frontiers | Research Topics

Economic evaluation of mental health interventions

Edited by
Huajie Jin, Ali Jalali, Ben Wijnen and Yuhua Bao

Published in
Frontiers in Psychiatry
Frontiers in Public Health

ISSN 1664-8714
ISBN 978-2-8325-3692-6
DOI 10.3389/978-2-8325-3692-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Economic evaluation of mental health interventions

Topic editors

Huajie Jin — King's College London, United Kingdom

Ali Jalali — Joan & Sanford I. Weill Medical College of Cornell University, United States

Ben Wijnen — Trimbos Institute, Netherlands

Yuhua Bao — Department of Population Health Sciences, Weill Cornell Medicine, United States

Citation

Jin, H., Jalali, A., Wijnen, B., Bao, Y., eds. (2023). *Economic evaluation of mental health interventions*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3692-6

Table of contents

frontiers | Frontiers in Psychiatry

TYPE Editorial
PUBLISHED 28 September 2023
DOI 10.3389/fpsyt.2023.1294245

OPEN ACCESS

EDITED AND REVIEWED BY
Wulf Rössler,
Charité University Medicine Berlin, Germany

*CORRESPONDENCE
Huajie Jin
huajie.jin@kcl.ac.uk

RECEIVED 14 September 2023
ACCEPTED 18 September 2023
PUBLISHED 28 September 2023

CITATION
Jin H, Jalali A, Wijnen B and Bao Y (2023) Editorial: Economic evaluation of mental health interventions. *Front. Psychiatry* 14:1294245. doi: 10.3389/fpsyt.2023.1294245

Editorial: Economic evaluation of mental health interventions

Huajie Jin[1]*, Ali Jalali[2], Ben Wijnen[3] and Yuhua Bao[2]

[1]King's Health Economics (KHE), Institute of Psychiatry, Psychology and Neuroscience at King's College London, London, United Kingdom, [2]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, United States, [3]Center for Economic Evaluations, Trimbos Institute, Utrecht, Netherlands

KEYWORDS

economic evaluation, editorial, mental health, cost-effectiveness analysis, cost-of-illness

Editorial on the Research Topic
Economic evaluation of mental health interventions

Introduction

Mental health disorders affected more than 1 billion people globally and were responsible for 7% of the global burden of disease as measured in Disability Adjusted Life Years (DALYs) and 19% of all years lived with disability (1). Economic evaluation is increasingly employed to guide resource allocation decisions and policymaking in mental health.

Our Research Topic showcases cutting-edge economic evaluations in mental health to inform public and private resource decisions. It includes nine papers: two partial economic evaluations focusing solely on the costs of interventions/diseases, three full economic evaluations assessing both costs and consequences of interventions, two systematic reviews, one perspective paper, and one protocol paper.

Partial economic evaluations

Based on data from two RCTs, Paterson et al. estimated that delivering a mental health recovery narrative web application costs £349 per user for those with psychosis and £241 per user for those without psychosis. Crucially, this study accounts for intervention development costs, which is important but often overlooked in costing studies. The results of this study can be used to estimate the cost of delivering NEON at scale and improve consistency in reporting of cost for similar digital health interventions.

Based on multiple sources of information, Sousa et al. estimates the total disease burden of Treatment-Resistant Depression and Major Depression with Suicide Risk in Portugal at 66.3 thousand DALYs. Direct costs were €30.8 million, mainly from medical appointments and medication. Adding productivity losses, the total cost reached €1.1 billion. This study emphasized the need for prioritizing health promotions for both disorders.

Full economic evaluations

Of the three full economic evaluations included in this topic, Le Novere et al. conducted a trial-based economic evaluation, in which the trial provides the main source of input data; while Liu et al.(a) and Kleijburg et al. employed a decision-analytic modeling approach.

Le Novere et al. assessed the cost-effectiveness of peer-supported self-management for people discharged from mental health crisis teams in England, from a mental health service perspective. Compared to usual care, the intervention had a 57% chance of being cost-effective at £20,000 per QALY gained. The main methodological challenge in this study is the significant (nearly 50%) missing data for the utility outcome, while data regarding resource use were nearly complete. This study illustrates how common strategies to address missingness or distributional features of cost and utility data may or may not mitigate biases.

Both Liu et al.(a) and Kleijburg et al. employed a Markov modeling approach, dividing the disease into distinct states with assigned transition probabilities for movement over discrete time periods, known as "Markov cycles" (2). By attaching costs and health outcomes to each of these states and using intervention-specific transition probabilities, a Markov model can be used to estimate the long-term costs and outcomes for the interventions of interest.

Using a previously published model, Liu et al.(b) found that lurasidone was a dominant treatment compared to olanzapine and risperidone in the first-line treatment of schizophrenia in China, resulting in greater QALY gains at lower costs. Kleijburg et al. reported the development of TiBipoMod—A model which can simulates the lifetime costs and health outcomes for various interventions in the treatment of bipolar disorders type I and II, from a societal perspective. A case study conducted based on TiBipoMod showed that mindfulness-based cognitive therapy dominates standard care.

Systematic reviews

Systematic reviews of economic evaluations of healthcare programs and interventions can synthesize crucial data to inform healthcare decision-making and highlight research priorities (3). However, synthesizing evidence from such studies is challenging given inconsistencies in cost-effectiveness research designs (e.g., synthesizing evidence from simulation and trial-based analyses) and reporting guidelines (4, 5). The two systematic reviews included in this Research Topic are notable in that they both follow the Consensus on Health Economic Criteria (CHEC) list for assessing the methodological quality of economic evaluations in systematic reviews (6), and both provide a thorough discussion of the limitations in consolidating evidence across the included studies.

Kugener et al. provide a timely update on the economic evidence for prevention and treatment interventions for child maltreatment, abuse and neglect in high-income countries (US, Australia, UK, Canada). Their study evaluated a total of 11 studies, 7 of which were model-based economic evaluations while 4 were conducted alongside a clinical trial. All studies demonstrated improved outcomes at common cost-effectiveness value thresholds, with two demonstrating cost-savings in addition to effectiveness gains. Kugener et al. noted that cross study comparisons and/or pooling was made difficult due to limited comparability of measures across studies, including lack of commonly applied terminology for child maltreatment, as well as variation in the methodological rigor such as hand lined missing data, which continues to be an issue in the field (7).

Hannah et al. reviewed economic evidence pertaining to interventions for treatment-resistant depression. Their review encompassed 31 studies-−11 conducted alongside clinical trials, and 20 used modeling methods. Similar to Kugener et al., Hannah et al. identified heterogeneity in methodological quality, most notably finding that fewer than half of the model-based evaluations conducted a comprehensive sensitivity analysis of model parameters. An important feature of Hannah et al.'s review is their in-depth discussion on the divergences between the model vs. trial approaches to economic evaluations.

Others

The successful implementation of mental health interventions, particularly behavioral health interventions, frequently demands significant stakeholder engagement. Nonetheless, the expenses associated with this involvement are commonly overlooked in current economic evaluations. In their perspective paper, Raciborski et al. delve into the integration of stakeholder engagement with established economic analysis methods, aiming to enhance decision-making regarding the implementation of behavioral health interventions.

Shah et al. reports on a protocol for a return-on-investment analysis of system-wide service transformation for young people experiencing mental health problems in Canada. Novelties of the proposed study lie in two aspects: economic evaluation of a system transformation (rather than a particular health technology) and assessing population-wide implications of the system intervention (thus capturing complex links between intervention and outcomes and spillovers). Findings of the proposed studies will inform decisions regarding large scale, system transformation initiatives designed to benefit population health.

Author contributions

HJ: Conceptualization, Data curation, Investigation, Methodology, Writing—original draft, Writing—review and editing. AJ: Conceptualization, Data curation, Investigation, Methodology, Writing—original draft, Writing—review and editing. BW: Conceptualization, Data curation, Investigation, Methodology, Writing—original draft, Writing—review and editing. YB: Conceptualization, Data curation, Investigation, Methodology, Writing—original draft, Writing—review and editing.

Acknowledgments

We thank authors of the papers published in this Research Topic for their valuable contributions and the referees for their rigorous review.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Rehm J, Shield KD. Global burden of disease and the impact of mental and addictive disorders. *Curr Psych Rep.* (2019) 21:997. doi: 10.1007/s11920-019-0997-0

2. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics.* (1998) 13:4. doi: 10.2165/00019053-199813040-00003

3. Jalali A, Ryan DA, McCollister KE, Marsch LA, Schackman BR, Murphy SM. Economic evaluation in the national drug abuse treatment clinical trials network: past, present, and future. *J Subst Abuse Treat.* (2020) 112:18–27. doi: 10.1016/j.jsat.2020.02.002

4. Jacobsen E, Boyers D, Avenell A. Challenges of systematic reviews of economic evaluations: a review of recent reviews and an obesity case study. *Pharmacoeconomics.* (2020) 38:259–67. doi: 10.1007/s40273-019-00878-2

5. Anderson R. Systematic reviews of economic evaluations: utility or futility? *Health Econ.* (2010) 19:350–64. doi: 10.1002/hec.1486

6. Evers S, Goossens M, De Vet H, Van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. *Int J Technol Assess Health Care.* (2005) 21:240–5. doi: 10.1017/S0266462305050324

7. Jalali A, Tamimi RM, McPherson SM, Murphy SM. Econometric issues in prospective economic evaluations alongside clinical trials: combining the nonparametric bootstrap with methods that address missing data. *Epidemiol Rev.* (2022) 44:67–77. doi: 10.1093/epirev/mxac006

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 11 August 2022
DOI 10.3389/fpubh.2022.898491

OPEN ACCESS

EDITED BY
Huajie Jin,
King's College London,
United Kingdom

REVIEWED BY
Nerea Almeda,
Universidad Loyola Andalucía, Spain
Giulio Castelpietra,
Direzione Centrale Szlute, Italy
Mencia R. Gutiérrez-Colosía,
Universidad Loyola Andalucía, Spain

*CORRESPONDENCE
Rute Dinis Sousa
rutedesousa@gmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Public Mental Health, a section of the journal Frontiers in Public Health

RECEIVED 17 March 2022
ACCEPTED 11 July 2022
PUBLISHED 11 August 2022

CITATION
Sousa RD, Gouveia M, Nunes da Silva C, Rodrigues AM, Cardoso G, Antunes AF, Canhao H and de Almeida JMC (2022) Treatment-resistant depression and major depression with suicide risk—The cost of illness and burden of disease.
Front. Public Health 10:898491.
doi: 10.3389/fpubh.2022.898491

Treatment-resistant depression and major depression with suicide risk—The cost of illness and burden of disease

Rute Dinis Sousa[1,2,3*†], Miguel Gouveia[4†], Catarina Nunes da Silva[1,2,3], Ana Maria Rodrigues[1,2,3], Graça Cardoso[5,6], Ana Filipa Antunes[5,6], Helena Canhao[1,2,3,7] and José Miguel Caldas de Almeida[5,6]

[1]Comprehensive Health Research Centre, NOVA Medical School, Universidade Nova de Lisboa, Lisboa, Portugal, [2]NOVA Medical School, Universidade Nova de Lisboa, Lisboa, Portugal, [3]Episaúde – Associação Científica, Évora, Portugal, [4]Católica Lisbon School of Business and Economics, Universidade Católica Portuguesa, Lisboa, Portugal, [5]Comprehensive Health Research Centre, Lisbon Institute of Global Mental Health, Lisbon, Portugal, [6]Department of Mental Health, NOVA Medical School, Nova University of Lisbon, Lisbon, Portugal, [7]National School of Public Health, UNL, Lisboa, Portugal

Introduction: Treatment-Resistant Depression (TRD) and Major Depression with Suicide Risk (MDSR) are types of depression with relevant effects on the health of the population and a potentially significant economic impact. This study estimates the burden of disease and the costs of illness attributed to Treatment-Resistant Depression and Major Depression with Suicide Risk in Portugal.

Methods: The disease burden for adults was quantified in 2017 using the Disability-Adjusted Life Years (DALYs) lost. Direct costs related to the health care system and indirect costs were estimated for 2017, with indirect costs resulting from the reduction in productivity. Estimates were based on multiple sources of information, including the National Epidemiological Study on Mental Health, the Hospital Morbidity Database, data from the Portuguese National Statistics Institute on population and causes of death, official data on wages, statistics on the pharmaceutical market, and qualified opinions of experts.

Results: The estimated prevalence of TRD, MDSR, and both types of depression combined was 79.4 thousand, 52.5 thousand, and 11.3 thousand patients, respectively. The disease burden (DALY) due to the disability generated by TRD alone, MDSR alone, and the joint prevalence was 25.2 thousand, 21 thousand, and 4.5 thousand, respectively, totaling 50.7 thousand DALYs. The disease burden due to premature death by suicide was 15.6 thousand DALYs. The estimated total disease burden was 66.3 thousand DALYs. In 2017, the annual direct costs with TRD and MDSR were estimated at € 30.8 million, with the most important components being medical appointments and medication. The estimated indirect costs were much higher than the direct costs. Adding work

productivity losses due to reduced employment, absenteeism, presenteeism, and premature death, a total cost of € 1.1 billion was obtained.

Conclusions: Although TRD and MDSR represent relatively small direct costs for the health system, they have a relevant disease burden and extremely substantial productivity costs for the Portuguese economy and society, making TRD and MDSR priority areas for achieving health gains.

KEYWORDS

treatment-resistant depression, major depression with suicide risk, cost of illness, burden of disease, disability-adjusted life years (DALYs) lost

Background

Depressive disorders, which can be lasting or recurrent, are characterized by sadness, loss of interest or pleasure, feelings of guilt or low self-esteem, disturbances in sleep or appetite, feeling tired, and low concentration (1). Depression is associated with deficits in individuals' professional, social, and personal functioning, contributing to decreased patients' quality of life. In its most severe form, the depressive disorder can lead to suicide (1, 2).

Globally, it is estimated that more than 300 million individuals, equivalent to 4.4% of the world population, are affected by depression (1). A recent study based on Global Burden of Disease (GBD) estimates on subjects aged 10–24 years, shows that Portugal has the highest prevalence of mental disorders in Europe during the period 1990–2019. It showed also that YLDs due to mental disorders are the first cause of disability at this age in Portugal, as in other European countries (3). Portugal is the second country in Europe with the highest prevalence of psychiatric diseases, and mood disorders had a prevalence of 7.9% in 2009 (4).

Although several drugs are indicated for the treatment of depression, studies reveal that one to two-thirds of patients will not respond to the first prescription, and 15–33% will fail to respond to multiple interventions (5).

According to the literature, there is no exact definition for treatment-resistant depression (6). However, it has been established that the term refers to an inadequate response to at least one antidepressant, with adequate dose and duration, in patients with depression (7). Nevertheless, what constitutes an inadequate response is still the subject of debate. Nowadays, for many specialists, the goal of treatment is to achieve remission (8). Despite the lack of consensus, the duration of response assessment is generally defined as a minimum of 6 weeks of treatment (9).

There are few estimates of the prevalence of treatment-resistant depression. Available data points to a global life-time prevalence of Major Depression (MD) of 10 to 15%, and some studies estimate this prevalence to be between 8.1 and 11.2% in low/medium income countries and 13% in high-income countries (10, 11). The epidemiology of Treatment-Resistant Depression's (TRD) is not so well studied and characterized due to the heterogeneity of criteria and methodologies and scarcity of epidemiology-related studies. However, it is estimated that more than one-third of treated patients with MD progress to TRD (7, 11, 12), a proportion that has been corroborated by the study "Sequenced Treatment Alternatives to Relieve Depression" (study STAR*D), one of the largest and most important studies that evaluated the burden of treatment-resistant depression (13). There has been an increase in the number of people diagnosed with this type of depression. These patients tend to have higher comorbidities with other psychiatric disorders, incapacity for work, absenteeism, more frequent hospitalizations, which consequently generates higher costs for the health system (14).

Much of the cost and disability associated with depression are explained by resistance to treatment (15). Depressive disorders are responsible for an overall loss of more than 50 million years of life adjusted for disability (1). Estimates indicate that depression will be the leading cause of global disease burden worldwide by the year 2030 (1, 2, 16).

Management of treatment-resistant depression requires a multimodal approach, which includes pharmacological and non-pharmacological intervention. It may include cognitive-behavioral therapy, electroconvulsive therapy, vagus nerve stimulation, and transcranial magnetic stimulation (12, 14, 17–20). However, pharmacological therapy remains the main component of treatment (14).

Abbreviations: AF, Attributable Fraction; ATC, Anatomical Therapeutic Chemical; DALY, Disability Adjusted Life Years; HDG, Homogeneous Diagnostic Groups; hmR, Health Market Research; ICD 10-CM, International Classification of Diseases 10, Clinical Modification; ICD 9-CM, International Classification of Diseases 9, Clinical Modification; MDSR, Major Depression with Suicide Risk; NESMH, National Epidemiological Study of Mental Health; RR, Relative Risk; TRD, Treatment-Resistant Depression; WHO, World Health Organization; WMH-CIDI, World Health Organization World Mental Health Composite International Diagnostic Interview; WMHSI, World Mental Health Surveys Initiative; YLD, Years Lived with Disability; YLL, Years of Life Lost.

Suicidal ideation can occur in several psychopathological contexts, namely in depressive pathology (21). In the pharmacological treatment of patients with depression with suicide risk, the drugs must have a rapid and early effect. Even in patients who respond to conventional antidepressant medication, obtaining a response always implies a latency time of action that can take up to 3 weeks after treatment initiation. This delay can be fatal in severe cases, hence the need for therapeutic alternatives with a rapid onset of action, which justifies the evaluation and therapeutic indication of specific drugs for this purpose.

The endeavor of the present study is to contribute to a comprehensive perspective of the costs endured by the Portuguese health system, generated by TRD and MDSR.

Our research on the burden and costs of TRD and MDSR in Portugal had two main objectives. The first was to uncover the disease burden by estimating the component attributed to treatment-resistant depression and major depression with suicide risk based on the following indicators: mortality, morbidity, and Disability-Adjusted Life Years (DALYs) for the year 2017. Our second objective was to estimate the economic cost of diagnosing and treating TRD and MDSR, including direct costs (medical and non-medical costs) and indirect costs (loss of productivity—relevant to the perspective of society) for the year 2017.

Methods

The primary source of the epidemiological information used in this study is the "National Epidemiological Study of Mental Health" (NESMH) (4), which took place between 2008 and 2009. This study is the most accurate and robust national data available, assessing primary data collected in a representative sample of the adult population in mainland Portugal. The results of the NESMH were adjusted in our study regarding the demographic composition of the population due to aging, and all estimates were calculated using the population and demographic structure of the year 2017. Portugal's NESMH was part of the World Mental Health Surveys Initiative (WMHSI), promoted by the World Health Organization and Harvard University. WMHSI was coordinated internationally by Prof. Ronald Kessler and is fully described elsewhere (22–27). The methodology and implementation of NESMH are described in Xavier et al. (16). The NESMH questionnaire (16) was divided into two parts to reduce the time taken to answer. Part I used a total sample number of 3,849 participants to represent the general population and included an initial baseline assessment to diagnose major mental disorders. Part II of the questionnaire was answered by 2,060 individuals which included all participants with a mental disorder diagnosis and a random sample of 25% of participants without psychiatric disorder. Two weights were created to accommodate the stratification of the sample. The data in Part I were weighted for the differential probability of selection (between and within households), non-response bias, and discrepancies between the sample and the geographic and sociodemographic distribution of the Portuguese population assessed in the *census*. The data of Part II was additionally weighted for the differential sampling of participants from Part I to Part II.

Psychiatric disorders were assessed using the World Health Organization World Mental Health Composite International Diagnostic Interview version 3.0 (WMH-CIDI 3.0), a comprehensive and fully structured interview designed to assess mental disorders according to the definitions and criteria of the DSM-IV (28) and ICD-10 (29).

Estimates of the prevalence of treatment-resistant depression

Before obtaining epidemiological information related to TRD, the concept of TRD was operationalized within the framework of available epidemiological data, adopting a "proxy" as collected data does not strictly respect any of the possible definitions of TRD. The criteria used to define a TRD case included the presence of a major depressive disorder in the last 12 months assessed objectively with a validated scale, and a negative answer to the question "Have you ever received treatment that you considered to be useful or effective?", present in the depression module of WMH-CIDI 3.0.

Results show that the global prevalence rate of treatment-resistant depression in the population aged 18 and over in Portugal should be 1.1%. This prevalence corresponds to about 14.9% of the population with major depressive disorder. These values are obtained with a small number of observations ($n = 43$) in a representative sample of the population, using the previously described weights. Combining this prevalence rate, estimated using the referred weights, with the estimates of the Mainland's adult population in 2017 from Portugal's National Institute of Statistics (INE), a total prevalence for TRD of ∼90 thousand patients is obtained.

Estimates of the prevalence of major depression with suicide risk

In the case of MDSR, the operationalization of the concept was achieved by combining major depressive disorder with suicidal ideation. Both situations were reported in the 12 months before the moment of the survey. The results obtained indicate that the prevalence of MDSR is expected to be

TABLE 1 Distribution of estimated TRD and MDSR cases by the level of severity.

Severity level	TRD	MDSR
Severe cases	27.7%	64.4%
Moderate cases	60.7%	31.0%
Mild cases	11.6%	4.6%

TRD, Treatment-resistant depression; MDSR, Major Depression with suicide risk.

around 0.8% of the adult population in Mainland Portugal. Multiplying the Mainland adult population by this rate, a total prevalence of MDSR of ∼64 thousand patients was obtained in 2017.

Estimates of combined prevalence of treatment-resistant depression and major depression with suicide risk

To fully understand the data presented above, it should be highlighted that some patients check both criteria for identifying the different types of depressive pathology and, therefore, appear in both prevalence estimates. This reality reinforces the adequacy and need to study the cost and burden of the two clinical entities together.

The crossing of individual information in the survey with objective and representative data from the Portuguese population shows that about 11.2% of patients with TRD also suffer from MDSR. Alternatively, it appears that about 19.3% of patients with MDSR suffer from TRD. This information implies that the characterization of these patient populations should consider the prevalence of TRD without MDSR, the prevalence of MDSR without TRD, and the joint prevalence of TRD and MDSR. The aggregate results present a combined prevalence of Treatment-Resistant Depression and Major Depression with Suicide Risk of 11,283 patients who simultaneously meet the criteria for both types of depression (7.9% of the total study patients).

Severity level distribution

The National Epidemiological Study on Mental Health (NESMH) contains information on the severity levels of the disease. The classification of patients by severity levels was based on the criteria adopted in the World Mental Health Survey (WMHS) (see the paper by Xavier et al.). For patients with TRD and MDSR, the distribution estimated by levels of severity can be seen in Table 1.

Table 1 shows that the distribution by severity levels indicates a difference between TRD and MDSR, with the latter pathology having significantly higher levels. In the case of the joint prevalence of TRD and MDSR, it is assumed that the distribution by severity levels is that of the MDSR, as it is the most severe condition.

The matter of duration

An important component necessary to estimate the burden of TRD and MDSR is the duration of the episodes and, consequently, the fraction of the time affected by the disease in the reference year for the analysis. The distribution of the duration of episodes of major depression appears to be very heterogeneous. Spijker et al. (30) retrospectively used data from the Dutch mental health survey to estimate the duration of an episode of major depression. They found that in 50% of cases the duration was 3 months or less, in 63% of cases was 6 months or less, in 76% of cases was 12 months or less, but in 20% of cases the duration was found to be 24 months or more. Spijker et al. (30) reported that the median duration was 3 months, but the estimated average duration was 8.4 months. No specific references were found for the case of TRD and MDSR. In the present study, the assumption is made that the duration of TRD and MDSR can be approximated by the duration of episodes of major depression. The study by Ferrari et al. (31), which synthesizes information on estimates of the various parameters necessary to calculate the disease burden, indicated that the average duration used in the Global Burden of Disease, resulting from a synthesis of the literature, was 37.7 weeks. The present study has used this value, i.e., it is assumed that depression affects the health of patients at an average of 72.3% of the time in the year in which an episode occurs.

Burden of disease

The burden of disease was estimated through the Disability-Adjusted Life Years (DALYs). The most recent version of the methodology introduced by the World Bank and the World Health Organization (WHO) was adopted in this study (32).

DALYs are a measure, expressed in time, of the amount of health lost due to the disability generated by disease or premature death. The measure includes two indicators: (1) the years lost due to premature death (Years of Life Lost - YLL), the lost time being operationalized as the difference between age at the time of death and the standard life expectancy for that age; and, (2) the Years lived with Disability (YLD), where the time spent suffering a disability is considered (33).

The equation used to estimate the number of DALYs lost by an individual is as follows:

$$DALY(c, s, a, t) = YLL(c, s, a, t) + YLD(c, s, a, t)$$

Where c is cause, s is sex, a is age, and t is time.

Disability is measured by a coefficient with values between 0 (without any disability, perfect health) and 1 (total disability or death). Standard life expectancy results from a reference mortality table designed to have universal applicability.

Years of life lost due to premature death

The years of life lost due to premature death (YLL) are calculated by multiplying the number of deaths caused by the disease under analysis and the years of life lost, which are a function of the age at which death occurs. In the case of depression, it is not usually taken as a direct cause of death. However, a significant fraction of suicide deaths can be statistically attributed to depression. In this context, YLL were estimated considering that a fraction of the overall suicide mortality is attributable to depression. Following the approach that uses the concept of attributable fraction (34), the fraction of total mortality attributable to MDSR was determined by the equation:

$$Attributable\ Fraction\ (AF) = \frac{p(RR-1)}{p\,(RR-1)+1}$$

where RR is the Relative Risk of death in patients with the disease under study and p is the prevalence of the disease. The RR of suicide mortality in major depression considered by Ferrari et al. (31) was 19.9. In the present study, TRD, and especially MDSR, would be expected to have higher RRs than those of major depression in general. In the literature, specific estimates were not found. The adopted methodology was to estimate the attributable fraction of suicides to major depression and, later, through the opinion of experts to obtain the proportion of this attributable fraction (AF) that applies to the two types of depressive pathology under study.

According to NESMH, the prevalence of major depression in the year prior to the survey interview was 6.8% (4). The AF that results from the above equation is therefore 56.24%.

The next step was to estimate the proportion of these suicides attributable to the prevalence of TRD and MDSR. Our research assumed that, as in these data, suicide did occur, so ex post it is a case of MDSR. Considering the diagnostic criteria for major depressive disorder in its different configurations, the possibility of some suicides occurring in patients who would not be diagnosed with MDSR should be considered. Bearing this is mind, it was conservatively assumed that 90% of suicides attributable to major depression in general can be more specifically attributed to MDSR or TRD. Assuming this rate of 90% (out of the previous AF of 56.24%), the final attributable fraction is 50.6%. This is the percentage of the disease burden and costs generated by suicide deaths that will be attributed to MDSR and TRD in the present study.

TABLE 2 Disability weights considered in the burden of disease estimates.

Severity level (1)	Disability weight (2)	Proportion in the prevalence of TRD (3)	Proportion in the prevalence of MDSR (4)
Mild	0.1451	11.6%	4.6%
Moderate	0.396	60.7%	31%
Severe	0.658	27.7%	64.4%
Average weight	–	0.439	0.553

Sources: Salomon et al. (35) regarding (2), and experts' qualified opinions and author's calculations to (3) and (4).

Years lived with disability

DALY indicator, as a metric of disease burden, estimates, in addition to mortality, the disease burden generated by morbidity, considering that the time lived with a disease contributes to the years of life lost as that such a disease is disabling. The equation used to estimate the YLD number is as follows:

$$YLD(c,s,a,t) = P(c,s,a,t) \times DW(c,s,a)$$

Where P is Prevalence of cause (c), by age (a) and sex (s), in year (t); and DW is Disability Weight specific to the cause (c), age (a) and sex (s).

The disease burden was estimated from the indicators: prevalence, mortality, disease duration.

The estimation of YLDs requires the use of disease-specific weighting or disability coefficients and is calibrated according to the different levels of disease severity. The most current version of the weights was published by Salomon et al. (35) and the weights by severity level for the case of depression are shown in Table 1.

The use of these weights in the case of TRD and MDSR depends on the distribution of patients by severity levels. This information, from NESMH is shown in Table 1, and can be reviewed in columns (3) and (4) of Table 2. The average values of the weights are very high compared to other pathologies, probably because an expressive proportion of patients with major depression are not actually being treated, which in the case of MDSR is reinforced by the fact that even in the target patients of treatment they do not evaluate it as being effective.

Costs of illness

Direct costs

The direct costs of TRD and MDSR resemble the monetary appreciation of the resources consumed in treating these diseases. The study of direct costs is based on the previously presented estimates of disease prevalence and the information on the pattern of use of resources contained in the "National Epidemiological Study of Mental Health" (NESMH) (4) and in the opinion of their experts. Microdata available in the 2017 Hospital Morbidity Database regarding inpatient and outpatient episodes registered using the International Classification of Diseases, tenth version (ICD-10 CM) and billing in Homogeneous Diagnostic Groups (HDG), were considered. Microdata was used to estimate the number of relevant hospitalizations and outpatient episodes, as well as the respective costs. The study also used aggregated data on the consumption of drugs associated with the treatment of depression, from IQVIA and hmR, and expert opinions were used to estimate costs in areas where databases or other quantifiable sources of information are not known officially or academically recognized. Finally, in this study, the unit costs of hospitalizations, hospital consultations, and complementary means of diagnosis and therapy were obtained from the prices defined in Portuguese Law (Ordinance No. 207/2017, of 11 July).

Costs of hospitalizations and ambulatory hospital visits

Estimates of hospital activity related to hospitalization episodes associated with TRD and MDSR are reported. The estimates are based on an analysis of the Hospital Morbidity Database for 2017. In this database, the use of ICD 10 hinders a finer analysis with the separation of TRD and MDSR. Thus, for greater accuracy, the hospitalization episodes associated with TRD and MDSR are presented together.

Registered and coded episodes are included in this database, including coded and registered hospital outpatient episodes. The identification of relevant episodes associated with TRD and MDSR was based on the International Classification of Diseases. The selection of relevant cases was made by the clinical team and experts. The episodes on which the subsequent analysis is based have all been classified with the ICD 10 CM. The use of the ICD 10 - CM classification was evaluated to guarantee the compatibility between the selected episodes and the diseases under study. The starting point was given by the ICD 10 - CM encodings. Major depressive pathology, single episode, and ICD 10 - CM. 0-9 Major, recurrent, depressive pathology. It was also necessary to add some episodes that were considered to be relevant and that were not part of the preliminary analysis. Thus, a set was added to the selected episodes in which the main diagnosis was "suicidal ideation" (ICD 10-CM R45851) provided that secondary diagnoses (from d2 to d50) were included or a sub-item of diagnoses F32 (Pathology major depression, single episode), or a sub-item of diagnoses F33 (major, recurrent depressive pathology). The use of episodes was further refined by considering additional information on the GDHs of the selected episodes.

Costs of pharmacological therapy

Estimates of drug costs in the treatment of TRD and MDSR are based on the intersection information from the EENSM regarding the consumption of medicines, information on the drug market from IQVIA and hmR, and finally, information on the drug market required selection and quantification criteria designed by experts.

The following classes of drugs were studied, following the terminology of the Anatomical Therapeutical Chemicals classification (ATC): N6A Antidepressants and Mood Stabilizers, N5A Antipsychotics, N5C Tranquilizers/Anxiolytics, and N5B Hypnotics/Sedatives.

Detailed data were obtained at the level of pills or equivalent since it was not possible to have access to market statistics specifying the quantities based on Defined Daily Doses (DDD). The available data made it necessary to make assumptions about the average consumption for each class of drugs. Average consumption patterns of one, two, or three tablets per day were chosen, depending on the class of medication.

Costs with complementary means of diagnosis and therapeutics

According to the experts, in the context of the usual follow-up, patients with TRD and MDSR tend to do routine tests twice a year. In addition to these consumptions, about 5% of patients undergo thyroid tests. Finally, about 5% of patients are submitted to imaging tests whose main objective is to overlook the possibility of other somatic pathologies. In two-thirds of the cases, the exam is a Computed Axial Tomography (CT) and, in the remaining cases, a Magnetic Resonance Imaging (MRI).

Based on data from NESMH, it is possible to obtain the proportion of patients who had any contact with the health system during the year prior to the study, both in mental health and primary care. According to that information, 26.3% of the patients with TRD had some contact with mental health services, and 39% had some contact with primary health care services. Assuming that the contact probabilities in the two areas are independent, the probability of having at least one contact for patients with TRD is given by 1- (1–0.263) × (1–0.39) = 0.55. Consequently, this result will calibrate the estimates that follow, as it is assumed that only patients in contact with the health system generate consumption of complementary means of diagnosis and therapeutics. Specifically, for the case of patients with TRD (not including joint prevalence with MDSR), the following analysis assumes that the pattern of resource use of complementary means of diagnosis and therapeutics applies to 55% of patients.

In regard to patients with MDSR, estimates based on NESMH indicate that 43.2% will have some contact with mental

health services and 57.7% with primary health care services. As in the previous case, independence of the probabilities of contact is assumed, resulting in an estimate of the percentage of patients with some type of contact with the health system of 76%. Weighting the two percentages by the proportion of patients with TRD only (55.5%) and patients with MDSR (44.5%), an average percentage of patients with contact of 64.4% is obtained. Applying this percentage to the prevalence in 143,163 patients results in 92,147 patients who generate a consumption of complementary means of diagnosis and therapeutics. It should be noted that the estimates presented treat the cases in which the patients have TRD and MDSR together as equivalent to those of the patients with the most severe situation, a methodology already adopted in other parts of this study.

We also considered the routine analyses that patients with TRD and MDSR would do twice a year, on average. A second set of analyses, related to the thyroid test, is carried out annually by about 5% of the patients. In addition, about 5% of patients undergo a CT scan (2/3 of the cases) or an MRI (remaining 1/3) to screen for other pathologies. In 50% of these cases, it is necessary to use contrast, increasing costs.

Costs of emergency department visits

According to the billing rules of the Portuguese NHS, episodes of urgency followed by hospitalization are integrated into hospitalization prices. It is then assumed that the Homogenous Diagnostic Groups (HDG) hospitalization prices are estimates of the overall costs of hospitalization, including the costs of the previous emergencies that generated these hospitalizations. For this reason, the costs of emergency department visits will estimate only the costs of emergency episodes without hospitalization.

Indirect costs

Indirect costs result from the loss of productivity of patients and are defined in the present study as the value of production losses attributed to treatment-resistant depression and major depression with suicide risk. These may include absenteeism as short-term disability, premature exit from the labor market as long-term disability, and productivity lost by premature death.

The sources of information to identify these costs include academic literature, Portuguese databases, observational studies, and surveys conducted in Portugal. Other variables, such as the average wages by sex and age, will be estimated based on data from the 2017 Personnel Tables of the Portuguese Ministry of Labor, Solidarity, and Social Security.

Labor costs, given by gross wages and employers' social security contributions, are the best measure of the productivity of potential workers, following the Human Capital theory. The average salary, by gender and age group, is added by the employer's contribution to Social Security (23.75%). The resulting value is multiplied by 14 to obtain an estimate of annual productivity.

Long-term indirect costs: Effects on employment

The performed analysis takes into account employment until the age of 65 to consider a better approximation to the effective age of leaving the labor market. This effective age reflects that not all workers retire at the official retirement age, given the existence of multiple exceptions: the receipt of disability pensions, early retirements after long-term unemployment, and other situations of an early exit from the labor market.

