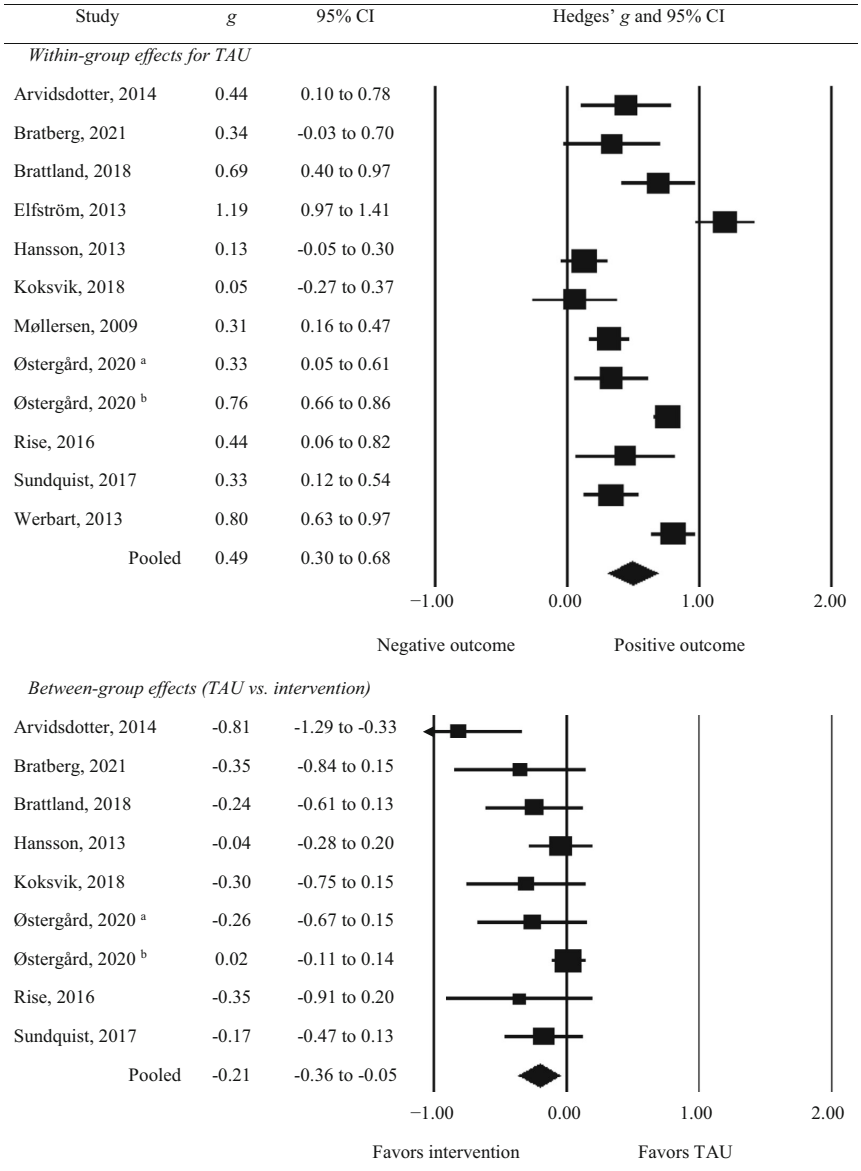
Study	Therapists profession (*n*)	Format of therapy	Psychological Intervention	Sessions *M* (*SD*)	Mean duration (Frequency)	Medication
Arvidsdotter *et al.* (2014)	Unknown (Unknown)	Psychological & pharmacological	Unknown	Unknown	Unknown	Unknown
Bratberg *et al.* (2021)	Psychologists, psychiatrists, psychiatric nurses, milieu therapists and social working clinicians (Unknown)	Psychological & pharmacological	CBT, MCT, PDT, and/or support therapy	13.8 (Unknown)	7 months (Unknown)	33%
Brattland *et al.* (2018)	Psychologists (11), psychiatrists (6), and other mental health care professions (3) (Total 20)	Psychological	PDT, humanistic/ existential, CBT	13.01 (10.92)	(Weekly/Biweekly)	Unknown
Elfström *et al.* (2013)	Psychologists (Unknown)	Psychological	Unknown	Unknown	Unknown	Unknown
Hansson *et al.* (2013)	Psychiatrists, nurses, psychologists, social workers, physiotherapists and occupational therapists (Total 56)	Psychological & pharmacological	Unknown	Unknown	10 weeks (Unknown)	Unknown
Koksvik *et al.*(2018)	Unknown (Unknown)	Not specified	Unknown	Unknown	Unknown	Unknown
Møllersen *et al.* (2009)	Psychiatrists (11), Psychologists (9), Psychiatric nurse (7), Clinical social workers (6) (Total 33)	Psychological & pharmacological	Unknown	9.0 (8.0)	7.2 months (1.4 times per month)	34.9%
Østergård *et al.* (2020)	Psychologists and social workers (Total 33)	Psychological	CBT, PDT, systemic/ humanistic	9.3 (4.78)[a] 4.12 (2.46)[b]	56 days (10 days)[a] 83 days (27 days)[b]	Unknown
Ramirez *et al.* (2008)	Unknown (Unknown)	Not specified	Unknown	Unknown	Unknown	Unknown
Rise *et al.* (2016)	Psychiatric nurse (3), psychologists (10) (Total 17)	Not specified	Unknown	Unknown	Unknown	47.4%
Sundquist *et al.* (2017)	Psychologists and social counsellors (Unknown)	Psychological & pharmacological	80% CBT, 20% unknown	6.3 (Unknown)	Unknown	35% Antidep. 16% Anx.
Werbart *et al.* (2013)	Unknown (Total 75)	Psychological & pharmacological	17% CBT, 66% PDT, 17% INT	27.4 (Unknown)	Unknown (intervention dependent)	41%

Notes: Anx = Anxiolytics/tranquilizer; Antidep = Antidepressive medication; CBT = Cognitive therapy or cognitive behavioral therapy; INT = Interpersonal therapy; MCT = Metacognitive therapy; PDT = Psychodynamic therapy.

[a]Group treatment,

[b]Individual treatment.

Forest plots depicting random effect sizes at post-treatment for TAU and comparisons with other interventions



[a] Group treatment. [b] Individual treatment.

Fig. 2. Forest plots depicting random effect sizes at post-treatment.

and Tweedie's (2000) method suggested one condition to be trimmed (adjusted $g = 0.53$). However, potential publication bias was found for between-group analysis, with significant Egger's regression intercept ($t = -3.49$, $p = 0.010$), and Duval and Tweedie's procedure indicated five conditions to be trimmed left of the mean (adjusted $g = -0.06$). Thus, the difference between the intervention and TAU groups could be close to zero when adjusted for publication bias.

Moderator analysis. Only the subgroup analysis for difference between open trials and RCTs was statistically significant (see Table 3; $Q = 4.60$, $p = 0.032$). Open trials demonstrated both

Table 3. *Subgroup analysis for clinically representative treatment (within-group random effects)*

Variable	*k*	*g*	95% CI	*Q*	*p*
Country				1.402	0.496
Denmark	2	0.563	0.144–0.982		
Norway	5	0.365	0.173–0.556		
Sweden	5	0.579	0.192–0.965		
Study design				4.598	0.032*
Open trial	5	0.680	0.407–0.953		
RCT	7	0.331	0.166–0.496		
Level of care				3.087	0.079
Primary	5	0.620	0.323–0.917		
Secondary	6	0.312	0.140–0.485		

Notes: *k* = number of comparisons.
*$p < 0.05$.

Table 4. *Meta-regression analysis of potential moderators of treatment outcome (within-group analysis)*

Variable	*k*	Point est.	*z*	*p*
Quality	12	0.003	0.09	0.926
N start	12	0.001	0.82	0.412
Female %	12	−0.175	−0.17	0.864
Age mean	12	0.003	0.16	0.876
Attrition	11	0.161	0.18	0.855
Excluded	10	0.033	0.04	0.964
Weeks after start of treatment	11	−0.001	−0.14	0.890
Publication year	12	−0.007	−0.22	0.823

Notes: *k* = number of comparisons. Publication year = years after first study, namely, 2013.

higher ES ($g = 0.68$, $p < 0.001$) and also greater heterogeneity (df[4], $Q = 53.10$, $p < 0.001$, $I^2 = 92\%$) compared to RCTs ($g = 0.33$, $p < 0.001$, df[6], $Q = 14.57$, $p = 0.024$, $I^2 = 59\%$). When one outlier was removed (Elfström, Evans, Lundgren, Johansson, Hakeberg & Carlsson, 2013) the results demonstrated the same tendency, but were non-significant (open trials $g = 0.56$, RCT $g = 0.33$, $Q = 2.17$, $p = 0.141$). Statistically significant differences were not found for country, level of care, or any of the continuous variables in the meta-regression analysis including study quality (see Table 4).

DISCUSSION

This review demonstrated a great variability in what constituted clinically representative therapy, TAU, and resulted in a small to moderate within-group ES ($g = 0.49$). It also showed high heterogeneity, which partially was explained by research design (higher effects in open trials, smaller in RCTs). Compared to a broad set of interventions, TAU was only marginally less effective ($g = -0.21$, adjusted for potential publication bias $g = -0.06$). Further, follow-up scores were not significantly different from post-treatment scores. The results should be interpreted with caution and not as the true effect of clinically representative therapy in the Nordic countries, as there is a need for more studies of higher methodological quality.

The results demonstrated a variety of methodological qualities, and a plethora of patient characteristics, professional backgrounds and treatments were reported in these, assumed to be, clinically representative therapies within the Nordic countries. However, more than half of the studies did not provide information on what intervention the psychological treatment consisted of, and half of the comparative studies did not provide mean number of sessions of both intervention and TAU. In addition, assessment for publication bias indicated potentially missing studies of TAU with higher ES. Further, high heterogeneity of ES is a finding in line with previous meta-analyses of TAU (Wampold *et al*., 2011; Watts *et al*., 2015). Still, no study in the present review examined TAU on its own as the primary intervention, which is noteworthy, considering the widely recognized knowledge gap on routine treatment (Hewlett & Moran, 2014). However, this review used strict inclusion criteria, which excluded articles examining other routine treatments, and this may have biased the results.

TAU in the Nordic countries demonstrated weaker within-group ES ($g = 0.49$) than EBT in IAPT ($d = 0.87$–0.88; Wakefield *et al*., 2021). Thus, TAU seems to be less effective than evidence-based treatments, but compared to a broad set of interventions, the difference could be negligible, especially when adjusting for a potential publication bias. One excluded but highly relevant study (Nordmo, Sønderland, Havik, Eilertsen, Monsen & Solbakken, 2020) reported a larger effect size ($d = 0.85$) for patients treated between 1995–2008, who received a considerably higher number of sessions (mean of 51) than the included studies (range 4–27). This could indicate that a larger dose of psychotherapy is associated with increased effects. As noted by Shadish *et al*. (2000), study outcomes may vary depending on treatment dose as well as sample- and treatment characteristics. The effect size reported by Shadish *et al*. (2000), $d = 0.41$, which resembles the effect found in our study, may thus be representable for routine outpatient psychiatric facilities in the Scandinavian countries today, but as illustrated by the Nordmo *et al*. study, this effect could vary depending upon treatment duration.

Adding to the already acknowledged efficacy-effectiveness gap, this meta-analysis demonstrated a significant difference in ES in favor of studies conducted in open trials, in contrast to other meta-analyses (e.g., Shadish *et al*., 2000). It has been said that the ES of effectiveness studies could easily be overestimated, for example, due to regression towards the mean and spontaneous remissions (Cuijpers, Weitz, Cristea & Twisk, 2017). However, using within-group comparisons, the same confounding factors apply to efficacy studies.

Limitations

The heterogeneity of TAU, as demonstrated both in the systematic review and meta-analysis could arguably undermine the certainty of the results. Although this meta-analysis included more studies than median number of studies found in the Cochrane library, the I^2 statistics is both prone to be imprecise and biased in small meta-analyses (von Hippel, 2015). While our results had a comparable between-size I^2 statistics compared to other TAU meta-analyses, it also indicates something else: Nordic mental health care could be very unequal, which contradicts the very presumption of egalitarian health care systems.

All Nordic countries were represented in the search strategy, but we identified no matching articles from Iceland and Finland.

Thus, only Scandinavian studies were synthesized. The database SveMed+ was only available for Scandinavian gray literature and was no longer updated as of January 2020. Although no gray literature was found eligible in the present study, there was a risk of not retrieving potential non-peer reviewed literature.

Additionally, due to low power, moderator analysis of many potentially confounding variables was not conducted, such as for instrument information (e.g., language of instruments), patient characteristics (e.g., diagnosis), or treatment information (e.g., use of pharmacotherapy or number of sessions). The latter has been demonstrated to be significantly associated with ES in one of the few meta-analyses on clinically representative therapies (Shadish *et al*., 2000). Also, many variables were presented in different ways, such as education, working status, and medication, which resulted in insufficient information for subgroup comparisons. Moreover, the subgroup analysis that was conducted may have been underpowered, which poses a risk of both alpha and beta errors. Therefore, non-significant findings of moderators in the current study must be interpreted with caution.

The present meta-analysis included studies with generic measures of symptom severity, which may have affected the results. Generic measures have been suggested justified for comparisons across diagnostic groups but are also associated with less precision and lower estimates of ES than specific measures (de Beurs, Vissers, Schoevers, Carlier, van Hemert & Meesters, 2019; Shadish *et al*., 2000).

CONCLUSION

To the best of our knowledge this is the first review to systematically assess clinically representative TAU in a restricted geographical area. Unlike previous meta-analyses, this study applied a stricter definition of TAU, and only studies with heterogeneous clinical samples and treatments were included. Although limited to the Nordic countries, this study demonstrated the ambiguity of TAU, but also its effect compared to other comparable meta-analyses.

Although with limitations, this review and meta-analysis may not only serve as a benchmarking study of clinical effect in mental health treatment for CMHP within the Nordic region, but also an in-depth examination of the nature of TAU within a region that shares many commonalities within the mental health sector. Considering the widespread use of TAU in clinical practice, and also the lack of research on it, there is a need for a pivotal change in research attitude toward routine treatment. Further research is warranted, to increase the understanding of the most commonly delivered treatment by the majority of mental health professionals, to the majority of patients.

ACKNOWLEDGMENT

Data is available on reasonable request. The study was conducted without external funding. The authors declare that they have no competing interests. MB, ML, SS, OH and AH designed the study and the search strategy. MB, ML & JL extracted studies, together with their information. MB & ML did quality assessment. MB and LGÖ were responsible for statistical analyses. All authors has been involved writing this paper, making revisions and accepted it in its final form. Special thanks to Katrine Aronsen (Head librarian at library section for Medicine and Health Sciences at NTNU University Library) who helped designing the search strategy..

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

*Arvidsdotter, T., Marklund, B., & Taft, C. (2014). Six-month effects of integrative treatment, therapeutic acupuncture and conventional treatment in alleviating psychological distress in primary care patients – follow up from an open, pragmatic randomized controlled trial. *BMC Complementary and Alternative Medicine*, *14*, 210.

Bear, H.A., Edbrooke-Childs, J., Norton, S., Krause, K.R. & Wolpert, M. (2020). Systematic review and meta-analysis: Outcomes of routine specialist mental health care for young people with depression and/or anxiety. *Journal of the American Academy of Child and Adolescent Psychiatry*, *59*(7), 810–841.

Bergmark, M., Sundberg, L.R., Markström, U. & Rosenberg, D. (2022). Evidence-based methods in rural areas – knowledge and national guideline utilization in mental health service development. *Journal of Evidence-Based Social Work*, *19*(2), 161–184.

Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. (2014). *Comprehensive meta-analysis* (Version 3.3.070) [Computer software]. Biostat. Retrieved October 3, 2023, from https://www.meta-analysis.com/

*Bratberg, G., Leira, K., Granan, L-P., Jonsbu, E., Fadnes, B. L., Thuland, S. F., & Myklebust, T. Å. (2021). Learning oriented physiotherapy (LOP) in anxiety and depression: An 18 months multicentre randomised controlled trial (RCT). *European Journal of Physiotherapy*, *23*(5), 295–304.

*Brattland, H., Koksvik, J. M., Burkeland, O., Gråwe, R. W., Klöckner, C., Linaker, O. M., Ryum, T., Wampold, B., Lara-Cabrera, M. L., & Iversen, V. C. (2018). The effects of routine outcome monitoring (ROM) on therapy outcomes in the course of an implementation process: A randomized clinical trial. *Journal of Counseling Psychology*, *65*(5), 641–652.

Cuijpers, P., Karyotaki, E., Reijnders, M. & Ebert, D.D. (2019). Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. *Epidemiology and Psychiatric Sciences*, *28*(1), 21–30.

Cuijpers, P., Quero, S., Papola, D., Cristea, I.A. & Karyotaki, E. (2021). Care-as-usual control groups across different settings in randomized trials on psychotherapy for adult depression: A meta-analysis. *Psychological Medicine*, *51*(4), 634–644.

Cuijpers, P., Weitz, E., Cristea, I. & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, *26*(4), 364–368.

de Beurs, E., Vissers, E., Schoevers, R., Carlier, I., van Hemert, A.M. & Meesters, Y. (2019). Comparative responsiveness of generic versus disorder-specific instruments for depression: An assessment in three longitudinal datasets. *Depression and Anxiety*, *36*(1), 93–102.

Downs, S.H. & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, *52*(6), 377–384.

Duval, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634.

*Elfström, M. L., Evans, C., Lundgren, J., Johansson, B., Hakeberg, M., & Carlsson, S. G. (2013). Validation of the Swedish version of the clinical outcomes in routine evaluation outcome measure (CORE-OM). *Clinical Psychology & Psychotherapy*, *20*(5), 447–455.

Follmann, D., Elliott, P., Suh, I. & Cutler, J. (1992). Variance imputation for overviews of clinical trials with continuous response. *Journal of Clinical Epidemiology*, *45*(7), 769–773.

Freedland, K.E., Mohr, D.C., Davidson, K.W. & Schwartz, J.E. (2011). Usual and unusual care: Existing practice control groups in randomized controlled trials of behavioral interventions. *Psychosomatic Medicine*, *73*(4), 323–335.

*Hansson, H., Rundberg, J., Österling, A., Öjehagen, A., & Berglund, M. (2013). Intervention with feedback using outcome questionnaire 45 (OQ-45) in a Swedish psychiatric outpatient population. A randomized controlled trial. *Nordic Journal of Psychiatry*, *67*(4), 274–281.

Hewlett, E. & Moran, V. (2014). *Making mental health count: The social and economic costs of neglecting mental health care*. OECD Health Policy Studies, OECD Publishing. https://doi.org/10.1787/9789264208445-en.

Higgins, J.P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. *et al*. (2021). *Cochrane handbook for systematic reviews of interventions*. version 6.2 (updated February 2021). Retrieved October 3, 2023, from https://training.cochrane.org/handbook.

Hooper, P., Jutai, J.W., Strong, G. & Russell-Minda, E. (2008). Age-related macular degeneration and low-vision rehabilitation: A systematic review. *Canadian Journal of Ophthalmology. Journal Canadien D'ophtalmologie*, *43*(2), 180–187.

Kazdin, A.E. (2015). Treatment as usual and routine care in research and clinical practice. *Clinical Psychology Review*, *42*, 168–178.

*Koksvik, J. M., Linaker, O. M., Gråwe, R. W., Bjørngaard, J. H., & Lara-Cabrera, M. L. (2018). The effects of a pretreatment educational group programme on mental health treatment outcomes: A randomized controlled trial. *BMC Health Services Research*, *18*(1), 665.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863.

Leichsenring, F., Steinert, C., Rabung, S. & Ioannidis, J. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *21*(1), 133–145.

Löfholm, C.A., Brännström, L., Olsson, M. & Hansson, K. (2013). Treatment-as-usual in effectiveness studies: What is it and does it matter? *International Journal of Social Welfare*, *22*, 25–34.

Lyttkens, C.H., Christiansen, T., Häkkinen, U., Kaarboe, O., Sutton, M. & Welander, A. (2016). The core of the Nordic health care system is not empty. *Nordic Journal of Health Economics*, *4*(1), 7–27.

*Møllersen, S., Sexton, H. C., & Holte, A. (2009). Effects of client and therapist ethnicity and ethnic matching: A prospective naturalistic study of outpatient mental health treatment in northern Norway. *Nordic Journal of Psychiatry*, *63*(3), 246–255.

National Institute for Health and Care Excellence (NICE). (2011). *Common mental health problems: identification and pathways to care*. (NICE Clinical Guideline CG123). Retrieved October 3, 2023, from https://www.nice.org.uk/guidance/CG123.

Nordmo, M., Sønderland, N.M., Havik, O.E., Eilertsen, D.E., Monsen, J.T. & Solbakken, O.A. (2020). Effectiveness of open-ended psychotherapy under clinically representative conditions. *Frontiers in Psychiatry*, *11*, 384.

*Østergård, O. K., O'Toole, M. S., Svendsen, M. M., & Hougaard, E. (2020). The effect of implementing the partners for change outcome management system in a nationwide student counseling service. *Journal of Counseling Psychology*, *67*(1), 90–103.

Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210.

*Ramirez, A., Ekselius, L., & Ramklint, M. (2008). Axis V – global assessment of functioning scale (GAF), further evaluation of the self-report version. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *23*(8), 575–579.

*Rise, M. B., Eriksen, L., Grimstad, H., & Steinsbekk, A. (2016). The long-term effect on mental health symptoms and patient activation of using patient feedback scales in mental health out-patient treatment. A randomised controlled trial. *Patient Education and Counseling*, *99*(1), 164–168.

Setkowski, K., Boogert, K., Hoogendoorn, A.W., Gilissen, R. & van Balkom, A. (2021). Guidelines improve patient outcomes in specialised mental health care: A systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, *144*(3), 246–258.

Shadish, W.R., Matt, G.E., Navarro, A.M. & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin*, *126*(4), 512–529.

Shadish, W.R., Matt, G.E., Navarro, A.M., Siegle, G., Crits-Christoph, P., Hazelrigg, M.D. *et al*. (1997). Evidence that therapy works in clinically representative conditions. *Journal of Consulting and Clinical Psychology*, *65*(3), 355–365.

*Sundquist, J., Palmér, K., Johansson, L. M., & Sundquist, K. (2017). The effect of mindfulness group therapy on a broad range of psychiatric symptoms: A randomised controlled trial in primary health care. *European Psychiatry: the Journal of the Association of European Psychiatrists*, *43*, 19–27.

Sundquist, J., Palmér, K., Memon, A.A., Wang, X., Johansson, L.M. & Sundquist, K. (2019). Long-term improvements after mindfulness-based group therapy of depression, anxiety and stress and adjustment disorders: A randomized controlled trial. *Early Intervention in Psychiatry*, *13*(4), 943–952.

Tarescavage, A.M. & Ben-Porath, Y.S. (2014). Psychotherapeutic outcomes measures: A critical review for practitioners. *Journal of Clinical Psychology*, *70*(9), 808–830.

Valentine, J.C., Pigott, T.D. & Rothstein, H.R. (2010). How many studies do you need?: A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247.

von Hippel, P.T. (2015). The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Medical Research Methodology*, *15*, 35.

Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A. & Delgadillo, J. (2021). Improving access to psychological therapies (IAPT) in the United Kingdom: A systematic review and meta-analysis of 10-years of practice-based evidence. *The British Journal of Clinical Psychology*, *60*(1), 1–37.

Wampold, B.E., Budge, S.L., Laska, K.M., Del Re, A.C., Baardseth, T.P., Fluckiger, C. *et al*. (2011). Evidence-based treatments for depression and anxiety versus treatment-as-usual: A meta-analysis of direct comparisons. *Clinical Psychology Review*, *31*(8), 1304–1312.

Watts, S.E., Turnell, A., Kladnitski, N., Newby, J.M. & Andrews, G. (2015). Treatment-as-usual (TAU) is anything but usual: A meta-analysis of CBT versus TAU for anxiety and depression. *Journal of Affective Disorders*, *175*, 152–167.

*Werbart, A., Levin, L., Andersson, H., & Sandell, R. (2013). Everyday evidence: Outcomes of psychotherapies in Swedish public health services. *Psychotherapy (Chicago, Ill.)*, *50*(1), 119–130.

World Health Organization. (2021). *Mental health atlas 2020*. Geneva: World Health Organization. Retrieved October 3, 2023, from. https://apps.who.int/iris/handle/10665/345946.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Search strategy in PsycINFO with an Ovid interface.

Appendix S2. PRISMA checklist.

Appendix S3. Excluded after full text reading.

Appendix S4. Study design, attrition rate, and methodological quality of included studies

Appendix S5. Downs and Black checklist of quality assessment.

Appendix S6. Forest plots depicting random effect sizes at follow-up.

Received 15 November 2022, Revised 3 October 2023, accepted 5 October 2023

Brattmyr *et al. BMC Psychiatry* (2022) 22:461
https://doi.org/10.1186/s12888-022-04101-z

BMC Psychiatry

RESEARCH

Open Access

Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the Generalized Anxiety Disorder scale-7 in a Norwegian psychiatric outpatient sample

Martin Brattmyr*, Martin Schevik Lindberg, Stian Solem, Odin Hjemdal and Audun Havnen

Abstract

Objective: The aim of this study was to test factor structure, measurement invariance, and concurrent validity of the nine item Patient Health Questionnaire-9 (PHQ-9) and the seven item Generalized Anxiety Disorder scale-7 (GAD-7) in a heterogeneous outpatient sample.

Method: Outpatients completed the PHQ-9, GAD-7, and the Working Social Adjustment Scale (WSAS) before starting treatment. Study design was cross-sectional, with convenience sampling. The total sample consisted of 831 participants (61% women).

Results: Both PHQ-9 and GAD-7 demonstrated better fit statistics with two-factor and bifactor solutions consisting of a cognitive and somatic factor. Omega hierarchical was .78 for PHQ-9 and .81 for GAD-7. Both instruments achieved scalar invariance across gender, diagnosis, and comorbidity. However, the somatic factors demonstrated poor discriminant validity. These factors are not well separatable and risks being too similar if used together. The general factors of both instruments were most associated with functional impairment, although PHQ-9 demonstrated a stronger association with WSAS ($\gamma = .74$, $r^2 = .62$) than GAD-7 ($\gamma = .54$, $r^2 = .32$). Using latent mean difference, women and patients with comorbidity had significantly higher scores of both depression and anxiety.

Conclusion: This study shows that the PHQ-9 and GAD-7 may be used as one-dimensional instruments in clinical settings. Tests for measurement invariance supported that both measures are understood and interpreted comparably across gender and diagnostic subgroups.

Keywords: PHQ-9, GAD-7, Factor structure, Measurement invariance, Reliability, Validity

Standardized outcome measures have been promoted for at least half a century in the mental health field [1]. Two instruments currently at the center of attention are the Patient Health Questionnaire-9 (PHQ-9) [2] measuring depression, and the Generalized Anxiety Disorder scale-7 (GAD-7) [3] measuring anxiety. These instruments have been proposed to be included in core-sets of measures in clinical research [4, 5]. However, these recommendations has also been criticized, amongst other reasons due to conflicting results regarding factor structures, uncertainties about how well the results generalize across groups, and little available knowledge on their transferability to

*Correspondence: martin.brattmyr@ntnu.no

Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

clinical contexts [6]. As a result, there is limited evidence on the adequacy of using these instrument with clinical heterogenous populations, where they also are used the most [6].

Others acknowledge that these instruments are becoming frequently more applied in research and clinical contexts, but emphasizes the importance of measuring other aspects of mental health as well, such as level of functioning [1]. Therefore, factor structures, generalizability across different patient groups, and relationship with functional impairment for PHQ-9 and GAD-7 in adult outpatients with mixed psychiatric disorders will be in focus for this study.

Many different factor structures have been suggested for PHQ-9 [7]. However, the inconsistencies in research findings can be a product of sample properties [8] and methodology [9]. Results from confirmatory factor analysis (CFA) using psychiatric outpatient populations with mixed disorders are sparse. For example, only one out of 33 articles in a recent systematic review included such heterogenous psychiatric outpatient sample [7]. In that particular study, the proposed factor-solution was a two-factor model of the PHQ-9, comprising a cognitive factor and a somatic factor [10]. Still, the usefulness of such two-factor solution has been disputed, amongst others due to a strong correlation between the factors [11]. Therefore, PHQ-9 have been suggested suitable with a bifactor-(S – 1) model assessing patients at risk, or with diabetes in India [12]. This modification of the classic symmetric bifactor model has been proposed as a solution for anomalous results due to single-level sampling and it also increases the interpretability due to using a reference domain [13].

Discussions have been similar regarding GAD-7. For heterogenous outpatient samples, both unitary models constrained with correlated residuals [14, 15], and two-factor solutions have been suggested [16]. The latter study demonstrated a two-factor model of GAD-7 using exploratory factor analysis (EFA), which consisted of a cognitive and a somatic factor, just like previous research on PHQ-9 [16]. Further, GAD-7 has also been suggested suitable with a bifactor-(S – 1) model but limited to the population mentioned above [12].

To justify comparisons between patient groups, tests of measurement invariance (MI) should demonstrate equality of indicator thresholds, or so-called scalar invariance [17]. MI implies restrictions in a hierarchical manner of a model, to point out whether and where properties of an instrument differ across groups. For example, if crying is more strongly associated with depression for women than men, an instrument measuring a latent construct of depression with an item about crying could risk biased results, and assumably not achieve scalar invariance [18]. A systematic review of MI of PHQ-9 presented support for scalar invariance across gender in several studies [7], including a psychiatric outpatient population with mixed disorders [10]. This has also been proposed for GAD-7, in a study with an heterogenous outpatient population [15]. Thus, with heterogenous psychiatric outpatients, both instruments have demonstrated scalar invariance for gender, or so-called gender invariance. However, there is still limited evidence for the Norwegian versions.

In addition to MI, it is important to evaluate the association between symptoms of depression and anxiety with functional impairment, as a way to test their usefulness in clinical contexts. A close relationship between symptoms of depression and anxiety with functional impairment is often implicitly assumed, but rarely tested [19, 20]. However, one review reported a moderate correlation between symptoms of depression and functional impairment [19] and another review reported a weak association between symptoms of anxiety and functional impairment [20]. Accordingly, symptoms of depression seem to be more associated with functional impairment than symptoms of anxiety. One commonly used instrument that measures functional impairment is the Work and Social Adjustment Scale (WSAS) [21]. It has been demonstrated with a unitary factor structure and scalar invariance across gender [22]. Studies have reported higher correlation between WSAS and PHQ-9 than WSAS and GAD-7, even when these were specified with a cognitive and a somatic factor each [23]. However, such relationships have rarely been investigated using Structural Equation Modeling (SEM).

In the current study, the factor structures of PHQ-9 and GAD-7 will be examined using CFA, where both one-, two- and bifactor models will be tested. Measurement properties across gender, diagnosis, and comorbidity will be evaluated with respect to MI, and the concurrent validity with WSAS will be investigated using SEM. Based on previous research, we hypothesize that two-factor models composed of a cognitive and a somatic factor will fit both instruments best. We expect to achieve scalar invariance across different patient groups for both instruments and that symptoms of depression will predict functional impairment to a greater extent than symptoms of anxiety.

Method

Sample

This study was based on data from a psychiatric outpatient clinic in Trondheim, Norway. Patients was referred by general practitioners, or other mental health clinics. Patients completed all instruments before starting treatment. Data was collected using a digital platform from February to November 2020 and informed consent was

given electronically. There were no exclusion criteria, but patients diagnosed with some specific disorders (e.g. psychosis and obsessive-compulsive disorder) received outpatient treatment elsewhere and was not represented in this sample. A total of 857 patients consented to participate, 145 declined. Fifteen patients completed the forms twice and the most recent was removed.