The employment rates of the population with TRD were approximated by the employment rates of the population with Major Depression, and the employment rates in the population with MDSR were approximated by the employment rates in the group of people with suicidal ideation in the last 12 months. Following the principle of considering people with both types of depression have the most serious disease, joint cases of TRD and MDSR were included in the estimates for TRD, as it exhibits a greater impact on employment rates.

To monetize the lost production due to the lower employment levels, the Human Capital approach was employed, and lost production was approximated by the wage costs that workers would receive.

Short term indirect costs: Absenteeism and presenteeism

To estimate the daily productivity lost due to absenteeism, the average annual salary was divided by 230, corresponding to the number of working days per year, given that absenteeism, by definition, only occurs on these days.

The next step in estimating the indirect costs of absenteeism and presenteeism is to estimate the employment of patients with TRD and MDSR, which is achieved by combining estimates of disease prevalence and employment rates by gender, age group, and disease used in the previous section. For the reasons previously indicated, patients with TRD and patients with TRD and MDSR are linked together.

A viable way to identify the incremental effect of the diseases under analysis on absenteeism was to consider the difference between the days of absenteeism in the population with the diseases under study and the days of absenteeism in the general population.

The methodology adopted to estimate the cost of presenteeism assumed that 1 day of presenteeism has a weight of 0.25 days of absenteeism. There is no single convention on estimating the cost of presenteeism (36). The literature and the sources available do not provide unambiguous estimates. Drummond et al. (37) mention explicitly that "productivity may be lost even though the worker remains at work. This is often called 'presenteeism' and has been argued to be a major proportion of the productivity lost through mood disorders (p. 248)." A reference in that textbook (38) formulates an idea that justifies giving some

TABLE 3 Years lived with disability—global results.

	TRD	MDSR	TRD + MDSR	Total
Prevalence	79,401	52,479	11,283	143,162
YLD	25,228	20,989	4,513	50,730

TRD, Treatment-resistant depression; MDSR, Major depression with suicide risk; YLD, Years lived with disability; rounding to units.

attention to presenteeismpresentism in depression: "The relative importance of presenteeism compared with absenteeism in this disease area is likely because individuals with depression or anxiety tend to stay at work and perform suboptimally rather than take sick leave (p. 1148)."

A review from 2017 (39) shows numerous instruments, surveys and evaluation methodologies. One of this methodologies is based on the conversion of presenteeismpresentism days in proportional reductions in productivity compared with absenteeismabsentism days. If one absenteeismabsentism day is the unit, what should be the fraction to inpute to a presenteeismpresentism day? In some surveys in the literature that fraction can be estimated based on the inneficiency rates self-reported by the patients. However, that type of information was not avaliable in our case. Some contributions in the literature allow us to calculate the ratio of days lost to absenteeismabsentism to equivalent days lost due to presenteeismpresentism [examples Smit et al. (40), in a Dutch context; Uribe et al. (41), in a Colombian context]. However, these papers were heterogeneous in their results, which they reported as total days lost and they were not explicit on the ratio that equalized absenteeism days and presenteeismpresentism days. Using our assumption of 0.25 days of absenteeism per presenteeismpresentism day we obtained conservative results but not totally outside the ballparks of the results in the studies mentioned.

Thus, an estimate is obtained for the total effect of each disease measured by additional equivalent days of absenteeism per year.

Indirect costs of premature suicide mortality

The indirect costs generated by suicide match the current value of all future production that would have been carried out by the deceased if he/she had survived. The updated rate used in this analysis is 4%, as outlined in the current guidelines for conducting health technology assessment studies in Portugal (42). The convention of estimating future values of employment rates and wages by age and gender according to the statistics for 2017 was followed. The population's probabilities of survival are also used to calculate the expected value of future productivity. It is assumed that patients deceased due to suicide would have the survival probabilities given by the 2016–2018 Mortality Tables for men and women in general (43).

Results

Burden of disease

Years lived with disability

Using data on prevalence, disability weights, and the fraction of the year corresponding to the duration of the disease, variables, and parameters presented in methods, the estimates of YLD obtained are shown in Table 3.

The global data of YLD can be broken down by type of depression, sex, and age group. Figure 1 summarizes the detailed information on the years of life lived with disability.

Years of life lost due to premature death

YLL are associated with deaths that are officially designated as "Intentionally self-inflicted injuries and sequelae".

Using a reference mortality table defined by the Global Burden of Disease, almost 31 thousand years of life were prematurely lost due to suicide in 2017, with 73.2% of this total attributed to men. The years of life lost due to suicide (including all ages) constitute 2% of the YLL due to premature mortality, this proportion being 2.5% for men and 1.3% for women.

However, only a fraction of these events is attributed to TRD and MDSR. This topic was studied in the Section Methods, where an attributable fraction of 50.6% was estimated. In total, these estimates are 11,440 YLL for men and 4,187 for women making a total of 15,627 YLL due to premature death attributed to TRD and MDSR.

Disability-adjusted life years

Adding years lived with a disability to the years lost due to premature death attributed to TRD and MDSR, separately and together, we obtain the total DALY generated by the two types of depression under study. The results obtained can be seen in Table 4, where the total burden of the disease of TRD and MDSR is measured by the loss of 66,357 years of life adjusted for disability in 2017.

Although there is an imbalance between men and women in the YLL due to premature death, with men losing 2.7 years for every year lost by women, the opposite is true in the case of YLD (Table 5). Thus, with regards to the results, women have a greater number of YLD than men, corresponding to 58.5% of total years lost.

Despite the enormous burden of premature suicide death disease attributed to the two types of depression understudy, the YLL constitute less than a quarter of the disease burden. The high prevalence of TRD and MDSR, and the high levels of disability that these depressive pathologies instigate, make the YLD amounting to more than three times the years lost due to suicide attributed to the depressive disorders under study.

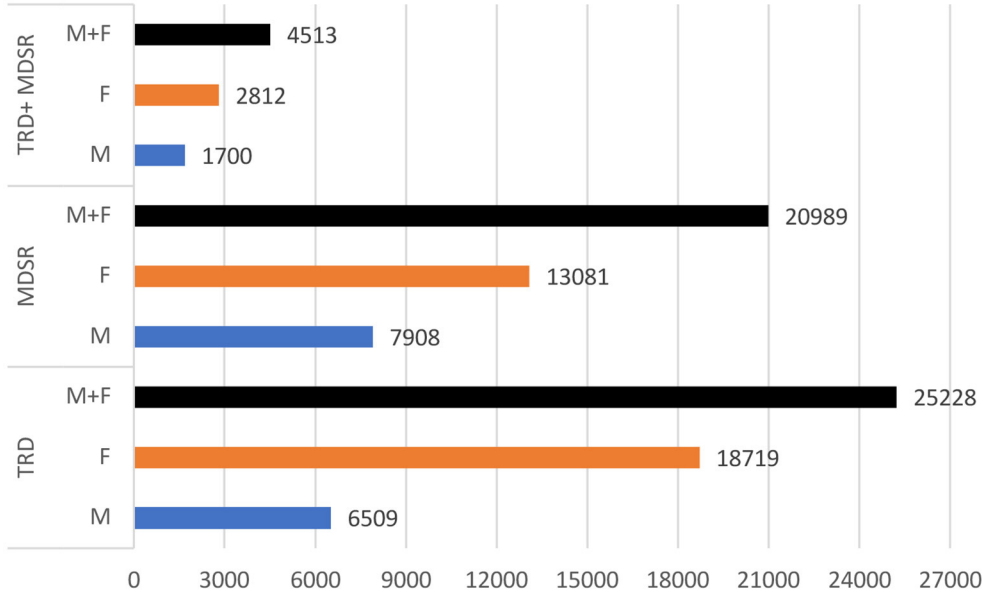


FIGURE 1
Years lived with disability. TRD, Treatment-resistant depression; MDSR, Major depression with suicide risk; YLD, Years lived with disability; M, Male; F, Female; rounding to units.

TABLE 4 Years of life lost due to premature death (YLL).

	Attributable YLL (at YLL)	All suicide YLL	All YLL
M	11,440	22,602	903,374
F	4,187	8,272	664,243
Total	15,627	30,874	1,567,617

YLL, Years of Life Lost; M, Male; F, Female.

TABLE 5 Disability-adjusted life years attributable to TRD and MDSR, 2017.

	YLD	YLL	DALYs	%
M	16,117	11,440	27,557	41.5%
F	34,613	4,187	38,800	58.5%
M+F	50,730	15,627	66,357	
%	76.4%	23.6%		

YLD, Years lived with disability; YLL, Years of Life Lost due to premature death; DALYs, Disability-adjusted life years; TRD, Treatment-resistant depression; MDSR, Major depression with suicide risk; M, Male; F, Female; rounding to units.

Cost of illness

Direct costs

Hospitalization costs

The result of the final analysis led to the identification of 1,696 relevant episodes. A subsequent analysis identified some patients with multiple episodes, so the estimated number of hospitalized patients was 1,502. The total cost of these episodes was estimated at € 3,083,109.

Pharmacological therapy costs

The collective costs of medication for TRD in 2017 were estimated to be € 4,244,921. In 2017, the estimates made pointed to a drug expenditure for the MDSR treatment of around € 6.9 million. Globally, the collective expenditure on drugs for the treatment of TRD and MDSR amounts to € 11 million.

Costs with complementary means of diagnosis and therapy

The overall results regarding the costs of complementary means of diagnosis and therapy indicated an expense of € 4,049,524 in 2017. An estimated separation of these costs with complementary means of diagnosis and therapy indicates that costs for TRD alone would have been € 1,920,662 and € 2,128,862 for MDSR.

Medical appointments costs

The final result obtained is that TRD and MDSR generated 398,511 annual medical visits in 2017. The reference price in Ordinance no. 207/2017, Art. 15, no. 1 is € 31, which applies to ambulatory psychiatry visits and general and family medicine visits, which results in an estimated expense total of € 12,353,851. This amount can be broken down into the cost attributed to TRD (€ 3,913,675) and the cost attributed to MDSR (€ 8,440,176).

TABLE 6 Summary of direct costs.

	Total	(%)
Hospitalizations	3,083,109 €	10.0%
Pharmacological therapy	10,468,895 €	34.0%
Complementary means of diagnosis and therapy	4,049,524 €	13.2%
Medical appointments	12,353,851 €	40.1%
Emergency episodes	824,364 €	2.7%
Total	30,779,743 €	100%

TABLE 7 Costs with presenteeism and absenteeism (€), by disease, age group, and sex.

	TRD + TRD and MDSR		MDSR		Total
Age group	M	F	M	F	
<34	5,989,558	16,112,571	26,180,113	17,500,500	65,782,741
35–49	8,566,711	20,760,965	32,607,277	19,636,170	81,571,123
50–64	4,909,932	10,531,366	11,417,642	6,085,475	32,944,415
Total	19,466,201	47,404,902	70,205,032	43,222,144	180,298,279

TRD, Treatment-Resistant Depression; MDSR, Major Depression with Suicide Risk; M, Male; F, Female; rounding to units; results presented in € (euro).

Emergency episodes costs

An analysis of the 1,696 inpatient episodes studied in the previous sections showed that 799 of these episodes occurred at the emergency room. It is estimated that the number of emergency episodes due to TRD and MDSR in 2017, without hospitalization, was $N = 799 * (100\text{-}7.94\%)/7.94\% = 9{,}264$. The average price obtained for an emergency in 2017 was € 88.94. Multiplying this value by the estimated number of emergencies without hospitalization results in a collective cost of € 824,364.

Total direct costs

The global results obtained are summarized in Table 6. The global direct costs of TRD and MDSR to health system is 31 million euros. Medical visits (40% of the expenditure calculated) and pharmacological therapy (34%) are the components with the most relevant costs, while the remaining components of direct costs having a substantially less weight.

Indirect costs

Long-term indirect costs—Effects on employment

The overall estimate for the costs of the lowest employment rate in the population with Treatment-Resistant Depression or Major Depression with Suicide Risk is € 834,786,764

Short term indirect costs—Absenteeism and presenteeism

Human Capital methodology was used, with daily labor costs to obtain the costs of absenteeism and presenteeism from TRD and MDSR. The global estimate of the costs of absenteeism and incremental presenteeism generated by TRD and MDSR is shown in Table 7. An estimated € 180.3 million is divided into approximately equal parts between men and women. The 35–49 years age group generates a greater fraction of the costs than the other age groups.

TABLE 8 Direct and Indirect costs (€) attributable to TRD and MDSR.

Direct costs	
Hospitalizations	€ 308,310,900
Pharmacological therapy	€ 1,046,889,500
Complementary means of diagnosis and therapy	€ 404,952,400
Medical appointments	€ 1,235,385,100
Emergency episodes	€ 82,436,400
Total direct costs	€ 3,077,974,300
Indirect costs	
Absenteeism and presenteeism	€ 18,029,827,900
Employment reduction	€ 83,478,676,400
Premature death	€ 5,660,441,500
Total indirect costs	€ 107,168,945,800
Total costs	€ 110,246,920,100

TRD, Treatment-Resistant Depression; MDSR, Major Depression with Suicide Risk; rounding to units; results presented in € (euro).

Indirect costs of premature mortality due to premature death

Total indirect costs due to premature death attributed to TRD and MDSR is estimated at € 56,604,415. Eighty-one percent of this total (€ 45,615,760) is attributed to men and 19% (€ 10,988,655) to women.

Table 8 shows total Direct and Indirect Costs Attributed to TRD and MDSR: The total indirect costs related to TRD and MDSR collectively reached € 1.1 billion, with men accounting for 36.7% of this cost. According to the type of costs, absenteeism/presenteeism was found to be responsible for 16.8% of the total costs, while the reduction of employment and the costs of premature mortality were responsible for 77.9 and 5.3% of the total indirect costs, respectively.

Finally, it should be noted that the direct costs supported by the health system are very small when compared to indirect costs. The direct costs calculated are only 2.7% of the total costs, that is, the sum of all types of estimated costs.

Discussion

This study focused on estimating the years of life lost attributed to treatment-resistant depression and major

depression with suicide risk, the burden of the disease, and the direct and indirect costs of these diseases. These are the traditional dimensions of disease burden studies and cost of illness studies. The estimated values reveal the colossal negative impact that treatment-resistant depression and major depression with suicide risk have on health and economic resources.

The prevalence of TRD only, MDSR only, and the combined prevalence of the two types of depression were estimated at 79.4 thousand, 52.5 thousand and 11.3 thousand patients, respectively. The disease burden (DALY) due to the disability generated by TRD alone, by MDSR alone, and by the joint prevalence was 25.2 thousand, 21 thousand, and 4.5 thousand, respectively, totaling 50.7 thousand DALY. The disease burden due to premature death by suicide, attributed to TRD and MDSR, was 15.6 thousand DALY. The estimated total disease burden was 66.3 thousand DALY. This figure can be compared with estimates available for other diseases in Portugal. Henriques et al. (44) estimated that ischemic heart disease in Portugal generated 95,413 DALY, which means that TRD and MDSR are responsible for a disease burden that represents ∼70% of the ischemic heart disease burden. On the other hand, Gouveia et al. (45) estimated that heart failure in Portugal carried a burden of 21,162 DALY, less than half of the estimated disease burden for TRD and MDSR.

Direct costs of TRD and MDSR were estimated at € 30.8 million, with the most important components being consultations and medication. The estimated indirect costs are much higher than the direct costs. Adding the productivity losses due to the reduction in the level of employment, absenteeism and presenteeism, and the productivity lost due to premature death, a total cost of € 1 billion was calculated. A possible comparison term is given by the costs of asthma in adults in Portugal, estimated by Barbosa et al. (46) at € 386.3 million, 93% of which are direct costs. The direct costs of asthma would thus be almost 12 times higher than those of TRD and MDSR, but the total costs of asthma would only be about 35% of the costs of TRD and MDSR, showing the great indirect costs that these pathologies generate.

The basis for this study was the National Epidemiological Study of Mental Health, that was performed between 2008 and 2009. Despite the NESMH's quality, the lack of more recent data is a limitation, and thus an extrapolation of results to the 2017 population was performed.

Another limitation associated with NESMH is that in this study, the diagnoses were not validated as they were not made by clinicians. However, psychiatric disorders were assessed through comprehensive and fully structured interviews designed by the World Mental Health Surveys Initiative, the World Health Organization, and the Harvard University.

The results obtained did not estimate all consequences of TRD and MDSR on the wellbeing of the Portuguese adult population. A limitation of the present study is that it was not possible to estimate hospitalizations and emergency episodes in private hospitals, as no source of information similar to the Hospital Morbidity Database was available to researchers.

Conclusions

In addition to the years of life lost and the direct and indirect costs, treatment-resistant depression and major depression with suicide risk have very negative effects in various dimensions relevant to the wellbeing and health of the affected population. These additional dimensions include the impact of depression on the educational outcome, the formation and stability of marital unions, fertility, or even on the quality of parental care (11, 47–49). In the economic area, in addition to the estimated effects on the labor market, there are indications of impact on other areas of financial performance of affected individuals (11).

Although TRD and MDSR represent relatively small direct costs to the health system, they have a significant disease burden and productivity costs on the Portuguese economy and society that are highly relevant, making TRD and MDSR priority areas for obtaining health gains.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the data that support the findings of this study are available from World Mental Health Survey Initiative and Health Ministry, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission at www.hcp.med.harvard.edu/wmh and at www.acss.min-saude.pt.

Ethics statement

This research did not use human participants and no patient data was presented. National Epidemiological Study of Mental Health was conducted according to all ethical requirements and its ethical procedures are described elsewhere.

Author contributions

MG, AR, and HC designed the study. MG, RS, AR, GC, AA, HC, and JA contributed with scientific knowledge and data, developed the study, and interpreted the data. MG carried out data analysis. RS written the manuscript. MG, CN-d-S, JA, GC, and AA revised the manuscript.

All authors contributed to the article and approved the submitted version.

Funding

The present work received support from Janssen-Cilag Farmacêutica, Lda, as an Unrestricted Grant. Janssen-Cilag had no role in the execution of the study, including analyses and interpretation of results.

Acknowledgments

We would like to thank Ana Matos Pires for her inputs as an expert in psychiatry to debate psychiatric services, and routes to treatment and care in Portugal.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. World Health Organization. *Depression and Other Common Mental Disorders: Global Health Estimates*. (2017). Available online at: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf (accessed May 03, 2021).

2. de Sousa JSP. *Epidemiologia, etiopatogenia, diagnóstico e tratamento farmacológico da depressão em Portugal*. Porto: Universidade Fernando Pessoa (2015).

3. Castelpietra G, Knudsen AK, Agardh EE, Armocida B, Beghi M, Iburg KM, et al. The burden of mental disorders, substance use disorders and self-harm among young people in Europe, 1990–2019: findings from the global burden of disease. *Lancet Regional Health Europe*. (2019) 16:100341. doi: 10.1016/j.lanepe.2022.100341

4. Caldas-de-Almeida J, Xavier M, Cardoso G, Gonçalves Pereira M, Gusmão R, Corrêa B. *Estudo epidemiológico nacional de saúde mental: 1º relatório* [National Mental Health Epidemiological Study - 1st Report]. (2013).

5. Little A. Treatment-resistant depression. *Am Fam Physician*. (2009) 80:167–72.

6. Gaynes BN, Lux L, Gartlehner G, Asher G, Forman-Hoffman V, Green J, et al. Defining treatment-resistant depression. *Depress Anxiety*. (2020) 37:134–45. doi: 10.1002/da.22968

7. Fava M, Davidson KG. Definition and epidemiology of treatment-resistant depression. *Psychiatr Clin North Am*. (1996) 19:179–200. doi: 10.1016/S0193-953X(05)70283-5

8. Fava M. Diagnosis and definition of treatment-resistant depression. *Biol Psychiatry*. (2003) 53:649–59. doi: 10.1016/S0006-3223(03)00231-2

9. Nierenberg AA, McLean NE, Alpert JE, Worthington JJ, Rosenbaum JF, Fava M. Early nonresponse to fluoxetine as a predictor of poor 8-week outcome. *Am J Psychiatry*. (1995) 152:1500–3. doi: 10.1176/ajp.152.10.1500

10. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry*. (2018) 75:336–46. doi: 10.1001/jamapsychiatry.2017.4602

11. Kessler RC. The costs of depression. *Psychiatr Clin North Am*. (2012) 35:1–14. doi: 10.1016/j.psc.2011.11.005

12. Rush AJ, Marangell LB, Sackeim HA, George MS, Brannan SK, Davis SM, et al. Vagus nerve stimulation for treatment-resistant depression: a randomized, controlled acute phase trial. *Biol Psychiatry*. (2005) 58:347–54. doi: 10.1016/j.biopsych.2005.05.025

13. Kubitz N, Mehra M, Potluri RC, Garg N, Cossrow N. Characterization of treatment resistant depression episodes in a cohort of patients from a US commercial claims database. *PLoS ONE*. (2013) 8:e76882. doi: 10.1371/journal.pone.0076882

14. Kolar D. Addictive potential of novel treatments for refractory depression and anxiety. *Neuropsychiatr Dis Treat*. (2018) 14:1513–9. doi: 10.2147/NDT.S167538

15. Wiles N, Thomas L, Abel A, Barnes M, Carroll F, Ridgway N, et al. Clinical effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: the CoBalT randomised controlled trial. *Health Technol Assess*. (2014) 18:1–167, vii–viii. doi: 10.3310/hta18310

16. Xavier M, Baptista H, Mendes JM, Magalhães P, Caldas-de-almeida JM. Implementing the World Mental Health Survey Initiative in Portugal – rationale, design and fieldwork procedures. *Int J Ment Health Syst*. (2013) 7:1–10. doi: 10.1186/1752-4458-7-19

17. Bretlau LG, Lunde M, Lindberg L, Undén M, Dissing S, Bech P. Repetitive transcranial magnetic stimulation (rTMS) in combination with escitalopram in patients with treatment-resistant major depression: a double-blind, randomised, sham-controlled trial. *Pharmacopsychiatry*. (2008) 41:41–7. doi: 10.1055/s-2007-993210

18. O'Reardon JP, Solvason HB, Janicak PG, Sampson S, Isenberg KE, Nahas Z, et al. Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: a multisite randomized controlled trial. *Biol Psychiatry*. (2007) 62:1208–16. doi: 10.1016/j.biopsych.2007.01.018

19. Thase ME, Friedman ES, Biggs MM, Wisniewski SR, Trivedi MH, Luther JF, et al. Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: a STAR*D report. *Am J Psychiatry*. (2007) 164:739–52. doi: 10.1176/ajp.2007.164.5.739

20. UK ECT Review Group. Efficacy and safety of electroconvulsive therapy in depressive disorders: a systematic review and meta-analysis. *Lancet*. (2003) 361:799–808. doi: 10.1016/S0140-6736(03)12705-5

21. Roca M, Del Amo AR-L, Riera-Serra P, Pérez-Ara MA, Castro A, Roman Juan J, et al. Suicidal risk and executive functions in major depressive disorder: a study protocol. *BMC Psychiatry*. (2019) 19:253. doi: 10.1186/s12888-019-2233-1

22. Alonso J, Petukhova M, Vilagut G, Chatterji S, Heeringa S, Üstün TB, et al. Days out of role due to common physical and mental conditions: results from the WHO World Mental Health surveys. *Mol Psychiatry*. (2011) 16:1234–46. doi: 10.1038/mp.2010.101

23. Alonso J, Liu Z, Evans-Lacko S, Sadikova E, Sampson N, Chatterji S, et al. Treatment gap for anxiety disorders is global: results of the World Mental Health Surveys in 21 countries. *Depress Anxiety*. (2018) 35:195–208. doi: 10.1002/da.22711

24. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. *Br J Psychiatry*. (2017) 210:119–24. doi: 10.1192/bjp.bp.116.188078

25. Antunes A, Frasquilho D, Azeredo-Lopes S, Neto D, Silva M, Cardoso G, et al. Disability and common mental disorders: results from the World Mental Health Survey Initiative Portugal. *Eur Psychiatry.* (2018) 49:56–61. doi: 10.1016/j.eurpsy.2017.12.004

26. Cardoso G, Xavier M, Vilagut G, Petukhova M, Alonso J, Kessler RC, et al. Days out of role due to common physical and mental conditions in Portugal: results from the WHO World Mental Health Survey. *BJPsych Open.* (2017) 3:15–21. doi: 10.1192/bjpo.bp.115.002402

27. Cardoso G, Antunes A, Silva M, Azeredo-Lopes S, Xavier M, Koenen K, et al. Trauma exposure and ptsd in portugal: findings from the world mental health survey initiative. *Psychiatry Res.* (2020) 284:112644. doi: 10.1016/j.psychres.2019.112644

28. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., Text Revision). Washington, DC: American Psychiatric Association (2000).

29. World Health Organization. *ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision*, 2nd ed. Geneva: World Health Organization (2004).

30. Spijker J, de Graaf R, Bijl R V, Beekman ATF, Ormel J, Nolen WA. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Br J Psychiatry.* (2002) 181:208–13. doi: 10.1192/bjp.181.3.208

31. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* (2013) 10:e1001547. doi: 10.1371/journal.pmed.1001547

32. World Health Organization Regional Office for Europe. *European Detailed Mortality Database 2017.* (2017). Available online at: https://www.euro.who.int/en/data-and-evidence/archive/mortality-database-updated (accessed May 03, 2021).

33. Murray CJL. Quantifying the burden of disease : the technical basis for disability-adjusted life years. *Bull World Heal Organ.* (1994) 72:429–45.

34. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health.* (1998) 88:15–9. doi: 10.2105/AJPH.88.1.15

35. Salomon JA, Haagsma JA, Davis A, de Noordhout CM, Polinder S, Havelaar AH, et al. Disability weights for the Global Burden of Disease 2013 study. *Lancet Glob Heal.* (2015) 3:e712–23. doi: 10.1016/S2214-109X(15)00069-8

36. Mattke S, Balakrishnan A, Bergamo G, Newberry SJ. A review of methods to measure health-related productivity loss. *Am J Manag Care.* (2007) 13:211–7.

37. Despiégel N, Danchenko N, François C, Lensberg B, Drummond M. The use and performance of productivity scales to evaluate presenteeism in mood disorders. *Value Health.* (2012) 15:1148–61. doi: 10.1016/j.jval.2012.08.2206

38. Drummond M, Sculpher M, Torrance G, O'Brien, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*, 4th ed. Oxford: Oxford University Press (2015).

39. Kigozi J, Jowett S, Lewis M, Barton P, Coast J. The estimation and inclusion of presenteeism costs in applied economic evaluation: a systematic review. *Value in Health.* (2017) 20:496–506. doi: 10.1016/j.jval.2016.12.006

40. Smit F, Willemse G, Koopmanschap M, Onrust S, Cuijpers P, Beekman A. Cost-effectiveness of preventing depression in primary care patients. *Br J Psychiatry.* (2006) 188:330–6. doi: 10.1192/bjp.188.4.330

41. Uribe JM, Pinto DM, Vecino-Ortiz AI, Gómez-Restrepo C, Rondón M. Presenteeism, absenteeism, and lost work productivity among depressive patients from five cities of Colombia. *Value Health Regional Issues.* (2017) 14:15–9. doi: 10.1016/j.vhri.2017.03.001

42. Perelman J, Soares M, Mateus C, Duarte A, Faria R, Ferreira L, et al. Methodological guidelines for economic evaluation studies. *INFARMED – National Authority of Medicines and Health Products*, Lisbon: I.P. (2019). Available online at: www.infarmed.pt

43. Statistics Portugal (2019). Available online at: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=354096866&DESTAQUESmodo=2 (accessed May 03, 2021).

44. Henriques A, Araújo C, Viana M, Laszczynska O, Pereira M, Bennett K, et al. Disability-adjusted life years lost due to ischemic heart disease in mainland Portugal, 2013. *Rev Port Cardiol.* (2017) 36:273–81. doi: 10.1016/j.repc.2016.08.011

45. Gouveia M, Ascenção R, Fiorentino F, Costa J, Caldeira D, Broeiro-Gonçalves P, et al. The current and future burden of heart failure in Portugal. *ESC Heart Fail.* (2019) 6:254–61. doi: 10.1002/ehf2.12399

46. Barbosa JP, Ferreira-Magalhães M, Sá-Sousa A, Azevedo LF, Fonseca JA. Cost of asthma in Portuguese adults: a population-based, cost-of-illness study. *Rev Port Pneumol.* (2017) 23:323–30. doi: 10.1016/j.rppnen.2017.07.003

47. Butterworth P, Rodgers B. Mental health problems and marital disruption: is it the combination of husbands and wives' mental health problems that predicts later divorce? *Soc Psychiatry Psychiatr Epidemiol.* (2008) 43:758–63. doi: 10.1007/s00127-008-0366-5

48. Kessler RC, Walters EE, Forthofer MS. The social consequences of psychiatric disorders, III: probability of marital stability. *Am J Psychiatry.* (1998) 155:1092–6. doi: 10.1176/ajp.155.8.1092

49. Kessler RC, Foster CL, Saunders WB, Stang PE. Social consequences of psychiatric disorders, I: educational attainment. *Am J Psychiatry.* (1995) 152:1026–32. doi: 10.1176/ajp.152.7.1026

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fpubh.2022.987408

OPEN ACCESS

EDITED BY
Huajie Jin,
King's College London,
United Kingdom

REVIEWED BY
Kai G. Kahl,
Hannover Medical School, Germany
Takefumi Suzuki,
University of Yamanashi, Japan

*CORRESPONDENCE
Jing Wu
jingwu@tju.edu.cn

SPECIALTY SECTION
This article was submitted to Public Mental Health, a section of the journal Frontiers in Public Health

RECEIVED 06 July 2022
ACCEPTED 30 August 2022
PUBLISHED 15 September 2022

CITATION
Liu J, Cao L and Wu J (2022) Cost-utility analysis of lurasidone for the first-line treatment of schizophrenia in China. *Front. Public Health* 10:987408. doi: 10.3389/fpubh.2022.987408

Cost-utility analysis of lurasidone for the first-line treatment of schizophrenia in China

Jia Liu[1,2], Lidan Cao[1,2] and Jing Wu[1,2]*

[1]School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, China, [2]Center for Social Science Survey and Data, Tianjin University, Tianjin, China

Objective: To evaluate the cost-effectiveness of lurasidone compared with olanzapine and risperidone in the first-line treatment of patients with schizophrenia from a Chinese healthcare system perspective.

Methods: A Markov model with 6-week cycle was constructed to reflect the disease progression of schizophrenia patients in the acute and maintenance phase. Probabilities of treatment discontinuation and adverse events in the acute phase were derived from the 6-week lurasidone clinical trial and a published network meta-analysis; long-term risks of relapse and discontinuation were estimated based on the 12-month lurasidone clinical trial and other treatment comparison studies. Cost inputs were derived from published literature and Chinese official documents, supplemented by expert opinions when necessary. Utility values were taken from published literature. Costs and quality-adjusted life-years (QALYs) were assessed over 15 years with a discount rate of 5% per year.

Results: Over a 15-year time horizon, lurasidone yielded an improvement of 0.197 QALYs with a cost saving of CN¥12,093 (US$1,753) vs. olanzapine and an improvement of 0.116 QALYs with a cost saving of CN¥6,781 (US$983) vs. risperidone. One-way sensitivity analyses demonstrated robust base-case results since all analyses yielded net monetary benefits >0 at a willingness-to-pay threshold of CN¥72,447.00 (US$10,499.57)/QALY. Probabilistic sensitivity analyses suggested that lurasidone had 99.7, 99.9, and 100% probability of being cost-effective vs. olanzapine and risperidone at the conventional decision thresholds of 1, 2, and 3 times the Chinese per capita gross domestic product [namely CN¥72,447.00 (US$10,499.57)/QALY, CN¥1,44,894.00 (US$20,999.13)/QALY, and CN¥2,17,341.00 (US$31,498.70)/QALY in 2020], respectively.

Conclusion: Treatment with lurasidone was predicted to improve health outcomes and be a dominant strategy for patients with schizophrenia, compared with olanzapine and risperidone, in China.

KEYWORDS

cost-utility, lurasidone, olanzapine, risperidone, schizophrenia

Introduction

Schizophrenia is a chronic and severely debilitating mental disorder with unknown etiology, which is characterized by high morbidity, high recurrence rate, high disability rate and heavy socio-economic disease burden (1). This disease affected approximately 23.6 million people worldwide and generated a humanity burden with a total of 15.1 million years lived with disability (YLDs) in 2019 (2). A systematic review informed that annual costs for the schizophrenia population were estimated to be varied between US$94 million (Puerto Rico) and US$102 billion (US), and indirect costs accounted for more than 50% of the total costs (3). In China, according to a recent national epidemiological investigation for mental disorders, the lifetime prevalence of schizophrenia was estimated at 0.6% (about 8.4 million people) (4). The YLDs caused by schizophrenia accounted for 2.35% of total YLDs in China in 2019 (5). A questionnaire-based investigation showed that the annual costs per case of schizophrenia in China amounted to US$2,586.21, which could be seen as a significant economic burden for chronic schizophrenic patients and their families (6).

Antipsychotics are the mainstay of pharmacological treatment for schizophrenia patients to alleviate psychotic symptoms and improve prognosis. First-generation antipsychotics (FGAs), such as chlorpromazine and haloperidol, have been shown to be effective; but their adverse effects, such as extrapyramidal symptoms (EPS) and tardive dyskinesia in some cases, often limit long-term adherence (7). Second-generation antipsychotics (SGAs), including clozapine, risperidone, olanzapine and aripiprazole, have been recommended as first-line treatment by national guidelines (8, 9) as having equal or better efficacy, and lower risk of EPS and tardive dyskinesia comparing to FGAs. However, SGAs have also been demonstrated to be associated with an increased risk of weight gain and other metabolic abnormalities (10–12), which frequently lead to discontinuation and/or cycling between different therapies (13–16).

Lurasidone, a new SGA, was approved by China National Medical Products Administration (NMPA) for the treatment of schizophrenia in January 2019. To date, several multicenter, double-blind, phase III studies have demonstrated that lurasidone was associated with significant improvements in symptom reduction and minimal changes in weight, body mass index, and metabolic outcomes vs. placebo and quetiapine (17–20). Moreover, indirect comparison studies evaluating the efficacy and safety profile of atypical antipsychotics indicated that lurasidone was associated with significant improvements in terms of weight gain, metabolic outcomes, relapse rates, hospitalizations, and rates of all-cause discontinuation compared with olanzapine, risperidone, and aripiprazole (21, 22).

Although the clinical effectiveness of lurasidone in the treatment of schizophrenia has been demonstrated, the cost-effectiveness of lurasidone vs. alternative therapies remains to be established. This study aims to assess the cost-effectiveness of lurasidone compared with olanzapine and risperidone, which are the most prescribed SGAs in China and have been incorporated into the National Reimbursement Drug List (NRDL), in patients with schizophrenia from a Chinese healthcare system perspective. Considering that this study was the first to evaluate the cost-effectiveness of lurasidone for the treatment of schizophrenia after the drug pricing negotiation conducted by the National Healthcare Security Administration (NHSA) in 2020 (the latest negotiated price was used), the results may help inform updated clinical decisions related to schizophrenia in China.