Forty-three of the patients did not answer all items. Out of these, 26 did not answer at least one question on one of the three instruments (mean age 33.44 years, 18 women), and were removed. The final sample consisted of 831 patients, with a mean age of 30.03 years ($SD = 9.99$, median = 27, range = 18–72), and 510 were women (61%).

Data for ICD-10 diagnoses was extracted in November 2020. This led to no available diagnosis for some patients that just started therapy. In this sample, 638 (77%) of the patients were diagnosed with an ICD-10 Mental and behavioral diagnosis at the time of data extraction. More women than men had been diagnosed (see Table 1). The most frequent diagnoses were mood disorders (37%) and anxiety disorders (34%). A total of 193 (23%) had comorbid diagnoses (with two or more ICD-10, chapter 5 subsections diagnosis), and of these, 99 (12%) were diagnosed with both a mood disorder (F30-F39) and an anxiety or stress disorder (F40-F49).

A majority of the patients scored over cut-off for depression and anxiety (≥ 10 for sum-score of PHQ-9 and GAD-7; see Table 1). Women scored statistically significantly higher on GAD-7 and were more associated with scoring greater than cut-off for both PHQ-9 and GAD-7.

Patients with a mood disorder and not an anxiety disorder ($n = 211$) scored significantly higher and more often over cut-off on PHQ-9, and higher on WSAS, than patients with an anxiety disorder and not a mood disorder ($n = 185$; PHQ-9 $t = 3.35$, $p < .001$, $\chi^2 = 6.27$, $p = .012$; WSAS $t = 4.05$, $p < .001$). Patients with an anxiety disorder and not mood disorder scored higher on GAD-7, although not significantly more often over cut-off (GAD-7 $t = -2.26$, $p = .024$, $\chi^2 = 1.72$, $p = .189$).

Patients with comorbid diagnosis ($n = 193$) scored significantly higher, and more often over cut-off on all instruments compared with patients diagnosed with only one diagnosis ($n = 445$; PHQ-9 $t = -4.95$, $p < .001$, $\chi^2 = 15.88$, $p < .001$; GAD-7 $t = -4.02$, $p < .001$, $\chi^2 = 13.61$, $p < .001$; WSAS $t = -2.60$, $p = .001$).

Instruments

The nine item Patient Health Questionnaire-9 (PHQ-9) measures severity of depression and can also be used as a diagnostic tool [2]. It comes with a diagnostic algorithm but using sum-score and applying a cut-off ≥10 has been suggested to be more sensitive for detecting depression [24]. PHQ-9 uses a 4-point Likert scale ranging from 0 (*not at all*) to 3 (*almost every day*). Its psychometric properties have been widely tested [25–27], and it has demonstrated good properties as a severity

Table 1 Characteristics of 831 patients on diagnostic, symptoms, and functioning including comparisons between women and men

	Total ($n = 831$)	Women ($n = 510$)	Men ($n = 321$)	t/χ^2	*p*
Demographics					
Age	30.03 (9.99)	29.53 (9.78)	30.81 (10.28)	−1.79	.072
Single	430 (52%)	240 (47%)	190 (59%)	11.61	<.001***
Sick leave	211 (25%)	130 (25%)	81 (25%)	0.01	.934
ICD-10 diagnoses					
Undiagnosed	193 (23%)	97 (19%)	96 (30%)	13.10	<.001***
Mood disorders, F30-F39	310 (37%)	188 (37%)	122 (38%)	0.11	.740
Anxiety/stress disorders, F40-F48	284 (34%)	194 (38%)	90 (28%)	8.76	.003**
Hyperkinetic disorders, F90-F98	134 (16%)	75 (15%)	59 (18%)	1.97	.161
Personality disorders, F60-F69	84 (10%)	61 (12%)	23 (7%)	4.99	.026**
Two sections or more	193 (23%)	119 (23%)	74 (23%)	0.01	.926
Sum-score					
PHQ-9	15.82 (5.71)	16.12 (5.61)	15.35 (5.85)	1.89	.059
≥ 10	700 (84.24%)	442 (86.67%)	258 (80.37%)	5.87	.015*
GAD-7	12.14 (4.89)	12.66 (4.85)	11.30 (4.83)	3.97	<.001***
≥ 10	566 (68.11%)	366 (71.76%)	200 (62.31%)	8.12	.004**

Note. Results presented include four of the most common ICD-10, chapter 5 sections from the sample. Age, and sum-score are presented as mean (SD). Single, sick leave, ICD-10 diagnoses and over cut-off are presented with number (%)

* $p < .05$, ** $p < .01$, *** $p < .001$

measure in a large psychiatric sample [10]. Psychometric properties of the Norwegian version have been tested with adolescents and adult women with and without eating disorders [28, 29].

The seven item Generalized Anxiety Disorder scale-7 (GAD-7) [3] was developed to detect and measure severity of generalized anxiety disorder. However, it has been demonstrated to perform well as a measure of other anxiety symptoms as well [16, 30]. The GAD-7 uses an identical 4-point Likert scale as the PHQ-9. It is considered to be a reliable and valid measure of anxiety symptoms in heterogenous psychiatric outpatients, amongst others in Norway and the U.S. [14, 16]. Both PHQ-9 and GAD-7 are available in several languages [31].

The Work and Social Adjustment Scale (WSAS) [21] measures functional impairment. It consists of five items that assess impairment of daily functioning (work, home chores, social leisure, private leisure, and relationships) that are rated on a 9-point Likert scale from 0 (*not at all impaired*) to 8 (*very severely impaired*). The psychometric properties of WSAS have been demonstrated in various studies, in a Norwegian outpatient setting [22] and in England, where it is suggested to be a good complement to PHQ-9 and GAD-7 [32].

Statistical analysis

Stata [33] was used for data preparation and testing group differences. Mplus version 8.4 [34] was used for CFA, MI and SEM. Missing items were less than 0.01% on all variables. Little's MCAR test showed non-significant results (PHQ-9 $p=.88$, GAD-7 $p=.78$, WSAS $p=.73$), indicating that data were missing completely at random. No imputations were done.

Weighted Least Squares Means and Variance adjusted (WLSMV) estimator was used [35], as it is less prone to bias than other estimators for ordinal data [36]. Several fit indices were used [17]: χ^2 as a measure of absolute fit, Root Mean Square Error of Approximation (RMSEA) for parsimony correction, and the comparative fit indices Comparative fit index (CFI) and Tucker-Lewis index (TLI) [37]. Thresholds close to or below .06 for RMSEA and above .95 CFI and TLI were used to indicate good fit [38].

A bifactor model was specified using the bifactor-(S – 1) modification, specified with a specific factor, and a reference domain [13]. Bifactor-(Sc – 1) was estimated with a specific cognitive group factor and by using the somatic domain as reference. Bifactor-(Ss – 1) was estimated with a specific somatic group factor and by using the cognitive domain as reference.

Internal consistency was measured with composite reliability, which has been proposed as a superior alternative to other measures [39]. A value between .7 and .9 was used for satisfactory internal consistensy. Discriminant validity was calculated with confidence intervals in CFA, using standardized Upper Limit 95% confidence intervals (UL) for correlation between the factors. UL < 0.8 indicates no problem, 0.8–0.9 indicates marginal problems, 0.9–1.0 indicates moderate problem and above 1.0 indicates severe problems [39].

Omega hierarcical was estimated [40], and omega hierarchical above .8 was interpreted to indicate a primarily one-dimensional construct [41]. Additionally, one-dimensionality was also interpreted if omega hierarchical for the general factor was over .7, percent of uncontaminated correlations (PUC) was lower than .8 and explained common variance (ECV) of the general factor was over .6 [41].

Measurement Invariance (MI) was evaluated sequentially, for configural, metric and scalar invariance, where each step implied more equality constraints. Configural invariance was achieved if the pattern of free and fixed loadings across gender was equivalent, i.e. number of factors and indicator-factor patterns were considered the same across men and women [17]. If configural invariance was supported, metric invariance was tested next, where factor loadings were constrained equally. If metric invariance was achieved, scalar invariance was evaluated by constraining item thresholds to be equal across the groups. Scalar invariance implies that differences in latent means are not biased and may be considered to be true differences between genders. We followed the recommendations by Millsap and Yun-Tein [42] and Pendergast with colleagues [43] for testing MI with ordered-categorical measures. The Mplus DIFFTEST function was used for comparison of model fit [33]. However, using ΔCFI $\geq -.01$ and ΔRMSEA <.015 has been suggested to be superior for evaluate MI, than relying on non-significant $\Delta\chi^2$ [44]. Thus, ΔCFI and ΔRMSEA was used for threshold guidance. For concurrent validity, latent path modeling with SEM was used with bifactor-(S – 1).

Results

Factor structure

Unitary factor solution of the PHQ-9 resulted in non-satisfactory fit statistics (model 1 in Table 2). PHQ-9 demonstrated better fit statistics with a two-factor solution and was accepted without modifications (model 2 in Table 2). The two-factor solution of PHQ-9 consisted of a cognitive factor of depression: PHQc (items 1, 2, 6, & 9), and a somatic factor of depression: PHQs (items 3, 4, 5, 7, & 8). Both PHQ-9 bifactor-(S – 1) models resulted in similar goodness of fit as the two-factor solution (model 3 and 4 in Table 2).

A unitary factor solution for GAD-7 showed poor model fit (model 5 in Table 2). GAD-7 was also tested

Table 2 Goodness of fit for Confirmatory factor analysis of PHQ-9, GAD-7 and WSAS ($n = 831$)

Model	χ^2	df	RMSEA [90% CI]	CFI	TLI
Total					
1. PHQ-9 single factor	341.080***	27	.118 [.107–.130]	.937	.916
2. PHQ-9 two-factor	105.070***	26	.060 [.049–.073]	.984	.978
3. PHQ-9 bifactor-(Sc – 1)	101.667***	23	.064 [.052–.077]	.984	.975
4. PHQ-9 bifactor-(Ss – 1)	103.436***	22	.067 [.054–.080]	.984	.973
5. GAD-7 single factor	183.117***	14	.121 [.105–.136]	.976	.964
6. GAD-7 single factor mod.[1]	50.288***	11	.066 [.048–.084]	.994	.989
7. GAD-7 two-factor	61.920***	13	.067 [.051–.085]	.993	.989
8. GAD-7 two-factor mod.[2]	45.815***	12	.058 [.041–.077]	.995	.991
9. GAD-7 bifactor-(Sc – 1)	42.805***	10	.063 [.044–.083]	.995	.990
10. GAD-7 bifactor-(Ss – 1)	50.288***	11	.066 [.048–.084]	.994	.989
11. WSAS single factor	138.321***	5	.179 [.154–.205]	.953	.906
12. WSAS mod.[3]	14.235***	4	.055 [.026–.088]	.996	.991
13. WSAS mod.[3] & PHQ-9 bifactor-(Sc – 1)	274.640***	70	.059 [.052–.067]	.976	.968
14. WSAS mod.[3] & PHQ-9 bifactor-(Ss – 1)	274.386***	69	.060 [.053–.067]	.975	.968
15. WSAS mod.[3] & GAD-7 bifactor-(Sc – 1)	180.710***	47	.059 [.050–.068]	.985	.979
16. WSAS mod.[3] & GAD-7 bifactor-(Ss – 1)	195.707***	48	.061 [.052–.070]	.983	.977

Note. df = degrees of freedom. Bifactor-(Sc – 1): cognitive group factor, with somatic domain as reference. Bifactor-(Ss – 1): somatic group factor, with cognitive domain as reference. [1]Items 4, 5, and 6 correlated residuals. [2]Items 2 and 3 correlated residuals. [3]Items 3 and 5 correlated residuals. ***$p < .001$

for a unitary factor solution, with a proposed somatic factor (items 4, 5, & 6) as correlated residuals (model 6 in Table 2). This latter solution provided acceptable model fit, although over the RMSEA treshold of ≤ .06. A two-factor solution yielded similar model fit as model 6 (model 7 in Table 2). Modification indices indicated a substantial residual covariance between item 2 and item 3 (Standardized Expected Parameter Change index [Stdyx E.P.C] .492) of the two-factor solution. Allowing these residuals to covary ($\delta = .34$, $p < .001$) resulted in an overall good fit, and this model was accepted (model 8 in Table 2). The model consisted of a cognitive factor of anxiety: GADc (items 1, 2, 3, & 7; with correlated residuals between item 2 & 3) and a somatic factor of anxiety: GADs (items 4, 5, & 6). Both GAD-7 bifactor-(S – 1) resulted in similar goodness of fit as the two-factor solution (model 9 and 10 in Table 2).

WSAS was also tested with CFA, to assess its suitability to evaluate concurrent validity of PHQ-9 and GAD-7. A unitary factor model resulted in unsatisfactory fit statistics (model 11 in Table 2). Modification indices indicated a substantial residual covariance between item 3 & item 5; Stdyx E.P.C .51). Allowing error terms to correlate (Stdyx total $\delta = .37$, $p < .001$) yielded a good fit (model 12 in Table 2). CFA with WSAS correlated with the PHQ-9 and GAD-7 bifactor-(S – 1) demonstrated good fit statistics for the total sample (model 13–16 in Table 2).

Standardized factor loadings for PHQc were between $\lambda = .91$ (item 2) and $\lambda = .70$ (item 9), and for PHQs between $\lambda = .77$ (item 4) and $\lambda = .60$ (item 8). For GADc it varied between $\lambda = .88$ (item 1) and $\lambda = .73$ (item 7), and for GADs it varied between $\lambda = .85$ (item 4) and $\lambda = .54$ (item 6). Composite reliability for PHQc was .87 and .80 for PHQs. For GADc it was .90 and for GADs .73. All factor loadings were above .5 and composite reliability were greater than .7, thus demonstrating acceptable loadings and internal consistensy reliability between indicator variables. The correlation between the factors in PHQ-9 and GAD-7 were all strong (PHQc with PHQs: $\phi = .74$, S.E. = .03, UL = .79; GADc with GADs: $\phi = .80$, S.E. = .03, UL = .85). The cognitive factors demonstrated weaker correlation with each other ($\phi = .67$, S.E. = .03, UL = .72) than the somatic factors with each other ($\phi = .84$, S.E. = .03, UL = .90). The weakest correlations were between the PHQc with GADs ($\phi = .57$, S.E. = .04, UL = .64), and PHQs with GADc ($\phi = .67$, S.E. = .03, UL = .73).

Test for dimensionality resulted in mainly one-dimensional results for the general factors, with some minor issues (see Table 3). Omega hierarchical for PHQ-9 bifactor-(Sc – 1) were below .8, but the PUC and ECV-values justified a one-dimensional interpretation, albeit with some indication of multidimensionality (omega hierarchical = .78, PUC = .83, ECV = .76). Comparable results were found for PHQ-9 bifactor-(Ss – 1) (omega hierarchical = .77, PUC = .72, ECV = .78), and for GAD-7 bifactor-(Sc – 1) (omega hierarchical = .76, PUC = .71, ECV = .75). For GAD-7 bifactor-(Ss – 1) the omega hierarchical was above .8, and thus interpreted as mainly one-dimensional

Table 3 Standardized factor loadings and omega hierarchical for PHQ-9 and GAD-7

Items		GeneralC	SpecificC	GeneralS	SpecificS	General mean
phq1	Little interest or pleasure […]	.600	.509	.798		.699
phq2	Feeling down, depressed, or hopeless	.599	.657	.425		.512
phq3	Trouble falling […] asleep, or sleeping too much	.775		.591	.500	.683
phq4	Feeling tired or having little energy	.661		.913	.457	.787
phq5	Poor appetite or overeating	.631		.441	.375	.536
phq6	Feeling bad about yourself […]	.667	.450	.519		.593
phq7	Trouble concentrating on things […]	.587		.755	.464	.671
phq8	Moving or speaking slowly […] or the opposite[…]	.631		.454	.460	.543
phq9	Thoughts that you would be better off dead […]	.492	.507	.695		.594
	PHQ-9 Omega Hierarchical	.784	.392	.770	.302	.777
gad1	Feeling nervous […]	.712	.452	.851		.782
gad2	Not able to stop worrying	.662	.605	.887		.775
gad3	Worrying too much about different things	.665	.613	.894		.780
gad4	Having trouble relaxing	.855		.674	.360	.765
gad5	Being so restless that it is hard to sit still	.643		.484	.683	.564
gad6	Becoming easily annoyed or irritable	.542		.442	.249	.492
gad7	Feeling afraid […]	.604	.371	.716		.660
	GAD-7 Omega Hierarchical	.761	.338	.850	.294	.806

Note. GeneralC General factor using somatic domain as reference, *SpecificC* Specific cognitive factor, *GeneralS* General factor using cognitive domain as reference, *SpecificS* Specific somatic factor

(omega hierarchical = .85, PUC = .86, ECV = .85). The mean omega hierarchical was .78 for PHQ-9, and .81 for GAD-7.