Materials and methods

Model overview

In this study, a Markov model was constructed to simulate costs and health outcomes in a hypothetical cohort of patients with schizophrenia, which had been previously developed to compare the cost-effectiveness of lurasidone with aripiprazole for the treatment of schizophrenia in the Scotland and Wales setting (23) (Figure 1). Given the chronic nature of schizophrenia, a 15-year horizon was used in the model as it was considered sufficient and recommended by Chinese clinicians to assess the long-term impact of treatment. A model cycle length of 6-week was used to reflect the clinically meaningful amount of time for the progression of schizophrenia and align with the short-term clinical trial design of lurasidone (17).

The model consisted of five health states: (1) non-stable/Rx trial, (2) stable/adherent, (3) stable/non-adherent, (4) relapse, and (5) death. Patients entered the model in the 'non-stable/Rx trial' health state (an acute phase of relapse undergoing trials of antipsychotic agents). After 6 weeks, patients who have not discontinued treatment were assumed to enter the 'stable/adherent' health state (the maintenance phase), while those who have discontinued treatment for any reason were assumed to switch therapy and re-enter the 'non-stable/Rx trial' health state to continue the process of trialing alternative antipsychotic agents. Patients in the 'stable/adherent' health state in the maintenance phase were further subject to risks of all-cause discontinuation and relapse. Patients discontinuing treatment in the maintenance phase were assumed to receive no therapy, and reside in the 'stable/non-adherent' health state until the onset of relapse, at which point they enter the 'relapse' health state. Patients who relapse were assumed to discontinue current therapy and switch to the next therapy, and most relapse patients were hospitalized. Patients may also die from any health state within the model.

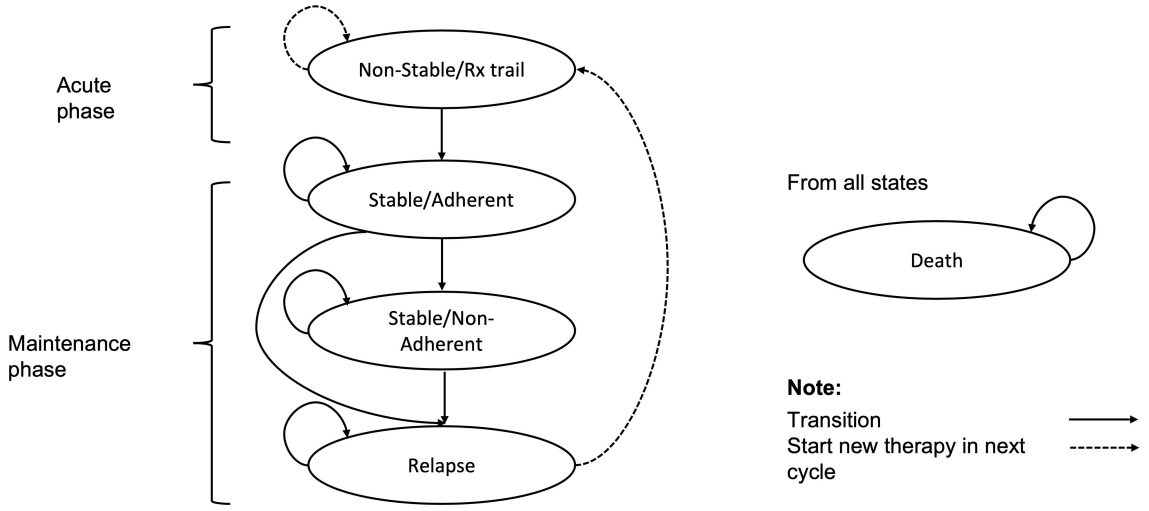


FIGURE 1
Markov model structure.

This analysis was conducted from a Chinese healthcare system perspective. Costs and health outcomes were discounted at a rate of 5% per year in accordance with the recommendation of the China Guidelines for Pharmacoeconomic Evaluations (2020) (24).

Patients and treatment sequences

The population for the model included adult patients diagnosed with schizophrenia. Patient characteristics were specified to reflect the average schizophrenia patient enrolled in the 6-week, randomized, double-blind, placebo- and active-controlled lurasidone clinical trial (17): 68.3% of patients were male, with age of 37.2 years old and weight of 74.5 kg.

The model compared three alternative treatment sequences. Based on the Chinese guidelines for the prevention and treatment of schizophrenia and the opinions of Chinese clinical experts, simplified treatment sequences were constructed. The first strategy consisted of lurasidone, followed by aripiprazole, clozapine and, finally, an augmented clozapine strategy (clozapine combined with risperidone). The subsequent treatment sequences of olanzapine and risperidone strategies were the same as lurasidone.

Clinical inputs

A 2019 published systematic review and network meta-analysis (NMA) of 32 oral antipsychotics (22), including lurasidone, olanzapine, and risperidone vs. placebo, was used to inform estimates of short-term efficacy (probability of all-cause discontinuation) in the acute phase. The probability of all-cause discontinuation of placebo arm was derived from the 6-week lurasidone clinical trial (17). The published NMA did not report long-term clinical outcomes, and no other comparative clinical data were available for lurasidone vs. olanzapine and risperidone. Therefore, for the maintenance phase of the model, long-term risks of relapse and all-cause discontinuation for lurasidone were taken from a 12-month, randomized, double-blind, active-controlled study vs. quetiapine (20). To inform the olanzapine and risperidone data, the quetiapine arm of the lurasidone trial was used as the common comparator of the indirect comparison, with hazard ratios (HR) for risks of treatment discontinuation taken from a published observational study (25) and for risks of relapse taken from a mixed treatment comparison conducted by National Institute for Health and Care Excellence (NICE) (26).

Additionally, in the acute phase, patients cycled through a variety of treatment regimens until they reached a stable disease state. The efficacy data of subsequent therapies (aripiprazole, clozapine, and augmented clozapine) were taken from the published NMA (22). Data for augmented clozapine were assumed to equal the data for clozapine. For the maintenance phase, the HR of aripiprazole vs. quetiapine on the risk of discontinuation was taken from the published observational study (25), and the HR of aripiprazole vs. quetiapine on the risk of relapse was taken from the NICE mixed treatment comparison (26). In the absence of data, the risk of relapse and discontinuation of clozapine and augmented clozapine were assumed to be equal to quetiapine in the maintenance phase. The proportion of relapse attributed to adherent patients was derived from a Chinese real-world study (27).

With respect to the safety data, weight gain (defined as a ≥7% change in weight from baseline), EPS and diabetes were taken into account in this study. Incidences of weight gain and EPS for diverse antipsychotics were derived from the short-term lurasidone clinical trial (the placebo arm was used) (17) and the published NMA (22). The incidence of diabetes was estimated based on a cost-effectiveness analysis of eleven antipsychotics in Singapore (the olanzapine arm was used) (28) and an economic evaluation conducted by NICE (the relative effect of developing diabetes was assumed to be equal to the relative effect of experiencing weight gain) (26).

Mortality was based on the Chinese life table of the general population (29) and adjusted by the standard mortality rates of Chinese schizophrenia patients (30). A summary of model clinical inputs is provided in Table 1.

Costs and resource utilization inputs

From the perspective of Chinese healthcare system, resource use included drug acquisition, schizophrenia related outpatient visits, schizophrenia related inpatient visits, and adverse events (AEs) treatment.

Due to the lack of data, a face-to-face survey of clinical experts was conducted to understand the healthcare resource utilization related to standard schizophrenia treatment and AEs treatment. To be eligible, clinical experts had to be working in tertiary hospital (where the majority of schizophrenia patients are treated), have more than 5 years of practical experience and be providing treatments for individual patients. A total of 5 clinical experts were selected, one each from Shenyang, Beijing, Chengdu, Shanghai, and Changsha. This was done to consider the different geographic areas and economic development in China.

The unit costs of antipsychotics were the most recent average bidding prices in all available provinces in China, which could be queried through the Chinese open-source Yaozh website (31). The daily dosages were consistent with the instructions of each drug (Table 2). According to the expert survey, patients with either non-stable disease, stable disease or relapse disease were required to take regular outpatient visits including tests for liver function, kidney function, blood routine, blood biochemistry, electrocardiogram, etc. The unit cost of those healthcare resources was acquired from the governmental publications in five cities where the clinical experts come from (32–36). Patients who experienced non-stable disease, stable disease and relapse disease were 52.0, 0.6, and 41.0% possible to be hospitalized, and the average hospital stay of those patients was 26.4, 3.8, and 31.0 days, respectively. Inpatient costs were then calculated via average inpatient days multiplied by inpatient daily cost, which could be found in Table 2. The treatment costs of AEs, including weight gain and EPS, were estimated by the expert survey. Specifically, the use of healthcare resources was described by clinicians, and the unit price of those healthcare resources was obtained from the governmental publications (32–36). The average annual cost for diabetes treatment was derived from a multicenter, prospective cohort study in China (37), and adjusted to 6-week cost to fit the model cycle length. All costs were expressed in 2020 Chinese Yuan (CN¥) and US$ [average exchange rate in 2020: US$1 = CN¥6.90 (38)]. A summary of the cost data in the model is presented in Table 2.

Utility inputs

Utility values of schizophrenia states and utility decrements associated with AEs were mainly obtained from a direct utility elicitation study (39). The specific utility and disutility values adopted in the model are shown in Table 3.

Base-case and sensitivity analyses

In the base-case analysis, total costs, and total numbers of quality-adjusted life years (QALYs) associated with lurasidone, olanzapine and risperidone over 15 years were estimated. Incremental cost-effectiveness ratios (ICERs) were also calculated, presented as incremental cost per QALY gained. Conventionally, the willingness-to-pay threshold was 1–3 times of Chinese per capita gross domestic product (GDP), namely CN¥72,447.00 (US$10,499.57)–CN¥2,17,341.00 (US$31,498.70) in 2020 (38).

Robustness of the results of this analysis was tested by one-way sensitivity analyses (OWSA) and probabilistic sensitivity analyses (PSA). In OWSA, the discount rates for costs and health outcomes were varied between 0 and 8% per annum (24), while other key parameters were varied by 95% confidence intervals or ±25% of the base-case values (when confidence intervals were not available). The net monetary benefit (NMB), assuming the willingness-to-pay threshold of CN¥72,447.00 (US$10,499.57) per QALY (one time of Chinese per capita GDP), was calculated at the upper and lower parameter values and was used to plot a tornado diagram. Monte Carlo simulation was used to conduct the PSA. All key parameters were assigned distributions and varied simultaneously over 5,000 iterations. The results of PSA were plotted on a cost-effectiveness acceptability curve. The specific values of parameters used in OWSA, and parameter distributions used in PSA are presented in Tables 1–3.

Results

Base-case analysis

Table 4 presents the results of the base-case analysis. Compared with olanzapine and risperidone, lurasidone was

TABLE 1 Summary of clinical data used in the model.

Variable	Base-case value	OWSA		PSA distribution	Source
		Lower value	Upper value		
All-cause discontinuation in non-stable/Rx trial state					
Placebo	39.34%	NA[a]	NA[a]	NA[a]	(17)
Lurasidone (RR vs. placebo)	0.88	0.80	0.96	Log-normal	(22)
Olanzapine (RR vs. placebo)	0.69	0.65	0.74	Log-normal	(22)
Risperidone (RR vs. placebo)	0.83	0.80	0.85	Log-normal	(22)
Aripiprazole (RR vs. placebo)	0.80	0.73	0.86	Log-normal	(22)
Clozapine (RR vs. placebo)	0.75	0.59	0.91	Log-normal	(22)
Augmented clozapine (RR vs. placebo)	0.75	0.59	0.91	Log-normal	Assumption
All-cause discontinuation in stable/adherent state					
Quetiapine	Weibull	NA[a]	NA[a]	NA[a]	(20)
Lurasidone (HR vs. quetiapine)	0.72	0.52	1.02	Log-normal	(20)
Olanzapine (HR vs. quetiapine)	0.74	0.55	0.92	Log-normal	(25)
Risperidone (HR vs. quetiapine)	1.16	0.87	1.45	Log-normal	(25)
Aripiprazole (HR vs. quetiapine)	0.87	0.65	1.09	Log-normal	(25)
Clozapine (HR vs. quetiapine)	1.00	0.75	1.25	Log-normal	Assumption
Augmented clozapine (HR vs. quetiapine)	1.00	0.75	1.25	Log-normal	Assumption
Relapse in stable state					
Quetiapine	Gompertz	NA[a]	NA[a]	NA[a]	(20)
Lurasidone (HR vs. quetiapine)	0.70	0.39	1.24	Log-normal	(20)
Olanzapine (HR vs. quetiapine)	0.69	0.52	0.87	Log-normal	(26)
Risperidone (HR vs. quetiapine)	1.00	0.75	1.25	Log-normal	(26)
Aripiprazole (HR vs. quetiapine)	0.99	0.75	1.24	Log-normal	(26)
Clozapine (HR vs. quetiapine)	1.00	0.75	1.25	Log-normal	Assumption
Augmented clozapine (HR vs. quetiapine)	1.00	0.75	1.25	Log-normal	Assumption
Proportion of relapse from adherent patients	38.20%	28.65%	47.75%	Beta	(27)
AE of weight gain					
Placebo	3.29%	NA[a]	NA[a]	NA[a]	(17)
Lurasidone (RR vs. placebo)	1.29	0.97	1.61	Log-normal	(22)
Olanzapine (RR vs. placebo)	6.10	4.58	7.63	Log-normal	(22)
Risperidone (RR vs. placebo)	2.83	2.12	3.54	Log-normal	(22)
Aripiprazole (RR vs. placebo)	1.50	1.13	1.88	Log-normal	(22)
Clozapine (RR vs. placebo)	10.91	8.18	13.64	Log-normal	(22)
Augmented clozapine (RR vs. placebo)	10.91	8.18	13.64	Log-normal	Assumption
AE of EPS					
Placebo	3.00%	NA[a]	NA[a]	NA[a]	(17)
Lurasidone (RR vs. placebo)	1.92	1.43	2.50	Log-normal	(22)
Olanzapine (RR vs. placebo)	1.02	0.79	1.28	Log-normal	(22)
Risperidone (RR vs. placebo)	1.79	1.41	2.38	Log-normal	(22)
Aripiprazole (RR vs. placebo)	1.33	0.90	1.82	Log-normal	(22)
Clozapine (RR vs. placebo)	0.46	0.19	0.88	Log-normal	(22)
Augmented clozapine (RR vs. placebo)	0.46	0.19	0.88	Log-normal	Assumption
AE of diabetes					
Olanzapine	0.69%	NA[a]	NA[a]	NA[a]	(28)
Lurasidone (RR vs. olanzapine)	0.21	0.16	0.26	Log-normal	(22)

(Continued)

TABLE 1 (Continued)

Variable	Base-case value	OWSA		PSA distribution	Source
		Lower value	Upper value		
Risperidone (RR vs. olanzapine)	0.46	0.35	0.58	Log-normal	(22)
Aripiprazole (RR vs. olanzapine)	0.25	0.19	0.31	Log-normal	(22)
Clozapine (RR vs. olanzapine)	1.79	1.34	2.24	Log-normal	(22)
Augmented clozapine (RR vs. olanzapine)	1.79	1.34	2.24	Log-normal	Assumption
SMR male	10.17	7.63	12.71	Log-normal	(30)
SMR female	12.42	9.32	15.53	Log-normal	(30)

[a]Variable not included in the sensitivity analysis.
AE, adverse event; EPS, extrapyramidal symptoms; HR, hazard ratio; OWSA, one-way sensitivity analysis; PSA, probabilistic sensitivity analysis; RR, risk ratio; SMR, standardized mortality ratio.

TABLE 2 Summary of cost data used in the model.

Variable	Base-case value	OWSA		PSA distribution	Source
		Lower value	Upper value		
Drug acquisition costs					
Daily dosage, mg					
Lurasidone	60.00	NA[a]	NA[a]	NA[a]	Drug instruction
Olanzapine	12.50	NA[a]	NA[a]	NA[a]	Drug instruction
Risperidone	5.00	NA[a]	NA[a]	NA[a]	Drug instruction
Aripiprazole	20.00	NA[a]	NA[a]	NA[a]	Drug instruction
Clozapine	150.00	NA[a]	NA[a]	NA[a]	Drug instruction
Unit cost per dosage, CN¥ (US$)/mg					
Lurasidone	0.240 (0.035)	0.180	0.300	Gamma	(31)
Olanzapine	1.549 (0.224)	1.162	1.936	Gamma	(31)
Risperidone	0.635 (0.092)	0.476	0.794	Gamma	(31)
Aripiprazole	0.787 (0.114)	0.590	0.984	Gamma	(31)
Clozapine	0.001 (0.0001)	0.001	0.001	Gamma	(31)
Schizophrenia related outpatient costs, CN¥ (US$)/6-week					
Non-stable state and relapse state	615.88 (89.26)	461.91	769.85	Gamma	(32–36)
Stable state	312.48 (45.29)	234.36	390.60	Gamma	(32–36)
Schizophrenia related inpatient costs					
Duration, days					
Non-stable state and relapse state	26.40	19.80	33.00	Log-normal	Expert survey
Stable state	3.80	2.85	4.75	Log-normal	Expert survey
Relapse state	31.00	23.25	38.75	Log-normal	Expert survey
Daily cost, CN¥ (US$)/day					
Non-stable state and relapse state	520.00 (75.36)	390	650	Gamma	(32–36)
Stable state	240.00 (34.78)	180	300	Gamma	(32–36)
Relapse state	520.00 (75.36)	390	650	Gamma	(32–36)
AEs management costs, CN¥ (US$)/6-week					
Weight gain	78.62 (11.39)	58.97	98.28	Gamma	(32–36)
EPS	100.98 (14.63)	75.74	126.23	Gamma	(32–36)
Diabetes	1,544.83 (223.89)	1,158.62	1,931.04	Gamma	(37)

[a]Variable not included in the sensitivity analysis.
AE, adverse event; EPS, extrapyramidal symptoms; OWSA, one-way sensitivity analysis; PSA, probabilistic sensitivity analysis.

TABLE 3 Summary of utility data used in the model.

Variable	Base-case value	OWSA		PSA distribution	Source
		Lower value	Upper value		
Health state utility values					
Stable	0.919	0.874	0.964	Beta	(39)
Non-stable/relapse	0.604	0.522	0.686	Beta	(39)
AE-related disutility values					
Weight gain	0.089	0.052	0.126	Beta	(39)
EPS	0.256	0.227	0.285	Beta	(39)
Diabetes	0.151	0.135	0.167	Beta	(39)

AE, adverse event; EPS, extrapyramidal symptoms; OWSA, one-way sensitivity analysis; PSA, probabilistic sensitivity analysis.

the dominant strategy associated with reduced costs and increased QALYs. Over a 15-year time horizon, the total cost of patients treated with lurasidone was CN¥128,662 (US$18,647), CN¥12,093 (US$1,753) lower than that of patients treated with olanzapine, and CN¥6,781 (US$983) lower than that of patients treated with risperidone. Total QALYs of patients treated with lurasidone were 8.147, 0.197 higher than those of patients treated with olanzapine, and 0.116 higher than those of patients treated with risperidone.

Sensitivity analyses

The OWSA revealed that the model parameter with the most impact on the cost-effectiveness of lurasidone vs. olanzapine was the relapse HR for lurasidone vs. quetiapine, with the NMB ranging from CN¥16,355 (US$2,370) to CN¥38,968 (US$5,648). Other influential parameters were the relapse HR for olanzapine vs. quetiapine and the discount rate of utilities. For all OWSA results, NMBs remained >0. Similar results were observed when assessing the cost-effectiveness of lurasidone compared with risperidone. The NMB ranged from CN¥2,38 (US$34) to CN¥32,790 (US$4,752) when the relapse HR for lurasidone vs. quetiapine varied by the 95% confidence interval. The results of OWSA comparing lurasidone with olanzapine and lurasidone with risperidone are shown in Figure 2, with the top 10 influential parameters presented in the tornado diagram.

The PSA of 5,000 simulations also showed lurasidone to be cost-effective compared with either olanzapine or risperidone at all willingness-to-pay thresholds. The probabilities that lurasidone was the cost-effective strategy were 99.7, 99.9, and 100% at the willingness-to-pay thresholds of 1, 2, and 3 times of Chinese per capita GDP in 2020 [namely CN¥72,447.00 (US$10,499.57)/QALY, CN¥1,44,894.00 (US$20,999.13)/QALY, and CN¥2,17,341.00 (US$31,498.70)/QALY], respectively. The results of PSA are presented in the cost-effectiveness acceptability curve (Figure 3).

Discussion

In recent years, the NHSA of China has been incorporating drugs into NRDL through the drug pricing negotiation, to improve the availability and affordability of patented drugs for patients and optimize the structure of NRDL. Lurasidone was incorporated into China NRDL through the drug pricing negotiation in 2020, with the drug price decrease of 82.7%. To the best of our knowledge, this study, using the latest NRDL-negotiated price of lurasidone, is the first economic evaluation of lurasidone in treating patients with schizophrenia in China.

In this study, a published Markov model was applied to assess the cost-effectiveness of lurasidone vs. olanzapine and risperidone in China from a healthcare system perspective. Findings of this analysis suggested that, compared with these two commonly prescribed antipsychotics, lurasidone was found to be a dominant strategy associated with greater QALY gains at lower costs. The results were mainly attributed to the lower risk of weight gain of lurasidone than olanzapine and risperidone, which led to a lower risk of developing diabetes and a lower cost of AEs treatment. A variety of OWSA and PSA demonstrated the robustness of base-case results, and all sensitivity analyses yielded NMBs >0 at the strictest willingness-to-pay threshold of CN¥72,447.00 (US$10,499.57)/QALY.

Economic evaluations evaluating lurasidone vs. other available atypical antipsychotics have been conducted in a few countries. One study from a US payer perspective evaluated the cost-effectiveness of lurasidone compared with risperidone, olanzapine, ziprasidone, aripiprazole, and quetiapine through a 5-year Markov model (40). Health states included in the model were patients: on an initial atypical antipsychotic; switched to a second atypical antipsychotic; and on clozapine after failing a second atypical antipsychotic. The results showed olanzapine, ziprasidone, aripiprazole, and quetiapine were dominated by other comparators and removed from the comparative analysis, and lurasidone was cost-effective at willingness-to-pay thresholds of >US$25,844 per hospitalization avoided compared with risperidone. Another study from the perspective

TABLE 4 Results of the base-case analysis.

Treatment	Total costs, CN¥ (US$)	Total QALYs	Incremental costs, CN¥ (US$)	Incremental QALYs	ICER, CN¥ (US$)/ QALY
Lurasidone	128,662 (18,647)	8.147	—	—	—
Olanzapine	140,755 (20,399)	7.950	−12,093 (-1,753)	0.197	Lurasidone dominant
Risperidone	135,443 (19,629)	8.031	−6,781 (-983)	0.116	Lurasidone dominant

QALY, quality-adjusted life-years; ICER, incremental cost-effectiveness ratio.

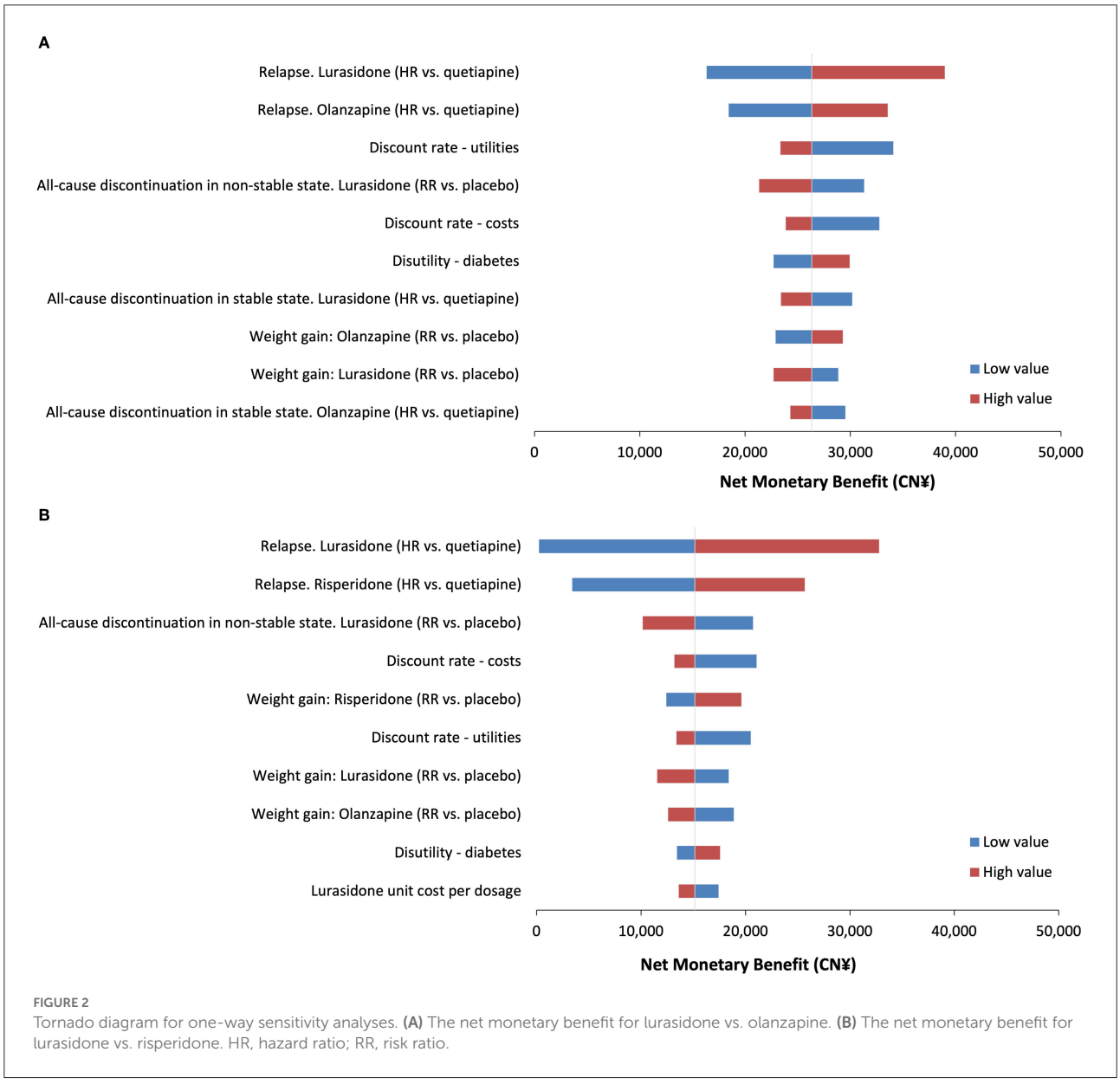


FIGURE 2
Tornado diagram for one-way sensitivity analyses. **(A)** The net monetary benefit for lurasidone vs. olanzapine. **(B)** The net monetary benefit for lurasidone vs. risperidone. HR, hazard ratio; RR, risk ratio.

of Scotland and Wales healthcare services evaluated the cost-effectiveness of lurasidone vs. aripiprazole through a 10-year Markov model (23), the structure of which was adopted in the present study. The findings of the prior study suggested that lurasidone was a dominant strategy, with an increase of 0.005 QALYs and cost savings of £3,383 in Scotland

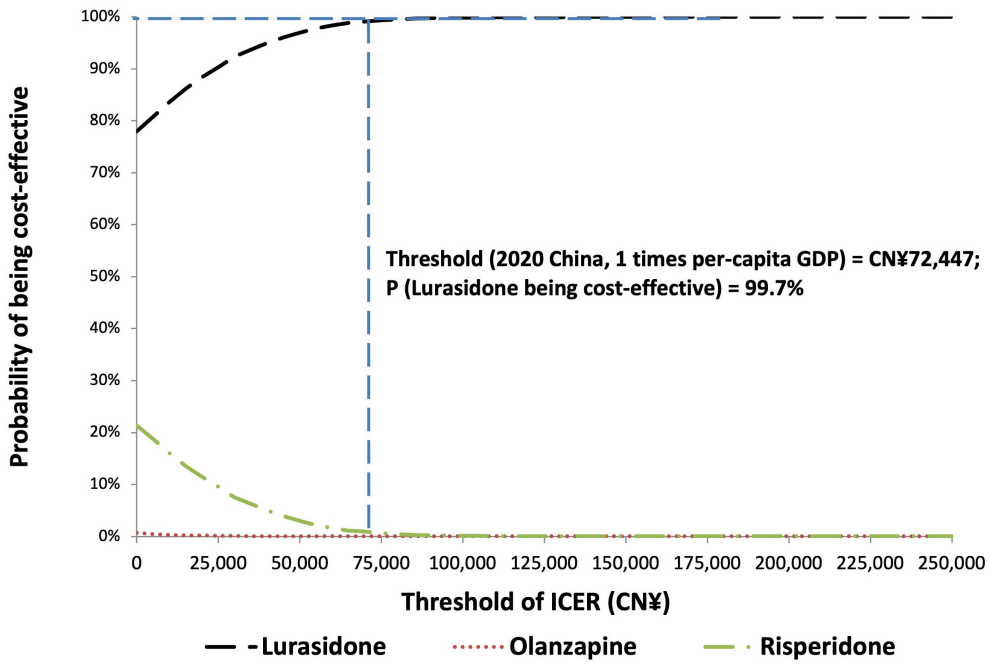


FIGURE 3
Cost-effectiveness acceptability curve for probabilistic sensitivity analysis. GDP, gross domestic product; ICER, incremental cost-effectiveness ratio.

and £3,072 in Wales. Thus, previous studies and our study are consistent in demonstrating the economic advantages of lurasidone compared with other atypical antipsychotics in treating patients with schizophrenia over a variety of time horizon.

There are some limitations to this study that should be considered when interpreting its results. First, to compare lurasidone vs. olanzapine and risperidone in the model, indirect comparisons were used to inform the clinical efficacy and safety. While healthcare decision-makers increasingly recognize indirect comparisons as an acceptable alternative method of comparison in the absence of real-world parallel-group data, differences in study populations may limit their comparability. Therefore, future studies evaluating the cost-effectiveness of lurasidone compared with olanzapine and risperidone based on direct comparison data are needed to verify the findings of this study. Second, due to the lack of data, a face-to-face survey of clinical experts was conducted to understand the healthcare resource utilization related to standard schizophrenia treatment and AEs treatment, which may lead to the uncertainty associated with schizophrenia-related outpatients, inpatient, and AEs treatment costs. Nonetheless, sensitivity analyses showed that changes in these costs had limited effect on the ICER value. Third, utility values used in this study were obtained from foreign studies as we did not identify available data on Chinese schizophrenia patient. As discussed in a recent publication, applying utility values derived from the previous studies to cost-utility analyses may result in the heterogeneity among results, which might be impacted by the differences in survey responders, elicitation methods, and regions (41). We therefore tested model utility parameters in sensitivity analyses and found that these values did not have a major impact on the study results. However, caution should be taken when extrapolating our findings to other health systems, as all model inputs in this study were specific to the Chinese healthcare setting. Finally, one limitation of our analysis is that it relies on the *post-hoc* analysis of clinical trials, in which assessing economic value is rarely the primary purpose. Since the results of this study could be regarded as preliminary, it will be important to further explore the cost-effectiveness of lurasidone in China based on the real-world evidence or to conduct an economic evaluation alongside the clinical trial of lurasidone.

As far as this study was concerned, compared with olanzapine and risperidone, lurasidone was a dominant strategy that yield more QALY gains with lower costs for the first-line treatment of schizophrenia in China. The robustness of the results was verified by sensitivity analyses. As the first analysis accessing the cost-effectiveness of lurasidone in China, the results may assist to fill gaps in clinical decisions regarding pharmacotherapies of schizophrenia.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

JL, LC, and JW contributed to conception and design of the study, collection of data, and development of decision analytical model. JL and LC conducted the data analysis. All authors participated in critically reviewing and interpreting the data, reviewed the manuscript for intellectual content and approved the submitted version, and agree to be accountable for all aspects of the work.

Funding

The authors declare that this study received funding from Sumitomo Pharma (Suzhou) Co., Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

The authors thank the health economics and outcome research team of Sumitomo Pharma Co., Ltd. for their support and advice on the modeling methodology.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Phillips MR, Yang G, Li S, Li Y. Suicide and the unique prevalence pattern of schizophrenia in mainland China: a retrospective observational study. *Lancet.* (2004) 364:1062–8. doi: 10.1016/S0140-6736(04)17061-X

2. The Lancet. GBD cause and risk summaries. *Lancet.* (2020). Available online at: https://www.thelancet.com/pbassets/Lancet/gbd/summaries/diseases/schizophrenia.pdf. (accessed December 1st, 2021).

3. Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou CF, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. *Neuropsychiatr Dis Treat.* (2016) 12:357–73. doi: 10.2147/NDT.S96649

4. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. *Lancet Psychiatry.* (2019) 6:211–24. doi: 10.1016/S2215-0366(18)30511-X

5. The Institute for Health Metrics and Evaluation (IHME). *GBD Compare. IHME; [Cited 2021 Jul 31].* Available online at: https://vizhub.healthdata.org/gbd-compare/. (accessed December 1st, 2021).

6. Zhai J, Guo X, Chen M, Zhao J, Su Z. An investigation of economic costs of schizophrenia in two areas of China. *Int J Ment Health Syst.* (2013) 7:26. doi: 10.1186/1752-4458-7-26

7. McDonagh MS, Dana T, Selph S, Devine EB, Cantor A, Bougatsos C, et al. *Treatments for Schizophrenia in Adults: A Systematic Review [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US).* (2017). Report No.: 17(18)-EHC031-EF. Available online at: https://www.ncbi.nlm.nih.gov/books/NBK487628/?report=reader. (accessed December 1st, 2021).

8. Zhao J P. *The Second Edition of Guidelines for the Prevention and Treatment of Schizophrenia.* Beijing: China Medical Electronic Audio & Video Press. (2015).

9. Scottish Intercollegiate Guidelines Network (SIGN). *Management of schizophrenia. Edinburgh: Scottish Intercollegiate Guidelines Network.* (2013). Available online at: http://www.sign.ac.uk. (Accessed December 1st, 2021).