Measurement invariance

Scalar invariance was achieved across genders, diagnoses, and comorbidity for all bifactor-(Sc – 1) solutions of PHQ-9 and GAD-7 (Table 4). Thus, with cut-off values of ΔCFI $\geq -.01$ and ΔRMSEA <.015, this demonstrated equality of factor loadings, equality of indicator tresholds, and equality of indicator residuals. PHQ-9 for patients with a diagnosis of depression versus patients with an anxiety disorder diagnosis demonstrated issues with achieving configural invariance according to the RMSEA value. However, the CFI-value was above the treshold and interpreted as supporting configural invariance. Latent mean differences (LMD) using bifactor-(Sc – 1) resulted in significantly higher scores on PHQ-9 for women (LMD = .38, SE = .09, $p < .001$), and patients with comorbidity (LMD = .40, SE = .11, $p < .001$), but no significant differences between depression and anxiety diagnoses were found (LMD = .21, SE = .12, $p = .083$). Comparable results were found for GAD-7, with significantly higher scores for women (LMD = .37, SE = .09, $p < .001$), patients with comorbidity (LMD = .37, SE = .11, $p < .001$), with non-significant results for depression vs. anxiety (LMD = −.22, SE = .17, $p = .115$).

Concurrent validity with WSAS

WSAS regressed on bifactor-(S – 1) models of PHQ-9 and GAD-7 each resulted in significant coefficients for the full sample (see Fig. 1). The general factors demonstrated stronger associations with functional impairment than the cognitive and somatic factors, and PHQ-9 demonstrated a stronger association with functional impairment than GAD-7 (WSAS regressed on general factor mean PHQ-9 $\gamma = .74$, $r^2 = .62$; WSAS regressed on general factor mean GAD-7 $\gamma = .54$, $r^2 = .32$). WSAS regressed on the general bifactor-(Sc – 1), resulted in higher associations with PHQ-9 (women $\gamma = .82$, $r^2 = .78$, men $\gamma = .70$, $r^2 = .53$; anxiety $\gamma = .52$, $r^2 = .65$, depression $\gamma = .41$, $r^2 = .49$; no comorbidity $\gamma = .74$, $r^2 = .61$, comorbidity $\gamma = .62$, $r^2 = .53$) than GAD-7 (women $\gamma = .54$, $r^2 = .39$, men $\gamma = .50$, $r^2 = .28$; anxiety $\gamma = .67$, $r^2 = .46$, depression $\gamma = .44$, $r^2 = .24$; no comorbidity $\gamma = .52$, $r^2 = .31$, comorbidity $\gamma = .39$, $r^2 = .21$).

Discussion

The aim of this study was to test the factor structure and measurement invariance of PHQ-9 and GAD-7 in a heterogeneous psychiatric outpatient sample. We also examined the concurrent validity of PHQ-9 and GAD-7 with functional impairment, measured with WSAS, across gender. Firstly, the results supported a two-factor solution for both PHQ-9 and GAD-7, consisting of a cognitive and a somatic factor for each measure. This

Table 4 Measurement invariance using bifactor-(Sc – 1) solution of PHQ-9 and GAD-7

	χ^2 (df)	CFI	RMSEA [90% CI]	Δ χ^2 (df)	*p*	ΔCFI	ΔRMSEA
Gender							
PHQ-9							
Configural	120.690 (46)	.985	.063 [.049–.076]	–	–	–	–
Metric	117.193 (57)	.988	.050 [.037–.063]	8.210 (11)	.694	.003	−.013
Scalar	130.459 (73)	.989	.044 [.031–.055]	18.611 (16)	.289	.001	−.006
GAD-7							
Configural	59.039 (20)	.994	.069 [.049–.089]	–	–	–	–
Metric	70.058 (29)	.994	.058 [.041–.076]	17.209 (9)	.046	.000	−.011
Scalar	68.998 (41)	.996	.041 [.023–.057]	4.492 (12)	.973	.002	−.017
Depression vs. *Anxiety*							
PHQ-9							
Configural	108.401 (46)	.969	.083 [.063–.103]	–	–	–	–
Metric	128.598 (57)	.980	.063 [.048–.077]	26.805 (11)	.005	.011	−.020
Scalar	144.209 (73)	.980	.055 [.042–.069]	20.402 (16)	.203	.000	−.008
GAD-7							
Configural	26.786 (20)	.998	.041 [.000–.079]	–	–	–	–
Metric	38.262 (29)	.997	.040 [.000–.072]	12.163 (9)	.204	−.001	−.001
Scalar	58.513 (41)	.994	.047 [.012–.072]	20.473 (12)	.059	−.003	.007
Comorbid vs. *single diagnsosis*							
PHQ-9							
Configural	105.079 (46)	.984	.063 [.047–.080]	–	–	–	–
Metric	128.598 (57)	.980	.063 [.048–.077]	26.805 (11)	.005	−.004	.000
Scalar	144.209 (73)	.980	.055 [.042–.069]	20.402 (16)	.203	.000	−.008
GAD-7							
Configural	44.238 (20)	.996	.062 [.037–.086]	–	–	–	–
Metric	47.574 (29)	.997	.045 [.019–.067]	8.708 (9)	.465	.001	−.017
Scalar	59.560 (41)	.997	.038 [.012–.057]	13.513 (12)	.333	.000	−.007

Note. ΔCFI ≥ − .01 and ΔRMSEA < .015 indicates established MI. Gender ($n = 831$), depression/anxiety ($n = 396$), comorbidity/no comorbidity ($n = 638$)

finding corresponds with previous research with heterogenous outpatient samples [11, 16]. However, tests for dimensionality of the instruments indicated a general factor, which demonstrated acceptable fit statistics, in accordance with previous studies [12].

Secondly, the bifactor solutions PHQ-9 and GAD-7 achieved scalar invariance across gender, diagnosis, and comorbidity which supports that both instruments measure the same construct for different patient groups, and hence are suitable for comparing differences across these.

Thirdly, all factors were significantly associated with functional impairment, with the general factors accounting for most of the variance compared to the cognitive and somatic factors. However, symptoms of depression demonstrated stronger associations with functional impairment than symptoms of anxiety. Thus, PHQ-9 and GAD-7 demonstrate support for a general factor, albeit with cognitive and somatic subcomponents, when used in heterogenous psychiatric outpatients.

The background of this study was limited research regarding properties of the PHQ-9 and GAD-7 in heterogenous clinical populations. Non-clinical populations may display greater variance in item scores and therefore load on a single factor [8]. In contrast, patients in the present study were assessed prior to psychiatric treatment, and therefore the sample represents a more heterogeneous population. Previous research has advised against multidimensional solutions of these instruments, due to strong factor correlations [11]. Other studies have justified using a sum-score for PHQ-9 and GAD-7 using the extracted factors from an EFA in a bi-factor model [9]. However, such model may create a risk of overfitting the data, and the results could be seriously affected by captured noise [45].

A strength in present study was examining the factor structure a-priori, using the same factor structure specified using a similar population [7]. Additionally, we specified these underlying subdimensions using a modified bifactor, well suitable to our data [13]. However,

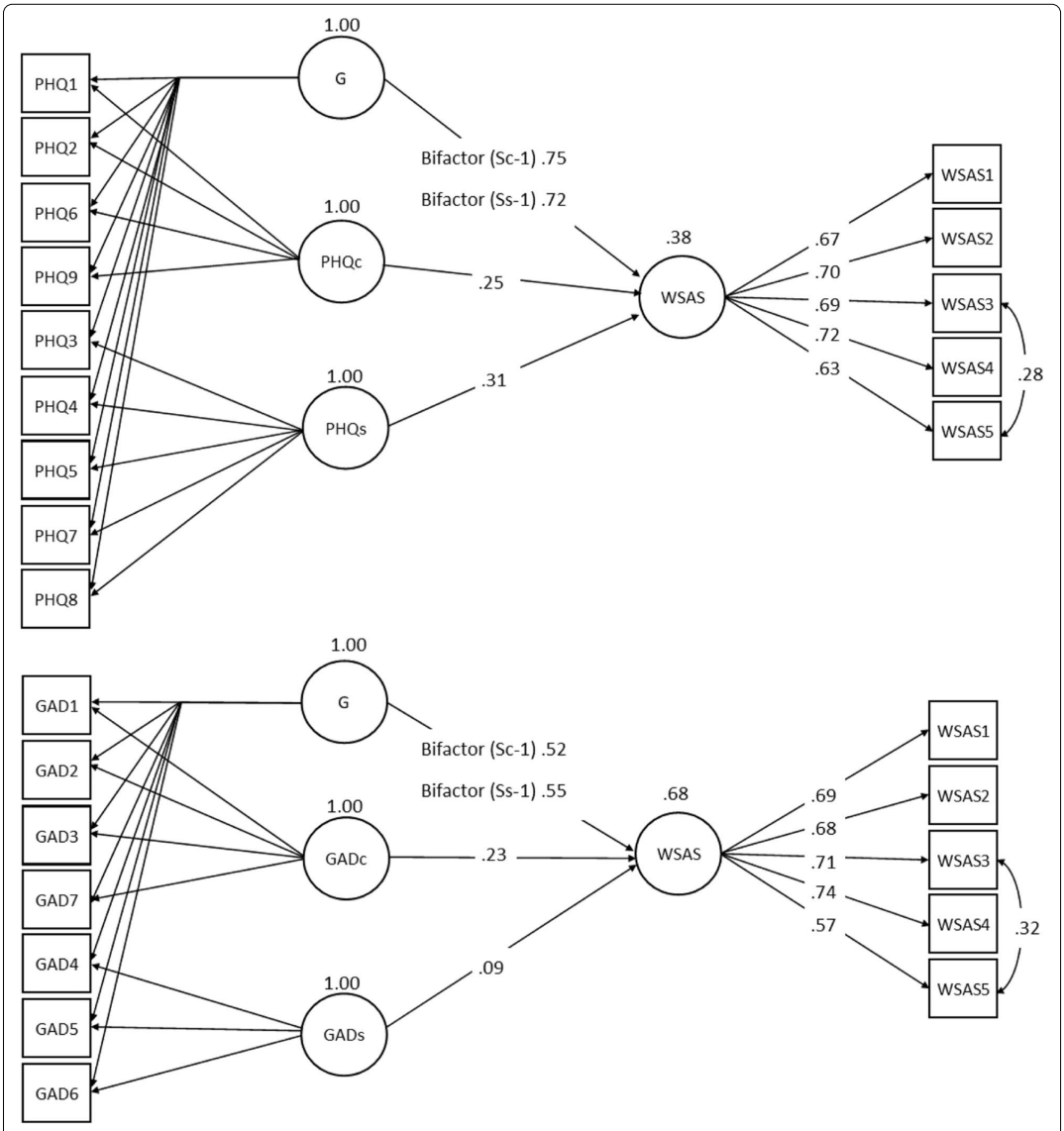


Fig. 1 Standardized results from Latent path models ($n = 831$), where all loadings and paths are significant at $p < .001$, except WSAS regressed on GADs ($p = .038$)

in the present study patients completed assessment before treatment, and we therefore examined a more heterogenous population. Thus, the present study adds to the knowledge of how to properly specify a bifactor model in studies with heterogenous patients initiating treatment.

Some modifications were made to the two-factor solutions, based on both statistical properties and theoretical justifications. We decided to let the residuals (item 2 and 3 covering *Not being able to stop/control worrying, and Worrying too much*) in GAD-7 covary due to their similarities, and let residual covary (item 3 and 5, covering

Impaired social activities, and Impaired close relationships) in WSAS, which corroborates with previous results from Norwegian outpatients [22]. The suggested unitary factor solution with correlated residuals regarding GAD-7 [14, 15] could be criticized for overlooking theoretical reasoning. We argue that the correlations between these (items 4, 5, and 6 covering *Trouble relaxing*, *Being restless* and *Being easily annoyed*) are essential parts of the latent anxiety construct (i.e. a somatic factor), hence, not to be viewed as misfits in the two-factor model. But the moderate problem with discriminate validity between this somatic factor of anxiety and the somatic factor of depression indicate that these constructs are not very well separatable. And the low factor loadings, and a potential crossloading (i.e. GAD-7 item 5 and PHQ-9 item 8 both deal with restlessness), mean that these factors must be handled cautiously. The high correlations can potentially lead to multicollinearity problems if used simultanously, e.g. in multiple regression. If these instruments would be further revised, our recommendation would be to investigate GAD-7 item 4, 5, 6, i.e. the somatic factor of anxiety. Regarding the cognitive factors, the weaker correlations between PHQc and GADc implies that these two factors explains two different constructs, i.e. a cognitive aspect of depression and anxiety each.

To the best of our knowledge, no previous studies have to the same extent examined the association of the factor structure of PHQ-9 and GAD-7 on functional impairment across patient groups in a heterogenous psychiatric outpatient population. The results indicate justification of using these instruments as one-dimensional in clinical settings for measuring symptom severity. However, the results suggest the importance of specifying the underlying factor structure when precise estimates are needed. Further, factorization of these instruments will assess symptom severity measured by a latent general factor. These factors are more robust for comparisons across groups, but the instruments may also be valuable as diagnostic tools, or for single item assessment. For example, we found that PHQ-9 item 9 which assesses suicidal thoughts loaded the general factor below .6, which still has a high clinical value.

Several limitations to this study should be noted. The results are limited by the observational nature of the study. Although few patients declined participation, we were not able to control their reasons nor background data due to research ethical concerns for patients who did not consent to participation. Furthermore, patients were diagnosed in a non-controlled environment, hence, no inter-rater reliability was available, and follow-up assessment is not reported.

Another noteworthy point is that when estimating the bifactor-(S – 1), the general factor was defined by the reference domain. MI and LMD was estimated using somatic domain as reference, thus the scores of the general factor could be interpreted as somatic symptoms corrected for measurement error. Thus, MI and LMD could also be calculated with the cognitive domain as a reference. It is suggested for further studies, to do multiple sampling for overcoming the problems with anomalous results using symmetric bifactors if such solution are preferred. However, a symmetrical bifactor will also create ambiguous interpretations [13, 45].

Additionally, using a longitudinal design could determine the suitability of using the instruments over time. Examining for example individual differences and clinical subgroups over time would improve the clinical utility of these instruments in treatment of mental illness.

Conclusion

The results of this study show that PHQ-9 and GAD-7 may be conceptualized as one-dimensional instruments, with underlying subdimensions of both cognitive and somatic factors. We found support for measurement invariance across gender, diagnostic subgroups and comorbidity, which means that the instruments are interpreted equally among these groups of patients. The higher associations between functional impairment and symptoms of depression highlights the importance with this relation.

Thus, one-dimensionality was supported, and an aggregated score can be justified in clinical settings. However, when precise estimation is needed, such as in psychometric studies with heterogeneous psychiatric populations, our results suggest that the underlying subdimensions should be specified. In conclusion, our study lends further support for the use of PHQ-9 and GAD-7 for assessment of symptoms of depression and anxiety in patients with mental illness.

Abbreviations

PHQ-9: The nine item Patient Health Questionnaire-9; GAD-7: The seven item Generalized Anxiety Disorder scale-7; EFA: Exploratory Factor Analysis; CFA: Confirmatory Factor Analysis; MI: Measurement Invariance; WSAS: The Working Social Adjustment Scale; SEM: Structural Equation Modeling; WLSMV: Weighted Least Squares Means and Variance adjusted; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; UL: Upper Limit 95% confidence interval; Stdyx E.P.C: Standardized Expected Parameter Change index; PUC: Percent of Uncontaminated Correlations; ECV: Explained Common Variance; LMD: Latent mean differences.