10. Correll CU, Frederickson AM, Kane JM, Manu P. Metabolic syndrome and the risk of coronary heart disease in 367 patients treated with second-generation antipsychotic drugs. *J Clin Psychiatry.* (2006) 67:575–83. doi: 10.4088/JCP.v67n0408

11. Italien GJ, Casey DE. Kan HJ, Carson WH, Marcus RN. Comparison of metabolic syndrome incidence among schizophrenia patients treated with aripiprazole vs olanzapine or placebo. *J Clin Psychiatry.* (2007) 68:1510–6. doi: 10.4088/JCP.v68n1006

12. Sernyak MJ, Leslie DL, Alarcon RD, Losonczy MF, Rosenheck R. Association of diabetes mellitus with use of atypical neuroleptics in the treatment of schizophrenia. *Am J Psychiatry.* (2002) 159:561–6. doi: 10.1176/appi.ajp.159.4.561

13. American Diabetes Association A. American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diab Care.* (2004) 27:596–601. doi: 10.2337/diacare.27.2.596

14. Lindenmayer JP, Czobor P, Volavka J, Citrome L, Sheitman B, McEvoy JP, et al. Changes in glucose and cholesterol levels in patients with schizophrenia treated with typical or atypical antipsychotics. *Am J Psychiatry.* (2003) 160:290–6. doi: 10.1176/appi.ajp.160.2.290

15. Mackin P, Bishop D, Watkinson H, Gallagher P, Ferrier IN. Metabolic disease and cardiovascular risk in people treated with antipsychotics in the community. *Br J Psychiatry.* (2007) 191:23–9. doi: 10.1192/bjp.bp.106.031716

16. Nasrallah HA, Meyer JM, Goff DC, McEvoy JP, Davis SM, Stroup TS, et al. Low rates of treatment for hypertension, dyslipidemia and diabetes in schizophrenia: data from the CATIE schizophrenia trial sample at baseline. *Schizophr Res.* (2006) 86:15–22. doi: 10.1016/j.schres.2006.06.026

17. Loebel A, Cucchiaro J, Sarma K, Xu L, Hsu C, Kalali AH, et al. Efficacy and safety of lurasidone 80 and 160 mg/day in the treatment of schizophrenia: a randomized, double-blind, placebo- and active-controlled trial. *Schizophr Res.* (2013) 145:101–9. doi: 10.1016/j.schres.2013.01.009

18. Meltzer HY, Cucchiaro J, Silva R, Ogasa M, Phillips D, Xu J, et al. Lurasidone in the treatment of schizophrenia: a randomized, double-blind,

placebo- and olanzapine-controlled study. *Am J Psychiatry.* (2011) 168:957–67. doi: 10.1176/appi.ajp.2011.10060907

19. Nasrallah HA, Silva R, Phillips D, Cucchiaro J, Hsu J, Xu J, et al. Lurasidone for the treatment of acutely psychotic patients with schizophrenia: a 6-week, randomized, placebo-controlled study. *J Psychiatr Res.* (2013) 47:670–7. doi: 10.1016/j.jpsychires.2013.01.020

20. Loebel A, Cucchiaro J, Xu J, Sarma K, Pikalov A, Kane JM. Effectiveness of lurasidone vs. quetiapine XR for relapse prevention in schizophrenia: A 12-month, double-blind, noninferiority study. *Schizophr Res.* (2013) 147:95–102. doi: 10.1016/j.schres.2013.03.013

21. Leucht S, Cipriani A, Spineli L, Mavridis D, Orey D, Richter F, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet.* (2013) 382:951–62. doi: 10.1016/S0140-6736(13)60733-3

22. Huhn M, Nikolakopoulou A, Schneider-Thoma J, Krause M, Samara M, Peter N, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. *Lancet.* (2019) 394:939–51. doi: 10.1016/S0140-6736(19)31135-3

23. Rajagopalan K, Trueman D, Crowe L, Squirrell D, Loebel A. Cost-utility analysis of lurasidone vs. aripiprazole in adults with schizophrenia. *Pharmacoeconomics.* (2016) 34:709–21. doi: 10.1007/s40273-016-0405-0

24. Liu GG, Hu SL, Wu JH, Wu J, Yang L, Li HC, et al. *China Guidelines for Pharmacoeconomic Evaluations*. China Market Press. (2020).

25. Dima L, Vasile D, Voicu VA. Comparative discontinuation rates in patients treated with antipsychotics. *Therapeutics.* (2012) 16:181–9. Available online at: https://www.researchgate.net/publication/287605954_Comparative_discontinuation_rates_in_patients_treated_with_antipsychotics/stats

26. National Collaborating Centre for Mental Health, National Institute for Health and Clinical Excellence. *Schizophrenia: core Interventions in the Treatment and Management of Schizophrenia in Primary and Secondary Care (Update), Chapter 7.2. 2009*. Available online at: http://www.ncbi.nlm.nih.gov/books/NBK11681/pdf/TOC.pdf. (accessed December 1st, 2021).

27. Xiao J, Mi W, Li L, Shi Y, Zhang H. High relapse rate and poor medication adherence in the Chinese population with schizophrenia: results from an observational survey in the People's Republic of China. *Neuropsychiatr Dis Treat.* (2015) 11:1161–7. doi: 10.2147/NDT.S72367

28. Lin L, Zhao YJ, Zhou HJ, Khoo AL, Teng M, Soh LB, et al. Comparative cost-effectiveness of 11 oral antipsychotics for relapse prevention in schizophrenia within Singapore using effectiveness estimates from a network meta-analysis. *Int Clin Psychopharmacol.* (2016) 31:84–92. doi: 10.1097/YIC.0000000000000111

29. Statistical Yearbook of the People's Republic of China on Population and Employment in 2018 [EB/OL]. *National Bureau of Statistics, [2020-07-14]*. Available online at: https://data.cnki.net/trade/Yearbook/Single/N2019030259?z=Z001. (accessed December 1st, 2021).

30. Liu T, Song X, Chen G, Paradis AD, Zheng X. Prevalence of schizophrenia disability and associated mortality among Chinese men and women. *Psychiatry Res.* (2014) 220:181–7. doi: 10.1016/j.psychres.2014.04.042

31. YaoZh. *Inquiry of Drug Bid Information*. Available online at: https://www.yaozh.com. (accessed December 1st, 2021).

32. Beijing Municipal Medical Insurance Bureau. *Medical Service Price Inquiry*. Available online at: http://ybj.beijing.gov.cn/2020_zwfw/2020_bmcx/. (accessed December 1st, 2021).

33. Shanghai Municipal Development and Reform Commission. *Medical Service Price Inquiry*. Available online at: http://fgw.sh.gov.cn/jggl/. (accessed December 1st, 2021).

34. Changsha Healthcare Security Administration. *Price List of Current Medical Service Price in Changsha*. (2020). Available online at: http://www.tianxin.gov.cn/zwgk8/bmxxgkml/qylbzj/zcwj15/gzzd51/202009/P020200915588540967098.pdf. (accessed December 1st, 2021).

35. Chengdu Healthcare Security Administration; Chengdu Municipal Health Commission. *Price List of Medical Service Price Structure Adjustment Projects of Municipal Public Medical Institutions in Chengdu*. Available online at: http://cdyb.chengdu.gov.cn/ylbzj/c128998/2020-12/02/content_ab17741660454da38fa5181ae88b2877.shtml. (accessed December 1st, 2021).

36. Liaoning Provincial Development and Reform Commission; Health Commission of Liaoning Province; Liaoning Provincial Department of Human Resources and Social Security; Liaoning Province Finance Department. *Notice on Printing and Distributing Price of Medical Service Projects of Public Medical Institutions in Liaoning Province*. Available online at: http://wsjk.ln.gov.cn/wst_wjgg/201710/t20171027_3092708.html. (accessed December 1st, 2021).

37. Li X, Xu Z, Ji L, Guo L, Liu J, Feng K, et al. Direct medical costs for patients with type 2 diabetes in 16 tertiary hospitals in urban China: a multicenter prospective cohort study. *J Diabetes Investig.* (2019) 10:539–51. doi: 10.1111/jdi.12905

38. National Bureau of Statistics of China. *Statistical Bulletin of National Economic and Social Development of the People's Republic of China in 2020*. Available online at: http://www.stats.gov.cn/tjsj/zxfb/202102/t20210227_1814154.html. (accessed December 1st, 2021).

39. Briggs A, Wild D, Lees M, Reaney M, Dursun S, Parry D, et al. Impact of schizophrenia and schizophrenia treatment-related adverse events on quality of life: direct utility elicitation. *Health Qual Life Outcomes.* (2008) 6:105. doi: 10.1186/1477-7525-6-105

40. O'Day K, Rajagopalan K, Meyer K, Pikalov A, Loebel A. Long-term cost-effectiveness of atypical antipsychotics in the treatment of adults with schizophrenia in the US. *Clinicoecon Outcomes Res.* (2013) 5:459–70. doi: 10.2147/CEOR.S47990

41. Wang L, Shi F, Guan X, Xu H, Liu J, Li H, et al. Systematic review of methods and study quality of economic evaluations for the treatment of schizophrenia. *Front Public Health.* (2021) 9:689123. doi: 10.3389/fpubh.2021.689123

frontiers | Frontiers in Public Health

TYPE Correction
PUBLISHED 09 November 2022
DOI 10.3389/fpubh.2022.1059495

OPEN ACCESS

EDITED AND REVIEWED BY
Huajie Jin,
King's College London,
United Kingdom

*CORRESPONDENCE
Jing Wu
jingwu@tju.edu.cn

SPECIALTY SECTION
This article was submitted to Public Mental Health, a section of the journal Frontiers in Public Health

RECEIVED 01 October 2022
ACCEPTED 27 October 2022
PUBLISHED 09 November 2022

CITATION
Liu J, Cao L and Wu J (2022) Corrigendum: Cost-utility analysis of lurasidone for the first-line treatment of schizophrenia in China. *Front. Public Health* 10:1059495. doi: 10.3389/fpubh.2022.1059495

Corrigendum: Cost-utility analysis of lurasidone for the first-line treatment of schizophrenia in China

Jia Liu[1,2], Lidan Cao[1,2] and Jing Wu[1,2]*

[1]School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, China, [2]Center for Social Science Survey and Data, Tianjin University, Tianjin, China

KEYWORDS

cost-utility, lurasidone, olanzapine, risperidone, schizophrenia

A corrigendum on
Cost-utility analysis of lurasidone for the first-line treatment of schizophrenia in China

by Liu, J., Cao, L., and Wu, J. (2022). *Front. Public Health* 10:987408. doi: 10.3389/fpubh.2022.987408

In the published article, there was an error. The name of the drug "risperidone" was mistakenly written as "aripiprazole" in the Abstract and Results sections.

A correction has been made to the Abstract section, Results sub-section, Paragraph 1. This sentence previously stated: "Over a 15-year time horizon, lurasidone yielded an improvement of 0.197 QALYs with a cost saving of CN¥12,093 (US$1,753) vs. olanzapine and an improvement of 0.116 QALYs with a cost saving of CN¥6,781 (US$983) vs. aripiprazole." The corrected sentence appears below: "Over a 15-year time horizon, lurasidone yielded an improvement of 0.197 QALYs with a cost saving of CN¥12,093 (US$1,753) vs. olanzapine and an improvement of 0.116 QALYs with a cost saving of CN¥6,781 (US$983) vs. risperidone."

Three corrections have been made to the Results section, Base-case analysis sub-section, Paragraph 1. One sentence previously stated: "Compared with olanzapine and aripiprazole, lurasidone was the dominant strategy associated with reduced costs and increased QALYs." The corrected sentence appears below: "Compared with olanzapine and risperidone, lurasidone was the dominant strategy associated with reduced costs and increased QALYs." Another sentence previously stated: "Over a 15-year time horizon, the total cost of patients treated with lurasidone was CN¥1,28,662 (US$18,647) and CN¥12,093 (US$1,753) lower than that of patients treated with olanzapine, and CN¥6,781 (US$983) lower than that of patients treated with aripiprazole." The corrected sentence appears below: "Over a 15-year time horizon, the total cost of patients treated with lurasidone was CN¥128,662 (US$18,647), CN¥12,093 (US$1,753) lower than that of patients treated with olanzapine, and CN¥6,781 (US$983) lower than that of patients treated with risperidone." The other sentence previously stated: "Total QALYs of patients

treated with lurasidone were 8.147, 0.197 higher than those of patients treated with olanzapine, and 0.116 higher than those of patients treated with aripiprazole." The corrected sentence appears below: "Total QALYs of patients treated with lurasidone were 8.147, 0.197 higher than those of patients treated with olanzapine, and 0.116 higher than those of patients treated with risperidone."

Two corrections have been made to the Results section, Sensitivity analyses sub-section, Paragraph 1. One sentence previously stated: "Similar results were observed when assessing the cost-effectiveness of lurasidone compared with aripiprazole." The corrected sentence appears below: "Similar results were observed when assessing the cost-effectiveness of lurasidone compared with risperidone." The other sentence previously stated: "The results of OWSA comparing lurasidone with olanzapine and lurasidone with aripiprazole are shown in Figure 2, with the top 10 influential parameters presented in the tornado diagram." The corrected sentence appears below: "The results of OWSA comparing lurasidone with olanzapine and lurasidone with risperidone are shown in Figure 2, with the top 10 influential parameters presented in the tornado diagram."

A correction has been made to the Results section, Sensitivity analyses sub-section, Paragraph 2. This sentence previously stated: "The PSA of 5,000 simulations also showed lurasidone to be cost-effective compared with either olanzapine or aripiprazole at all willingness-to-pay thresholds." The corrected sentence appears below: "The PSA of 5,000 simulations also showed lurasidone to be cost-effective compared with either olanzapine or risperidone at all willingness-to-pay thresholds."

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 07 November 2022
DOI 10.3389/fpsyt.2022.1028156

OPEN ACCESS

EDITED BY
Huajie Jin,
King's College London,
United Kingdom

REVIEWED BY
Yuhua Bao,
Weill Cornell Medicine, United States
Tariq Munshi,
University of Toronto, Canada

*CORRESPONDENCE
Rachel A. Elliott
rachel.a.elliott@manchester.ac.uk

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Public Mental Health, a section of the journal Frontiers in Psychiatry

RECEIVED 25 August 2022
ACCEPTED 17 October 2022
PUBLISHED 07 November 2022

CITATION
Paterson L, Rennick-Egglestone S, Gavan SP, Slade M, Ng F, Llewellyn-Beardsley J, Bond C, Grundy A, Nicholson J, Quadri D, Bailey S and Elliott RA (2022) Development and delivery cost of digital health technologies for mental health: Application to the Narrative Experiences Online Intervention. *Front. Psychiatry* 13:1028156. doi: 10.3389/fpsyt.2022.1028156

Development and delivery cost of digital health technologies for mental health: Application to the Narrative Experiences Online Intervention

Luke Paterson[1†], Stefan Rennick-Egglestone[2†], Sean P. Gavan[1], Mike Slade[2,3], Fiona Ng[2], Joy Llewellyn-Beardsley[2], Carmel Bond[4], Andrew Grundy[5], Joe Nicholson[6], Dania Quadri[7], Sylvia Bailey[8] and Rachel A. Elliott[1*]

[1]Manchester Centre for Health Economics, Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, United Kingdom, [2]School of Health Sciences, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom, [3]Health and Community Participation Division, Faculty of Nursing and Health Sciences, Nord University, Namsos, Norway, [4]Nottingham University Business School, The University of Nottingham, Nottingham, United Kingdom, [5]School of Health Sciences, University of Manchester, Manchester, United Kingdom, [6]School of Humanities, The University of Nottingham, Nottingham, United Kingdom, [7]GKT School of Medical Education, King's College London, London, United Kingdom, [8]Narrative Experiences Online Intervention (NEON) Lived Experience Advisory Panel, Nottingham, United Kingdom

Background: The increasing development and use of digital health interventions requires good quality costing information to inform development and commissioning choices about resource allocation decisions. The Narrative Experiences Online (NEON) Intervention is a web-application that delivers recorded mental health recovery narratives to its users. Two randomized controlled trials are testing the NEON Intervention in people with experience of psychosis (NEON) and people experiencing non-psychosis mental health problems (NEON-O).

Aim: This study describes and estimates the cost components and total cost of developing and delivering the NEON Intervention.

Materials and methods: Total costs for the NEON Trial (739 participants) and NEON-O Trial (1,024 participants) were estimated by: identifying resource use categories involved in intervention development and delivery; accurate measurement or estimation of resource use; and a valuation of resource use to generate overall costs, using relevant unit costs. Resource use categories were identified through consultation with literature, costing reporting standards and iterative consultation with health researchers involved in NEON Intervention development and delivery. Sensitivity analysis was used to test assumptions made.

Results: The total cost of developing the NEON Intervention was £182,851. The largest cost components were software development (27%); Lived Experience Advisory Panel workshops (23%); coding the narratives (9%); and researchers' time to source narratives (9%). The total cost of NEON Intervention delivery during the NEON Trial was £118,663 (£349 per NEON

Intervention user). In the NEON-O Trial, the total delivery cost of the NEON Intervention was £123,444 (£241 per NEON Intervention user). The largest cost components include updating the narrative collection (50%); advertising (19%); administration (14%); and software maintenance (11%). Uncertainty in the cost of administration had the largest effect on delivery cost estimates.

Conclusion: Our work shows that developing and delivering a digital health intervention requires expertise and time commitment from a range of personnel. Teams developing digital narrative interventions need to allocate substantial resources to curating narrative collections.

Implications for practice: This study identifies the development and delivery resource use categories of a digital health intervention to promote the consistent reporting of costs and informs future decision-making about the costs of delivering the NEON Intervention at scale.

Trial registration: NEON Trial: ISRCTN11152837, registered 13 August 2018, http://www.isrctn.com/ISRCTN11152837. NEON-O Trial: ISRCTN63197153, registered 9 January 2020, http://www.isrctn.com/ISRCTN63197153.

KEYWORDS

narrative, psychosis, mental health, recovery, healthcare costs, digital health intervention, online

Introduction

In recent years, there has been a rapid increase in digital health technologies (DHTs) to deliver mental health interventions remotely, either to replace or supplement face-to-face healthcare, with that increase accelerating during the COVID-19 pandemic (1, 2).

Digital health technologies range from predefined functions such as remote diagnosis or disease monitoring, to more complex functions around supporting behavior change through interactive and personalized interactions. A sub-type of DHTs are digital health interventions (DHIs), which provide users with remote access to treatment through text messaging services, smartphone apps, and web-based resources. Recent reviews of the growing number of trials have shown promising but varied effectiveness for DHIs in a wide range of mental health conditions, including but not limited to depression (3, 4), eating disorders (5), post-traumatic stress disorder (6), and schizophrenia (7).

Investment in mental health DHIs is increasing, partly to improve access to care in overstretched health systems and partly due to emerging patient and user preference for DHI delivery of mental health care (4). This means that decisions are being made about effectiveness and affordability of individual DHIs. Economic evaluations inform healthcare resource allocation decisions and treatment recommendations by comparing the costs and health benefits of alternative ways to treat patients (8). Demonstrating cost-effectiveness is an important factor for delivering DHIs into healthcare systems across Europe (9). Economic evaluations of mental health DHIs have been conducted for a range of conditions such as anxiety (10, 11), depression (10–12), eating disorders (10, 13), and substance abuse disorders (10, 14). The quality of economic evaluations has been suggested to be very variable including heterogeneous reporting of costs, outcomes and comparators (15). Such inconsistencies may be a barrier to the adoption of effective DHIs for mental health within healthcare systems or may lead policymakers to invest in services that are not cost-effective (16–18).

There is a lack of standardization in how costs are reported and described for economic evaluations of DHIs (11, 15). The National Institute of Health and Care Excellence's (NICE) evidence standards framework for DHTs recommends reporting all costs associated with the intervention and costs relevant to a health and social care decision maker (19). The framework focuses on delivery (implementation) costs, as these are relevant to health and social care providers looking to commission a service, and includes initial investment costs such as training, as well as operation, and maintenance costs (19). Apart from the initial fixed set-up costs, including infrastructure and training costs, evolution of the DHI over time (interface design, software updates, and content updates) requires more flexible data collection tools that can keep pace with these changes. For example, the marginal cost is the additional cost incurred when one more person uses the product. DHIs tend to have a low marginal cost (one more user accessing an app tends to a zero

marginal cost) up to a certain threshold, beyond which there is a large increase due to the need to re-engineer components of technology (such as centralized servers maintaining user account records and delivering web-based content) to provide additional capacity. Other challenges for DHI costing include having to respond to rapidly changing prices (such as exchange rate fluctuations) and short life-cycles of technology, short depreciation periods, and attribution of cost from a shared resource such as a wireless network (20).

Development costs are not typically considered in economic evaluations as, by definition, they have already been incurred prior to delivery. Historically, these development costs are sunk costs, and may include cross-subsidization of product development where there is a failure to reach product launch. Development costs can be included in a market price, which will fall as the market share, or the number of users increases. A recent systematic review reported development costs were only reported in four out of 24 economic evaluations of internet-based interventions for anxiety and depression (11). There are no guidelines for estimating the development cost of DHIs, resulting in non-standardized approaches and a lack of comparability across studies. However, development costs are relevant to publicly funded research groups, research funders, university enterprise offices and private companies and by informing decisions to develop DHIs or scale an existing technology. DHI development incurs significant research and development costs prior to launch, and the resources consumed during development can be very different from pharmaceuticals and physical medical devices. DHIs have other unique characteristics that can affect development processes and thus costs, including faster evolution, active and more dynamic user input (21). DHI development costs are usually incurred for content and software design, website and graphic design, digital platform development and regulatory approval processes. A recent study focusing on the development of a mental health DHI proposed a costing framework for development costs and we have applied this framework to our development costs methods (22).

The aim of this study is to describe, estimate, and present the associated cost components and total cost of developing and delivering the Narrative Experiences Online (NEON) Intervention, using current recommendations for costing DHIs (11, 19, 21, 22).

Digital health intervention under investigation: NEON Intervention

The NEON Intervention is a web-based application that delivers recorded recovery narratives to its users. A systematic review (23) and qualitative validation study (24) defined mental health recovery narratives as first-person lived experience accounts of recovery from mental health problems, which include elements of adversity or struggle and of self-defined strengths, successes or survival. The impact of recovery narratives was then investigated in a systematic review (25), qualitative interviews (26), and experimental studies (27, 28). Approaches to curation of recorded recovery narrative collections were developed through systematic review (29), stakeholder consultation (30) and best practice guidelines development (31). These studies, together with related work on post-traumatic growth (32, 33), non-service user perspectives (34), institutional injustice (35), and clinician perspectives on use of narratives in practice (36), provided the theory base.

The NEON Intervention was then developed (37). The NEON Intervention allows recipients to engage with recovery narratives by watching, reading, or listening to narrator-led stories on the website. Access to these narratives is provided through different avenues: a hybrid recommender system using collaborative and content-based filtering to recommend appropriate stories; self-selected stories from the entire collection of narratives (referred to as the NEON Collection); randomly selected stories; recommendations sent to users in emails, to serve as a mechanism for engaging people with the intervention; and re-requested narratives that have been previously seen. The recommender system uses feedback data from stories received by participants, characteristics of the participants, and characteristics of each recorded recovery narrative assessed using the Inventory of Characteristics of Recovery Stories (INCRESE) measure (38), to match participants to narratives intended to be of benefit. The website was engineered to work on personal computers, mobile devices, or communal computers such as in a public library to enable participation by people experiencing digital exclusion (39).

An economic evaluation is being conducted as part of two definitive randomized controlled trials (RCTs) to evaluate the cost-effectiveness of offering the NEON Intervention to individuals with experience of psychosis (NEON Trial, $n = 739$) and people experiencing non-psychosis mental health problems (NEON-O Trial, $n = 1{,}024$) (40, 41).

This study reports the estimation of development and delivery costs for the NEON Intervention, which is used in both trials.

Materials and methods

This study was conducted as part of the Narrative Experiences Online (NEON) Programme (researchintorecovery.com/neon), which is investigating whether receiving recorded mental health recovery narratives improves the quality of life in people who experience mental health issues.

Design of costing study

Estimated total cost and individual cost components for the development and delivery of the NEON Intervention are reported. The development costs included those incurred from the perspective of the research body who funded the development and testing of the NEON Intervention. The delivery costs included those incurred from the perspective of the health and social care provider (National Health Service, NHS England) (19). The length of time over which delivery costs were collected was from the beginning of delivery (9th March 2020) to the end of the trial periods (13th May 2022 for NEON Trial, 23rd June 2022 for NEON-O Trial). The NEON Intervention was identical for both NEON and NEON-O trials, since both trials provided participants with access to the same narrative collection. A separate cost of delivery of the NEON Intervention is presented for both trials. Downstream costs (those incurred as a result of using the intervention from the perspective of the NHS) will be reported in a subsequent full economic evaluation.

The development costs begin from the start of the current research program funding, comprising software and intervention development, through feasibility testing and up to the starting point of the definitive trial. Early conceptual development work, and research-related tasks including the randomized controlled trials which contribute to a large part of the program grant, are excluded in the development cost estimates. The duration of the development of the NEON Intervention was considered over the period January 2017 to January 2021.

The total cost for the development stage of the NEON Intervention was constructed with reference to a development costing checklist for a digital program for training community health workers to deliver treatment for depression (22). The costing strategy consisted of: identifying resource use categories involved in the development and delivery of the intervention; accurate measurement or estimation of resource use; and a valuation of resource use to generate overall costs (42). Identifying resource use categories was carried out through consultation with published literature relevant to DHIs, costing reporting standards and expert consultation with team members involved in intervention development and delivery.

The product of resource use and unit costs generated total cost estimates, and this is referred to as the base-case analysis. In this study, the resource use data obtained for development and delivery was obtained for the whole user cohort (top-down costs), rather than data relevant to a specific user (bottom-up costs). Therefore, the approach taken was to generate total development and delivery costs and then apportion to individual users as a "mean cost per user." The number of users was defined as the number of people randomized to the intervention group (370 in the NEON Trial and 512 in the NEON-O Trial).

In the base-case analysis, costs for human resources were obtained by multiplying the personnel's midpoint hourly salary plus on-costs (pension contributions and payroll taxes) by the proportion of hours spent on the task. For external experts and consultants their costs were recorded as invoices to the NEON study. In some cases where records were not kept, the proportion of hours spent on a task was estimated through expert consultation to produce the maximum and minimum plausible duration; with the midpoint (average) selected during the base-case analysis. For information technologies used in the development and delivery of the NEON Intervention, financial records were used to calculate the cost of the components purchased. In one case, an estimate was derived from the NEON budget proposal.

Deriving estimates of certain cost components required some assumptions to be made with uncertainty further associated with the true values of several components. To assess the impact of our assumptions and parameter uncertainty, all input parameters were adjusted to their extreme values individually in a one-way sensitivity analysis. Tornado diagrams (43) were used to illustrate which input parameters had the most impact on cost estimates. Structural uncertainty resulting from the assumptions made were examined through scenario analysis.

This costing study was developed and reported in accordance to standard validation and reporting criteria (44). A team member not involved in the analysis used the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 checklist to ensure that relevant items were reported completely (see Appendix 1). Face validity of cost categories and costing methods was ascertained through continuous feedback from clinical and patient experts. All costs are presented in UK Sterling (£) for the costing period 2020–2021.

Costing the NEON Intervention

We identified several resource use categories in intervention development and delivery, summarized in Table 1. This section provides detail on how resource use and unit costs were obtained for each category.

NEON Intervention development resource use

Four components that incurred resources were needed to develop the NEON Intervention:

- building a collection of recovery narratives (the NEON Collection) that would be used in the intervention (consisting of sourcing narratives, securing ethical approval to use those narratives, liaising with collection organizers to source narratives from existing collections, Lived Experience Advisory Panel (LEAP) workshops to develop recommendations on (1) the ethical issues around narrative

TABLE 1 Summary of NEON Intervention development and delivery resource use categories.

Development resource use categories
Curation of the NEON Collection:
• Staff time to source recovery narrative
• Staff time to secure ethical approval
• Collection organizers
• Lived Experience Advisory Panel workshops
• Collection Steering Group meetings (including preparation for meetings)
• Training researchers
• Coding narratives
Web application development:
• Database specification
• Software development
• Recommender system and integrating into the codebase
• Graphic design
• Interaction design
Communication:
• Task meetings
• Advisory Board meetings
• Design Group meetings
Intervention:
• Intervention testing
• Feasibility study
Delivery resource use categories
Web hosting
Personnel training
Periodic updates to the narrative collection
Web-application maintenance
Administrative support
Advertisement
Intervention engagement
Safeguarding

curation and (2) the initial curation procedures, Collection Steering Group (CSG) meetings to make decisions on (1) the inclusion of individual narratives and (2) refinement of the curation procedures, training researchers to use the INCRESE tool, coding narratives using INCRESE);

- developing the web-application as a platform to deliver the NEON Intervention [consisting of reporting a database specification to support the development of the web-application, developing source code for the web-application (software development), conceptualizing and developing a recommender system that matches users with the most appropriate narratives, integrating the recommender system into the web-application codebase, and designing the intervention to ensure it was appealing to both operate aesthetically (graphic design) and practically (interaction design)];
- additional communications necessary to develop the intervention [consisting of NEON study team task meetings for general discussions, International Advisory Board (IAB) meetings to advise on safety and ethical concerns, and Intervention Development Group (IDG) meetings to provide (feedback on the intervention)];
- testing the intervention (consisting of testing the intervention's performance on the web-application including the collection of outcome data, and a feasibility study evaluating a prototype of the intervention in a small sample of mental health service users (baseline: $n = 25$; follow-up: $n = 22$) with experience of mental health problems);

Resource use during the development of the NEON Intervention was measured through examinations of records and/or derived estimates with assistance from expert consultation. Tables 2–5 summarize resource use and unit cost categories and sources for the four components of NEON intervention development.

An essential component of the development of the NEON Intervention was building the NEON Collection (37). The recovery stories used were sourced based on the objective of maximizing the diversity of the types of stories within the collection based on different diversity domains (such as narrator ethnicity, sexuality, gender identity, neurodiversity, etc.). There were two routes in which narratives were sourced for the collection: individual donations (∼7% of the collection) and existing collections (∼93% of the collection). Both routes required work to secure permission to use the narratives within the NEON Collection. For donated narratives, this involved liaising directly with the narrator to secure permission for the re-use of the story. For narratives sourced from existing collections, the curator either had prior permission for the narratives to be redistributed or they approached the narrators for whom they did not already have the appropriate permission for the re-use of the narratives. The eligibility for a narrative to be adopted within the collection was assessed based on the inclusion/exclusion criteria identified through NEON task meetings and Lived Experience Advisory Panel (LEAP) workshops with any ethical uncertainties resolved during Collection Steering Group (CSG) meetings.

The NEON study LEAP workshops were chaired 13 times with attendance from 10 LEAP members, a LEAP meeting chair, and members of the NEON team. It was estimated that 60% of these workshops were necessary for the development of NEON as opposed to other research-related activities by examining the workshop agendas. The LEAP members had personal experiences of mental health problems and advised on the ethical principles of curating narratives, categorizing narrative content warnings, general issues raised by the research team, and the types of narratives to be included/excluded from the NEON Collection (28). As

TABLE 2 Development resource use and unit cost input parameters: NEON Collection curation.

Resource item	Staff members and details	Quantity consumed (range)	Method	Cost per unit* (range)	Source
Staff time to source recovery narrative	APM3	12.5 days (10–15 days)	Expert estimation	£23.90 per hr (£19.92–£28.72)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	RA	65 days (50–85 days)		£28.80 per hr (£23.18–£34.47)	
Staff time to secure ethical approval	CI	2.5 days (2–3 days)		£69.91 per hr (£46.49–£102.26)	[Glassdoor March 2022]
	SRF	12.5 days (10–15 days)		£41.34 per hr (£34.47–£49.54)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
Collection organizers	N/A	N/A	Direct observation	£2.79 per collection organizer source narrative	[Invoice to the NEON study]
LEAP workshop	APM3	23.4 hr (13.7–33.2 hrs)	Direct observation with assumption	£23.90 per hr (£19.92–£28.72)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	RA			£28.80 per hr (£23.18–£34.47)	
	SRF			£41.34 per hr (£34.47–£49.54)	
	CI			£69.91 per hr (£46.49–£102.26)	[Glassdoor March 2022]
	10 LEAP members			£20 per hr	[Internal communication with NEON study team]
	Travel/venue	13 workshops	Direct observation	£128 per LEAP member meeting	[Invoice to the NEON study]
Collection Steering Group Meetings	RA	14 hr	Direct observation	£28.80 per hr (£23.18–£34.47)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	SRF			£41.34 per hr (£34.47–£49.54)	
	4 LEAP members			£20 per hr	[Internal communication with NEON study team]
	APM3; preparation for meetings	7.25 hr		£23.90 per hr (£19.92–£28.72)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
Training researchers	APM3	5 hr (3.75–6.25 hr)	Expert estimation	£23.90 per hr (£19.92–£28.72)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	RA	5 hr (3.75–6.25 hr)		£28.80 per hr (£23.18–£34.47)	
	SRF	5 hr (3.75–6.25 hr)		£41.34 per hr (£34.47–£49.54)	
Staff time to code the narratives	Researchers	1,092 hr (819–1,638 hr)	Direct observation with assumption	£20 per hr	[Internal communication with NEON study team]

APM3, Administrative, Professional and Managerial level 3 (APM3); CI, Clinical Investigator; hr, hour; LEAP, Lived Experience Advisory Panel; RA, Research Associate; SRF, Senior Research Fellow.
As per the University of Nottingham's "Normal Full-time Working Week," it was assumed that the number of hours worked per week was 36.25 h for all professionals. Given there were 215 working days in 2021 (including 30 days annual leave, 8 public holidays, and 7 university holidays/closures), the number of hours worked per year was assumed to be 1,559 h.
*The unit costs used in the base case model contain direct salary plus on-costs [employer's national insurance contributions, employer's Universities Superannuation Scheme (USS) pension contribution, and the apprenticeship levy].

with all in-person meetings/workshops run by the NEON study all participants were paid for their attendance. Additionally, before the COVID-19 pandemic, travel expenses and hospitality (including venues and refreshments) were covered as expenses and recorded through invoices.

A CSG had the authority to make all final ethical decisions regarding the approval of narratives into the collection where the research team expressed uncertainty surrounding whether all inclusion criteria were met and/or whether an exclusion criterion was met. Moreover, the steering group could make recommended updates to the inclusion and exclusion criteria (see researchintorecovery.com/research/neon/neoncollection). This group was comprised of four LEAP members, a senior research fellow, and a research assistant. In total, there were eight meetings lasting 2 h per session with preparation for the meetings completed by an administrator; referred to as an Administrator, Professional, Managerial level 3 (APM3) within the host university. All the CSG meetings were necessary for the development of the NEON Intervention.

Once permission to use the stories were granted, the narratives were characterized using a standardized 77-item INCRESE tool (38). Researchers were trained to rate narratives to identify latent characteristics (e.g., the stage of recovery, genre) and manifest characteristics (e.g., narrator gender and

TABLE 3 Development resource use and unit cost input parameters: Web-application development.

Resource item	Staff members and details	Quantity consumed	Method	Cost per unit*	Source
Database specification	SRF	4.5 days (4–5 days)	Expert estimation	£41.34 per hr (£34.47–£49.54)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
Software development	N/A	N/A	Direct observation	100% of invoice	[Invoice to the NEON study]
Recommender system development	Work package 3.1	N/A	Direct observation with assumption	30% of budget	[NEON study budget proposal]
	SRF; integrating the algorithm into site	4 days (3–5 days)	Expert estimation	£41.34 per hr (£34.47–£49.54)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
Graphic design	N/A	N/A	Direct observation	100% of invoice	[Invoices to the NEON study]
Interaction design	SRF	4 days (3–5 days)	Expert estimation	£41.34 per hr (£34.47–£49.54)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]

hr, hour; SRF, Senior Research Fellow.
As per the University of Nottingham's "Normal Full-time Working Week," it was assumed that the number of hours worked per week was 36.25 h for all professionals. Given there were 215 working days in 2021 (including 30 days annual leave, 8 public holidays, and 7 university holidays/closures), the number of hours worked per year was assumed to be 1,559 h.
*The unit costs used in the base case model contain direct salary plus on-costs [employer's national insurance contributions, employer's Universities Superannuation Scheme (USS) pension contribution, and the apprenticeship levy].

TABLE 4 Development resource use and unit cost input parameters: Additional communication.

Resource item	Staff members and details	Quantity consumed	Method	Cost per unit* (range)	Source
Task meetings	CI	56.8 hrs (13.76–99.8 hrs)	Expert estimation	£69.91 per hr (£46.49–£102.26)	[Glassdoor March 2022]
	RA			£28.80 per hr (£23.18–£34.47)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	SRF			£41.34 per hr (£34.47–£49.54)	
	APM3			£23.90 per hr (£19.92- £28.72)	
International Advisory Board	CI	1 hr (0.75–1.25 hrs)	Expert estimation	£69.91 per hr (£46.49–£102.26)	Glassdoor March 2022]
	4 Profs			£69.91 per hr (£46.49–£102.26)	
	SRF			£41.34 per hr (£34.47–£49.54)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	2 Profs; further consultations	4 hrs (3–5 hrs)		£69.91 per hr (£46.49–£102.26)	[Glassdoor March 2022]
Intervention Design Group meetings	SRF	1.5 hrs	Direct observation	£41.34 per hr (£34.47–£49.54)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	2 LEAP members			£20 per hr	[Internal communication with NEON study team]

APM3, Administrative, Professional and Managerial level 3 (APM3); CI, Clinical Investigator; hr, hour; LEAP, Lived Experience Advisory Panel; Profs, Professors; RA, Research Associate; SRF, Senior Research Fellow.
As per the University of Nottingham's "Normal Full-time Working Week," it was assumed that the number of hours worked per week was 36.25 h for all professionals. Given there were 215 working days in 2021 (including 30 days annual leave, 8 public holidays, and 7 university holidays/closures), the number of hours worked per year was assumed to be 1,559 h.
*The unit costs used in the base case model contain direct salary plus on-costs [employer's national insurance contributions, employer's Universities Superannuation Scheme (USS) pension contribution, and the apprenticeship levy].

content warnings). Each narrative was double rated for the content warning section of INCRESE. Training to code the recovery narratives using INCRESE was conducted as a part of a 10-day pilot study to test the validity of the tool, in which 100 narratives were rated. The cost of training researchers to use the INCRESE tool only accounts for the trainers'

TABLE 5 Development resource use and unit cost input parameters: Testing the intervention.

Resource item	Staff members and details	Quantity consumed	Method	Cost per unit*	Source
Intervention testing	RA	8.5 days (7–10 days)	Expert estimation	£28.80 per hr (£23.18–£34.47)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	SRF	13.5 days (12–15 days)		£41.34 per hr (£34.47–£49.54)	
	CI	20 hrs (15–25 hrs)		£69.91 per hr (£46.49–£102.26)	
	STAT	2.5 days (2–3 days)		£50.74 per hr (£40.64–£60.46)	Queen Mary University of London Aug 2021 Salary Scales
Feasibility study Baseline	RA	3 hrs (2.25–3.75 hrs)	Expert estimation	£28.80 per hr (£23.18–£34.47)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	25 participants	3 hrs (2.25–3.75 hrs)		£20 per hr	[Internal communication with NEON study team]
	Transcript	25 transcripts	Direct observation	£72.13 per transcript	[Invoice to the NEON study]
	Travel	25 participants	Expert estimation	£5.15 per participant	[Internal communication with NEON study team]
Follow-up	RA	3hrs (2.25–3.75 hrs)	Expert estimation	£28.80 per hr (£23.18–£34.47)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	22 participants	3hrs (2.25–3.75 hrs)		£20 per hr	[Internal communication with NEON study team]
	Transcript	22 transcripts	Direct observation	£41.22 per transcript	[Invoice to the NEON study]
	Travel	22 participants	Expert estimation	£5.15 per participant	[Internal communication with NEON study team]
Analysis	RA	30 hrs (22.5–37.5 hrs)	Expert estimation	£28.80 per hr (£23.18–£34.47)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	SRF	10 hrs (7.5–12.5 hrs)		£41.34 per hr (£34.47–£49.54)	

CI, Clinical Investigator; hr, hour; RA, Research Associate; SRF, Senior Research Fellow; STAT, Senior Statistician.
As per the University of Nottingham's "Normal Full-time Working Week," it was assumed that the number of hours worked per week was 36.25 h for all professionals. Given there were 215 working days in 2021 (including 30 days annual leave, 8 public holidays, and 7 university holidays/closures), the number of hours worked per year was assumed to be 1,559 h.
*The unit costs used in the base case model contain direct salary plus on-costs [employer's national insurance contributions, employer's Universities Superannuation Scheme (USS) pension contribution, and the apprenticeship levy].

costs with the trainees' costs captured within the total coding time estimate.

For the process of coding the narratives, the researchers recorded the time taken to read, watch, or listen to the narratives. Although this provides an accurate estimate of the time taken to code the narratives, it does not contain a record of the length of breaks the researchers used during the process. Apart from rest periods being a standard part of a working day, a selection of the recovery stories may have caused an emotional or distressing impact on the researchers. Therefore, including breaks may more closely reflect the practical reality of coding the narratives (37). It was assumed that for every 6 h spent coding the coder had a 2-h break. After the recovery stories had been rated with INCRESE, they were added to the NEON Collection.

The NEON Intervention was delivered through a web-application to its users. A database specification document was required to support database implementation and source code development. Fundamental to the NEON Intervention is a recommender system that matches users with narratives intended to be of benefit. The researchers' time to conceptualize and develop the recommender system was challenging to estimate without records. The proposed budget line from the NEON trials was used as an informed approximation of the cost involved in developing the recommender system. Following this, the recommender system was integrated into the web-application codebase. As both developing and integrating the recommender system into the web-application are related tasks, they were combined into one cost.

An important component of developing the NEON Intervention was ensuring the web-application was appealing to use. Graphic designers made improvements to the aesthetical appeal of the web-application interfaces following feedback during the feasibility study and from LEAP members. The costs of keeping users engaged in the intervention (e.g., gamification, testimonials from other users, etc.) are spread between both graphic design and software development costs. Interaction design was also important

to ensure the web-application's functions operated as intended (i.e., navigations).

Throughout the NEON study, task meetings were used to communicate and raise general discussions surrounding the interventions development. An assumption was made, through inspection of meeting agendas, that 33% of task meetings were necessary for the purpose of developing the intervention (i.e., excluding research costs). A total of 86 meetings were held with recorded attendance from members of the NEON team.

The international advisory board was used to provide expert consultation on safety strategies for the development of the NEON Collection and to provide advice on intervention engagement strategies. The meeting was attended by four professors and members of the NEON study team. During the intervention development stage, there was a single Intervention Design group meeting aimed at discussing particular features of the intervention and how they could be improved. This meeting was attended by two LEAP members and a senior research fellow. As before, all meetings had expenses covered.

To ensure that the evolving intervention operated as intended, the web-applications functions were tested by researchers using dummy accounts as well as data collection tools. This included testing of forms used to collect demographics and outcome data, and interactive features included to provide access to recovery narratives. During the analysis, the full intervention testing costs were deemed necessary for the development of the NEON Intervention on the basis that monitoring routine outcomes and usage data, necessary for the NEON trials, is commonplace in clinical practice (45).

Finally, a feasibility study to evaluate a prototype delivery of the NEON Intervention in a population of people with experience of mental health problems was conducted (37). The user feedback from this feasibility evaluation led to improvements in the intervention, for example, updating the color scheme to resemble the UK NHS website less closely. Digital technologies generally test a prototype of their technology on a sample of potential users for their feedback. Therefore, the full cost of the study was deemed relevant to development of NEON.

NEON Intervention delivery resource use

Table 6 summarizes resource use and unit cost categories and sources for the components required to deliver the NEON Intervention.

The web hosting capacity of the site, together with associated cyber-security features, is supplied by Amazon Web Services Lightsail (https://aws.amazon.com/lightsail/). This service allows websites to host a specific number of users for a publicly advertised price. The current intervention is designed to host 2,000 users. Therefore, a specific level of resource use was predetermined. The invoices from the supplier were made over a monthly billing period in US Dollars (USD); the exchange rates were determined by the credit card issuer.

The cost of personnel training for the administrator, within the host university, is referred to as Administrative, Professional and Managerial level 2 (APM2). The number of days spent training the administrator was recorded for each personnel conducting the training (see Table 5). Training costs would need to be incurred every time there is an administrative staff turnover. Therefore, an assumption was made that a new administrator would need to be trained every 1.5 years. Similarly, the cost of personnel training for the researchers who rated recovery narratives using the INCRESE tool is included as a delivery cost. It was assumed that training costs would need to be incurred every year.

The NEON Intervention required new narratives to be introduced into the collection over time to maintain diversity and relevance to users. Firstly, we consulted with the NEON team to estimate that an additional 200 narratives per year would be needed based on preliminary work looking at the diversity of the current NEON Collection. The cost per narrative was calculated from the current collection size of 659 recovery stories then re-scaled to 200 narratives. Although the cost per narrative approach can provide an estimate of the cost to update the narrative collection, it explicitly assumes a linear relationship between the cost and the narrative collection size. There is uncertainty about whether the cost of updating the narrative collection will increase or decrease for newly sourced narratives. In reality, the process of updating the narratives may become more streamlined and productivity gains can be made in coding the narratives. On the other hand, the cost of updating the narrative collection may be greater if sourcing new narratives becomes more cumbersome, e.g., exhausting the number of existing collections to source narratives.

The web-application requires ongoing maintenance to ensure the NEON Intervention can be delivered as intended for its users. There are challenges in costing for web-application maintenance due to the variability in need for maintenance and the broad definition of what maintenance means in practice. In this case, we define web-application maintenance as any change, modification, or update to the web-application codebase to correct faults, to improve performance, or to update the content on the web-application.

To deliver the NEON Intervention, it was assumed that 9.75 h per week of administrative support by an APM2 is required to conduct operational tasks (e.g., intervention engagement support tasks). The hours per week was estimated by assuming a smaller proportion (50%) of the administrative support observed during the NEON trials (19.5 h per week) would be required in a routine operational setting.

Operational activities that may have influenced the NEON Intervention's effectiveness (i.e., Advertising and Engagement strategies) were also considered to be a delivery cost.

TABLE 6 Resource use and unit cost input parameters for the delivery stage.

Resource item	Staff members and details	Quantity consumed (range)	Method	Cost per unit[a] (range)	Source
Web hosting	Amazon Web Services Lightsail	Maximum capacity: 2,000 users	Direct observation	£2.40 per day[b]	[Invoice to the NEON study]
Personnel training:	APM2	2 days	Direct observation	£18.25 per hr (£15.61–£21.80)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
Administrator	APM3	2 days		£23.90 per hr (£19.92–£28.72)	
	SRF	1 day		£41.34 per hr (£34.47–£49.54)	
Personnel training: Researchers	APM3	2 hrs (1.5–2.5 hrs)	Expert estimation	£23.90 per hr (£19.92–£28.72)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	RA	2 hrs (1.5–2.5 hrs)		£28.80 per hr (£23.18–£34.47)	
	SRF	1 hr (0.75–1.25 hrs)		£41.34 per hr (£34.47–£49.54)	
	Researchers	2 days (1.5–2.5 days)		£20 per hr	[Internal communication with NEON study team]
Periodically updating narrative collection	Scaling the cost per narrative	200 narratives per year (100–300 per year)	Expert estimation	£135 per narrative[c] (£127–£142)	[Authors' calculations]
Web-application maintenance	Technician	20 days per year (20–30 days)	Expert estimation	£300 per day	[Invoice to the NEON study]
Administrative support	APM2	9.8 hrs per week (4.9–14.6 hrs)	Direct observation with assumption	£18.25 per hr (£15.61–£21.80)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
Advertisement	Paid adverts	N/A	Direct observation	£19.22 per day (duration: NEON Trial) £18.29 per day (duration: NEON-O Trial)	[Invoice to the NEON study]
	APM3	20 days (15–25 days)	Expert estimation	£23.90 per hr (£19.92–£28.72)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	RA	18 days (14–22 days)		£28.80 per hr (£23.18–£34.47)	
	SRF	2 days (1–3 days)		£41.34 per hr (£34.47–£49.54)	
Intervention engagement	RA	58 hrs (44–72 hrs)	Expert estimation	£28.80 per hr (£23.18–£34.47)	[The University of Nottingham's Aug 2021 Clinical Salary Scales]
	SRF	20 hrs (15–25 hrs)		£41.34 per hr (£34.47–£49.54)	
Safeguarding	CI	5 hrs (3.75–6.25 hrs)	Expert estimation	£69.91 per hr (£46.49–£102.26)	[Glassdoor March 2022]

APM2, Administrative, Professional and Managerial level 2 (APM2); APM3, Administrative, Professional and Managerial level 3 (APM3); CI, Clinical Investigator; hr, hour; RA, Research Associate; SRF, Senior Research Fellow.

As per the University of Nottingham's "Normal Full-time Working Week," it was assumed that the number of hours worked per week was 36.25 h for all professionals. Given there were 215 working days in 2021 (including 30 days annual leave, 8 public holidays, and 7 university holidays/closures), the number of hours worked per year was assumed to be 1,559 h.

[a]The unit costs used in the base case model contain direct salary plus on-costs (employer's national insurance contributions, employer's Universities Superannuation Scheme (USS) pension contribution, and the apprenticeship levy).

[b]The web hosting cost per day is derived by dividing the total cost from the invoiced bills (converted $USD to £GBP) by the number of days hosting services were supplied as of the time of this research (1st March 2020 to 31st March 2022). The total cost of web hosting during the 760-day period was £1,823.92 giving a cost per day at £2.40.

[c]The cost to periodically update the narratives was calculated by dividing the total cost of curating the narratives by the number of narratives in the collection at the time of this research, then rescaling to the selected narrative per year amount. The total cost of curating the narratives was £88,640.09 (low: £83,954.30; high: £93,325.89) and the number of narratives currently in the collection was 659 giving a cost per narrative at £135.

Advertisement for the study following agreed advertising principles (46) to recruit eligible participants for both NEON trials. Since the effectiveness and types of adverts may have impacted upon the effectiveness of the intervention it is included as a delivery cost. To ensure users of the NEON Intervention were making use of the intervention, engagement strategies (such as message prompting) were used to encourage use. As these strategies encourage the use of the intervention, they may impact the effectiveness of the intervention and are therefore a part of the delivery cost.

TABLE 7 Sensitivity and scenario analysis plan.

Resource item	Baseline assumption(s)	Sensitivity analysis
Development stage		
Curation of the narrative collection		
Staff time to source recovery narrative	Midpoint resource use	Max/Min resource use
Staff time to secure ethical approval	Midpoint resource use	Max/Min resource use
LEAP workshops	60% resource use	+(-) 25% resource use
Training researchers	Midpoint resource use	Max/Min resource use
Web-application development		
Database specification	Midpoint resource use	Max/Min resource use
Recommender system	30% of budget	+(-) 25% resource use
Interaction design	Midpoint resource use	Max/Min resource use
Additional communication		
Task meetings	33% resource use	+(-) 25% resource use
International advisory board	Resource use estimate	+(-) 25% resource use
Testing the intervention		
Intervention testing	Midpoint resource use	Max/Min resource use
Feasibility study	Resource use estimate	+(-) 25% resource use
Delivery stage		
Personnel training (administrator)	Turnover period every 1.5 years	0.5-2.5 years
Personnel training (researcher)	Turnover period every 1 year	0.5-1.5 years
Periodically updating the narrative collection	200 narratives per year	100-300 narratives per year
Web-application maintenance	Midpoint resource use	Max and min resource use
Administrative support (APM2)	50% resource use	+(-) 25% resource use
Advertising	Midpoint resource use	Max/Min resource use
Intervention engagement	Midpoint resource use	Max/Min resource use
Safeguarding	Midpoint resource use	Max/Min resource use
Scenario analysis		
Resource use	Midpoint resource use	Max and min resource use
Wage per hour	Mid-spline salary	Max and min spline salary
Salary	Direct salary plus on-costs*	Direct salary only and direct salary plus on-costs* and overheads
Hours worked per week	36.25	31.25–41.25
Staff time to code the narrative	2-h breaks	0–4-h breaks
Impact of the number of users	No. of users in the intervention arms of NEON and NEON-O Trials	500, 1,000, 2,000 users
Best/worst case	Baseline assumptions	Optimistic/pessimistic assumptions

LEAP, Lived Experience Advisory Panel.
*Wages contain direct salary plus on-costs [employer's national insurance contributions, employer's Universities Superannuation Scheme (USS) pension contributions, and the apprenticeship levy].

Throughout the delivery of the NEON Intervention, the NEON study was responsible for the wellbeing of those using the NEON Intervention. Safeguarding concerns were dealt with during the trials by the clinical principal investigator.

Sensitivity analysis plan

To assess the impact of our assumptions and parameter uncertainty, all input parameters were varied to their extreme values in one-way sensitivity analysis. To examine the impact of the structural assumptions, scenario analysis was used. The assumptions that were made during the base-case analysis for both the development and delivery of the NEON Intervention and the sensitivity and scenario analyses are shown in Table 7.

Results

Costs of developing the NEON Intervention

In the base-case, the total number of hours to develop the intervention (excluding resource use external to the NEON team) was 2,709 h (45.2 days). The resource items that required the most personnel time were the staff time to code the narrative

TABLE 8 Cost of developing the NEON Intervention (base-case).

Resource item	Costs (£, 2020/21)*
Curation of the NEON Collection	
Staff time to source recovery narrative	15,740
Staff time to secure ethical approval	5,013
Collection organizers	1,650
LEAP workshops	41,372
CSG meetings including preparation for meetings	2,555
Training researchers	470
Coding narratives using INCRESE	21,840
Web-application development	
Database specification	1,349
Software development	49,279
Recommender system and integrating into the codebase	7,531
Graphic design	4,560
Interaction design	1,349
Additional communication	
Task meetings	12,575
International Advisory Board meetings	670
Intervention Design Group meetings	131
Testing the intervention	
Intervention testing	8,139
Feasibility study	8,629
Total cost	**182,851**

CSG, Collection Steering Group; INCRESE, inventory of characteristics of recovery stories; LEAP, Lived Experience Advisory Panel.
*Costs have been rounded.

collection (1,092 h); the staff time to source recovery narratives (562 h); task meetings (341 h); and LEAP workshops (328 h). The members of staff that contributed the most hours include the coding researchers (1,092 h), the research assistants (895 h), and the senior research fellows (394 h). The cost per unit of staff time varied from £18.25 to £69.91 per hour.

A summary of the estimated costs of developing the NEON Intervention is provided in Table 8. The total cost of developing the NEON Intervention was £182,851. The largest cost components include software development (27%); LEAP workshops (23%); coding the narratives using the INCRESE tool (9%); and researchers' time to source narratives (9%). The total cost of curating the narrative collection was £82,710. The majority of this cost is attributed to the LEAP meetings (50%), coding the narratives using the INCRESE tool (20%), and the researchers' time to source narratives (19%). The total cost of developing the web-application was £64,067. The largest contributions to the development cost of the web-application were the software development (77%); developing the recommender system (12%); and the graphic design (7%).

Costs of delivering the NEON Intervention

In the base-case, the total number of personnel hours to deliver the intervention during the NEON trial was 1,708 h (28.5 days). In the NEON-O trial, the total number of personnel hours to deliver the intervention was 1,776 h (29.6 days). The resource items that require the most personnel time were the administrative support (54%); maintenance (19%); advertising (17%). The cost per unit of resource use varied between £2.40 and £125.

A summary of NEON Intervention delivery costs is provided in Table 9. The total cost of delivering the NEON Intervention during the NEON trial was £118,663 (£321 per user). In the NEON-O trial, the total delivery cost of the NEON Intervention was £123,444 (£241 per user). Therefore, the total delivery cost during the NEON trial was 4% lower than during the NEON-O trial. However, the cost per user during the NEON trial was 33% higher than the NEON-O trial. The proportion of fixed costs (advertising, engagement, and safeguarding) was 22% during the NEON trial compared to 21% during the NEON-O trial. The largest cost components include updating the narrative collection (50%); advertising (19%); administration (14%); and software maintenance (11%). The cost of delivering the intervention for a 1-year period is £68,521.

Sensitivity analysis

The one-way sensitivity analysis of the most sensitive cost components for NEON Intervention development are presented as a Tornado diagram in Figure 1. Uncertainty in the cost of the LEAP meetings had the greatest impact on the base-case estimate. Specifically, the total cost may be 5.8% higher or lower than the base-case estimate. The difference between the extreme values of the LEAP meetings, task meetings and recommender system development are £21,041; £17,910; and £11,513, respectively. Varying cost components such as intervention testing and efforts to secure ethical approval had a relatively smaller effect on the total development cost. By varying these costs to their extreme values, the impact on total cost of intervention testing and effort to secure ethical approval was +/- £1,296 and +/-£1,003, respectively. Other components had comparatively little effect on the overall development cost such as the database specification (+/- £150).

The one-way sensitivity analysis of the cost components for the NEON Intervention delivery during the NEON trial are presented as a Tornado diagram in Figure 2. Uncertainty in the cost of administration was shown to have the largest effect on the base-case estimate. Specifically, the total cost may be 7% higher or lower than the base-case estimate. Cost components such as the web-application maintenance and updating the narrative

TABLE 9 Cost of delivering the NEON Intervention (base-case).

Resource item	NEON (£, 2020/21)*	NEON-O (£, 2020/21)*	Yearly (£, 2020/21)*
Web hosting	1,907	2,005	875
Personnel training (administrator)	1,515	1,593	696
Personnel training (researchers)	951	1,000	437
Periodic updates to the narrative collection	58,593	61,615	26,901
Web-application maintenance	13,068	13,742	6,000
Administrative support	16,669	17,529	7,653
Advertisement	23,111	23,111	23,111
Intervention engagement	2,497	2,497	2,497
Safeguarding	350	350	350
Total cost*	**118,663**	**123,444**	**68,521**

*Costs have been rounded.

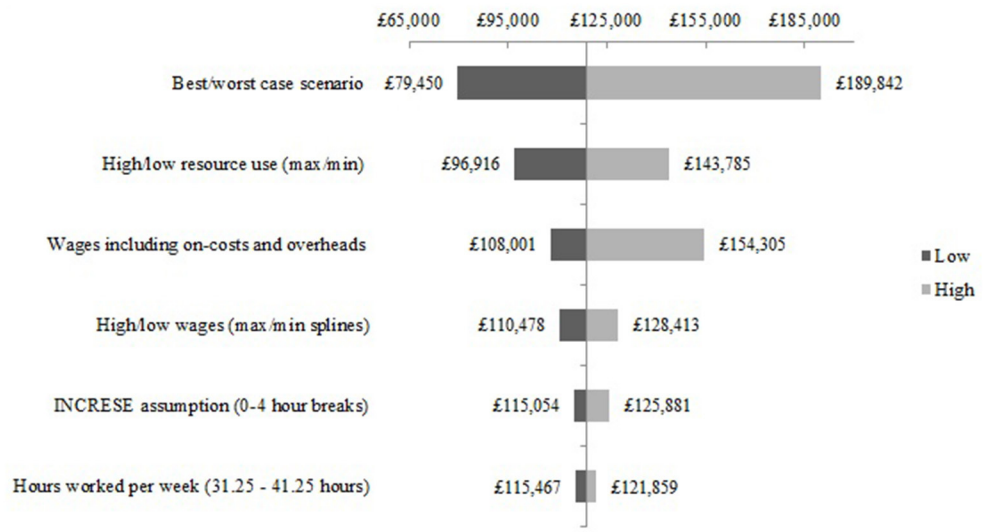


FIGURE 1
One-way sensitivity analysis - NEON Intervention development costs.

collection had a relatively smaller effect on delivery costs. For example, the impact on the total delivery cost of varying web-application maintenance to its extreme values was £6,534 higher or lower than the base case estimate. Cost components such as engagement had a relatively small effect on the delivery of the NEON Intervention during the NEON trial. The impact on the total delivery cost of varying the engagement component to its extreme values was 0.5% higher or lower than the base case estimate. Similar results can be seen in the one-way sensitivity analysis of the cost components for the NEON Intervention delivery during the NEON-O trial (see Figure 3).

The scenario analysis of the structural assumptions for developing the NEON Intervention are presented as a Tornado diagram in Figure 4. The impact on the total cost of the best and worst case scenario analyses show a feasible total cost range to develop the NEON Intervention. The cost of developing the NEON Intervention given the best possible scenario is £54,724 lower than the base-case estimate of £181,851. Similarly, given the worst possible scenario, the cost of developing the NEON Intervention is £60,805 higher than the base-case estimate.

The scenario analysis of the structural assumptions for the NEON Intervention delivery during the NEON trial are presented in Figure 5. The best-case scenario is £39,213 lower than the base-case estimate, and the worst-case scenario cost is £71,180 higher than the base-case estimate. Similar results can be seen in the scenario analysis of the structural assumptions for the NEON Intervention delivery during the NEON-O trial (see Figure 6).

As expected, increasing the number of users, reduced the cost per user year, such that 500, 1,000, or 2,000 users cost £137.04, £68.52, £34.26, respectively.

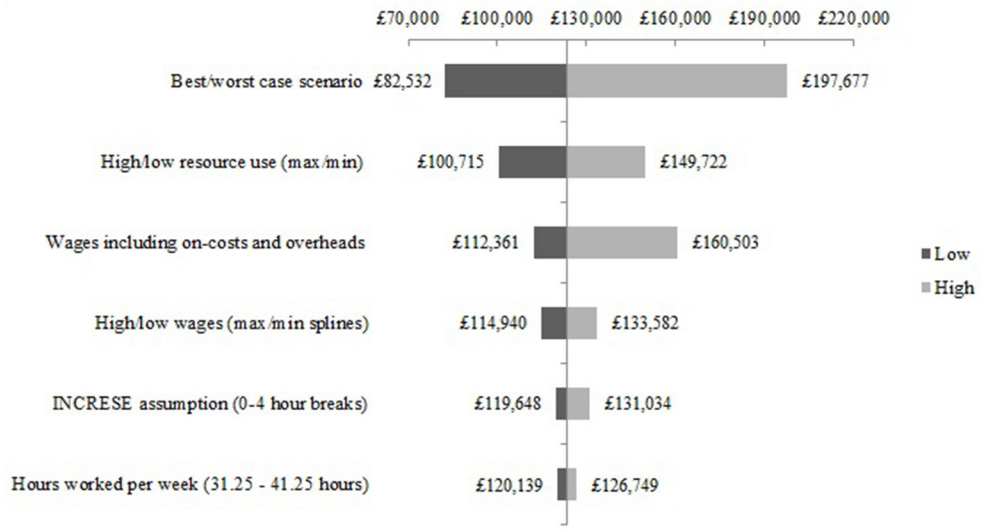


FIGURE 2
One-way sensitivity analysis - NEON Intervention delivery costs during the NEON trial.

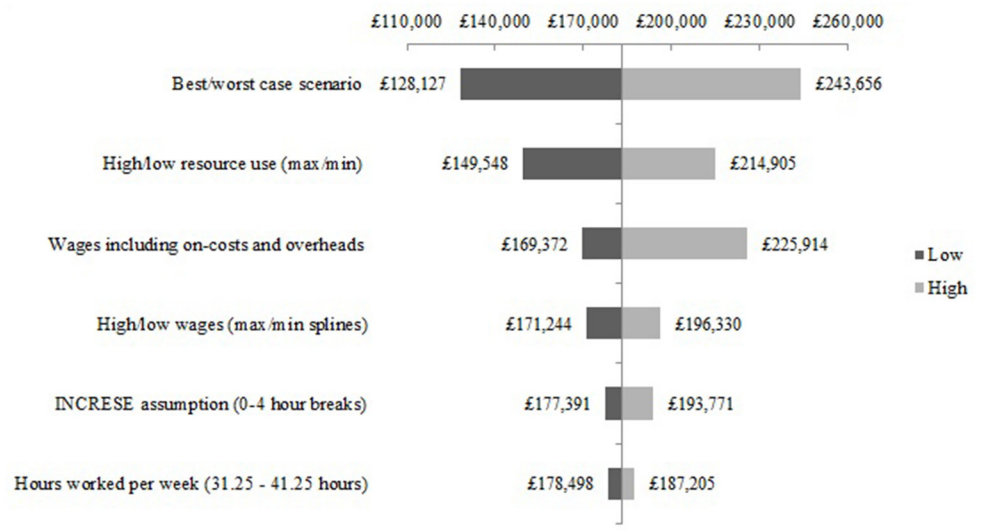


FIGURE 3
One-way sensitivity analysis - NEON Intervention delivery costs during the NEON-O trial.

Discussion

Key findings

We identified several resource use categories for intervention development and delivery through a literature review and work with experts on the research team. Developing the NEON Intervention cost £182,851, which was largely attributed to building the NEON Collection (£82,710) and developing the web-application (£64,067). The curation costs were mostly made up of LEAP meetings (50%); narrative coding using the INCRESE tool (20%); and the researchers' time to source recovery narratives (19%). The largest components of the web-application costs included software development (77%); developing the recommender system (12%); and the graphic design costs (7%).

Delivering the NEON Intervention during the NEON trial and NEON-O trial costed £118,663 and £123,444, respectively. This equates to £349 (NEON trial) and £241 (NEON-O trial) per user. Over an annual period, the NEON Intervention cost was £68,521. Delivery costs were driven by updating the narrative collection (50%); advertisement (19%); and administrative

FIGURE 4
Scenario analysis - NEON Intervention development costs.

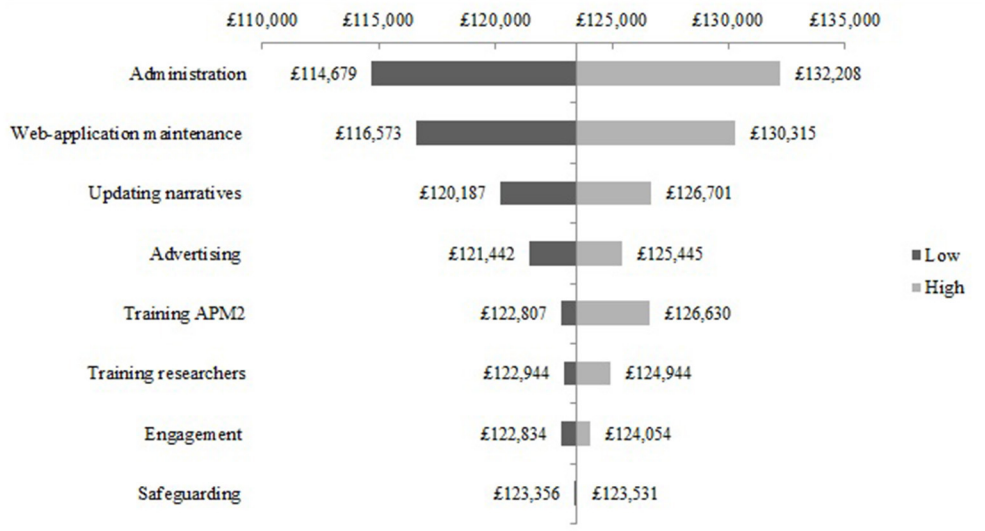


FIGURE 5
Scenario analysis - NEON Intervention delivery costs during the NEON trial.

support. Similarly to other studies (20), we found that total costs are dependent on usage rates, which are difficult to predict. Due to a high proportion of fixed costs, costs per user were high for the lower caseload in NEON-O but would be expected to decline with increasing volume of use.

Strengths and limitations

The strength of our study is the use of published guidelines where available to inform our costing methods. We have provided detailed and transparent reporting of cost components and we have utilized multi-disciplinary input into identifying categories, including strong service user involvement. We have carried out sensitivity analysis to make explicit which parameters are associated with uncertainty, and the direction and magnitude of that uncertainty.

The limitations of our study include the necessary use of expert opinion sources where primary data were not collected or available. Due to limitations in patient-level resource use data availability, we have had to use top-down costing methods rather than bottom-up, or micro-costing approaches, which arguably

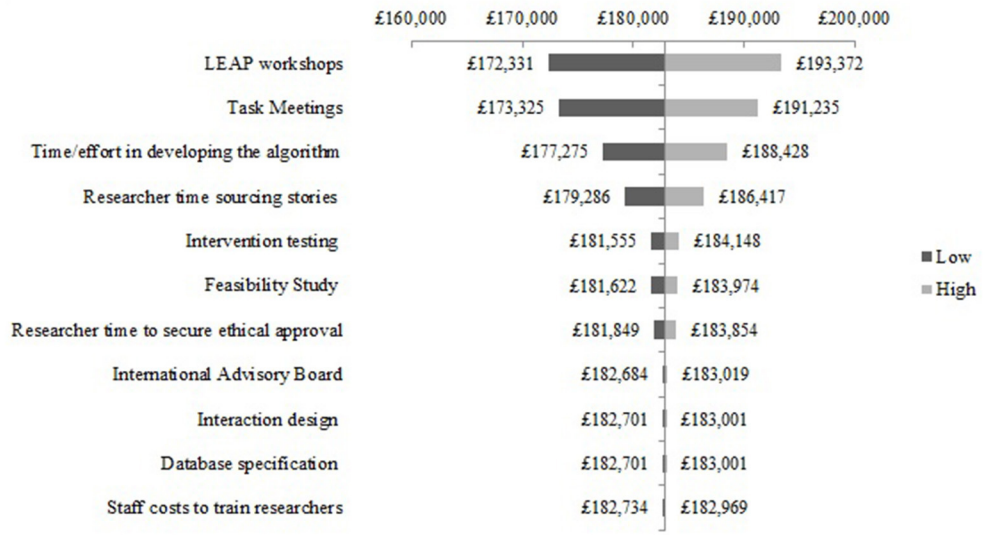


FIGURE 6
Scenario analysis - NEON Intervention delivery costs during the NEON-O trial.

does not provide a patient-level cost sensitive to different levels of usage by individual patients. This limitation is common to costing DHIs, due to the nature of the costs involved (47).

Defining a starting point for development costs necessarily meant excluding early development costs. As with other research-led endeavors, this work is supported by the accumulation of an existing knowledge base and the experience and expertise of the research team, as well as the opportunity cost of the researchers. This can be conceptualized using the health service research payback model which examines the complex interaction and costs of early and late publicly-funded research, and their effect on research outputs (48). There has been a substantial program of earlier research to support the development and testing of the NEON Intervention, which is outside the scope of this work to quantify. Other costs which were outside the scope include the intellectual work involved in the development of the NEON proposal for funding, the other types of resource such as existing collaboration networks which could be accessed for advice, and work conducted in other studies in the research group which may have cross-fertilized the NEON work. We also did not include the RCT costs in our development costs as we explicitly examined development costs up to the point of the beginning of the RCTs. The impact that the trial's findings will have upon the NEON Intervention is currently uncertain. As such, it is unclear what changes the NEON trials may inform to improve, scale, or discontinue the intervention in its current form. Therefore, trial costs were not considered as a development cost during the current development period.

The delivery costs provided here are derived from the delivery of the NEON Intervention during the trials, so it is likely that delivery costs in practice will be different, and health care payers need to be prepared for costs to vary once the intervention is implemented outside a clinical trial environment. How the NEON Intervention is implemented in the NHS will likely determine what resources will be needed in future versions of the intervention. As the NEON trials do not compare the NEON Intervention with a face-to-face version of the same intervention, we are not able to examine the differences in costs (or effectiveness) between these two scenarios.

Assumptions had to be made around tasks and roles taken up by researchers in the trial that would actually be covered by healthcare professionals, administrators, and other members of the healthcare team when the intervention is delivered in practice. Necessarily there are likely to be some differences between trial delivery and practice delivery. For example, the clinical principal investigator coordinated all safeguarding activities during NEON Intervention delivery as part of the RCT. If the NEON Intervention is rolled out in practice, at a substantial scale, then the safeguarding approaches and infrastructure needed will necessarily be different. Safeguarding approaches will need to scale with safeguarding demands, and to take into account different regulatory requirements in everyday practice compared to clinical trials. We have treated advertising costs in the trials as a proxy for accessing patients in practice. Recruitment for the NEON Trial was targeted carefully, for example through the production of several 100 online messages with content specific to a psychosis trial and the design and dissemination of targeted adverts (for example displayed as banners on websites used by health and social care professionals and potential participants). This involved human effort and expertise to design these materials. If the NEON Intervention is rolled out to a general mental health population on a larger scale through online mechanisms, then less targeting of recruitment

material might be needed, and individual messages or adverts can also be reused and redeployed. This means that the human cost of generating recruitment material might be less and the dominant cost might be spent on the services that social media companies provide to promote messages to a relevant audience. Especially at the point where the NEON Intervention is deployed on a wider scale, the cost of recruiting one person through social media promotion should be routine to estimate.

Implementation of the NEON Intervention beyond the trials will lie on a spectrum. On the one end, the implementation through statutory mental health services (costs associated with staff awareness and training, implementing safeguarding procedures etc.), through to the implementation *via* primary care or the voluntary sector, to "direct to consumer" (costs associated with advertising), and the value for money offered by these different implementation routes (which may be separate or additive) that could be evaluated in relation to reach and engagement. The most efficient way to implement DHIs like the NEON Intervention is very relevant in a resource-constrained system as implementation methods cost money and affect the effectiveness and cost-effectiveness of the DHI (49).

Other limitations in the delivery costs include the lack of regulatory or scale-up costs, such that these costs may be an underestimate of true costs once the NEON Intervention is implemented in practice. The derivation of costs per user from top-down costs necessarily means that the cost per user is related to the number of users assumed in the calculations. The number of users could be the number of people in each trial, the number of people eligible to use the NEON Intervention in the real world (21), or we would suggest, the maximum capacity of the current technology implementation to provide an appropriate level of quality of service (50) to its users (for example with sufficient responsiveness of interactive features to allow for a satisfying engagement). If we use the number of people in each trial to estimate cost per user, the mean cost per user could be overestimated as the maximum technological capacity consistent with acceptable service quality has not been reached. However, using the number of people eligible to use the NEON Intervention in the real world is not straightforward. This is for a number of reasons: the computational complexity of the algorithms used in the technology may not scale linearly with the size of the user base (meaning that needed server performance and hence server cost may not scale linearly), and at discrete points, technology re-engineering may be needed to maintain service quality, for example by replacing propriety content delivery system with commercially-provided Content Delivery Networks (51). This reflects the development cycle for web-based DHIs. It is similar to most start-ups that proceed through a series of versions of their systems, re-engineering them each time for a larger number of users. Costs associated with capacity-related engineering and enhanced server capacity should be spread across the anticipated number of users, and allocated equally per user. For example, additional costs required to scale the intervention from 2,000 to 5,000 users should be spread equally across those 3,000 extra users to prevent spikes in costs per user. In the final report for the NEON trials, we will examine this issue further and look at scenarios for future costs for the NEON Intervention as it is scaled up to a range of anticipated user base sizes.

Given the evolving nature of DHIs, input parameters into economic evaluations (like delivery costs) need to be re-examined at different stages of the intervention capacity. It is likely that certain cost components will change over time. Hosting costs are likely to reduce due to increases in process performance and reductions in storage cost per unit (Moore's Law) (52), and salaries may increase (given human resources have quite a large impact on the delivery, this is important).

Another issue beyond the scope of this study is the likelihood that the intervention will be delivered differently once it is implemented in practice. This will have an impact not only on resource use but also on effectiveness of the intervention. However, the data we have provided will provide healthcare providers with approximations of the resources required to deliver the intervention in practice.

Comparisons with existing evidence

We are not aware of a DHI that has focused on the use of narratives to support recovery in mental health, so it is not possible to compare our development and delivery costs with another equivalent intervention or directly relevant costing study. It is also unclear as to the utility of comparing the costs we have estimated with other mental health DHIs costs, due to the lack of comparability of costing methods between studies. The reporting of different development and delivery cost components of mental health DHIs varies, making meaningful comparison difficult (11, 15).

Development costs are reported in a minority of economic evaluations, as they are usually seen as sunk costs and not relevant to the health and social care provider. The list price of commercially developed healthcare products such as pharmaceuticals are perceived to include development costs and thus recoup those costs. Development costs are more difficult to determine for DHIs. Opinions differ on the inclusion of development costs, and it is the perspective that should determine whether they are included or not (47). When development costs are included, a judgement has to be made regarding when development costs are considered to begin and end, as we had to make in our study, and this judgment can have a significant effect on the overall costs derived.

When reported, development costs vary from as little as £19,000 (53) to £500,000 (54), but whether this variation stems from the varied nature of the DHIs or from the methods used to collect resource use data is challenging to untangle. However, they can influence whether development occurs, so explicit

methods for deciding on resource categories to include, methods of resource use data collection and sources of unit costs should be explicitly given in any study. We used a similar approach to Joshi and colleagues who reported explicit methods to derive development cost for a digital program for training community health workers to deliver treatment for depression in rural India (22). Similarly to our study, they reported that staffing costs constituted 61% of development costs.

In a recent review of methods used in economic evaluations of mental health DHIs, 16 out of 66 did not report staffing costs as part of delivery costs (15). Given the human resource intensive nature of DHI delivery as found in our study, this omission would lead to a significant underestimate of costs. Only 14 out of 66 studies included costs for website maintenance and hosting, again we found that this constituted a significant delivery cost.

We used the NICE guidance to support our costing methodology. However, there have been questions raised surrounding how fit for purpose this guidance is for costing DHIs in practice (21). Like other economic evaluation guidelines, although they provide a costing framework, there is little practical guidance on specific costing (47). We identified several resource use categories in intervention development and delivery, and future developers of DHIs should consider including these categories, and working with relevant experts to identify any further categories specific to their intervention. We also recommend keeping formal resource use records to allow easier derivation of costs. Since DHIs (in keeping with commercial web-applications) are typically subject to ongoing periods of maintenance punctuated by periodic redevelopment and reengineering work which will have an impact on the cost per user, we would advocate for the collection of a broad range of case studies providing evidence on the cost over the life of a DHI of ongoing development work. We have excluded conceptual work conducted before the NEON program was funded, and in the early stages of the NEON program. Future studies might consider approaches to costing in such activities.

Human resources constituted a large proportion of our development and delivery costs. These estimates were affected by uncertainties around which costs to include, other than salary, so we followed the PSSRU approach to include employers' costs, as well as estates and indirect costs for the organization employing that person (55). Our sensitivity analysis demonstrated that inclusion of these latter costs had a significant effect on overall costs.

Conclusion

This study makes two knowledge contributions. First, it provides a usable estimate of the cost of developing and implementing a DHI from software and intervention development, through feasibility testing and up to commencement of the definitive trial. This can be used to inform commissioning of new DHIs in general, giving explicit consideration to all the different types of resources required, and the quantity and cost, as well as specifically, the implementation of the NEON Intervention. Second, the costing challenges which have been identified indicate the need for updated best practice guidance for economic evaluation of DHIs by NICE and other clinical and funding agencies.

Relevance for clinical practice

The NEON Intervention is intended for widespread use as a low-cost self-management intervention. Two uses are envisaged: adjunctive to clinical treatment and direct access. This study identifies the costs associated with population-level roll-out of the NEON Intervention. In relation to use within services, the staff costs needed to support access are identified. In relation to direct access, the public health costs associated with maintaining and developing the intervention are estimated. We are currently evaluating the effectiveness of the NEON intervention based on the two RCTs described in this study. Data analysis is near completion and will be published separately building on the findings in this study. These findings will inform decision-making about whether, and how, to implement the NEON Intervention at scale.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Leicester Central Research Ethics Committee, 19/EM/0326. Informed consent to participate was obtained from all participants using an online form. The patients/participants provided their written informed consent to participate in this study.

Author contributions

RE, SG, LP, and SR-E contributed to the conception and design of the study. LP carried out the data collection and analysis and wrote the first draft of the manuscript. SR-E, SG, and RE wrote sections of the manuscript. JL-B, FN, CB, AG, JN, DQ, and SB contributed to the data collection. JL-B and FN provided feedback on the analysis. MS is the chief investigator

for this trial. RE is the lead health economist. All authors have read, commented, and reviewed the manuscript.

Funding

This study was funded by the NIHR [Personal experience as a recovery resource in psychosis: Narrative Experiences Online (NEON) Programme (RP-PG-0615-20016)]. The funder was not involved in the development of the methods, data collection, analysis or writing of this manuscript.

Acknowledgments

MS acknowledges the support of NIHR Nottingham Biomedical Research Center.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Author disclaimer

The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.1028156/full#supplementary-material

References

1. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. *Br J General Practice.* (2021) 71:e166–77. doi: 10.3399/BJGP.2020.0948

2. Neves AL, Li E, Serafini A, Jimenez G, Lingner H, Koskela TH, et al. Evaluating the impact of COVID-19 on the adoption of virtual care in general practice in 20 countries (inSIGHT): protocol and rationale study. *JMIR Res Protoc.* (2021) 10:e30099. doi: 10.2196/30099

3. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin Psychol Rev.* (2012) 32:329–42. doi: 10.1016/j.cpr.2012.02.004

4. Himle JA, Weaver A, Zhang A, Xiang X. Digital mental health interventions for depression. *Cogn Behav Pract.* (2022) 29:50–9. doi: 10.1016/j.cbpra.2020.12.009

5. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: digital health interventions for children and young people with mental health problems–a systematic and meta-review. *J Child Psychol Psychiatry.* (2017) 58:474–503. doi: 10.1111/jcpp.12663

6. Simblett S, Birch J, Matcham F, Yaguez L, Morris R. A systematic review and meta-analysis of e-mental health interventions to treat symptoms of posttraumatic stress. *JMIR Ment Health.* (2017) 4:e14. doi: 10.2196/mental.5558

7. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. *JMIR Ment Health.* (2020) 7:e16460. doi: 10.2196/16460

8. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press (2015). ISBN: 0191643580.

9. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, et al. Attitudes towards digital treatment for depression: a European stakeholder survey. *Internet Intervent.* (2017) 8:1–9. doi: 10.1016/j.invent.2017.01.001

10. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. *Psychol Med.* (2015) 45:3357–76. doi: 10.1017/S0033291715001427

11. Mitchell LM, Joshi U, Patel V, Lu C, Naslund JA. Economic evaluations of internet-based psychological interventions for anxiety disorders and depression: a systematic review. *J Affect Disord.* (2021) 284:157–82. doi: 10.1016/j.jad.2021.01.092

12. Paganini S, Teigelkotter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. *J Affect Disord.* (2018) 225:733–55. doi: 10.1016/j.jad.2017.07.018

13. Kass AE, Balantekin KN, Fitzsimmons-Craft EE, Jacobi C, Wilfley DE, Taylor CB. The economic case for digital interventions for eating disorders among United States college students. *Int J Eat Disord.* (2017) 50:250–8. doi: 10.1002/eat.22680

14. Buntrock C, Kählke F, Smit F, Ebert DD. A systematic review of trial-based economic evaluations of internet- and mobile-based interventions for substance use disorders. *Eur J Public Health.* (2021) 31(Suppl.1):i19-28. doi: 10.1093/eurpub/ckz022

15. Jankovic D, Saramago Goncalves P, Gega L, Marshall D, Wright K, Hafidh M, et al. Cost effectiveness of digital interventions for generalised anxiety disorder: a model-based analysis. *PharmacoEconomics.* (2021) 21:318. doi: 10.1007/s41669-021-00318-y

16. Bergmo TS. How to measure costs and benefits of ehealth interventions: an overview of methods and frameworks. *J Med Internet Res.* (2015) 17:e254. doi: 10.2196/jmir.4521

17. Neher M, Nygårdh A, Broström A, Lundgren J, Johansson P. Perspectives of policy makers and service users concerning the implementation of eHealth in Sweden: interview study. *J Med Internet Res.* (2022) 24:e28870. doi: 10.2196/28870

18. Lennon MR, Bouamrane M-M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. *J Med Internet Res.* (2017) 19:e42. doi: 10.2196/jmir.6900

19. NICE. *Evidence Standards Framework (ESF) for Digital Health Technologies*. National Institute for Health and Care Excellence (2018). Available online

at: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies (accessed July 12, 2022).

20. Huter K, Krick T, Rothgang H. Health economic evaluation of digital nursing technologies: a review of methodological recommendations. *Health Econ Rev.* (2022) 12:35. doi: 10.1186/s13561-022-00378-8

21. Gomes M, Murray E, Raftery J. Economic evaluation of digital health interventions: methodological issues and recommendations for practice. *Pharmacoeconomics.* (2022) 40:367–78. doi: 10.1007/s40273-022-01130-0

22. Joshi U, Naslund JA, Anand A, Tugnawat D, Vishwakarma R, Bhan A, et al. Assessing costs of developing a digital program for training community health workers to deliver treatment for depression: a case study in rural India. *Psychiatry Res.* (2022) 307:114299. doi: 10.1016/j.psychres.2021.114299

23. Llewellyn-Beardsley J, Rennick-Egglestone S, Callard F, Crawford P, Farkas M, Hui A, et al. Characteristics of mental health recovery narratives: Systematic review and narrative synthesis. *PLoS ONE.* (2019) 14:e0214678. doi: 10.1371/journal.pone.0214678

24. Llewellyn-Beardsley J, Rennick-Egglestone S, Bradstreet S, Davidson L, Franklin D, Hui A, et al. Not the story you want? Assessing the fit of a conceptual framework characterising mental health recovery narratives. *Social Psychiatry Psychiatric Epidemiol.* (2020) 55:295–308. doi: 10.1007/s00127-019-01791-x

25. Rennick-Egglestone S, Morgan K, Llewellyn-Beardsley J, Ramsay A, McGranahan R, Gillard S, et al. Mental health recovery narratives and their impact on recipients: systematic review and narrative synthesis. *Can J Psychiatry.* (2019) 2019:706743719846108. doi: 10.1177/0706743719846108

26. Rennick-Egglestone S, Ramsay A, McGranahan R, Llewellyn-Beardsley J, Hui A, Pollock K, et al. The impact of mental health recovery narratives on recipients experiencing mental health problems: qualitative analysis and change model. *PLoS ONE.* (2019) 14:e0226201. doi: 10.1371/journal.pone.0226201

27. Ng F, Charles A, Pollock K, Rennick-Egglestone S, Cuijpers P, Gillard S, et al. The mechanisms and processes of connection: developing a causal chain model capturing impacts of receiving recorded mental health recovery narratives. *BMC Psychiatry*. (2019) 19:2405. doi: 10.1186/s12888-019-2405-z

28. Ng F, Newby C, Robinson C, Llewellyn-Beardsley J, Yeo C, Roe J, et al. How do recorded mental health recovery narratives create connection and improve hopefulness? *J Mental Health.* (2022) 31:273–80. doi: 10.1080/09638237.2021.2022627

29. McGranahan R, Rennick-Egglestone S, Ramsay A, Llewellyn-Beardsley J, Bradstreet S, Callard F, et al. Curation of mental health recovery narrative collections: systematic review and qualitative synthesis. *JMIR Mental Health.* (2019) 6:e14233. doi: 10.2196/14233

30. Yeo C, Hare-Duke L, Rennick-Egglestone S, Bradstreet S, Callard F, Hui A, et al. The VOICES typology of curatorial decisions in narrative collections of the lived experiences of mental health service use, recovery, or madness: qualitative study. *JMIR Mental Health.* (2020) 7:e16290. doi: 10.2196/16290

31. Yeo C, Rennick-Egglestone S, Armstrong V, Borg M, Franklin D, Klevan T, et al. Uses and misuses of recorded mental health lived experience narratives in healthcare and community settings: systematic review. *Schizophr Bull.* (2022) 48:134–44. doi: 10.1093/schbul/sbab097

32. Slade M, Rennick-Egglestone S, Blackie L, Llewellyn-Beardsley J, Franklin D, Hui A, et al. Post-traumatic growth in mental health recovery: qualitative study of narratives. *BMJ Open.* (2019) 9:e029342. doi: 10.1136/bmjopen-2019-029342

33. Ng F, Ibrahim N, Franklin D, Jordan G, Lewandowski F, Fang F, et al. Post-traumatic growth in psychosis: a systematic review and narrative synthesis. *BMC Psychiatry*. (2021) 21:3. doi: 10.1186/s12888-021-03614-3

34. McGranahan R, Jakaite Z, Edwards A, Rennick-Egglestone S, Slade M, Priebe S. Living with psychosis without mental health services: a narrative interview study. *BMJ Open.* (2021) 11:e045661. doi: 10.1136/bmjopen-2020-045661

35. Hui A, Rennick-Egglestone S, Franklin D, Walcott R, Llewellyn-Beardsley J, Ng F, et al. Institutional injustice: implications for system transformation emerging from the mental health recovery narratives of people experiencing marginalisation. *PLoS ONE.* (2021) 16:e0250367. doi: 10.1371/journal.pone.0250367

36. Roe J, Brown S, Yeo C, Rennick-Egglestone S, Repper J, Ng F, et al. Opportunities, enablers, and barriers to the use of recorded recovery narratives in clinical settings. *Front Psychiatry*. (2020) 11:589731. doi: 10.3389/fpsyt.2020.589731

37. Slade M, Rennick-Egglestone S, Llewellyn-Beardsley J, Yeo C, Roe J, Bailey S, et al. Recorded mental health recovery narratives as a resource for people affected by mental health problems: development of the narrative experiences online (NEON) intervention. *JMIR Format Res.* (2021) 5:e24417. doi: 10.2196/24417

38. Llewellyn-Beardsley J, Barbic S, Rennick-Egglestone S, Ng F, Roe J, Hui A, et al. INCRESE: development of an inventory to characterize recorded mental health recovery narratives. *J Recovery Ment Health.* (2020) 3:25–44. Available online at: https://jps.library.utoronto.ca/index.php/rmh/article/view/34626

39. Helsper EJ. A corresponding fields model for the links between social and digital exclusion. *Commun Theory.* (2012) 22:403–26. doi: 10.1111/j.1468-2885.2012.01416.x

40. Rennick-Egglestone S, Elliott R, Smuk M, Robinson C, Bailey S, Smith R, et al. Impact of receiving recorded mental health recovery narratives on quality of life in people experiencing psychosis, people experiencing other mental health problems and for informal carers: Narrative Experiences Online (NEON) study protocol for three randomised controlled trials. *Trials.* (2020) 21:661. doi: 10.1186/s13063-020-04428-6

41. Rennick-Egglestone S, Elliott R, Newby C, Robinson C, Slade M. Impact of receiving recorded mental health recovery narratives on quality of life in people experiencing non-psychosis mental health problems (NEON-O Trial): updated randomised controlled trial protocol. *Trials*. (2022) 23:6027. doi: 10.1186/s13063-022-06027-z

42. Chapel JM, Wang G. Understanding cost data collection tools to improve economic evaluations of health interventions. *Stroke Vasc Neurol.* (2019) 4:214. doi: 10.1136/svn-2019-000301

43. Eschenbach TG. Technical note: constructing tornado diagrams with spreadsheets. *Eng Economist.* (2006) 51:195–204. doi: 10.1080/00137910600695676

44. Husereau D, Drummond M, Augustovski F, De Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *Br Med J.* (2022) 2022:e067975. doi: 10.1136/bmj-2021-067975

45. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. *Psychother Res.* (2015) 25:6–19. doi: 10.1080/10503307.2013.817696

46. Rennick-Egglestone S. Principles for the production and dissemination of recruitment material for three clinical trials of an online intervention. *Trials*. (2021) 22:4. doi: 10.1186/s13063-021-05412-4

47. Jacobs JC, Barnett PG. Emergent challenges in determining costs for economic evaluations. *Pharmacoeconomics.* (2017) 35:129–39. doi: 10.1007/s40273-016-0465-1

48. Buxton M, Hanney S. How can payback from health research be assessed? *J Health Serv Res Policy.* (1996) 1:35–43. doi: 10.1177/135581969600100107

49. Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: combined and uneven development. *Med Decis Making.* (2008) 28:21–32. doi: 10.1177/0272989X07308751

50. Zhao W, Olshefski D, Schulzrinne HG. *Internet Quality of Service: An Overview*. Columbia University Academic Commons. (2000). Availabole online at: https://academiccommons.columbia.edu/doi/10.7916/D8T15FTV (accessed January 10, 2022).

51. Zolfaghari B, Srivastava G, Roy S, Nemati HR, Afghah F, Koshiba T, et al. Content delivery networks: state of the art, trends, and future roadmap. *ACM Comput Surv*. (2020) 53:34. doi: 10.1145/3380613

52. Mack CA. Fifty years of Moore's law. *IEEE Trans Semiconduct Manufact.* (2011) 24:202–7. doi: 10.1109/TSM.2010.2096437

53. Lobban F, Dodd AL, Sawczuk AP, Asar O, Dagnan D, Diggle PJ, et al. Assessing feasibility and acceptability of web-based enhanced relapse prevention for bipolar disorder (ERPonline): a randomized controlled trial. *J Med Internet Res.* (2017) 19:e85. doi: 10.2196/jmir.7008

54. McNamee P, Murray E, Kelly MP, Bojke L, Chilcott J, Fischer A, et al. Designing and undertaking a health economics study of digital health interventions. *Am J Prev Med.* (2016) 51:852–60. doi: 10.1016/j.amepre.2016.05.007

55. Jones KC. *Burns, Amanda. Unit Costs of Health and Social Care 2021. (Unit Costs of Health and Social Care). ISBN: 9781911353140.* PSSRU; University of Kent (2021). Available online at: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-of-health-and-social-care-2021/~

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 10 November 2022
DOI 10.3389/fpsyt.2022.1030989

OPEN ACCESS

EDITED BY
Julian Schwarz,
Brandenburg Medical School Theodor Fontane, Germany

REVIEWED BY
Elisabeth Wells,
University of Otago, Christchurch, New Zealand
Cynthia Gong,
Children's Hospital of Los Angeles, United States

*CORRESPONDENCE
Anne Kleijburg
akleijburg@trimbos.nl

SPECIALTY SECTION
This article was submitted to Public Mental Health, a section of the journal Frontiers in Psychiatry

RECEIVED 29 August 2022
ACCEPTED 20 October 2022
PUBLISHED 10 November 2022

CITATION
Kleijburg A, Lokkerbol J, Regeer EJ, Geerling B, Evers SMAA, Kroon H and Wijnen B (2022) Designing and testing of a health-economic Markov model to assess the cost-effectiveness of treatments for Bipolar disorder: TiBipoMod.
Front. Psychiatry 13:1030989.
doi: 10.3389/fpsyt.2022.1030989

Designing and testing of a health-economic Markov model to assess the cost-effectiveness of treatments for Bipolar disorder: TiBipoMod

Anne Kleijburg[1,2]*, Joran Lokkerbol[2], Eline J. Regeer[3], Bart Geerling[4,5], Silvia M. A. A. Evers[1,2], Hans Kroon[6,7] and Ben Wijnen[2]

[1]Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, Netherlands, [2]Centre of Economic Evaluations & Machine Learning, Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands, [3]Altrecht Institute for Mental Health Care, Outpatient Clinic for Bipolar Disorder, Utrecht, Netherlands, [4]Dimence Mental Health Institute, Centre for Bipolar Disorder, SCBS Bipolaire Stoonissen, Deventer, Netherlands, [5]Department of Psychology, Health and Technology, Centre for eHealth and Wellbeing Research, University of Twente, Enschede, Netherlands, [6]Department of Social and Behavioural Sciences, Tranzo Scientific Center for Care and Welfare, Tilburg University, Tilburg, Netherlands, [7]Department of Reintegration and Community Care, Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands

Background: Bipolar disorder is an often recurrent mood disorder that is associated with a significant economic and health-related burden. Increasing the availability of health-economic evidence may aid in reducing this burden. The aim of this study is to describe the design of an open-source health-economic Markov model for assessing the cost-effectiveness of interventions in the treatment of Bipolar Disorders type I and II, TiBipoMod.

Methods: TiBipoMod is a decision-analytic Markov model that allows for user-defined incorporation of both pharmacological and non-pharmacological interventions for the treatment of BD. TiBipoMod includes the health states remission, depression, (hypo)mania and death. Costs and effects are modeled over a lifetime horizon from a societal and healthcare perspective, and results are presented as the total costs, Quality-Adjusted Life Years (QALY), Life Years (LY), and incremental costs per QALYs and LYs gained.

Results: Functionalities of TiBipoMod are demonstrated by performing a cost-utility analysis of mindfulness-based cognitive therapy (MBCT) compared to the standard of care. Treatment with MBCT resulted in an increase of 0.18 QALYs per patient, and a dominant incremental cost-effectiveness ratio per QALY gained for MBCT at a probability of being cost-effective of 71% when assuming a €50,000 willingness-to-pay threshold.

Conclusion: TiBipoMod can easily be adapted and used to determine the cost-effectiveness of interventions in the treatment in Bipolar Disorder type I and II, and is freely available for academic purposes upon request at the authors.

KEYWORDS

economic evaluation, Bipolar disorder, cost-effectiveness, open-source, manic or depressive episode, health economic modeling, Markov model

Introduction

Bipolar disorder (BD) is an often recurrent mood disorder that is characterized by episodes of depression and (hypo)mania alternated with periods of remission (1). In a largescale pooled analysis from the World Mental Health survey the lifetime prevalence's for BD type I (BD-I), type II (BD-II), and subthreshold were 0.6, 0.4, and 1.4%, respectively (2). During a depressive episode patients generally experience strong feelings of sadness and hopelessness, a loss of pleasure and interests in normal activities, and even suicidal thoughts. During episodes of mania patients may experience a strong increase in energy, feelings of excessive euphoria or agitation and a decreased ability to sleep and control impulsive behavior. Depending on the severity and duration of manic episodes BD can be classified according to four types, type I, type II, cyclothymia and unspecified or subthreshold BD. BD-I describes patients experiencing manic episodes with a duration of at least one week and a severity that significantly limits their functioning, and is more likely to result in psychosis or require admission. BD-II describes patients experiencing hypomanic episodes with a duration of at least four days where manic symptoms are less severe and functioning is affected but not limited. With cyclothymia patients experience milder forms of manic and depressive mood episodes, and patients with symptoms indistinctive of the other types are classified as unspecified or subthreshold (1, 3).

Treatment options for BD generally consist of both pharmacological and non-pharmacological interventions that aim to prevent the relapse of manic and depressive episodes, decrease the severity of manic and depressive symptoms, and improve inter-episodic functioning (4–6). As BD is most commonly diagnosed during early adulthood, its lifelong and highly variable nature often requires long-term treatment, whilst introducing significant detriments to the coping individual's quality of life and productivity (7, 8). Consequently, BD is associated with both substantial healthcare costs and productivity losses, incurred by patients as well as caregivers, introducing a significant economic burden on society (7, 9–12).

When aiming to create an efficient and sustainable healthcare system, policy-makers require not only information on the effectiveness of interventions but also their relative value for money, as this guides decisions ultimately impacting a finite healthcare budget (13). Such decisions can for example encompass whether or not to implement new (treatment) strategies in practice, to adopt certain treatments over others in new clinical guidelines, or to reimburse treatments by health insurers. Economic evaluations can provide decision-makers with such information by determining the relative efficiency and costs (or cost-effectiveness) of new interventions when compared to current interventions. When the evidence needed to perform an economic evaluation is not available from a single source, decision-analytic models allow combining data from various sources and its extrapolation over a sufficiently long time horizon while explicitly taking into account uncertainty (14, 15).

Given the significant economic and health-related burden of BD, as well as the wide variety of interventions that exist for the treatment of BD, we believe that a better understanding of their relative cost-effectiveness may aid to reduce this burden. Therefore, the aim of this study is to present and describe a flexible decision-analytic model, Trimbos institute's BipoMod (TiBipoMod), that can be used to examine the long-term cost-effectiveness of user-defined pharmacological and non-pharmacological interventions in the treatment of adults with Bipolar Disorder type I (BD-I) and type II (BD-II). The model will be made available for all researchers with interest upon request. Similarly, easily adaptable decision-analytic models aiming to increase the availability of cost-effectiveness evidence for treatment and prevention are already available for psychosis and depression (16, 17). To provide complete transparency toward its potential users, this paper describes (1) the process of developing a conceptual model, (2) the final structure of the model and its assumptions, (3) the parameters used by the model, and 4) a case study to illustrate the use and results generated by the model. Overall, the model's details described here may aid its users in the process of adapting the model and its parameters to match the context and research question at hand.

Materials and methods

Model development

As the aim of this study is to create a flexible decision-analytic model to examine the long-term cost-effectiveness for

treatment of Bipolar Disorder, TiBipoMod was developed as an easy-to-use Microsoft Excel-based Markov cohort model. In a Markov cohort model a cohort of patients is modeled over a predefined time period during which they transition between the various included health states, accumulating costs and health effects associated with each health state given the treatment condition over time (15).

The model was developed in line with the guidelines of the Professional Society for Pharmacoeconomics and Outcomes Research (ISPOR) for conceptualizing a model (18). First, the research problems to be answered with this model were formalized, providing the foundations for the conceptual model. To conceptualize the model structure, a scoping literature review on the disease progression of BD and existing cost-effectiveness studies was performed, after which its final structure was validated by an expert panel (see below). The expert panel for this study consisted of two healthcare professionals in the treatment of BD in the Netherlands who were consulted throughout the development process to validate assumptions and parameter values.

After finalizing the model structure, health state parameters and model assumptions were formulated using available treatment guidelines, national databases, published literature and expert opinions. This iterative process finally resulted in the following PICOT for TiBipoMod:

- Population: Adults with the diagnosis of Bipolar Disorder, type I and type II, as defined by the 2013 DSM-V[1] (1).
- Intervention: User-defined interventions are modeled in addition to the reference treatment(s), and compared to the standard of care (SOC) alone. In order to model an intervention, users are required to insert the relative risks of experiencing a manic and depressive episode given the intervention of interest, and its associated costs. The model is able to compare two interventions simultaneously using separate Markov traces, and present its outcomes. Interventions may be pharmacological and non-pharmacological.
- Comparator: The modeled comparator is the SOC, which by default has been parameterized based on clinical treatment guidelines and expert opinion. The SOC may be easily adapted to a user-defined SOC by adjusting parameter values. By default, the comparative scenario includes commonly prescribed pharmacotherapy and psychotherapy, outpatient mental specialist care, community treatment, and episode crisis care.
- Outcome: Costs per quality-adjusted life year (QALY) gained.
- Time Horizon: Given the lifelong nature of BD, costs and health effects are modeled over a lifetime horizon, but also a 5-year horizon when shorter horizons are preferred. The model uses a cycle-length of three months.

To provide in the varying demands of guidelines for health-economic evaluation both a healthcare and societal perspective can be applied, and future costs and effects are discountable by user-defined rates (by default: 4 and 1.5%, respectively) (19). A half-cycle correction is applied to account for the fact that transitions between states may occur at any time during the cycle (20).

For deciding on a willingness-to-pay (WTP) threshold to be applied, country-specific guidelines for economic evaluation can be used or, when absent, the WHO recommends using a threshold of three times the national GDP (21). For example, in the Netherlands the guidelines for Disease Burden in economic evaluations provides WTP thresholds based on disability weights (22). According to the Global Burden of Disease 2013 study BD disability weights are estimated at 0.40 and 0.49 for depressive and manic episodes, respectively, resulting in a recommended WTP threshold of €50,000 in the Netherlands (23). When using a GDP-based threshold this would result in a WTP of €147,300 (2021 GDP in the Netherlands: €49,100).

Model conceptualization

The first step in the development of TiBipoMod was to explore disease progression of BD and the conceptualization of BD in published health economic models by performing a scoping literature review. In this process ten model-based health-economic evaluations for the treatment of BD were identified (**Supplementary Material I**). In this and in clinical literature six potential health states became apparent that were to be considered for model inclusion; depression, mania, hypomania, rapid cycling, remission/euthymia, and death. Whereas mania, depression and remission were found in previous economic evaluations, hypomania and rapid cycling were not (6, 24–29). Reasons in the literature for the exclusion of hypomania as a separate health state are the lack of evidence surrounding parameters for hypomania, and that the burden imposed on the patient by hypomania is considered less severe than during a depressive or manic episode of BD-I. As for rapid cycling, clinical guidelines stated that depending on the polarity of the episode this is treated as either a manic or depressive episode. Finally, patients with BD experience an elevated risk of suicide and higher mortality rates due to comorbidity and poorer lifestyle choices throughout their life course, contributing to an overall reduction in life-expectancy, supporting the inclusion of death as health state (30–32).

Based on the considerations above, our conceptual model aimed to include the health states depression, (hypo)mania,

1 As little differences are observed in the classification of bipolar disorder between the DSM-IV and V, studies using the DSM-IV to inform model parameters were also included.

remission, and death. From the existing models identified during the literature review the schematic model structure published by Ekman et al. (33) best matched these health states and was considered to best fit the Markov modeling approach of this study. In the study of Ekman et al. (33), a discrete event simulation was used to simulate the occurrence of four health states to determine the cost-effectiveness of quetiapine in patients with acute bipolar depression and maintenance treatment (33). The possibility of treatment discontinuation in this model described by Ekman et al. is not included for the current purpose.

As a second step, an expert panel of healthcare professionals was consulted to validate the conceptual model based on Ekman et al. The panel confirmed the structure of the initial conceptual model, however, as the initial model only included a transition from depression to mania, the panel collectively recommended the inclusion of the transition from mania to depression, which may occur in response to the excitatory processes of mania (34, 35). In addition to this, the panel was consulted on the differences between manic and hypomanic episodes including its implications for treatment. The experts stated that despite similarities, important differences are often observed in the time spent in the mood episodes, quality of life, and experienced during episodes.

Given that the aim of the current model is to represent both BD-I and BD-II, the Markov model was built with two separate Markov traces for each type, estimating costs and effects for both subpopulations, which can then be combined into a single weighted ICER using the proportion in prevalence. The final model structure therefore includes the health states depression, (hypo)mania (i.e., mania for BD-I and hypomania for BD-II), remission, and death (**Figure 1**).

Model parameters

Mood state epidemiology and transition probabilities

To populate our conceptual model with health state transition probabilities, available literature on the epidemiological characteristics of BD and its longitudinal disease course was reviewed and compared. In this process, multiple studies were identified that report on the differences in long-term symptomatic status of BD-I and BD-II, the time spent in the various mood states, and recurrence rates (36–42). When looking at studies that present the percentages of time spent in various mood states, large variations can be observed in their findings (see discussion) (36, 39–41). Here, comparison and drawing conclusions is challenged by the highly heterogenous study designs and definitions of mood states. As such, a single study was selected to inform prevalence rates as our main guidance in selecting and verifying the modeled epidemiology. Based on the assessment frequency of reported symptoms, the study by Kupka et al. (36) was chosen, which describes the largest naturalistic cohort of patients (n: BD-I = 405, BD-II = 102), where patient's daily self-reported symptoms were assessed weekly to monthly for one year by their physicians, and translated to DSM-IV mood episodes (36). The following paragraphs describe studies (or study arms) which have been selected to inform model transition probabilities. In those studies all patients are provided with some form of pharmacological treatment typical to the respective mood episode studied, meaning the model does not simulate untreated disease progression, but rather progression given commonly prescribed or naturalistic pharmacotherapy.