Acknowledgements

Not applicable.

Authors' contributions

All authors participated in designing the study. MB prepared, analyzed and interpreted the data. MB prepared the first draft of the manuscript. All authors participated in revising the manuscript and all approved the final manuscript.

Funding
N/A. The study was conducted without external funding.

Availability of data and materials
Data are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Consent to participate was given electronically. This research complies with 7th revision of Declaration of Helsinki, laws and regulation in Norway. It was approved by the Regional committee for medical and health ethics, REK Midt-Norge (REK 2019/31836). Norwegian Centre for Research Data has approved the project (NSD 2020/605327). A data protection impact assessment has been conducted in collaboration with NSD.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 29 June 2021 Accepted: 29 June 2022
Published online: 11 July 2022

References

1. Barkham M. Towards greater bandwidth for standardised outcome measures. Lancet Psychiatry. 2021;8(1):17. https://doi.org/10.1016/S2215-0366(20)30488-0.
2. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
3. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092–7. https://doi.org/10.1001/archinte.166.10.1092.
4. Wolpert M. Funders agree first common metrics for mental health science. 2020. Available from: https://www.linkedin.com/pulse/funders-agree-first-common-metrics-mental-health-science-wolpert/. Accessed 15 June 2021.
5. Obbarius A, van Maasakkers L, Baer L, Clark DM, Crocker AG, de Beurs E, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM depression and anxiety working group. Qual Life Res. 2017;26(12):3211–25. https://doi.org/10.1007/s11136-017-1659-5.
6. Patalay P, Fried EI. Editorial perspective: prescribing measures: unintended negative consequences of mandating standardized mental health measurement. J Child Psychol Psychiatry. 2020;62(8):1032–6. https://doi.org/10.1111/jcpp.13333.
7. Lamela D, Soreira C, Matos P, Morais A. Systematic review of the factor structure and measurement invariance of the patient health questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings. J Affect Disord. 2020;276:220–33. https://doi.org/10.1016/j.jad.2020.06.066.
8. Petersen JJ, Paulitsch MA, Hartig J, Mergenthal K, Gerlach FM, Gensichen J. Factor structure and measurement invariance of the patient health Questionnaire-9 for female and male primary care patients with major depression in Germany. J Affect Disord. 2015;170:138–42. https://doi.org/10.1016/j.jad.2014.08.053.
9. Stochl J, Fried EI, Fritz J, Croudace TJ, Russo DA, Knight C, et al. On dimensionality, measurement invariance, and suitability of sum scores for the PHQ-9 and the GAD-7. Assessment. 2020;29(3):355–66. https://doi.org/10.1177/1073191120976863.
10. Beard C, Hsu KJ, Rifkin LS, Busch AB, Björgvinsson T. Validation of the PHQ-9 in a psychiatric sample. J Affect Disord. 2016;193:267–73. https://doi.org/10.1016/j.jad.2015.12.075.
11. Boothroyd L, Dagnan D, Muncer S. PHQ-9: one factor or two? Psychiatry Res. 2019;271:532–4. https://doi.org/10.1016/j.psychres.2018.12.048.
12. De Man J, Absetz P, Sathish T, Desloge A, Haregu T, Oldenburg B, et al. Are the PHQ-9 and GAD-7 suitable for use in India? a psychometric analysis. Front Psychol. 2021;12:676398. https://doi.org/10.3389/fpsyg.2021.676398.
13. Eid M, Geiser C, Koch T, Heene M. Anomalous results in G-factor models: explanations and alternatives. Psychol Methods. 2017;22(3):541–62. https://doi.org/10.1037/met0000083.
14. Johnson SU, Ulvenes PG, Øktedalen T, Hoffart A. Psychometric properties of the general anxiety disorder 7-item (GAD-7) scale in a heterogeneous psychiatric sample. Front Psychol. 2019;10:1713. https://doi.org/10.3389/fpsyg.2019.01713.
15. Rutter LA, Brown TA. Psychometric properties of the generalized anxiety disorder Scale-7 (GAD-7) in outpatients with anxiety and mood disorders. J Psychopathol Behav Assess. 2017;39(1):140–6. https://doi.org/10.1007/s10862-016-9571-9.
16. Beard C, Björgvinsson T. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. J Anxiety Disord. 2014;28(6):547–52. https://doi.org/10.1016/j.janxdis.2014.06.002.
17. Brown TA. Confirmatory factor analysis for applied research. 2nd ed. New York: The Guilford Press; 2015.
18. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev Rev. 2016;41:71–90. https://doi.org/10.1016/j.dr.2016.06.004.
19. McKnight PE, Kashdan TB. The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. Clin Psychol Rev. 2009;29(3):243–59. https://doi.org/10.1016/j.cpr.2009.01.005.
20. McKnight PE, Monfort SS, Kashdan TB, Blalock DV, Calton JM. Anxiety symptoms and functional impairment: A systematic review of the correlation between the two measures. Clin Psychol Rev. 2016;45:115–30. https://doi.org/10.1016/j.cpr.2015.10.005.
21. Mundt JC, Marks IM, Shear MK, Greist JH. The work and social adjustment scale: a simple measure of impairment in functioning. Br J Psychiatry. 2002;180:461–4. https://doi.org/10.1192/bjp.180.5.461.
22. Pedersen G, Kvarstein EH, Wilberg T. The work and social adjustment scale: psychometric properties and validity among males and females, and outpatients with and without personality disorders. Personal Ment Health. 2017;11(4):215–28. https://doi.org/10.1002/pmh.1382.
23. Boothroyd LDD, Muncer S. Psychometric analysis of the generalized anxiety disorder scale and the patient health questionnaire using Mokken scaling and confirmatory factor analysis. Health. Prim Care. 2018;2(4):1–4. https://doi.org/10.15761/HPC.1000145.
24. He C, Levis B, Riehm KE, Saadat N, Levis AW, Azar M, et al. The accuracy of the patient health Questionnaire-9 algorithm for screening to detect major depression: an individual participant data Meta-analysis. Psychother Psychosom. 2020;89(1):25–37. https://doi.org/10.1159/000502294.
25. Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the patient health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. Gen Hosp Psychiatry. 2015;37(1):67–75. https://doi.org/10.1016/j.genhosppsych.2014.09.009.
26. Mitchell AJ, Yadegarfar M, Gill J, Stubbs B. Case finding and screening clinical utility of the patient health questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. BJPsych Open. 2016;2(2):127–38. https://doi.org/10.1192/bjpo.bp.115.001685.
27. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the patient health questionnaire (PHQ-9): a meta-analysis. Gen Hosp Psychiatry. 2015;37(6):567–76. https://doi.org/10.1016/j.genhosppsych.2015.06.012.
28. Burdzovic Andreas J, Brunborg GS. Depressive symptomatology among Norwegian adolescent boys and girls: the patient health Questionnaire-9 (PHQ-9) psychometric properties and correlates. Front Psychol. 2017;8:887. https://doi.org/10.3389/fpsyg.2017.00887.
29. Wisting L, Johnson SU, Bulik CM, Andreassen OA, Rø Ø, Bang L. Psychometric properties of the Norwegian version of the patient health Questionnaire-9 (PHQ-9) in a large female sample of adults with and without eating disorders. BMC Psychiatry. 2021;21(1):6. https://doi.org/10.1186/s12888-020-03013-0.

30. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med. 2007;146(5):317–25. https://doi.org/10.7326/0003-4819-146-5-200703060-00004.
31. Pfizer. Patient Health Questionnaire (PHQ) Screeners. 2022. Available from: https://www.phqscreeners.com. Accessed 31 May 2022.
32. Zahra D, Qureshi A, Henley W, Taylor R, Quinn C, Pooler J, et al. The work and social adjustment scale: reliability, sensitivity and value. Int J Psychiatry Clin Pract. 2014;18(2):131–8. https://doi.org/10.3109/13651501.2014.894072.
33. StataCorp. Stata statistical software: release 16. StataCorp LLC 2019.
34. Muthén LK, Muthén BO. Mplus User's Guide (8th ed.). 1998-2017. Available from: https://www.statmodel.com/html_ug.shtml. Accessed 15 June 2021.
35. Asparouhov T, Muthén B. Weighted least squares estimation with missing data. 2010. Available from: https://www.statmodel.com/download/GstrucMissingRevision.pdf. Accessed 15 June 2021.
36. Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behav Res Methods. 2016;48(3):936–49. https://doi.org/10.3758/s13428-015-0619-7.
37. Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. Behav Res Methods. 2019;51(1):409–28. https://doi.org/10.3758/s13428-018-1055-2.
38. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118.
39. Rönkkö M, Cho E. An updated guideline for assessing discriminant validity. Organ Res Methods. 2020;25(1):6–14. https://doi.org/10.1177/1094428120968614.
40. Dueber, D. M. Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. 2017. https://doi.org/10.13023/edp.tool.01.
41. Reise SP, Bonifay WE, Haviland MG. Scoring and modeling psychological measures in the presence of multidimensionality. J Pers Assess. 2013;95(2):129–40. https://doi.org/10.1080/00223891.2012.725437.
42. Millsap RE, Yun-Tein J. Assessing factorial invariance in ordered-categorical measures. Multivar Behav Res. 2004;39(3):479–515. https://doi.org/10.1207/S15327906MBR3903_4.
43. Pendergast LL, von der Embse N, Kilgus SP, Eklund KR. Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. J Sch Psychol. 2017;60:65–82. https://doi.org/10.1016/j.jsp.2016.11.002.
44. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Eq Model. 2007;14:464–504. https://doi.org/10.1080/10705510701301834.
45. Bonifay W, Lane SP, Reise SP. Three concerns with applying a bifactor model as a structure of psychopathology. Clin Psychol Sci. 2017;5(1):184–6. https://doi.org/10.1177/2167702616657069.

Publisher's Note

Brattmyr *et al. BMC Psychiatry* (2023) 23:804
https://doi.org/10.1186/s12888-023-05314-6

BMC Psychiatry

RESEARCH Open Access

Symptoms and prevalence of common mental disorders in a heterogenous outpatient sample: an investigation of clinical characteristics and latent subgroups

Martin Brattmyr[1*], Martin Schevik Lindberg[1,2], Jakob Lundqvist[1], Stian Solem[1], Odin Hjemdal[1], Frederick Anyan[1] and Audun Havnen[1,3]

Abstract

Background Patient-reported outcome measures (PROM) provide clinicians with information about patients' perceptions of distress. When linked with treatment and diagnostic registers, new information on common mental health disorders (CMHD) and service use, may be obtained, which might be useful clinically and for policy decision-making. This study reports the prevalence of CMHD and their association with PROM severity. Further, subgroups of self-reported symptoms of depression and anxiety were examined, and their association with clinician-assessed mental disorders, functional impairment, and service use.

Methods In a cohort study of 2473 (63% female) outpatients, CMHD was examined with pre-treatment scores of self-reported depression and anxiety, and the number of assessments and psychotherapy appointments one year after treatment start. Factor mixture modelling (FMM) of anxiety and depression was used to examine latent subgroups.

Results Overall, 22% of patients with a CMHD had an additional comorbid mood/anxiety disorder, making the prevalence lower than expected. This comorbid group reported higher symptoms of anxiety and depression compared to patients with non-comorbid disorders. FMM revealed three classes: "anxiety and somatic depression" (33%), "mixed depression and anxiety" (40%), and "cognitive depression" (27%). The anxiety and somatic depression class was associated with older age, being single and on sick leave, higher probability of depressive-, anxiety-, and comorbid disorders, having more appointments and higher functional impairment. Although the cognitive depression class had less somatic distress than the mixed depression and anxiety class, they reported more functional impairment and had higher service use.

Conclusion The results show that higher levels of somatic symptoms of depression could both indicate higher and lower levels of functional impairment and service use. A group of patients with high somatic depression and anxiety was identified, with severe impairment and high service needs. By gaining insights into CMHD factors' relation with clinical covariates, self-reported risk factors of depression and anxiety could be identified for groups with different

*Correspondence:
Martin Brattmyr
martin.brattmyr@ntnu.no

Full list of author information is available at the end of the article

levels of aggravating life circumstances, with corresponding service needs. These could be important symptom targets in different groups of patients.

Keywords Patient-reported outcome measures, Latent class analysis, Factor mixture models, Common mental disorders, Comorbidity

Introduction

Patient-reported outcome measures (PROM) have been increasingly applied to encourage patient involvement [1]. PROM can be helpful in the diagnostic process of common mental health disorders (CMHD) [2] and is often implemented to facilitate service planning [3]. Since accessible quality indicators about patients' needs often are rudimentary and lack diagnostic and symptomatic information [4], PROM data could help providers to facilitate person-centred services [1, 2].

Although linking PROM with register data is underutilized, important knowledge about patients may be obtained by connecting these sources of information [3]. One potential outcome of combining PROM- and register data is to increase the knowledge about specific groups of patients, such as those with comorbid disorders. Compared to patients with non-comorbid CMHD, patients with comorbidity have been associated with higher service utilization, higher symptom severity, and higher levels of functional impairment [5]. One PROM study that examined comorbidity in outpatient treatment showed that 45% of patients with an anxiety disorder had comorbid depression, 64% with post-traumatic stress disorder (PTSD) had a depressive disorder, and that comorbid depression was associated with higher symptom severity both before and after treatment [6]. Studies like these demonstrate the additive value of PROM data linked with diagnostic registers of CMHD, to provide a better understanding of specific diagnostic groups of patients.

PROM-data has the additional capability of providing organizations with information about patients useful for decision-making purposes and allocation of resources, however, aggregated data at this level is seldom published [7]. Thus, a potential of PROM data is the possibility offered to analyse symptom severity across the whole sample, capturing symptom heterogeneity beyond diagnoses. However, the great heterogeneity in mental health populations is complicated to address, and various modelling procedures exist—all with strengths and limitations. In a review by Feczko et al. [8], they commented upon clinical subtypes, dimensional, and computational models. Although clinical subtypes in diagnostic manuals could account for some heterogeneity, their descriptive approach to establishing nosology has been criticized. Dimensional models have been proposed as more suitable in psychiatric research; however, subtypes are often inevitable when multiple dimensional constructs are measured simultaneously. Therefore, with technical developments in recent decades, computational models have been increasingly applied [8].

Mixture model analysis has become a frequently utilized computational statistical framework to aggregate symptom heterogeneity into a smaller set of homogeneous groups [9]. One model specification, latent class analysis (LCA), has been used to analyse symptoms of depression and anxiety simultaneously, and class indicators have identified important sub-samples in non-clinical adult populations [10–17] and clinical adult populations [18, 19]. In the latter two studies, three class solutions were identified, labelled by a gradient of symptom severity, where the higher severity classes were more strongly associated with CMHD, and with a decreased association with mild and low severity classes.

LCA, however, has been criticized when applied to detect homogeneous subgroups of CMHD [20] because of the strict methodological requirements of local independence: class indicators should only depend on the latent classes and not correlate. This is an unrealistic requirement for symptoms of depression and anxiety due to their frequent co-occurrence. The advantage of factor mixture modelling (FMM) over LCA is that FMM does not assume conditional independence of latent classes [21] and therefore has been suggested to be more suitable in psychiatric research [9]. Applying FMM with class indicators of depression and anxiety has been conducted within general adult populations [22], and for people with a lifetime episode of depression [23], but rarely within a clinical mental health setting.

The scope of the current study was two-folded: first, to examine the one-year prevalence of CMHD in a diagnostic register and their association with patient-reported symptom severity in a large cross-sectional heterogeneous psychiatric outpatient sample. In accordance with previous research, we hypothesized comorbid mood and anxiety disorders to be frequently occurring and associated with higher symptom severity.

Since examining the sample based on clinician-assessed diagnoses could obscure other clinically relevant associations of symptom-homogenous clusters of patients, we continued to assess latent subgroups using FMM. Thus, the second aim was to identify homogeneous subgroups based on self-reported symptoms of depression and anxiety and analyse their association with clinician-assessed mental disorders, functional impairment, and service

use. Due to the exploratory nature of FMM, the numbers of identified classes were not hypothesized a priori.

Method

Sample

Norwegian patients requiring non-urgent specialized mental health treatment are typically referred by their general practitioners to local psychiatric outpatient clinics for treatment. At one such clinic, electronic PROM data using the nine-item Patient Health Questionnaire (PHQ-9) [24], the seven-item Generalized Anxiety Disorder scale (GAD-7) [25] and the Work and Social Adjustment Scale (WSAS) [26] was collected for patients who started treatment between February 2020 and February 2022. All patients with a first assessment were invited to participate. Groups of patients with a primary diagnosis of obsessive-compulsive disorder (OCD), schizophrenia, substance abuse, and elderly patients, were treated at other specialized units. Out of 2519 patients who consented to participate, patients with missing information on all self-report questionnaires were excluded (n=46). The final sample consisted of 2473 outpatients (79% of the invited participants; see Table 1). Out of the participating patients, 97% also had available register data with diagnostic and treatment information.

Table 1 Sample characteristics of outpatients (n = 2473), and comparisons between comorbid and non-comorbid disorders

	n (%)	PHQ–9 sum [95% CI]	GAD–7 sum [95% CI]
Female	1563 (63)	15.69 [15.41–15.98]	12.33 [12.10–12.57]
Male	910 (37)	14.80 [14.41–15.19]	11.14 [10.81–11.47]
Single	1277 (57)	15.88 [15.57–16.19]	11.82 [11.55–12.09]
Not single	977 (43)	14.80 [14.47–15.13]	11.97 [11.69–12.26]
Sick leave	719 (32)	16.41 [15.99–16.84]	12.82 [12.46–13.19]
Not sick leave	1520 (68)	14.97 [14.71–15.24]	11.55 [11.32–11.77]
Diagnostic data			
Bipolar	125 (5)	16.50 [15.51–17.50]	12.21 [11.38–13.03]
Depression	865 (36)	17.57 [17.24–17.90]	12.77 [12.46–13.08]
PTSD	245 (10)	16.04 [15.34–16.74]	12.30 [12.37–13.55]
Anxiety	613 (25)	15.67 [15.22–16.11]	13.31 [12.96–13.67]
Somatization	92 (4)	15.10 [13.96–16.24]	12.41 [11.50–13.31]
Mixed anxiety depressive disorder	55 (2)	14.09 [12.66–15.52]	11.47 [10.36–12.59]
Comorbid	333 (14)	18.14 [17.63–18.66]	13.85 [13.37–14.32]
No CMHD or mixed anxiety depressive disorder	832 (35)	13.78 [13.39–14.18]	10.62 [10.28–11.00]
Comparisons between patients diagnosed with comorbid (n = 333) vs. non-comorbid disorders			
Bipolar	87 (70)	t = 2.16, p = 0.033	t = 2.17, p = 0.032
Depression	557 (64)	t = 2.31, p = 0.021	t = 5.53, p < 0.001*
PTSD	181 (74)	t = 4.45, p < 0.001*	t = 2.45, p = 0.026
Anxiety	359 (59)	t = 9.49, p < 0.001*	t = 3.30, p = 0.001*
Somatization	59 (64)	t = 3.60, p = 0.001*	t = 0.68, p = 0.499

Note: Frequency in valid percent. Comorbid disorder was defined as having a bipolar or depressive disorder and PTSD or anxiety disorder or somatization disorder. Diagnostic data were available for n = 2411 patients

* Statistically significant with Bonferroni adjusted p-value at 0.01

Measures

Demographic information was extracted for age, gender, self-reported relationship status, and work status before treatment started. Diagnostic information in accordance with the ICD-10 and the number of appointments was collected until one year after the self-reported symptom assessment (extraction date in March 2023). Diagnoses were manually clustered into bipolar, depressive, anxiety, PTSD, somatization, and comorbid mood/anxiety disorder. The comorbid mood/anxiety disorders were patients with bipolar/depressive disorder and PTSD/anxiety/somatization disorder. This is equivalent to the ICD-10 categorization of chapter F30 and F40-disorders. A diagnostic category was made for patients who were only diagnosed with mixed anxiety and depressive disorder (F412). If they had an additional CMHD, they were categorized into their respective CMHD.

PHQ-9 and GAD-7 were used to measure symptoms of depression and anxiety, respectively. Both instruments use a 4-point Likert scale ranging between 0 (not at all) and 3 (almost every day). Their psychometric properties have been widely tested, both internationally [27, 28] and in Norwegian outpatient populations [29, 30]. Although both instruments were created to screen for their respective disorders, they are commonly used to measure symptom severity and have been recommended for health outcomes measurement [31].

WSAS was used to examine the degree of functional impairment [26]. The scale has five items that are scored on a 9-point Likert scale from 0 (not at all) to 8 (very severely), with a maximum score of 40. Its psychometric properties are well established, including among Norwegian outpatients [32].

Statistical analysis

Skewness ≥ 2, and kurtosis ≥ 4 were used as thresholds for examining normal distributions. Sum-scores of PHQ-9 and GAD-7 showed acceptable skewness and kurtosis, thus means (M) with standard deviations (SD) were used together with one-way analysis of variance (ANOVA)

with Bonferroni post hoc test for comparisons between non-comorbid CMHD. Student t-tests were conducted for comparisons between non-comorbid and comorbid CMHD. For gender distribution between non-comorbid and comorbid disorders, Pearson's x^2 was used. Due to multiple comparisons, Bonferroni-adjusted p-values were applied.

Mplus version 8.8 was used as statistical software for mixture model analysis [33]. To make use of all available data, the full-information maximum likelihood was used with robust estimation (MLR). Missing items for the class indicators PHQ-9 and GAD-7 were below 0.1% and Little's MCAR test indicated missing completely at random ($p=0.361$). Owing to the overlap in assessing restlessness by one item each from the PHQ-9 and GAD-7, item 8 in the PHQ-9 was removed prior to the analyses. We retained a three-latent-dimensional structure, comprising a cognitive depressive-, a somatic depressive-, and an anxiety latent dimension, informed and validated in a previous study [29]. The analyses were performed in three stages following the recommendations by Clark et al. [21]. In the first stage, the latent dimensions of symptoms of depression and anxiety were verified through confirmatory factor analysis (CFA). Model fit was evaluated with the following indices: Standardized Root Mean Square Residual (SRMR) [34] and Root Mean Square Error of Approximation (RMSEA) [33] values less than 0.08 and values equal to or less than 0.06 (upper 90% CI close to or <0.08) respectively, a Comparative Fit Index (CFI) and a non-normed fit index, Tucker-Lewis index (TLI) greater than 0.95 [35].

In the second stage, LCA was used to identify progressively higher numbers of latent classes to determine patient clusters of class membership, based on self-reported symptoms of depression and anxiety. The purpose of the LCA was to investigate the degree of heterogeneity in the sample and determine the highest number of classes for the FMM. The LCA was examined for one to five classes to determine the optimal model, using maximum likelihood estimation. To avoid local maxima solutions, 1000 random starting sets with 250 final stage optimizations were specified. For model fit indices, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used together with the sample size adjusted BIC (aBIC), where lower values equal better fit. Entropy levels closer to 1 indicate greater classification accuracy. Significant levels of the Lo-Mendell-Rubin likelihood ratio test (LMR), adjusted LMR (aLMR), and Bootstrap likelihood ratio test (BLRT) were used as indicators for satisfying model fit. The optimal number of classes was determined according to the model fit indices and theoretical interpretability.

Finally, in the third stage, FMM was used to explore diagnostic class membership and the range of severity within and across diagnostic classes based on self-reported symptom data. Due to the exploratory nature of FMM, all model variations reported in the existing literature were estimated [21]. However, the third and fourth model variations recommended by Clark et al. [21] mostly produced inadmissible solutions as a result of estimating too many parameters compared to the first and second model variations, therefore only the first and second model variations are reported. In the first model variation, FMM-1, only the factor means were allowed to vary across classes while the factor loadings and item intercepts were constrained invariant across classes, indicating that symptoms of depression and anxiety are measured equally across classes. The factor covariance was fixed to zero to indicate no within-class heterogeneity for the symptoms. This model variation, FMM-1, is also called nonparametric factor analysis model [36]. Next, the factor variance in FMM-1 was freely estimated in each class for the estimation of FMM-2 and allowed for within-class heterogeneity in the levels of the symptoms. The factor means of the first class were fixed to zero for identification purposes, but freely estimated in the other classes. The FMM-2 is also called mixture factor analysis [36]. Once the best fitting model solution was identified, various covariates were included to (i) explain between-class heterogeneity by regressing class membership on age, gender, and relationship status, using the Mplus R3STEP option, and (ii) to determine how class membership predicts relevant outcomes using the DECAT option (i.e., work status, diagnosed with bipolar, depression, anxiety, PTSD, somatization, comorbid depression and anxiety, and mixed anxiety depression disorder), and the BCH option (i.e., number of assessment and psychotherapy appointments, and functional impairment). The effect sizes of differences between classes were calculated using Cramér's V.

Results

Overall, 63% ($n=1524$) of the patients were diagnosed with a CMHD (bipolar, depressive, PTSD, anxiety, or somatization disorder). Further, 2% ($n=55$) were registered with a mixed anxiety depressive disorder without a CMHD. Other frequent prevalent diagnoses for patients with no CMHD were attention deficit hyperactivity disorder (10%, $n=245$) and personality disorders (4%, $n=106$), and symptoms and signs involving emotional state (ICD-10 code R45, 21%, $n=496$). The 1-year prevalence of comorbid CMHD differed somewhat between diagnostic groups. For patients diagnosed with bipolar disorder ($n=125$), 25% also had a depressive disorder, 8% had PTSD, 17% had an anxiety disorder, and 6% had somatization disorder. For patients with a depressive disorder ($n=865$), 4% were also diagnosed with bipolar disorder, 6% with PTSD, 28% with anxiety disorder, and 3%

with somatization disorder. For PTSD ($n=245$), 4% were diagnosed with bipolar disorder, 23% with depressive disorder, 9% with anxiety disorder, and 2% a somatization disorder. For anxiety disorders ($n=613$), 3% had bipolar disorder, 39% had depressive disorder, 3% had PTSD, and 5% had somatization disorder. For patients with somatization disorder ($n=92$), 8% had bipolar disorder, 30% had a depressive disorder, 7% had PTSD, and 32% had an anxiety disorder.

Overall, 22% of patients diagnosed with a CMHD had a comorbid mood and anxiety disorder. Correspondingly, 14% of all patients with diagnostic data had a comorbid mood and anxiety disorder. Patients with a comorbid disorder reported more anxiety than non-comorbid depression and anxiety disorder, and more depression than non-comorbid PTSD, anxiety, and somatization disorder (Table 1).

There were statistically significant differences in symptoms of depression between non-comorbid CMHD [$F(4, 1019)=20.17$, $p<0.001$]. Bonferroni post hoc test showed that non-comorbid depression had a higher mean (17.33, SD=4.97) than PTSD (M=15.19, SD=5.90), anxiety (M=14.05, SD=5.64), and somatization disorder (M=14.39, SD=5.57). However, there were no significant differences in anxiety symptoms between non-comorbid CMHD [$F(4, 1033)=1.31$, $p=0.265$]. Gender distributions between comorbid and non-comorbid CMHD were not statistically significantly different when adjusted for multiple comparisons, with 33% being male.

Factor structure results

One factor CFA of depression and anxiety showed unsatisfactory model fit ($x^2=3205.815$, df=90, $p<0.001$; SRMR=0.075; RMSEA=0.118 [90% CI=0.115, 0.122]; CFI=0.762; TLI=0.722), as was a two-factor model, consisting of a depressive and anxiety factor (PHQ-9 and GAD-7 respectively) ($x^2=1637.922$, df=89, $p<0.001$; SRMR=0.057; RMSEA=0.084 [90% CI=0.080, 0.087]; CFI=0.882; TLI=0.860). Separate analyses of the factor structures of GAD-7 and PHQ-9 revealed a two-factor – cognitive and somatic – structure of the PHQ-9. Thus, a 3-factor structure comprising (i) cognitive and (ii) somatic depression, and (iii) anxiety reached acceptable model fit ($x^2=872.436$, df=85, $p<0.001$; SRMR=0.046; RMSEA=0.061 [90% CI=0.058, 0.065]; CFI=0.940; TLI=0.926) with two error covariances.

Latent class analysis results

Model fit indices for the LCA are presented in Table 2. The 1-Class model had the largest AIC, BIC, and aBIC, thus demonstrating the worst model fit. The LMR test, aLMR test, and BLRT in the 2-Class model solution all had p-values<0.01, indicating to reject the 1-Class model solution in favour of a 2-Class model solution. Statistically significant p-values for the LMR and BLRT indicated that the current (k-class) model fitted the data better than the model with one less class (k-1 class). Results from comparing the 3-Class to the 2-Class model solution favoured a 3-Class model solution, which had lower criterion indices than the 2-Class model solution. Similarly, the 4-Class model and 5-Class model all had smaller criterion indices. Although the criterion fit indices showed that there was an improvement in model fit when comparing the 3-Class model solution to the 4-, and 5-Class model solutions, the deterioration in the entropy fit statistic for the 4-Class model solution was more pronounced, followed by the 5-Class model solution, which indicates that the 4- and 5-Class model solutions contain

Table 2 Latent class, latent class factor and factor mixture model of depression symptoms (PHQ-9) and anxiety symptoms (GAD-7)

Model	LL	*k*	Entropy	AIC	BIC	aBIC	LMR	aLMR	BLRT
LCA									
One-class	-51091.173	30		102242.347	102416.743	102321.425			
Two-class	-46643.901	46	0.890	93379.801	93647.208	93501.055	< 0.001	< 0.001	< 0.001
Three-class	**-45609.141**	**62**	**0.867**	**91342.283**	**91702.701**	**91505.711**	**< 0.001**	**< 0.001**	**< 0.001**
Four-class	-44927.061	78	0.852	90010.123	90463.552	90215.726	< 0.001	< 0.001	< 0.001
Five-class	-44466.280	94	0.856	89120.559	89666.999	89368.338	< 0.001	< 0.001	< 0.001
FMM-1									
Three-factor, two-class	-45977.979	63	0.879	92081.959	92448.190	92248.023	< 0.001	< 0.001	< 0.001
Three-factor, three-class	-45124.728	82	0.848	90413.456	90890.137	90629.603	0.116	0.116	< 0.001
Three-factor, four-class	-44513.440	101	0.844	89228.881	89816.013	89495.111	0.239	0.239	< 0.001
Three-factor, five-class	-44124.081	120	0.844	88488.162	89185.745	88804.475	0.239	0.239	< 0.001
FMM-2									
Three-factor, two-class	-43660.406	69	0.911	87458.812	87859.921	87640.692	0.240	0.240	< 0.001
Three-factor, three-class	**-43260.611**	**88**	**0.993**	**86697.222**	**87208.782**	**86929.185**	**0.194**	**0.194**	**< 0.001**
Three-factor, four-class	-42800.962	107	0.998	85815.924	86437.935	86097.970	0.165	0.165	< 0.001
Three-factor, five-class	-43114.548	126	0.787	86481.095	87213.557	86813.224	0.221	0.221	1.000

Note: AIC=Akaike information criterion. BIC=Bayesian Information Criterion. aBIC=Sample size adjusted BIC. LMR=Lo-Mendell-Rubin likelihood ratio test. BLTR=Bootstrap likelihood ratio test. In bold is the selected model

classes that are not clearly separated. Higher entropy values indicate that classes are easily distinguishable and distinctive, and as such favoured the 3-Class model solution which had a relatively high entropy value. The three classes were labelled as; high distress class (43%), moderate distress class (41%), and low distress class (16%) since they only differed by the degree of symptom severity.