First, the probabilities of relapsing from remission to both depression and mania were informed by the literature review and meta-analysis of Vazquez et al. (42) (BD-I: 96% of patients), combined with the reported time spent in depression/mania ratio by Kupka et al. (36) (BD-I: 81% of patients). Vazquez et al. (42) report an annual recurrence rate for any mood relapse while treated with active medications of 21.9% based on 15 RCTs, which could be translated to the quarterly transition probability of 5.99% (42). To correct for differences seen in the time spent in depressive and manic mood episodes, and to match the prevalence of mood episodes seen in BD-I and BD-II epidemiology, days spent in depression and mania (excluding days with mild/subsyndromal symptoms) reported by Kupka et al. (36) were used to construct a depression/mania ratio (36). This resulted in depression/mania ratio for BD-I of 4.7 and BD-II of 10.7, which combined with the probability of recurrence by Vazquez et al. (42) resulted in the final probabilities for relapsing to mania or depression presented in **Table 1**. The probability of remaining in remission was found by subtracting the probabilities of leaving the health state.

Second, to determine the probabilities of remaining in a mood episode, time-to-recovery estimates provided by Solomon et al. (41) were used. This study was performed using data of an observational study where patients of predominantly BD-I patients (n = 219) who were not controlled for any somatic treatment received. They report 50% of patients remaining in a major depressive episode after 15 weeks, 25% of patients remaining in a manic episode after 15 weeks, and 25% of patients remaining in a hypomanic episode after 6 weeks, with a median duration for a mood episode of 13 weeks (41). Using R statistical software, exponential regression equations were applied to the reported number of weeks per quantile for patients to recover from each mood episode type to estimate the transition probabilities per model cycle (13 weeks) (43). This results in a probability of remaining in a depressive, manic and hypomanic episode after 13 weeks of 56.1, 29.8, and 4.8%, respectively.

Then, to inform the probabilities of transitioning between mood states, Weibull distributed parameters reported by Ekman et al. (33) (BD-I: 66% of patients) were combined with the relative risks for receiving pharmacotherapy with mood stabilizers and olanzapine. Given these conditions, the

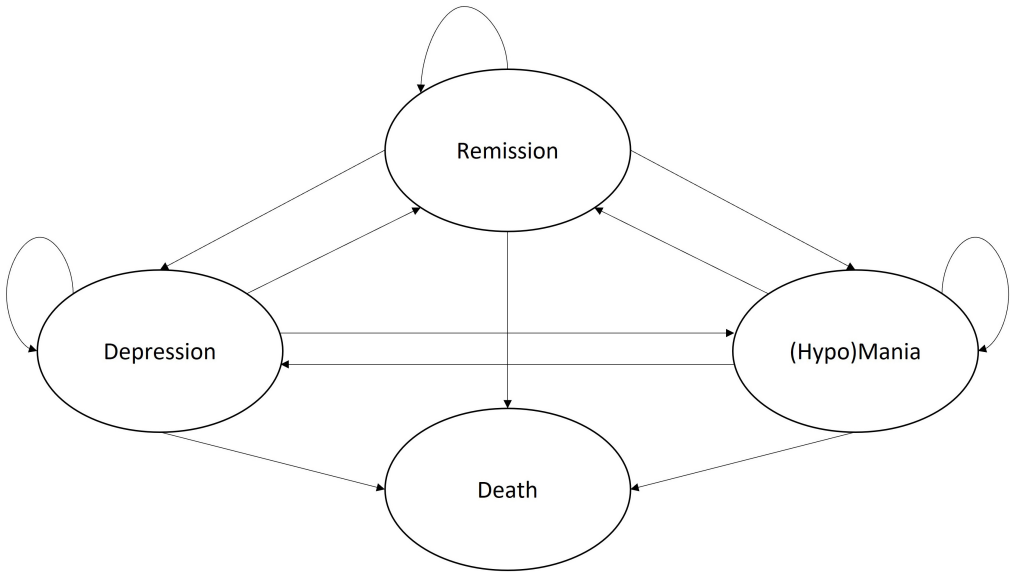


FIGURE 1
Final conceptual Markov model for the treatment of Bipolar Disorder. Adapted version of the model published by Ekman et al. (33).

TABLE 1 TiBipoMod state transition probabilities per (three-month) cycle and health state utilities.

	BD-I	BD-II	Distribution	References
Proportion BD-type population	0.600	0.400	Beta	(2)
Time spent mood states				
Ratio Depression/Mania	4.7	10.7		(36)
% Time in Depression	0.744	0.792	Beta	(36)
% Time in Mania	0.256	0.208	Beta	(36)
Health state transitions - SOC				
Remission to Remission	0.880	0.880	Dirichlet	
Remission to Depression	0.094	0.117	Dirichlet	(42)
Remission to Mania	0.026	0.003	Dirichlet	(42)
Depression to Depression	0.561	0.561	Dirichlet	(41)
Depression to Remission	0.365	0.365	Dirichlet	
Depression to Mania	0.074	0.074	Dirichlet	(33)
Mania to Mania	0.298	0.048	Dirichlet	(41)
Mania to Depression	0.074	0.074	Dirichlet	(33)
Mania to Remission	0.628	0.878	Dirichlet	
Mortality				
RR Premature death	2.060	2.060	Lognormal	(44)
RR Suicide	9.660	9.660	Lognormal	(44)
Intervention effect (see case study below)				
RR Mania	1.320	1.320	Lognormal	(55)
RR Depression	0.810	0.810	Lognormal	(55)
Utilities				
Remission	0.800	0.800	Beta	(52)
Depression	0.290	0.290	Beta	(52)
Mania	0.540	0.800	Beta	(52)

probability of transitioning from depression to mania was 7.4% (33). Despite being recognized frequently in clinical practice by the expert panel, little evidence was found on the transition probability of mania to depression. In consultation with the expert panel, the probability of this transition was set equal to the transition of depression to mania (7.4%). The probabilities of transitioning from mania to remission and depression to remission were found by subtracting the probabilities of leaving the health states.

Finally, transitions to death were based on general mortality statistics in the Netherlands as reported by Statistics Netherlands. To account for the increased risk of suicide and death by comorbidities associated with BD relative mortality rate ratios (MRR) from Westman et al. (44) were applied to the general mortality. For comorbidities and lifestyle effects of BD a MRR of 2.06 was applied independent of the health state, and for suicide an additional MRR of 9.65 was applied to the depressive state only, as suicide occurs less frequently during mania or remission (44, 45).

Validating modeled epidemiology

Combining the above mentioned transition probabilities in the Markov chain resulted in the modeled epidemiology presented in **Figure 2**. Here, from the patients not transitioned to death, around 78% of patients are in remission, 18% are experiencing a depressive episode, and 4% a manic episode.

To validate the modeled epidemiology with the empirical data by Kupka et al. (36), a comparison was made between the time spent in the various mood states. To this extent, distinctions were made between the severity of mood episodes, as the degree to which functional impairment occurs is an important factor in the increasing need of health services and reduced productivity. As such, we assigned time spent with mild or subsyndromal symptoms to the remission states, and only included time spent with moderate to severe symptoms to the respective health states. Based on the prevalences reported by Kupka et al. this assumption would translate into a guiding estimate of patients spending 74.0-76.2% of time in remission/euthymia, 4.6-7.5% in (hypo)mania and mood cycling, and 18.4-19.1% of time in depression. It is important to note that this distribution was used as a guidance during the process of informing state transition parameters. However, given the variation seen in the available evidence we chose not to calibrate parameters to match these guiding estimates exactly.

Based on the above mentioned heterogeneity and uncertainty (also discussed later), the modeled prevalence estimates were considered in agreement with the guiding estimates reported by Kupka et al. (36).

Defining the standard of care

Similar to the heterogeneity in clinical presentation of patients suffering from BD, the amount of health services that patients use is also highly heterogenous (3, 46). In addition to this, a wide variety of treatment options are available for the treatment of BD, including various forms of pharmacotherapy, psychological therapy, community-based treatments, and inpatient treatments. As a result, defining the "standard of care" for the treatment of BD is challenging. Given that this current model aims to serve as a health-economic tool that is easily adaptable by its users to match their needs, by default the model is designed as such that the SOC during each health state included all most commonly provided treatment options identified in the treatment guidelines and by expert opinion (3, 46–48). This resulted in the identification of five major care components that are generally included in treatment regimens for BD and are included in the model. These components are:

1. Pharmacotherapy: For patients with BD pharmacotherapy most commonly consists of either monotherapy or polypharmacy with mood stabilizers, antipsychotics, anticonvulsants or antidepressants, depending on the health state and patient-specific preferences.
2. Outpatient mental specialist care: Routine treatment and monitoring by, e.g., psychiatrists and a mental health nurse (practitioner) is recommended to promote relapse prevention, stimulate self-management, and adjust treatment during episodes.
3. Psychotherapy: Consensus exists on the importance of psychotherapy programs for patients and relatives to stimulate successful long-term management and relapse prevention. Commonly recommended outpatient psychotherapy programs are psycho-education, cognitive behavioural therapy (CGT) and interpersonal and social rhythm therapy (IPSRT).
4. Community-based care: For patients with a more severe form of BD, a form of community-based treatment may be indicated where multidisciplinary teams provide continuous, flexible and outreaching treatment and monitoring. Examples of such service models are collaborative care models, (flexible) Assertive Community Treatment models, and crisis models such as Crisis Resolution Teams or Intensive Home-based Treatment.
5. Inpatient care: During severe mood episodes patients may also be admitted to a psychiatric hospital. Depending on the local availability of alternative crisis treatments, the incidence and duration of hospitalization may vary.

Based on these individual components, the SOC for each of the model health states is assumed to consist of the treatment options presented in **Table 2**. Here, pharmacotherapy and outpatient specialist care is included for both the remission and mood episode states. Despite commonly being provided during periods of remission, in consultation with the expert panel the majority of psychotherapy sessions has been assigned to the mood episode health states to account for the fact that patients

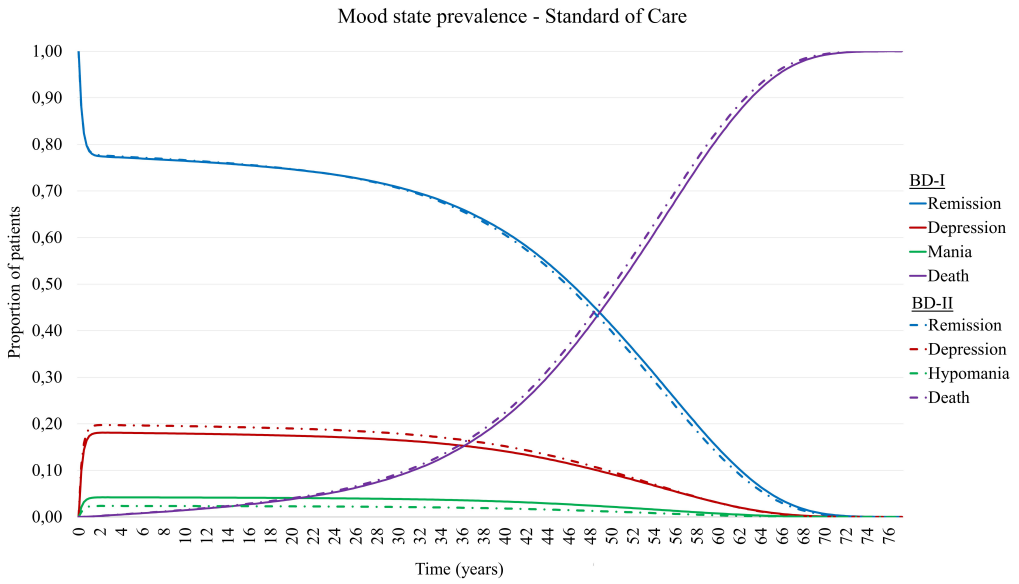


FIGURE 2
Mood state prevalence of simulated patient cohort with Bipolar Disorder type I (BD-I) and type II (BD-II) in TiBipoMod, consisting of ~ 4% of live patients in mania, ~ 18% in depression, and ~78% in remission.

often participate in these upon remitting from an episode. Inpatient care is included only for the mood episode states.

To account for the fact that not all patients may need or want to make use of these treatment components during a depressive or manic episode, use of care can be weighted by means of percentages based on three categories for treatment intensity; "outpatient low intensity," "outpatient high intensity," and "inpatient care." For example, patients with a less severe mood episode may receive additional treatment with low intensity, patients with more severe episodes may receive additional treatment with high intensity, and only patients with very severe episodes may be admitted. By default, the category "outpatient low intensity" constitutes of outpatient mental specialist care and community-based treatment with increased frequency as compared to remission care, and "outpatient high intensity" constitutes of outpatient mental specialist care, a psychotherapy program and community-based treatment. Pharmacotherapy is included for all patients. Both the assigned percentage weights and components of the treatment categories can be altered to match the local context of the user.

Valuation of cost components

TiBipoMod offers analysis from both a healthcare perspective as well as a societal perspective, including productivity costs and patient and family costs. From the healthcare perspective, direct medical costs consist of pharmaceutical costs, costs relating to outpatient specialist care, psychotherapy, community-based treatment and inpatient care. Indirect medical costs included in the model are the periodic costs for drug-induced renal failure testing and medical costs for unrelated diseases during other and the last year of life, calculated using the tool Practical Application to Include Disease Costs (PAID) (49, 50). Productivity loss estimates associated with BD for absenteeism and presenteeism are included in the model based on literature estimates (9). Costs included in the model for the patient and family are limited to the hours spent by caregivers on informal care (7, 51). All cost components are comprised of individual units for resource use and unit costs. Costs related to the intervention are included by a separate parameter, applied only to the intervention arm. Also, adjustable parameters are included allowing to adjust the number of cycles with which intervention costs and effects are experienced. By default, TiBipoMod will be populated with resource use and unit cost inputs based on the Dutch context (section 3.1), however, as all inputs can be adjusted, we encourage users to adjust accordingly to match their local context.

Quality of life

Quality of life experienced during each health state was described using utility scores published by Revicki et al. Utilities are based on both inpatient and outpatient treated patients suffering from BD type I (n = 96) and measured using the standard gamble (SG) method (52). Health state utility scores for patients suffering from BD type II were not found in the literature. However, based on the differences in clinical presentation of mood episodes seen with BD type I and type II, differences may be expected in quality of life experienced and thus utility scores. For example, differences in clinical presentation are especially significant for mania and hypomania,

where in general hypomania is shorter in duration but, most importantly, not associated with severe functional impairment (1, 3).Therefore, to include QoL estimates in our model better representable for both BD-I and BD-II, assumptions were made based on its clinical presentation and considering the model cycle time of 90 days. As such, we assumed QoL during a hypomanic episode for the patients with BD type II to be equal to the quality of life during remission (0.80), rather than that of mania (0.54), as measured by Revicki et al. Health state utilities for remission and depression (0.29) are assumed equal for BD type I and type II, as presented in **Table 1**.

Model outputs

Model outputs of TiBipoMod are expressed in costs, life years (LYs) and QALYs for both the intervention(s) and comparator (discounted and undiscounted). Outcomes are compared using the incremental cost-effectiveness ratio (ICER). The ICER is calculated as followed: (Costs intervention - Costs control)/(QALYs intervention - QALYs control). Here, the ICER represents the incremental costs per QALY gained.

Probabilistic sensitivity analysis

To assess the uncertainty surrounding the parameters included in the model probabilistic sensitivity analyses (PSA) can be performed. Probability distributions are assigned to each parameter in the model based on its characteristics, e.g., beta distributions for utilities with a value between 0 and 1, the skewed gamma distribution for costs, and Dirichlet distributions for transition probabilities that sum up to 1. For parameters of which its value was to remain between predefined bounds (e.g., in case of treatment guidelines stating a minimum and maximum amount of treatment sessions) a beta-PERT distribution was applied. Subsequently, the PSA can be run 5,000 times, each time drawing a random value from the distribution for each parameter. As incremental costs and effects are simulated 5,000 times they can be plotted in an incremental cost-effectiveness (CE) plane, with the incremental QALYs on the x-axis and the incremental costs on the y-axis, illustrating its uncertainty. In addition to this, a cost-effectiveness acceptability curve (CEAC) is constructed illustrating the likelihood of the intervention being considered cost-effective given a series of willingness-to-pay (WTP) thresholds (15).

TABLE 2 Treatment components of the standard of care per health state included in TiBipoMod.

	Patients assigned to treatment intensity (%)			
	Remission	**Depression**	**Mania**	**References**
Outpatient low intensity treatment ○ *Pharmacotherapy* ○ *Outpatient mental specialist care*	85%	90%	30%	(45, 53), expert opinion)
Outpatient high intensity treatment ○ *Pharmacotherapy* ○ *Outpatient mental specialist care* ○ *Community-based treatment* ○ *Psychotherapy/education*	15%	7%	40%	(45, 53), expert opinion)
Inpatient care ○ *Pharmacotherapy* ○ *Outpatient mental specialist care* ○ *Hospital admission*	NA	3%	30%	(45, 53), expert opinion)

Validation

To validate the final model, both internal and external validations have been performed. First of all, for external validation of conceptual ideas and input parameters the expert opinion panel played a key factor. In addition to this the Assessment of the Validation Status of Health-Economic decision models (AdViSHE) tool was used, a 13-item questionnaire assessing four typologies; conceptual validation, data validation, computerized model validation and operational validation (53). For internal validation the black box test TECH-VER checklist was applied, ensuring technical verification, completeness and consistency (54). The results of both tools are presented in **Supplementary Material II**.

Case study: Mindfulness based cognitive therapy in the Dutch context

To apply TiBipoMod to a real-world example, the effectiveness of a mindfulness-based cognitive therapy (MBCT) intervention, as described in the results of a randomized controlled trial performed, was combined with the costs of providing the intervention. The MBCT intervention aims to reduce the chance of relapse, and to reduce the severity of depressive symptoms during an episode. This effect was studied in the RCT by Perich et al. (55), where the intervention group received MBCT and the SOC, and the control group only received the SOC. The RCT's primary outcome was the 12-month recurrence rates of depressive and (hypo)manic episodes. Despite not being significantly different, 59% of the participants in the MBCT group had suffered a (hypo)manic episode in the past year and also 59% a depressive episode, while in the SOC group 48% of the participants had a (hypo)manic episode

and 68% a depressive episode (55). These annual recurrence percentages were calibrated to quarterly recurrence rates and used to determine the relative risks for a mood episode. This resulted in relative risks of 0.81 and 1.32 for transitioning to depression and (hypo)mania when treated with MBCT, respectively. This effect was modeled to persist for 4 cycles (separate parameter).

Valuation of unit costs

Cost parameters for the SOC in this case-study were determined using a bottom-up costing approach. Direct medical costs incurred to each patient by the use of included treatment options were determined using treatment guidelines (56–58), expert panel estimates, national cost databases (59, 60), and reference prices published in the Dutch manual for cost research (19, 51). For example, costs related to specialized mental healthcare are based on estimates for hours spent on consultations provided by the expert panel, and combined with the hourly reference physician rates. Productivity costs are included in the model up until the Dutch retirement age of 67, were informed by the literature (9). Costs of informal care for the patient and family are valued at the Dutch reference price for unpaid work (7).

Costs relating to the MBCT intervention were based on its 8 sessions in groups of 8 to 12 people offered by two mental health nurse practitioners. This resulted in average additional costs of €291 per person per quarterly cycle for MBCT, which was modeled for a single cycle. The relative risks for (hypo)manic and depressive episodes and its intervention costs were added to the model to determine interventional transition probabilities and costs for MBCT + SOC. All costs were expressed in 2021 Euros by indexing unit cost prices with the consumer price index when necessary (61).

Sensitivity and scenario analyses

To provide insight in the impact of changes in modeled epidemiology of BD-I and BD-II, two one-way sensitivity analyses were performed with alternative ratios for time spent in depression/mania. For these scenarios the studies of Joffe et al. (40) and Judd et al. (37, 38) were used which found significantly higher ratios for BD-I = 6 and BD-II = 14, and BD-I = 3.6 and BD-II = 38.7, respectively.

Results

Finally, the cost-effectiveness of MBCT + SOC compared to the SOC alone was determined from a healthcare and a societal perspective. From a societal perspective, MBCT + SOC resulted in an average per-patient increase of 0.017-0.019 QALYs and a decrease in costs of €339- €674 depending on the simulated BD subtype, resulting in a dominant ICER per QALY gained. All outcomes for modeled scenarios are presented in **Table 3**.

When running sensitivity analyses this resulted in the cost-effectiveness plane and CEAC presented in **Figures 3A,B**. When considering a WTP threshold of €50,000, there was a 71% probability that MBCT + SOC is cost-effective.

Discussion

We have presented TiBipoMod, a Markov model that is able to evaluate the (long-term) cost-effectiveness of both pharmacological and non-pharmacological interventions for patients suffering from both BD-I and BD-II. When provided with the necessary input parameters describing the intervention and local context (e.g., relative risks for a depressive and manic episode, intervention costs, expected duration of effect, unit costs), our model is able to present outcomes from a healthcare and societal perspective, a 5-year and lifetime time horizon, and includes various built-in parameters to adjust for the heterogeneity of BD, e.g., in terms of quality of life, functional impairment, healthcare resource use, and differences in epidemiology between BD-I and BD-II. In addition to this, because the model is Excel-based, its use does not require advanced health-economic modeling skills and the model is easily adjustable in its functionalities. The model was developed in line with (inter)national clinical treatment guidelines, available literature on its epidemiology, treatment, intervention effects and costs, and in consultation with Dutch healthcare professionals in the treatment of BD. Additionally, these professionals provided important input in the validation process of our input sources and model assumptions, supplemented by the AdViSHE and TECH-VER validation tools.

To illustrate the outcomes generated by this model a case study was performed assessing the cost-effectiveness of MBCT + SOC compared to the SOC, which found that MBCT + SOC is dominant over the

TABLE 3 Costs included in the model per patient per (three-month) cycle for each health state.

	Remission	Depression	Mania	Sources
Drugs	€ 39	€ 53	€ 53	(3, 60)
Medical services	€ 161	€ 786	€ 804	(51, 57)
Psychological treatment	€ 271	€ 126	€ 722	(51, 56, 57)
Home-based treatment	€ 131	€ 153	€ 874	(51, 57)
Indirect medical costs	€ 40	€ 40	€ 40	(49)
Productivity losses	€ 443	€ 3,637	€ 2,182	(9, 51)
Patient *and* family costs	-	€ 4,996	€ 2,997	(7, 51)
Admission costs	-	€ 617	€ 6,166	(51, 57)
Intervention costs	€ 291	-	-	(51, 55)

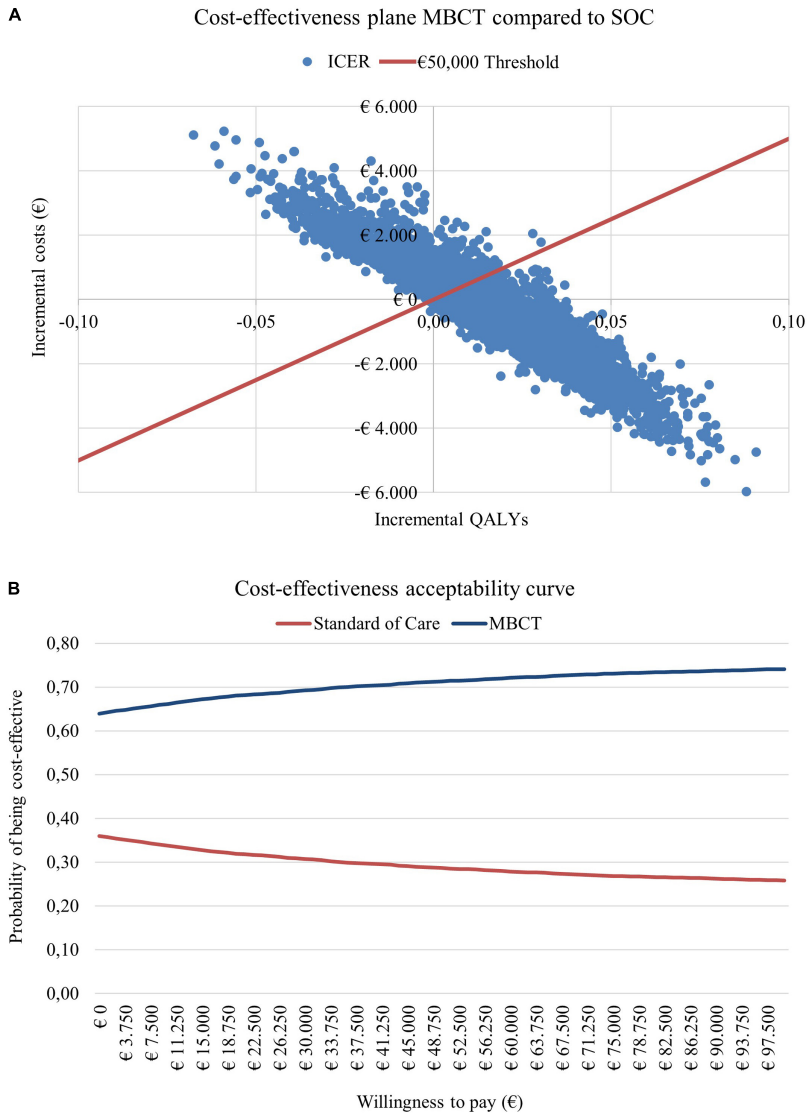


FIGURE 3
(A) Cost-effectiveness plane and **(B)** cost-effectiveness acceptability curve for adding mindfulness-based cognitive therapy (MBCT) to the standard of care (SOC) compared to the SOC alone.

SOC alone when considering a societal perspective, but is associated with an ICER of €15,993 - €28,987 per QALY gained from a healthcare perspective. This difference illustrates the impact of including societal costs such as productivity loss and caregiver costs, and their relevance for inclusion when evaluating interventions for BD.

Strengths and limitations

Strengths of our model are related to its potential to aid in the generation of cost-effectiveness evidence that is more easily comparable across intervention (compared to outcomes derived from different economic models), while also based on methods that are fully transparent. Though cost-effectiveness of both pharmacological and non-pharmacological interventions in the treatment of BD have been studied and systematically reviewed, common conclusions were drawn that 1) the number of available studies was relatively low and 2) the methods applied were heterogeneous, creating a need for more robust and better comparable (long-term) evidence to inform policy decisions (4–6, 62). In addition to that, increasing interest emerges toward health-economic models that are open source, i.e., available to anyone who wishes to access it (63–65). Important arguments

for this have been its potential for increasing knowledge-sharing, efficiency, consistency and, perhaps most importantly, transparency and credibility of evidence generation in cost-effectiveness research thereby reducing uncertainties. Here, transparency is achieved by providing full access and insight to all methods and assumptions made throughout the model (66, 67). In addition to that, TiBipoMod includes additional background information sheets in the Excel model to ensure full disclosure on all sources used and subsequent assumptions made.

The development of the model should be seen in light of some limitations that are important to acknowledge when considering to adopt our model. First, TiBipoMod is constructed as a Markov cohort model, a model type that is widely used in estimating cost-effectiveness resulting from its relative simplicity, transparency and useability whilst often maintaining sufficient accuracy depending on its application (68). However, limitations of Markov models that have been frequently identified in the literature are its lack of memory (i.e., the subsequent health state only depends on the present health state, and not the sequence of preceding states), fixed cycle length and state-transition probabilities, and its limited ability to model complex diseases better represented by a larger numbers of health states (68, 69). For example, the wide variation in duration of mood episodes is not well represented by the fixed blocks of time, and little distinction can be made in the severity of the mood episodes of the respective episodes when represented by a single health state (70–72). When interested in capturing time- or patient-specific effects, one should consider modeling approaches that allow for greater complexity and detail such as discrete-event simulation (DES) models. However, considering the limitations of a DES model, being that its complexity requires advanced modeling skills, resources, and the fact that it is more data heavy, we felt it did not align with our aim of creating an easily-adaptable model, opting for a Markov model with additional parameters attempting to correct for the heterogeneous nature of BD presentation and treatment (73).

A second potential limitation of our model relates to uncertainty following from the epidemiological parameters included in our model. First of all, our transition probabilities have been derived from multiple sources where the study populations existed of varying patient population characteristics, such as the proportion of patients included with a BD-I and BD-II diagnosis (ranging from 66% to 96% BD-I). Multiple studies report on the long-term symptomatic status and time spent in various mood states by patients suffering from BD-I and BD-II (36–41). Although comparability of studies is complicated by several factors, such as the use of different rating scales [i.e., the National Institute of Mental Health Life Chart Methodology (NIMH-LCM) or the Longitudinal Interval Follow-up Evaluation (LIFE) system (74, 75)] discrepancies present, stressing the importance of acknowledging the uncertainty underlying the epidemiology. For example, despite comparable findings on the amount of time spent in remitting phases (44%-54%) of BD-I and BD-II, some studies report on significant differences in time spent in depressive and (hypo)manic episodes (36, 39, 40). However, when looking at reported ratios for time spent in depression/mania per BD subtype, Kupka et al. (36) find relatively comparable ratios for BD-I and BD-II with 2.9 and 3.8 including mild symptoms (when excluding mild symptoms this becomes BD-I = 4.7 and BD-II = 10.7), respectively, and no differences in episode frequency which suggest similar tendencies in mood switching and symptomatic status. When comparing this to the depression/mania ratios found by Joffe et al. (40) (BD-I = 6 and BD-II = 14), and even more so to those found by Judd et al. (37, 38) (BD-I = 3.6 and BD-II = 38.7), these suggest significant differences in clinical course between BD-I and BD-II. Although these discrepancies can be partially explained by differences in mood state definition, study design and patient assessment frequency, favoring the outcomes by Kupka et al. (36), significant uncertainty surrounding the true clinical trajectories of BD subtypes remains. It is therefore important to emphasize that differences in modeled epidemiology for BD-I and BD-II should be subjected to sensitivity analyses, which is also why this feature has been implemented in TiBipoMod.

A third limitation of our model stems from a lack of available evidence to inform model parameters, e.g., for QoL, resource use, health-state transitions and societal losses, either in general or specifically for BD-II when only available for BD-I. For example, the available health state specific SG utilities published by Revicki et al. (52) were measured in BD-I patients only, requiring additional assumptions (52). Collectively, this lack of evidence for QoL, and the subsequent assumptions made introduce additional uncertainty, stressing the need for further research (i.e., especially in BD-II).

A fourth limitation that stems from this lack of evidence also relates to the studies used to inform transition probabilities in this model. Current transition probabilities are based on RCTs or observational studies in which (most) patients have received pharmacological treatment, which treatment(s) exactly, however, is not clear for each study. Therefore, our model simulates interventions that have been added to some form of best practice treatments, including pharmacotherapy, rather than untreated disease progression. As a result, the relative risk for experiencing a mood episode given the intervention considered for evaluation should, ideally, be measured in patients that receive some form of baseline pharmacotherapy.

Fifth, simplifying BD to a model with only four health states is a strong simplification of the true population heterogeneity.

In general, BD is characterized by its strongly heterogeneous mood swings, fluctuating somewhere between severe depression and extreme manic states, alternated with periods of remission. Even within the categorization of BD in type I and II or unspecified/subthreshold, the severity of mood episodes may vary per patient and per episode independent of the specific BD diagnosis (3, 46). Similarly, transitions between mood episodes, i.e., mania to depression or depression to mania, are frequently observed but often do not occur consecutively and may be separated by weeks to months of remission (35). However, given this Markov model is population-based it aims to describe the average probability for an event to occur and costs associated, rather than individual sequences of events.

A sixth limitation that stems from this heterogeneity is the wide availability of treatment options available to patients suffering from BD, and a lack of evidence regarding the use of these various options, as well as non-compliance to treatment over time which is currently not included in the model. As such, our model was limited to a selection of treatment options identified by (inter)national guidelines and expert opinion. With regards to the Dutch context, concordance with treatment guidelines assessed in the outpatient setting was found to be high (48). Moreover, as the main source for validating transition probabilities was based on empirical data stemming from the Dutch clinical setting, it is reasonable to assume that the modeled treatments (i.e., as part of SOC) are in line with the interventions provided in the Dutch study.

A seventh limitation concerns the generalizability of TiBipoMod's current model parameters, structure and assumptions across countries, for example in terms of locally available treatment options and the organization of care nationally. Currently, included treatment components (pharmacotherapy, psychotherapy, and community-based treatment etc.) are based on clinical guidelines published by the National Institute for Health and Care Excellence (NICE) in the United Kingdom and the Dutch National Health Care Institute, therefore likely better representing countries with similar health systems. In addition to that, by default the model is informed with healthcare resource use and unit costs representative of the Dutch context. Also, the model currently does not provide a detailed overview of the various accumulated costs carried across providers, which may be relevant for countries with a multiple payer system. Overall, depending on country-specific contexts, some future users may have to perform more model adaptations, or have limited information available to inform necessary parameters.

A final limitation of this model is that there remains room for further model development and implementation of novel concepts in health economic modeling. Examples of such novel concepts are the use of the expected value of (partial) perfect information (EV(P)PI), the value of hope, the inclusion of a broader societal perspective (i.e., costs related to public health, criminal justice, education, housing, or the environment), or alternative quality of life measures such as the Capabilities Approach, which contrasts the use of utilities in mental health by focusing on an individual's subjective wellbeing (76, 77). The use of EV(P)PI could, for example, provide insight in the expected costs of the decision uncertainty surrounding model input parameters, such as the transition probabilities. Outcomes of this analysis may identify if additional research is worthwhile, and what consequences could be when adopting the wrong treatment strategy (78, 79).

Conclusion

We presented TiBipoMod, a Markov model that is able to evaluate the long-term cost-effectiveness of pharmacological and non-pharmacological interventions in the treatment of adults with BD-I and BD-II from a healthcare and societal perspective. Overall, TiBipoMod aims to support researchers in adding conclusive knowledge to the limited health-economic evidence of treatments of BD in the clinical setting, supporting policy makers to make decisions considering the costs and effects of BD treatment. Moreover, TiBipoMod is freely available for academic purposes upon request from the authors. To support the development of this and other health-economic models for BD, future research should focus on increasing the availability of evidence to inform its parameters, and reduce related uncertainty for both BD-I and BD-II.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

JL and BW: conception and project design. BW, AK, and JL: model development. AK and BW: writing of the manuscript. AK, BW, SE, HK, JL, BG, and ER: critical appraisal of the manuscript. All authors read and approved the manuscript.