Factor mixture model results

Results from the factor mixture model analysis are presented in the bottom part of Table 2. The FMM with the lowest criterion indices was the 3-factor, 4-Class FMM-2 model. However, one of the classes in this model solution turned out to be spuriously extracted from the data as it contained no respondents, so we examined the next lowest criterion indices —the 3-factor, 5-Class FMM-2 model. The LMR test, aLMR test, and BLRT in the 3-factor, 5-Class FMM-2 model solution all had *p*-values greater than 0.05, which indicates that this model solution should be rejected despite lower criterion indices. Therefore, we then considered a 3-factor, 3-Class FMM-2 model which had the second lowest criterion indices. Although the LMR and the aLMR tests did not show unequivocal support for this model solution, the BLRT had a *p*-value < 0.001, indicating that this model solution significantly fitted the data. The entropy value for this model solution was also high, indicating that there is a clear separation between distinguishable classes. Furthermore, the 3-factor, 3-Class FMM-2 model replicates the combined results from the CFA and LCA analyses.

Selecting the 3-factor, 3-Class FMM-2 model implies that the underlying symptoms are conceptualized equivalently and normally distributed within classes. In other words, the three classes are represented by normally distributed patterns of symptoms of depression and anxiety such that individuals within classes can have quantitatively different ranges of symptom severity. The criterion indices for the FMM-1 solution were much higher and therefore unsuitable for model selection. Additionally, the assumptions of the FMM-1 imply that all patients within a class are having the same levels of distress and that there is no within-class heterogeneity in symptoms of depression and anxiety. This is unlikely to be correct as symptom variation exists as well as the range of severity, consistent with our FMM-2 model solution. Since we assumed variations in symptoms of depression and anxiety as well as the range of severity, the 3-factor, 3-Class FMM-2 model was chosen.

Interpretation of classes from the factor mixture model

The three distress classes were labelled according to differences in factor means (see Fig. 1). The reference Class two comprising 40% of the sample, was labelled "*mixed depression and anxiety*" whereas patients in Class one (33%), in comparison to the mixed depression and anxiety class reported lower levels of cognitive depression, but higher levels of somatic depression and anxiety symptoms and was thus labelled *"anxiety and somatic depression"*. Patients in Class three (27%) on the other hand reported lower somatic depression than mixed depression and anxiety class and Class three was thus labelled *"cognitive depression"*. The factor variances within the classes were all significant, which agreed with the interpretation that there are variations in diagnostic class membership and the range of severity of the patient's self-reported symptoms of depression and anxiety.

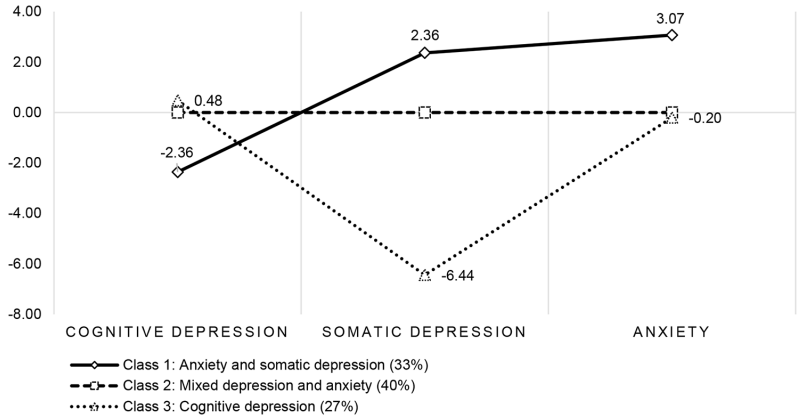


Fig. 1 Three-factor three-classes factor mixture model latent variable factor means

Note: Y-axis represents the factor mean in relation to the reference Mixed depression and anxiety class. The following dimension comprised class indicators; Cognitive depression: Loss of interest, sadness, worthlessness, and suicidal ideation. Somatic depression: Sleep problems, tiredness, appetite, and concentration. Anxiety: Nervous, not able to stop worrying, worrying too much, trouble relaxing, unable to sit still, annoyed and afraid

Table 3 Multinomial logistic regression parameters predicting class membership

Predictors	*OR*	*95% CI*	*OR*	*95% CI*
	Anxiety and somatic depression class		**Cognitive depression class**	
Reference Class: Mixed depression and anxiety class				
Females	0.92	[0.73–1.10]	1.05	[0.82–1.28]
Age	1.02**	[1.01–1.02]	1.00	[0.99–1.01]
Single	1.56**	[1.24–1.88]	1.27	[1.00–1.54]
	Anxiety and somatic depression class		**Mixed depression and anxiety class**	
Reference Class: Cognitive depression class				
Females	0.88	[0.68–1.07]	0.96	[0.75–1.17]
Age	1.01**	[1.00–1.02]	1.00	[0.99–1.01]
Single	1.23	[0.95–1.51]	0.79*	[0.62–0.96]

*$p<0.05$, **$p<0.01$

Predictors of class membership

See Table 3 for predictors of class membership. There were no findings of gender predicting any class membership. Older age predicted a higher probability of membership to the anxiety and somatic depression class, compared to both the mixed depression and anxiety and the cognitive depression classes. Being single predicted a higher probability of membership in the anxiety and somatic depression class compared to the mixed depression and anxiety class, and the cognitive depression class compared to the mixed depression and anxiety class.

Differences in relevant outcome variables

Outcomes across classes are presented in Table 4. Patients with a high probability of membership in the anxiety and somatic depression class were to a larger degree associated with being on sick leave ($x^2=48.378$, 6.569; V=0.147, 0.054) and being diagnosed with depression ($x^2=146.772$, 28.999; V=0.256, 0.113), anxiety ($x^2=12.417$, 1.651; V=0.074, 0.027) and comorbid mood/anxiety disorder ($x^2=53.645$, 7.788; V=0.155, 0.059) compared to the mixed depression and anxiety class, and cognitive depression class respectively. Patients with a high probability of belonging to the cognitive depression class were to a larger degree associated with being on sick leave ($x^2=14.800$; V=0.081) and being diagnosed with depression ($x^2=29.255$; V=0.114), and comorbid mood/anxiety disorder ($x^2=15.442$; V=0.083), compared to the mixed depression and anxiety class. They also had a higher probability of being diagnosed with mixed anxiety and depressive disorder, compared to the anxiety and somatic depression class ($x^2=10.361$, V=0.068). Regarding the mixed depression and anxiety class, they had a higher probability of being diagnosed with somatization disorder ($x^2=5.149$, V=0.048) compared to the cognitive depression class. Regarding functional impairment and service use, the anxiety and somatic depression class had more assessments ($x^2=9.778$, 0.652, V=0.066, 0.017) and psychotherapy appointments ($x^2=133.021$, 25.905; V=0.244, 0.108) and reported the highest degree

Table 4 Outcomes across classes

Outcomes (DECAT option)	1. Anxiety and somatic depression		2. Mixed depression and anxiety		3. Cognitive depression		Overall chi-square test	Sig. diff.	V
	P	95% CI	*P*	95% CI	*P*	95% CI			
Being on sick leave	0.414	[0.379–0.449]	0.252	[0.225–0.279]	0.313	[0.276–0.350]	48.383***	1 > 3 > 2	0.104
Bipolar disorder	0.047	[0.031–0.063]	0.050	[0.036–0.064]	0.060	[0.042–0.078]	1.145		0.017
Depressive disorder	0.546	[0.511–0.581]	0.193	[0.168–0.218]	0.375	[0.338–0.412]	270.687***	1 > 3 > 2	0.265
Anxiety disorder	0.289	[0.258–0.320]	0.239	[0.212–0.266]	0.235	[0.202–0.268]	7.133*	1 > 3,2	0.043
PTSD	0.103	[0.081–0.125]	0.106	[0.086–0.126]	0.093	[0.071–0.115]	0.732		0.014
Somatization	0.034	[0.022–0.046]	0.049	[0.035–0.063]	0.028	[0.016–0.040]	5.334	2 > 3	0.037
Comorbidity	0.228	[0.199–0.257]	0.075	[0.059–0.091]	0.121	[0.096–0.146]	79.783***	1 > 3 > 2	0.144
Mixed	0.011	[0.003–0.019]	0.022	[0.012–0.032]	0.038	[0.024–0.052]	11.173**	3 > 1	0.054
Outcomes (BCH option)	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**			
Appointments									
Assessment	3.951	[3.675–4.227]	3.376	[3.147–3.605]	3.241	[3.008–3.474]	15.823***	1 > 3,2	0.057
Psychotherapy	10.854	[10.223–11.485]	6.267	[5.812–6.722]	8.215	[7.619–8.811]	134.160***	1 > 3 > 2	0.165
Functional impairment									
WSAS score	26.077	[25.609–26.545]	16.749	[16.245–17.253]	22.077	[21.556–22.598]	735.858***	1 > 3 > 2	0.386

Note: DECAT option is the probability of endorsing yes compared to endorsing no

Sig. diff = statistically significant differences between classes at $p<0.05$ level in x^2. CI = Confidence interval. V = Cramér's V

*$p<0.05$, **$p<0.01$, ***$p<0.001$

of impairment (x^2=735.361, 206.953; V=0.573, 0.304) compared to the mixed depression and anxiety class, and cognitive depression class respectively. The cognitive depression class reported more psychotherapy appointments (x^2=35.389, V=0.126) and functional impairment (x^2=137.546, V=0.248) than the mixed depression and anxiety class.

Discussion

This study shows that linking PROM with register data on CMHD can provide supplemental information on the service needs of routine mental health outpatients. By applying several procedures, different aspects of CMHD may be studied. In turn, clinicians can be made more aware of risk symptoms for difficult to treat patients, and policy decision makers informed of potential treatment targets. Patients with comorbid mood and anxiety disorders differed from non-comorbid patients as a group characterized by higher symptom severity. However, the prevalence of comorbidity (22%) was lower than what has been reported in previous studies [6], making further examination of the sample warranted. Thus, FMM-aggregated PROM data was examined. Three classes distinguished themselves regarding levels of cognitive depression, somatic depression, and anxiety, together with clinically relevant covariates. Since these classes had great within-class variability, patients within each sub-group were highly diverse. Nevertheless, by applying FMM, the simultaneous conceptualization of psychiatric constructs as dimensional and categorical could be made, which may provide a more parsimonious perspective of these complex constructs.

One potential of combining these sources of information is to examine the one-year prevalence of comorbid CMHD in psychiatric treatment. Since comorbid mood and anxiety disorders in Norwegian diagnostic registers have been suggested to be both overestimated [37] and underestimated [38], further examination was warranted. In a study by Torvik et al. [37], Norwegian national health register data was cross-validated with diagnostic interviews, showing a higher prevalence of comorbid mood and anxiety disorders in specialized mental health care registers compared to diagnostic interviews. On the other hand, Øiesvold et al. [38] examined the main diagnoses from medical case reports. They used cross-examination with an expert who had diagnostic interviews and hospital records available while being blinded to the diagnosis given in the case reports. Due to the lower prevalence of comorbid disorders in the case reports compared to expert opinions, Øiesvold et al. [38] concluded that comorbidity could be underdiagnosed in psychiatric registers. Consequently, comorbid mood and anxiety disorders could be under-communicated to the service suppliers, who risk to under-estimate the service needs of the patients. Therefore, we applied a third procedure, using all available registered mood and anxiety diagnoses one year after treatment started. In theory, this procedure could have inflated the prevalence of comorbid mood and anxiety disorder compared to point-prevalence. However, the low prevalence of comorbid mood-anxiety disorders compared to previous PROM research with diagnostic data [6] indicates that this was not an issue. The reason behind the low prevalence of comorbidity is unknown. Since data was extracted from a routine clinical setting, where clinicians have high caseloads, clinicians may be more concerned with reporting the primary diagnosis and less concerned with coding secondary diagnoses. This is in line with the assumption made by Øiesvold et al. [38] that clinicians rely on heuristic principles, which make them prone to not following up on seemingly irrelevant questions.

The second objective was to examine the characteristics of symptom-homogeneous subgroups. Similar to previous clinical LCA studies [18, 19], we also extracted three LCA classes, labelled high- moderate- and low distress classes based on self-reported depression and anxiety symptom severity. Due to the strict model assumptions of LCA, analysis with FMM was conducted. Since we assumed variations in symptoms of depression and anxiety as well as the range of severity, the 3-factor, 3-Class FMM-2 model was chosen over FMM-1, which additionally had worse criterion indices. The classes identified by FMM-2 reflected qualitatively different aspects of patients' distress. The anxiety and somatic depression class had higher service use, higher levels of functional impairment, higher age, were more often single, and more often on sick leave, compared to the other classes. They had also higher rates of depression, anxiety, and comorbidity. It is therefore possible that this class, equal to one-third of the sample, was characterized by patients with persistent psychological problems of high complexity across a broad range of areas. The elevated levels of somatic-depressive problems together with anxiety reported by this class could therefore reflect a group of patients with a chronic course of depression. A longitudinal study on patients with depression showed higher cortisol levels and c-reactive proteins amongst patients with persistent problems, and these biomarkers had a considerably stronger relationship to somatic-, rather than cognitive symptoms of depression [39]. Thus, higher somatic symptoms of depression could indicate the chronicity of the disorder.

Levels of somatic symptoms of depression also separated the other two classes. The mixed depression and anxiety- were the largest class, compromising 40% of the patients. They were more often in a relationship and younger compared to patients in the other classes. Their levels of cognitive symptoms of depression and anxiety

were similar to the class labelled cognitive depression class. Their higher levels of somatic symptoms of depression were unexpected since they also had less service use, reported a lower degree of functional impairment and lower levels of sickness absence. A possible explanation is that patients in the mixed depression and anxiety class are highly distressed but still able to work or study, which means that they are more exposed to daily stressors. These patients may express more somatic depressive complaints as a result of trying to cope with difficult everyday challenges. Correspondingly, patients in the cognitive depression class may be less exposed to work or study-related daily stressors, and at the same time report more severe functional impairment, and higher treatment needs.

The relationship between symptoms and functioning is complex, and many times weaker than expected. In a review, McKnight and Kashdan [40] argued that the dimensionality of the instruments and sample properties could obscure the relationship. Thus, when these are accounted for, new information may be obtained. Results showed the accumulated value of conjoining clinician-rated and self-reported information, as this provides new perspectives of the patient population that are not readily available from each source of information alone. Thus, a clinical implication of these findings is that higher symptom severity does not necessarily imply more service use and higher functional impairment when latent subgroups were accounted for. At the clinician-patient level, this finding shows the wariness one must have when interpreting patients' symptoms of CMHD, since lack of somatic symptoms of depression could be associated with higher functional impairment. Further, by being informed of the severe impact high somatic symptom of depression together with anxiety could imply for difficult to treat patients, clinicians could become more aware of risk factors for resource demanding patients. The same information could also help policy decision makers evaluate treatments that target these factors for difficult to treat patients, and for patients at risk of deterioration. However, since the results are not unequivocal, alternative approaches for identifying subgroups might be better suited in clinical mental health contexts other than symptoms of CMHD, such as patients' perceptions of functional impairment.

Our 3-factor 3-class FMM solution differed from two previous FMM studies examining depression and anxiety [22, 23]. Since the aims of the studies varied, differences could be derived from the population studied and assumptions of the underlying factor structure of applied class indicators. In a study examining the validity of subthreshold mixed anxiety and depressive disorder, a 2-factor 4-class solution was found, consisting of comorbid, depressive, anxiety and low symptom classes [22]. Another study, examining participants' reports from their worst depressive episode in a general population subsample with a lifetime history of depression, specified a 1-factor 4-class solution, which was labelled severe depression with anxiety, moderate depression with anxiety, moderate depression without anxiety and mild depression [23]. Although objectives and results differed, these studies complement each other to show the additive value of examining depression and anxiety simultaneously.

Using FMM to aggregate PROM data, and to link such data with diagnostic registers could be used to combine two approaches to understanding psychopathology as dimensional (patients' self-reported symptoms), or categorical (clinicians' diagnostics) constructs. These perspectives may be viewed as complementary if analysed within an appropriate statistical framework. For example, Borsboom et al. [41] in a review of approaches to modelling the structure of psychiatric constructs, suggested that FMM may serve as a hybrid solution for modelling categorical and continuous approaches to measuring mental disorders at the same time. In line with this, we demonstrate that applying FMM to analyse PROM and register data may be justified in a context where dimensional phenomena such as symptoms of depression and anxiety are assumed to co-exist with categorical phenomena such as CMHD.