Funding

This work was supported by a grant from Netherlands Organization for Health Research and Development (Grant No. 843001803).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.1030989/full#supplementary-material

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5)*. Virginia: American Psychiatric Association (2013). doi: 10.1176/APPI.BOOKS.9780890425596

2. Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry.* (2011) 68:241–51. doi: 10.1001/ARCHGENPSYCHIATRY.2011.12

3. Kupka RW, Goossens P, van Bendegem M, Daemen P, Daggenvoorde T, Daniels M, et al. *Multidisciplinary Guidelines on Bipolar Disorder - Netherlands*. Utrecht: De Tijdstroom (2015). p. 343–54. doi: 10.2/JQUERY.MIN.JS

4. Mavranezouli I, Lokkerbol J. A systematic review and critical appraisal of economic evaluations of pharmacological interventions for people with bipolar disorder. *Pharmacoeconomics.* (2017) 35:271–96. doi: 10.1007/s40273-016-0473-1

5. Kraiss JT, Wijnen B, Kupka RW, Bohlmeijer ET, Lokkerbol J. Economic evaluations of non-pharmacological interventions and cost-of-illness studies in bipolar disorder: a systematic review. *J Affect Disord.* (2020) 276:388–401. doi: 10.1016/J.JAD.2020.06.064

6. Soares-Weiser K, Vergel YB, Beynon S, Dunn G, Barbieri M, Duffy S, et al. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder. *Health Technol Assess.* (2007) 11:3–4. doi: 10.3310/HTA11390

7. Cloutier M, Greene M, Guerin A, Touya M, Wu E. The economic burden of bipolar I disorder in the United States in 2015. *J Affect Disord.* (2018) 226:45–51. doi: 10.1016/J.JAD.2017.09.011

8. Vieta E, Berk M, Schulze TG, Carvalho AF, Suppes T, Calabrese JR, et al. Bipolar disorders. *Nat Rev Dis Prim.* (2018) 4:18008. doi: 10.1038/NRDP.2018.8

9. Hakkaart-van Roijen L, Hoeijenbos MB, Regeer EJ, Nolen WA, Veraart CPWM, Rutten FFH. The societal costs and quality of life of patients suffering from bipolar disorder in the Netherlands. *Acta Psychiatr Scand.* (2004) 110:383–92. doi: 10.1111/J.1600-0447.2004.00403.X

10. Blanthorn-Hazell S, Gracia A, Roberts J, Boldeanu A, Judge D. A survey of caregiver burden in those providing informal care for patients with schizophrenia or bipolar disorder with agitation: results from a European study. *Ann Gen Psychiatry.* (2018) 17:8. doi: 10.1186/S12991-018-0178-2

11. Perlick DA, Rosenheck RA, Miklowitz DJ, Chessick C, Wolff N, Kaczynski R, et al. Prevalence and correlates of burden among caregivers of patients with bipolar disorder enrolled in the Systematic Treatment Enhancement Program for Bipolar Disorder. *Bipolar Disord.* (2007) 9:262–73. doi: 10.1111/J.1399-5618.2007.00365.X

12. Miller S, Dell'osso B, Ketter TA. The prevalence and burden of bipolar depression. *J Affect Disord.* (2014) 169:3–11. doi: 10.1016/S0165-0327(14)70003-5

13. Rabarison KM, Bish CL, Massoudi MS, Giles WH. Economic evaluation enhances public health decision making. *Front Public Health.* (2015) 3:164. doi: 10.3389/FPUBH.2015.00164/BIBTEX

14. EUnetHTA. *Methods for Health Economic Evaluations-A Guideline based on Current Practices in Europe*. Diemen: EUnetHTA (2015).

15. Briggs A, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press (2006).

16. Wijnen BFM, Thielen FW, Konings S, Feenstra T, Van Der Gaag M, Veling W, et al. Designing and testing of a health-economic markov model for prevention and treatment of early psychosis. *Expert Rev Pharmacoecon Outcomes Res.* (2019) 20:269–79. doi: 10.1080/14737167.2019.1632194

17. Lokkerbol J, Wijnen B, Ruhe HG, Spijker J, Morad A, Schoevers R, et al. Design of a health-economic Markov model to assess cost-effectiveness and budget impact of the prevention and treatment of depressive disorder. *Expert Rev Pharmacoeconomics Outcomes Res.* (2021) 21:1031–42. doi: 10.1080/14737167.2021.1844566

18. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force-2. *Med Decis Making.* (2012) 32:678–89. doi: 10.1177/0272989X12454941

19. Dutch Institute National Health Care. *Guidelines for Health Economic Evaluations in The Netherlands*. (2016). Available online at: https://www.ispor.org/PEguidelines/source/NL-Economic_Evaluation_Guidelines.pdf (accessed February 29, 2016).

20. Naimark DMJ, Bott M, Krahn M. The half-cycle correction explained: two alternative pedagogical approaches. *Med Decis Making.* (2008) 28:706–12. doi: 10.1177/0272989X08315241

21. World Health Organization. *WHO–Choosing Interventions that are Cost Effective (WHO–CHOICE). Threshold Values for Intervention Cost-Effectiveness by Region*. Geneva: World Health Organization (2010).

22. Vijgen S, van Heesch F, Obradovic M. *Disease Burden in practice*. (2018). Available online at: https://www.zorginstituutnederland.nl/publicaties/rapport/2018/05/07/ziektelast-in-de-praktijk (accessed May 17, 2018).

23. Ferrari AJ, Stockings E, Khoo JP, Erskine HE, Degenhardt L, Vos T, et al. The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. *Bipolar Disord.* (2016) 18:440–50. doi: 10.1111/BDI.12423

24. Chisholm D, Van Ommeren M, Ayuso-Mateos JL, Saxena S. Cost-effectiveness of clinical interventions for reducing the global burden of bipolar disorder. *Br J Psychiatry.* (2005) 187:559–67. doi: 10.1192/BJP.187.6.559

25. Fajutrao L, Paulsson B, Liu S, Locklear J. Cost-effectiveness of quetiapine plus mood stabilizers compared with mood stabilizers alone in the maintenance therapy of bipolar I disorder: results of a markov model analysis. *Clin Ther.* (2009) 31:1456–68. doi: 10.1016/J.CLINTHERA.2009.06.009

26. Calvert NW, Burch SP, Fu AZ, Reeves P, Thompson TR. The cost-effectiveness of lamotrigine in the maintenance treatment of adults with bipolar I disorder. *J Manag Care Pharm.* (2006) 12:322–30. doi: 10.18553/jmcp.2006.12.4.322

27. McKendrick J, Cerri KH, Lloyd A, D'Ausilio A, Dando S, Chinn C. Cost effectiveness of olanzapine in prevention of affective episodes in bipolar disorder in the United Kingdom. *J Psychopharmacol.* (2007) 21:588–96. doi: 10.1177/0269881106068395

28. Woodward TC, Tafesse E, Quon P, Kim J, Lazarus A. Cost-effectiveness of quetiapine with lithium or divalproex for maintenance treatment of bipolar i disorder. *J Med Econ.* (2009) 12:259–68. doi: 10.3111/13696990903266612

29. Woodward TC, Tafesse E, Quon P, Lazarus A. Cost effectiveness of adjunctive quetiapine fumarate extended-release tablets with mood stabilizers in the maintenance treatment of bipolar i disorder. *Pharmacoeconomics.* (2010) 28:751–64. doi: 10.2165/11538350-000000000-00000

30. Ösby U, Brandt L, Correia N, Ekbom A, Sparén P. Excess mortality in bipolar and unipolar disorder in Sweden. *Arch Gen Psychiatry.* (2001) 58:844–50. doi: 10.1001/ARCHPSYC.58.9.844

31. Laursen TM. Life expectancy among persons with schizophrenia or bipolar affective disorder. *Schizophr Res.* (2011) 131:101–4. doi: 10.1016/J.SCHRES.2011.06.008

32. Hayes JF, Marston L, Walters K, King MB, Osborn DPJ. Mortality gap for people with bipolar disorder and schizophrenia: UK-based cohort study 2000-2014. *Br J Psychiatry.* (2017) 211:175–81. doi: 10.1192/BJP.BP.117.202606

33. Ekman M, Lindgren P, Miltenburger C, Meier G, Locklear JC, Lou Chatterton M. Cost effectiveness of quetiapine in patients with acute bipolar depression and in maintenance treatment after an acute depressive episode. *PharmacoEconomics.* (2012) 30:513–30. doi: 10.2165/11594930-000000000-00000

34. Koukopoulos A, Ghaemi SN. The primacy of mania: a reconsideration of mood disorders. *Eur Psychiatry.* (2009) 24:125–34. doi: 10.1016/J.EURPSY.2008.07.006

35. Koukopoulos A, Reginaldi D, Tondo L, Visioli C, Baldessarini RJ. Course sequences in bipolar disorder: depressions preceding or following manias or hypomanias. *J Affect Disord.* (2013) 151:105–10. doi: 10.1016/J.JAD.2013.05.059

36. Kupka RW, Altshuler LL, Nolen WA, Suppes T, Luckenbaugh DA, Leverich GS, et al. Three times more days depressed than manic or hypomanic in both bipolar I and bipolar II disorder. *Bipolar Disord.* (2007) 9:531–5. doi: 10.1111/J.1399-5618.2007.00467.X

37. Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch Gen Psychiatry.* (2002) 59:530–7. doi: 10.1001/ARCHPSYC.59.6.530

38. Judd LL, Akiskal HS, Schettler PJ, Coryell W, Endicott J, Maser JD, et al. A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder. *Arch Gen Psychiatry.* (2003) 60:261–9. doi: 10.1001/ARCHPSYC.60.3.261

39. Judd LL, Schettler PJ, Akiskal HS, Maser J, Coryell W, Solomon D, et al. Long-term symptomatic status of bipolar I vs. bipolar II disorders. *Int J Neuropsychopharmacol.* (2003) 6:127–37. doi: 10.1017/S1461145703003341

40. Joffe RT, MacQueen GM, Marriott M, Young LT. A prospective, longitudinal study of percentage of time spent ill in patients with bipolar I or bipolar II disorders. *Bipolar Disord.* (2004) 6:62–6. doi: 10.1046/J.1399-5618.2003.00091.X

41. Solomon DA, Leon AC, Coryell WH, Endicott J, Li C, Fiedorowicz JG, et al. Longitudinal course of bipolar I disorder duration of mood episodes. *Arch Gen Psychiatry.* (2010) 67:339–47. doi: 10.1001/archgenpsychiatry.2010.15

42. Vázquez GH, Holtzman JN, Lolich M, Ketter TA, Baldessarini RJ. Recurrence rates in bipolar disorder : systematic comparison of long-term prospective, naturalistic studies versus randomized controlled trials. *Eur Neuropsychopharmacol.* (2015) 25:1501–12. doi: 10.1016/j.euroneuro.2015.07.013

43. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team (2020).

44. Westman J, Hällgren J, Wahlbeck K, Erlinge D, Alfredsson L, Ösby U. Cardiovascular mortality in bipolar disorder: a population-based cohort study in Sweden. *BMJ Open.* (2013) 3:e002373. doi: 10.1136/BMJOPEN-2012-002373

45. Dome P, Rihmer Z, Gonda X. Suicide risk in bipolar disorder: a brief review. *Medicina.* (2019) 55:403. doi: 10.3390/MEDICINA55080403

46. National Institute for HealthCare and Excellence. *Bipolar Disorder: Assessment And Management - Clinical Guideline.* (2014). Available online at: www.nice.org.uk/guidance/cg185 (accessed February 21, 2022).

47. Goodwin GM, Haddad PM, Ferrier IN, Aronson JK, Barnes T, Cipriani A, et al. Evidence-based guidelines for treating bipolar disorder: revised third edition recommendations from the British Association for Psychopharmacology. *J Psychopharmacol.* (2016) 30:495–553. doi: 10.1177/0269881116636545

48. Renes JW, Regeer EJ, Hoogendoorn AW, Nolen WA, Kupka RW. A nationwide study on concordance with multimodal treatment guidelines in bipolar disorder. *Int J Bipolar Disord.* (2018) 6:22. doi: 10.1186/S40345-018-0130-Z

49. de Beus E, Rookmaker MB. *Guidelines on Renal Toxicity With Chronic Lithium Use [Dutch Report]*. Utrecht: Dutch Federation for Nefrology (2013).

50. Van Baal PHM, Wong A, Slobbe LCJ, Polder JJ, Brouwer WBF, De Wit GA. Standardizing the inclusion of indirect medical costs in economic evaluations. *Pharmacoeconomics.* (2011) 29:175–87. doi: 10.2165/11586130-000000000-00000/FGURES/TAB3

51. Hakkaart-van Roijen L, Van der Linden N, Bouwmans CAM, Kanters TA, Tan SS. *Costing Manual: Methodology for Costing Research and Reference Prices in Health Economic Evaluations [Dutch Report]*. Diemen: Dutch National Healthcare institute (2015).

52. Revicki DA, Hanlon J, Martin S, Gyulai L, Nassir Ghaemi S, Lynch F, et al. Patient-based utilities for bipolar disorder-related health states. *J Affect Disord.* (2005) 87:203–10. doi: 10.1016/j.jad.2005.03.017

53. Vemer P, Corro Ramos I, van Voorn GAK, Al MJ, Feenstra TL. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics.* (2016) 34:349–61. doi: 10.1007/S40273-015-0327-2

54. Büyükkaramikli NC, Rutten-van Mölken MPMH, Severens JL, Al M. TECH-VER: a verification checklist to reduce errors in models and improve their credibility. *Pharmacoeconomics.* (2019) 37:1391–408. doi: 10.1007/S40273-019-00844-Y/TABLES/4

55. Perich T, Manicavasagar V, Mitchell P, Ball J, Hadzi-Pavlovic D. A randomized controlled trial of mindfulness-based cognitive therapy for bipolar disorder. *Acta Psychiatr Scand.* (2013) 127:333–43. doi: 10.1111/A.12033

56. Mind National Platform for Mental Health, Trimbos Institute. *Bipolar Disorder | Treatment Options for Psychiological and Psychosocial Therapy - Roadmap to Dutch Treatment Guidelines*. Netherlands: Trimbos Institute (2017).

57. GGZstandaarden. *Budget Impact Analysis - Bipolar Disorder treatment guidelines [Dutch report]*. (2014). Available online at: https://www.ggzstandaarden.nl/uploads/side_products/3929e7b904812f504bc06cd7a6a6217c.pdf (accessed 10 May, 2022).

58. Ggz Standaarden. *Treatment guidelines for Bipolar Disorders in The Netherlands*. (2015). https://www.ggzstandaarden.nl/richtlijnen/bipolaire-stoornissen-richtlijn-2015/organisatie-van-zorg/inleiding (accessed February 11, 2022).

59. Healthcare Institute Netherlands. *Farmacotherapeutisch Kompas.* (2020). https://www.farmacotherapeutischkompas.nl/ (accessed January 20, 2020).

60. Healthcare Institute Netherlands. *National Database for Drug Prices*. (2020). Available online at: https://www.medicijnkosten.nl/ (accessed January 20, 2020).

61. Central Bureau for Statistics. *Consumer Price Indices - The Netherlands*. (2022). Available online at: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83131ned/table?fromstatweb (accessed February 21, 2022).

62. Shields GE, Buck D, Elvidge J, Hayhurst KP, Davies LM. Cost-effectiveness evaluations of psychological therapies for schizophrenia and bipolar disorder: a systematic review Europe PMC Funders Group. *Int J Technol Assess Heal Care.* (2019) 35:317–26. doi: 10.1017/S0266462319000448

63. Cohen JT, Neumann PJ, Wong JB. A call for open-source cost-effectiveness analysis. *Ann Intern Med.* (2017) 167:432–3. doi: 10.7326/M17-1153

64. Haji H, Afzali A, Karnon J. Addressing the challenge for well informed and consistent reimbursement decisions - the case for reference models. *Pharmacoeconomics.* (2011) 29:823–5. doi: 10.2165/11593000-000000000-00000

65. Poole C, Agrawal S, Currie CJ. Let cost effectiveness models be open to scrutiny. *BMJ.* (2007) 335:735–735. doi: 10.1136/BMJ.39360.379664.BE

66. Dunlop WCN, Mason N, Kenworthy J, Akehurst RL. Benefits, challenges and potential strategies of open source health economic models. *Pharmacoeconomics.* (2017) 35:125–8. doi: 10.1007/s40273-016-0479-8

67. Pouwels XGLV, Sampson CJ, Arnold RJG, Janodia MD, Henderson R, Lamotte M, et al. Opportunities and barriers to the development and use of open source health economic models: a survey. *Value Heal.* (2022) 25:473–9. doi: 10.1016/J.JVAL.2021.10.001

68. Standfield L, Comans T, Scuffham P. Markov modeling and discrete event simulation in health care: a systematic comparison. *Int J Technol Assess Health Care.* (2014) 30:165–72. doi: 10.1017/S0266462314000117

69. Karnon J, Haji Ali Afzali H. When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. *Pharmacoeconomics.* (2014) 32:547–58. doi: 10.1007/S40273-014-0147-9

70. Rosa AR, Reinares M, Michalak EE, Bonnin CM, Sole B, Franco C, et al. Functional impairment and disability across mood states in bipolar disorder. *Value Health.* (2010) 13:984–8. doi: 10.1111/J.1524-4733.2010.00768.X

71. Kapczinski F, Vieta E, Andreazza AC, Frey BN, Gomes FA, Tramontina J, et al. Allostatic load in bipolar disorder: implications for pathophysiology and treatment. *Neurosci Biobehav Rev.* (2008) 32:675–92. doi: 10.1016/J.NEUBIOREV.2007.10.005

72. Kapczinski F, Dias VV, Kauer-Sant'Anna M, Frey BN, Grassi-Oliveira R, Colom F, et al. Clinical implications of a staging model for bipolar disorders. *Expert Rev Neurother.* (2014) 9:957–66. doi: 10.1586/ERN.09.31

73. Caro JJ, Möller J. Advantages and disadvantages of discrete-event simulation for health economic analyses. *Expert Rev Pharmacoecon Outcomes Res.* (2016) 16:327–9. doi: 10.1586/14737167.2016.1165608

74. Denicoff KD, Leverich GS, Nolen WA, Rush AJ, McElroy SL, Keck PE, et al. Validation of the prospective NIMH-Life-Chart Method (NIMH-LCM-p) for longitudinal assessment of bipolar illness. *Psychol Med.* (2000) 30:1391–7. doi: 10.1017/S0033291799002810

75. Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, Mcdonald Scott P, et al. The Longitudinal Interval Follow-up Evaluation. A comprehensive method for assessing outcome in prospective longitudinal studies. *Arch Gen Psychiatry.* (1987) 44:540–8. doi: 10.1001/ARCHPSYC.1987.01800180050009

76. Drost RMWA, Paulus ATG, Ruwaard D, Evers SMAA. Valuing inter-sectoral costs and benefits of interventions in the healthcare sector: methods for obtaining unit prices. *Expert Rev Pharmacoecon Outcomes Res.* (2017) 17:77–84. doi: 10.1586/14737167.2016.1141679

77. White RG, Imperiale MG, Perera E. The capabilities approach: fostering contexts for enhancing mental health and wellbeing across the globe. *Global Health.* (2016) 12:1–10. doi: 10.1186/S12992-016-0150-3

78. Oostenbrink JB, Al MJ, Oppe M, Rutten-Van Mölken MPMH. Expected value of perfect information: an empirical example of reducing decision uncertainty by conducting additional research. *Value Heal.* (2008) 11:1070–80. doi: 10.1111/J.1524-4733.2008.00389.X

79. Grimm SE, Pouwels X, Ramaekers BLT, Wijnen B, Knies S, Grutters J, et al. Development and validation of the transparent uncertainty assessment (TRUST) tool for assessing uncertainties in health economic decision models. *Pharmacoeconomics.* (2020) 38:205. doi: 10.1007/S40273-019-00855-9

frontiers | Frontiers in Psychiatry

TYPE Perspective
PUBLISHED 21 December 2022
DOI 10.3389/fpsyt.2022.1031325

OPEN ACCESS

EDITED BY
Ali Jalali,
Joan and Sanford I. Weill Medical College of Cornell University, United States

REVIEWED BY
Cynthia Tschampl,
Brandeis University, United States

*CORRESPONDENCE
Jacob T. Painter
Jacob.Painter@va.gov

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Public Mental Health, a section of the journal Frontiers in Psychiatry

RECEIVED 29 August 2022
ACCEPTED 30 November 2022
PUBLISHED 21 December 2022

CITATION
Raciborski RA, Woodward EN and Painter JT (2022) Economic analyses of behavioral health intervention implementation: Perspective on stakeholder engagement. *Front. Psychiatry* 13:1031325. doi: 10.3389/fpsyt.2022.1031325

Economic analyses of behavioral health intervention implementation: Perspective on stakeholder engagement

Rebecca A. Raciborski[1,2†], Eva N. Woodward[1,3†] and Jacob T. Painter[1,2,4*]

[1]VA Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Healthcare System, North Little Rock, AR, United States, [2]Evidence, Policy, and Implementation Center, Central Arkansas Veterans Healthcare System, North Little Rock, AR, United States, [3]Department of Psychiatry, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, United States, [4]Division of Pharmaceutical Evaluation and Policy, College of Pharmacy, University of Arkansas for Medical Sciences, Little Rock, AR, United States

To provide full potential benefits to patients, behavioral health interventions often require comprehensive and systematic implementation efforts. The costs of these efforts should therefore be included when organizations decide to fund or adopt a new intervention. However, existing guidelines for conducting economic analyses like cost-effectiveness analyses and budget impact analyses are not well-suited to the complexity of the behavioral healthcare pathway and its many stakeholders. Stakeholder engagement, when used effectively with recent innovations in economic analysis, advance more equitable access to interventions for individuals living with behavioral health conditions. But early and ongoing stakeholder engagement has not yet been incorporated into best-practice guidelines for economic evaluation. We discuss our perspective, as researchers and clinicians in a large integrated health system, on how the integration of stakeholder engagement with existing economic analysis methods could improve decision-making about implementation of behavioral health interventions.

KEYWORDS

behavioral health, implementation, cost, economic analysis, cost-effectiveness, budget impact, stakeholder engagement, economic evaluation

Introduction

Treating behavioral health conditions is imperative. Mental health and substance disorders are the world's leading cause of disability and fifth highest cause of death (1). Treatment is also expensive; treatment of major depressive disorder in the U.S. alone exceeds $300 billion annually (2). If a new efficacious intervention is introduced, how much benefit will it provide in improved function or reduced mortality? What will it cost to implement it with fidelity so that it is effective in practice? Is it affordable? Economic analyses aim to answer these kinds of questions. Like clinical trials, they are

conducted with the goal of obtaining sufficient information to make a policy decision. These analyses are a standard method by which policymakers and payers decide if a new treatment should be available. However, existing pharmacologic-focused guidelines for conducting economic analyses do not easily extend to behavioral health interventions (2, 3) or their implementation (4).

While there are myriad challenges, many relate to the complexity of the behavioral healthcare pathway. Costs and benefits of behavioral health interventions are distributed unevenly and in meaningful ways to stakeholders beyond the payer and patient. When successful treatment depends on patients, their close contacts, and their clinicians, then decision makers are best served by analyses that incorporate these other perspectives. We believe that engaging all stakeholders increases the usefulness of economic analyses as a decision-making tool. We discuss our perspective on the challenges of applying economic analysis to behavioral health interventions, implementation thereof, and innovations in the field that may potentially improve equity in economic analyses.

Economic view of implementing behavioral health interventions

Implementing behavioral health interventions requires significant investment; trainings and manualized protocols for psychotherapies are not sufficient to ensure psychotherapies are used by clinicians (5). Other strategies are required to implement behavioral healthcare interventions, increase adoption by providers, and reach more patients. Such implementation strategies may include changing infrastructure (e.g., physical space alterations and re-organizing teams), increasing demand among patients through marketing, and engaging relevant personnel at multiple levels (e.g., leadership and frontline staff) (6). The economist sees a cost attached to not only new space or printed marketing materials, but the time spent training teams on new workflow and meetings to create buy-in.

When deciding to implement an intervention, economic analysis is one approach to inform organizational decision-making. In our work, economic analyses inform decisions about providing and funding behavioral health interventions in the US Department of Veterans Affairs (VA). Cost-effectiveness analyses have long been a feature of VA clinical trial research (7, 8); and research about implementation of new behavioral health interventions is incorporating these analyses too. VA's Quality Enhancement Research Initiative now requires a budget impact analysis before implementing new interventions. We focus in this article on two methods we use most frequently as they most often meet needs of payers and decision makers: cost-effectiveness and budget impact analyses. **Table 1** provides a brief overview of the two approaches. Other approaches (e.g., cost benefit analysis) also have applications to psychiatric care. Luyten et al. (9) provide an introduction and Knapp and Wong (10) provide a comprehensive review from a psychiatric perspective.

Cost-effectiveness analysis and budget impact analysis have different analytic goals but share some features. Cost-effectiveness analyses classically support decisions about whether an intervention should be made available to patients or to which patients it will be made available if there are heterogeneous treatment effects. Recently, they have examined the relative value of competing implementation strategy bundles to enhance uptake of behavioral health interventions (11, 12). By contrast, budget impact analyses support decisions about whether an intervention can be made available given the payer's budget or under what conditions it would be possible to do so (e.g., patient copayment). The "payer" is typically an insurer or national health service. When evaluating implementations, the payer is more likely to be the adopting organization alone

TABLE 1 Comparison of cost effectiveness and budget impact analysis.

	Cost effectiveness analysis	Budget impact analysis
Decision informed	Is the intervention worthwhile?	Can the intervention be afforded within current budget constraints?
Typical audience	Policymakers	Budget holders within an organization
Outcome of interest	Disease/condition-specific clinical endpoint *or* generic utility-based measure for quality of life	Changes in treatment mix/resource use after introduction of new intervention
Method of assigning value to benefits	Direct (or "natural") measurement of clinical endpoints *or* socially determined preference-weights for quality-of-life measures	Currency-denominated accounting cost that is avoided (if any)
Costs included	Health sector perspective: Formal health sector costs Additional cost for societal perspective: Informal caregiver costs, Non-health sector costs, Productivity losses	Costs incurred by the payer (typically only health sector costs are incurred)
Method of assigning value to costs	Currency-denominated economic cost, including value of opportunity cost	Currency-denominated accounting cost

because a new intervention's implementation is usually not directly reimbursable.

Care for behavioral health conditions is complex with many people involved in its delivery and use. An intervention incurs cost at each stage of the behavioral healthcare pathway. Some costs are easily identifiable (e.g., amount an insurer pays for a counseling session). But others are more complex. An office visit minimally involves scheduling clerks, screening technicians, and provider teams; inpatient and emergent care requires an even more diverse mix of staff and material resources. The immediate cost of time, measured by each staff member's wage and fringe benefit rate, and other operating expenses is borne by the organization providing care. Ideally, insurer payments are sufficient to compensate for these costs. Less common "costs" to providers are more qualitative in nature (e.g., managing higher severity conditions). Receiving care also requires time from patients and, in the case of conditions involving diminished capacity (e.g., severe posttraumatic stress disorder), their caregivers.

Finally, there are costs associated with the condition the intervention seeks to ameliorate. For payers, these costs are incurred because of disease complications (e.g., psychiatric hospitalization). For providers, cost may be increased time for patient disease management and care coordination or the added stress of caring for a patient in crisis. The greatest costs though tend to be borne by patients (e.g., lost wages when unable to work) and their informal caregivers (e.g., spouse's uncompensated time). Condition-related costs also include reduced quality of life for patients and, in the case of severe conditions (e.g., schizophrenia), for their families and close contacts. Some conditions reach even further into society (e.g., through cost of supportive housing).

Weighed against costs are the tangible and intangible benefits of treatment to patients, their families, close contacts, and the societies in which they live (e.g., improved health, jobs retained, and relationships stabilized). Providers benefit from seeing patient improvements and from implementation of interventions that improve workflow and reduce stress. Healthcare organizations benefit when implementation strategies are selected for cost-effectiveness and increased confidence in decisions. Payers benefit by avoiding the cost of disease complications.

The approach to measuring costs and benefits differs based on analytic goals. Cost-effectiveness analysis quantifies the benefit as a relative gain in some measure of health improvement (e.g., hospitalizations avoided) compared to the cost to health system or society. When benefits are measured in years added to a patient's life and weighted for the quality of life experienced in those additional years, i.e., *quality-adjusted life years* (QALYs), analyses may be called *cost-utility analyses*. Following others (13), we include analyses with benefits measured in QALYs in our use of "cost-effectiveness analysis." By contrast, budget impact analysis includes only cost incurred by the payer; benefits enter only if the payer avoids a cost. The two analyses differ in the time frame over which the measurement occurs as well. Cost-effectiveness analyses tend to be long-run projections while budget impact analyses are usually confined to a 1-to-5-year period based on the payer's budgetary planning cycle.

Regardless of method, the analytic team makes choices about what costs and benefits are relevant to the decision and how to measure them (e.g., where the data comes from, how detailed it should be, over what period it is gathered). These choices rely on assumptions about importance and magnitude. The assumptions set and calculations it leads to are referred to as the economic *model structure*. For both cost-effectiveness and budget impact analyses, the degree of uncertainty about costs and benefits increases the farther into the future projections occur. When this uncertainty is due to decisions about the assumptions themselves, it is called *structural uncertainty*.

Challenges analyzing behavioral health interventions and their implementation

The complexity of behavioral healthcare makes fundamental decisions that determine model structure particularly challenging. Challenges begin with defining which patients should be considered "treated" and what the intervention costs. Behavioral interventions like psychotherapy are tailored to patient needs; completion of treatment and likelihood of obtaining full benefits varies by patient. The amount of time, and thus labor, also depends on individual patient needs. Labor costs vary by provider type, geographic region, and other local factors. This contrasts with pharmaceuticals, which generally have a common "list price" from the payer's perspective.

Further, while efficacy and patient adherence to treatment mostly determine outcomes from pharmaceutical treatment, effectiveness of behavioral health interventions also depends on clinician fidelity. Measures of fidelity are sometimes gathered in the context of a trial but rarely collected otherwise. At the clinic or organizational level, proper implementation is necessary to ensure fidelity and thus patient improvements. The cost of the implementation effort is thus relevant to the payer but the implementation strategies, like the intervention itself, lack a list price (4, 14). Implementation cost data may be collected during a trial using existing methods (15) but they likely overstate cost in practice when cost per patient declines as each clinician treats more patients and as caseloads increase beyond typical trial size (3).

Another complication is the relative heterogeneity of the "usual care" comparator across geographic regions, healthcare systems, and individual providers. Standard treatment of some behavioral health conditions has been codified. But for many, a variety factors contribute to what is considered usual care, such as duration of sessions in scheduling grids, match between

patient severity and available services, or local norms regarding therapeutic orientation.

In cost-effectiveness analysis, these challenges are compounded when determining relevant costs and benefits. Recommendations are clear that an enhanced health-system perspective, and ideally a societal perspective, should provide the reference case (13). When non-health system costs and benefits are high, the choice between the health system and societal perspective can alter the conclusion about whether an intervention is cost-effective (16, 17). Regardless, it is common for analyses to omit costs and benefits that are time consuming to gather or that researchers do not conceptualize as relevant (17). Yet, many behavioral health interventions result in substantial non-health system costs (e.g., caregiver time) and benefits (e.g., opioid use disorder therapy reducing criminal justice involvement) (3, 10). Behavioral health interventions are subject to substantial uncompensated patient time costs (3), such as time spent integrating practices learned in psychotherapy into daily life. Neglecting these costs could skew the estimated cost-effectiveness if they ultimately change behavior of non-payer participants in the behavioral healthcare pathway.

Budget impact analysis confronts a similar challenge, even though defining costs from the payer perspective seems straightforward (18). The payer perspective may oversimplify how organizational change occurs, particularly for implementation of new behavioral health interventions with multiple intra-organization budget holders and stakeholders who incur disproportional costs (e.g., an integrated health system implements emergency department-based suicide screening that refers more patients to specialty mental health). Additionally, outside of integrated systems, budget impact analyses conducted from an insurer's perspective omit implementation costs incurred by providers (e.g., time spent training for required credentials) unless these costs are reimbursed.

Striking the right balance of detail for all stakeholders is critical. If the analytic team spends time gathering data on costs that ultimately have little effect on the conclusion, the timeliness of the analysis to the decision maker is reduced.

Discussion

Understanding needs of decision-makers and matching these to the knowledge of other stakeholders ensures the analysis provides useful information. Importantly, our operational and clinical partners often have concerns beyond cost, including health equity, when considering implementing an intervention. Understanding their concerns and successfully using other stakeholders' knowledge requires early and frequent engagement. Stakeholder engagement spans the continuum of intensity from unidirectional consultations (least intensive) to potentially co-produced evaluations (most intensive) (19).

As applied to economic analysis, stakeholder engagement can help specify relevant analytic goals. Applying principles of community-based participatory research (20) and implementation science (21) to engage those most affected by the decision being made may also make economic analyses more equitable. Stakeholder engagement can also help determine and refine the economic model structure when faced with the complexities of behavioral healthcare. Although economic analyses are increasingly acknowledging a role for stakeholder engagement, best-practice guidelines have not yet incorporated advice for when or how to do so.

Specifying analytic goals

Identifying and communicating with relevant stakeholders is essential for economic analyses. An analysis that omits key components needed for the decision leaves decision makers no better informed, and possibly worse, if the wrong conclusion is presented. While cost is a necessary consideration for budget holders, it may need to be weighed against other stakeholder-identified factors. If so, stakeholders can be involved when developing an economic analysis plan to specify goals. Despite its promise, this practice is not widely adopted (22) and has mostly involved pharmaceutical manufacturers requesting input from regulatory agencies (23).

More recently, implementation scientists are focusing on cost as a factor that enables or hinders uptake of an intervention and the value of cost information to decision-makers (4, 24, 25). What staff and providers stand to gain from offering or improving behavioral health interventions depends on their context (e.g., funding structures, existing capacity for quality improvement). Depending on their role, they may incur substantially different costs as well, which can influence the rapidity of the new intervention's diffusion in practice. Knowing such information in advance may help anticipate reluctance to engage during implementation.

Scoping and measuring costs and benefits with stakeholder engagement

Decision makers and other stakeholders exist all along the behavioral healthcare pathway. It is essential to ensure that senior leadership and core staff involved in quality improvement at the organization are involved in scoping and measuring cost and benefits (21, 26). This improves decision maker confidence and minimizes the risk that the model structure relies on poor assumptions about costs and benefits (22). For example, when conducting a budget impact analysis, stakeholders within an organization can rely on institutional knowledge to highlight areas where structural uncertainty may exist, particularly in later years. Consider the case of rapid organic practice change. If a

new behavioral intervention is being introduced when "usual care" patterns are already shifting, frontline staff will be best positioned to know this. Economists can then incorporate this knowledge with alternate scenarios to present more useful information.

As a technical matter, stakeholder engagement can reduce the analytic burden and ensure primary data collection efforts are focused where the value of information is highest. For example, in the absence of a standard list price for a behavioral health intervention, clinicians can identify the current local standard of practice and which components of usual care are most relevant. Clinical leadership can clarify which strategies and processes are required for implementation and how long it takes an organization to move from planning to offering a new intervention to patients. It is possible to quantify the value of additional information through a formal value of information analysis (27). Such analyses have become more common when conducting cost-effectiveness analyses.

Early and ongoing stakeholder engagement with those who will bear cost or stand to benefit from an intervention during implementation also helps identify if multiple perspectives are needed when scoping and measuring costs and benefits (25, 28, 29). Incorporating other stakeholder perspectives on costs and benefits offers insights to behaviors like treatment engagement, adherence, and clinician adoption of and fidelity to interventions. This may mean incorporating measures of departmental level cost within the organization to determine if cost is being shifted. Or, to the extent that provider-assessed value influences adoption of promising new interventions, it may mean adding non-cost