Limitations

Some service indicators of interest were not available. The lack of longitudinal assessment is a limitation since this would allow for analysis of between-group trajectories and group transitions over time. This study did not account for dropouts, which could have informed service providers of groups at more risk of not completing their treatment. A broader spectra of patient information, such as previous treatment attempts, economic- and ethnic background could have added further information about the class results. Due to ethical constraints, we could not collect data on patients who declined participation, and there is thus a risk of selection bias. The diagnostic process was not conducted with research in mind and reflects the prevalence of CMHD given by clinicians with ordinary caseloads. Thus, the "true" prevalence could be higher than found in this study.

It is considered a strength that the sample was heterogeneous and included patients with broad spectra of distress that are usually found in routine outpatient treatment. However, since the sample contained not only patients with depressive or anxiety disorders, the use of instruments that measure symptoms of depression and anxiety could have missed other aspects of distress that patients might endure.

Conclusion

PROM data is a valuable source of patient information, which has the potential to provide organizational knowledge on several levels when conjoined with patient data. However, due to the complexity of data, several procedures should be applied. This study used two procedures, first by making use of diagnostic data, showing prevalence of CMHD and associated clinical covariates, such as symptoms of depression and anxiety. Second, by aggregating PROM data with diagnostic and treatment registers using FMM, information about symptom-homogenous subgroups were obtained. This can be of use for understanding larger patient clusters, such as a group with high somatic symptom of depression and anxiety, with more severe functional impairment and higher service use. These patients were characterized by higher age and were more often single, indicating more aggravating life situations. This study also showed the caution that should be exercised when interpreting symptoms of CMHD, since their relationship with functional impairment and service use can be complex. Informed by class results and clinically relevant covariates, a group of patients had lower levels of somatic symptoms of depression, and at the same time higher levels of functional impairment, and higher service use than the least affected group.

However, since there is limited clinical research using FMM on anxiety and depression, and the research that's been conducted shows ambiguous findings, the translational value of this procedure is still unclear. Additionally, since different factors of CMHD could be important targets in various groups of patients, further research is warranted.

Abbreviations

PROM	Patient-Reported Outcome Measures
CMHD	Common Mental Health Disorders
PTSD	Post-Traumatic Stress Disorder
LCA	Latent Class Analysis
FMM	Factor Mixture Modelling
PHQ-9	The nine item Patient Health Questionnaire
GAD-7	The seven item Generalized Anxiety Disorder scale
WSAS	The Work and Social Adjustment Scale
OCD	Obsessive-Compulsive Disorder
SD	Standard Deviation
ANOVA	Analysis Of Variance
MLR	Maximum Likelihood with Robust standard errors
CFA	Confirmatory Factor Analysis
SRMR	Standardized Root Mean Square Residual
RMSEA	Root Mean Square Error of Approximation
CFI	Comparative Fit Index
TLI	Tucker-Lewis Index
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
aBIC	sample size adjusted Bayesian Information Criterion
LMR	Lo-Mendell-Rubin likelihood ratio test
aLMR	adjusted Lo-Mendell-Rubin likelihood ratio test
BLRT	Bootstrap Likelihood Ratio Test

Acknowledgements
Not applicable.

Authors' contributions
All authors participated in designing the study. MB and FA did statistical analyses. MB prepared the first draft of the manuscript, and all authors participated in revising and approving the final manuscript.

Funding
Open access funding is provided by Norwegian University of Science and Technology.

Data Availability
Data are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Informed consent was obtained electronically from all participants. This study complies with the 7th revision of the Declaration of Helsinki, and with the laws and regulations in Norway. It was approved by the regional committee for medical and health ethics, REK Midt-Norge (REK 2019/31836). Further, the Norwegian Centre for Research Data has approved the project (NSD 2020/605327). A data protection impact assessment has been conducted in collaboration with NSD.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Psychology, Norwegian University of Science and Technology (NTNU), Trondheim NO-7491, Norway
[2]Mental Health Care Services, Trondheim Municipality, Norway
[3]Division of Psychiatry, Nidaros Community Mental Health Centre, St. Olavs University Hospital, Trondheim, Norway

Received: 8 May 2023 / Accepted: 27 October 2023
Published online: 03 November 2023

References

1. OECD. Measuring what matters: the patient reported indicator surveys [Internet]. OECD Publishing. 2019. Available from: https://www.oecd.org/health/health-systems/Measuring-what-matters-the-Patient-Reported-Indicator-Surveys.pdf.
2. Pinho LGD, Lopes MJ, Correia T, Sampaio F, Arco HRD, Mendes A, et al. Patient-centered care for patients with depression or anxiety disorder: an integrative review. JPM. 2021;11(8):776.
3. Gelkopf M, Mazor Y, Roe D. A systematic review of patient-reported outcome measurement (PROM) and provider assessment in mental health: goals, implementation, setting, measurement characteristics and barriers. Int J Qual Health Care. 2022;34(Supplement1):ii13–27.
4. Hewlett E, Moran V. Making Mental Health Count: The Social and Economic Costs of Neglecting Mental Health Care [Internet]. OECD; 2014 [cited 2023 May 3]. (OECD Health Policy Studies). Available from: https://www.oecd-ilibrary.org/social-issues-migration-health/making-mental-health-count_9789264208445-en.
5. Roberts T, Miguel Esponda G, Krupchanka D, Shidhaye R, Patel V, Rathod S. Factors associated with health service utilisation for common mental disorders: a systematic review. BMC Psychiatry. 2018;18(1):262.
6. Klein Breteler J, Ikani N, Becker ES, Spijker J, Hendriks G. Comorbid depression and treatment of anxiety disorders, OCD, and PTSD: diagnosis versus severity. J Affect Disord. 2021;295:1005–11.
7. Al Sayah F, Lahtinen M, Bonsel GJ, Ohinmaa A, Johnson JA. A multi-level approach for the use of routinely collected patient-reported outcome measures (PROMs) data in healthcare systems. J Patient Rep Outcomes. 2021;5(S2):98.

8. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The heterogeneity problem: approaches to identify Psychiatric subtypes. Trends Cogn Sci. 2019;23(7):584–601.
9. Miettunen J, Nordström T, Kaakinen M, Ahmed AO. Latent variable mixture modeling in psychiatric research – a review and application. Psychol Med. 2016;46(3):457–67.
10. Curran E, Rosato M, Ferry F, Leavey G. Prevalence and Risk Factors of Psychiatric Symptoms Among Older People in England During the COVID-19 Pandemic: a Latent Class Analysis. Int J Ment Health Addiction [Internet]. 2022 Apr 26 [cited 2023 May 3]; Available from: https://link.springer.com/https://doi.org/10.1007/s11469-022-00820-2.
11. Curran E, Rosato M, Ferry F, Leavey G. Prevalence and factors associated with anxiety and depression in older adults: gender differences in psychosocial indicators. J Affect Disord. 2020;267:114–22.
12. Das-Munshi J, Goldberg D, Bebbington PE, Bhugra DK, Brugha TS, Dewey ME, et al. Public health significance of mixed anxiety and depression: beyond current classification. Br J Psychiatry. 2008;192(3):171–7.
13. Hettema JM, Aggen SH, Kubarych TS, Neale MC, Kendler KS. Identification and validation of mixed anxiety–depression. Psychol Med. 2015;45(14):3075–84.
14. Lei H, Chen C, Hagiwara K, Kusumi I, Tanabe H, Inoue T, et al. Symptom patterns of the occurrence of depression and anxiety in a Japanese General Adult Population Sample: a latent class analysis. Front Psychiatry. 2022;13:808918.
15. Liu Z, Liu R, Zhang Y, Zhang R, Liang L, Wang Y, et al. Latent class analysis of depression and anxiety among medical students during COVID-19 epidemic. BMC Psychiatry. 2021;21(1):498.
16. Rhebergen D, Van Der Steenstraten IM, Sunderland M, De Graaf R, Ten Have M, Lamers F, et al. An examination of generalized anxiety disorder and dysthymic disorder by latent class analysis. Psychol Med. 2014;44(8):1701–12.
17. Singham T, Saunders R, Brooker H, Creese B, Aarsland D, Hampshire A, et al. Are subtypes of affective symptoms differentially associated with change in cognition over time: a latent class analysis. J Affect Disord. 2022;309:437–45.
18. Eaton WW, McCUTCHEON A, Dryman A, Sorenson A. Latent class analysis of anxiety and depression. Sociol Methods Res. 1989;18(1):104–25.
19. Podlogar MC, Rogers ML, Stanley IH, Hom MA, Chiurliza B, Joiner TE. Anxiety, depression, and the suicidal spectrum: a latent class analysis of overlapping and distinctive features. Cogn Emot. 2018;32(7):1464–77.
20. Van Loo HM, Wanders RBK, Wardenaar KJ, Fried EI. Problems with latent class analysis to detect data-driven subtypes of depression. Mol Psychiatry. 2018;23(3):495–6.
21. Clark SL, Muthén B, Kaprio J, D'Onofrio BM, Viken R, Rose RJ. Models and strategies for factor mixture analysis: an Example concerning the structure underlying Psychological disorders. Struct Equation Modeling: Multidisciplinary J. 2013;20(4):681–703.
22. Shevlin M, Hyland P, Nolan E, Owczarek M, Ben-Ezra M, Karatzias T. ICD-11 'mixed depressive and anxiety disorder' is clinical rather than sub-clinical and more common than anxiety and depression in the general population. Br J Clin Psychol. 2022;61(1):18–36.
23. Ten Have M, Lamers F, Wardenaar K, Beekman A, De Jonge P, Van Dorsselaer S, et al. The identification of symptom-based subtypes of depression: a nationally representative cohort study. J Affect Disord. 2016;190:395–406.
24. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13.
25. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092.
26. Mundt JC, Marks IM, Shear MK, Greist JM. The work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry. 2002;180(5):461–4.
27. Beard C, Björgvinsson T. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. J Anxiety Disord. 2014;28(6):547–52.
28. Beard C, Hsu KJ, Rifkin LS, Busch AB, Björgvinsson T. Validation of the PHQ-9 in a psychiatric sample. J Affect Disord. 2016;193:267–73.
29. Brattmyr M, Lindberg MS, Solem S, Hjemdal O, Havnen A. Factor structure, measurement invariance, and concurrent validity of the Patient Health Questionnaire-9 and the generalized anxiety disorder scale-7 in a Norwegian psychiatric outpatient sample. BMC Psychiatry. 2022;22(1):461.
30. Johnson SU, Ulvenes PG, Øktedalen T, Hoffart A. Psychometric properties of the General anxiety disorder 7-Item (GAD-7) scale in a Heterogeneous Psychiatric Sample. Front Psychol. 2019;10:1713.
31. Obbarius A, Van Maasakkers L, Baer L, Clark DM, Crocker AG, De Beurs E, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and anxiety Working Group. Qual Life Res. 2017;26(12):3211–25.
32. Pedersen G, Kvarstein EH, Wilberg T. The work and Social Adjustment Scale: psychometric properties and validity among males and females, and outpatients with and without personality disorders: the work and social adjustment scale: psychometric properties and validity among males and females. Personality and Mental Health. 2017;11(4):215–28.
33. Muthén LK. B. O. Muthén. Mplus: Statistical analysis with latent variables: User's Guide (Version 8). 1998.
34. Browne MW, Cudeck R. Alternative ways of assessing model fit. Testing structural equation models. Sage; 1993. pp. 136–62.
35. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equation Modeling: Multidisciplinary J. 1999;6(1):1–55.
36. Wang J, Wang X. Structural equation modeling: applications using Mplus. John Wiley & Sons; 2012.
37. Torvik FA, Ystrom E, Gustavson K, Rosenström TH, Bramness JG, Gillespie N, et al. Diagnostic and genetic overlap of three common mental disorders in structured interviews and health registries. Acta Psychiatr Scand. 2018;137(1):54–64.
38. Øiesvold T, Nivison M, Hansen V, Skre I, Østensen L, Sørgaard KW. Diagnosing comorbidity in psychiatric hospital: challenging the validity of administrative registers. BMC Psychiatry. 2013;13(1):13.
39. Iob E, Kirschbaum C, Steptoe A. Persistent depressive symptoms, HPA-axis hyperactivity, and inflammation: the role of cognitive-affective and somatic symptoms. Mol Psychiatry. 2020;25(5):1130–40.
40. McKnight PE, Kashdan TB. The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. Clin Psychol Rev. 2009;29(3):243–59.
41. Borsboom D, Rhemtulla M, Cramer AOJ, Van Der Maas HLJ, Scheffer M, Dolan CV. Kinds *versus* continua: a review of psychometric approaches to uncover the structure of psychiatric constructs. Psychol Med. 2016;46(8):1567–79.

Publisher's Note

Appendices

Appendix A

Norwegian Version of The Patient Health Questionnaire, PHQ-9

SPØRRESKJEMA OM HELSEN DIN-9 (PHQ-9)

Hvor ofte har du vært plaget av ett eller flere av de følgende problemene i løpet av de siste 2 ukene,? *(Sett "✔" for å vise hvilket svar du velger)*	**Ikke i det hele tatt**	**Noen dager**	**Mer enn halvparten av dagene**	**Nesten hver dag**
1. Lite interesse for eller glede over å gjøre ting	0	1	2	3
2. Følt deg nedfor, deprimert eller fylt av håpløshet	0	1	2	3
3. Vansker med å sovne eller med å sove natten gjennom uten å våkne – eller med at du sover for mye	0	1	2	3
4. Følt deg trett eller slapp	0	1	2	3
5. Dårlig appetitt eller å spise for mye	0	1	2	3
6. Vært misfornøyd med deg selv eller følt deg mislykket - eller følt at du har sviktet deg selv eller familien din	0	1	2	3
7. Vansker med å konsentrere deg om ting, slik som å lese avisen eller se på TV	0	1	2	3
8. Beveget deg eller snakket så langsomt at andre kan ha merket det? Eller motsatt – følt deg så urolig eller rastløs at du har vært mye mer i bevegelse enn vanlig	0	1	2	3
9. Tanker om at du like gjerne kunne vært død eller på annen måte ville skade deg selv	0	1	2	3

FOR OFFICE CODING 0 + _____ + _____ + _____

=Total Score: _____

Hvis du har opplevd ett eller flere av de problemene som nevnes, i hvor stor grad har problemene gjort det vanskelig for deg å utføre arbeidet ditt, ordne med ting hjemme eller å komme overens med andre?

Ikke vanskelig i det hele tatt	**Litt vanskelig**	**Veldig vanskelig**	**Ekstremt vanskelig**
☐	☐	☐	☐

Appendix B

Norwegian Version of The Generalized Anxiety Disorder scale, GAD-7

GAD-7

Hvor ofte har du vært plaget av de følgende problemene i løpet av de siste 14 dagene? *(Sett "✔" for å vise hvilket svar du velger)*	**Ikke i det hele tatt**	**Noen dager**	**Mer enn halvparten av dagene**	**Nesten hver dag**
1. Følt deg nervøs, engstelig eller veldig stresset	0	1	2	3
2. Ikke klart å slutte å bekymre deg eller kontrolleren bekymringene dine	0	1	2	3
3. Bekymret deg for mye om ulike ting	0	1	2	3
4. Vansker med å slappe av	0	1	2	3
5. Vært så rastløs at det har vært vanskelig å sitte stille	0	1	2	3
6. Blitt lett sint eller irritert	0	1	2	3
7. Følt deg redd som om noe forferdelig kunne komme til å skje	0	1	2	3

(For office coding: Total Score T____ = ____ + ____ + ____)

Utviklet av Dr. Robert L. Spitzer, Dr. Janet B. Williams, Dr. Kurt Kroenke og medarbeidere med et utdanningsstipend fra Pfizer Inc. Det er ikke nødvendig med tillatelse til reproduksjon, oversettelse, fremvisning eller distribuering.

ISBN 978-82-326-8226-3 (printed ver.)
ISBN 978-82-326-8225-6 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology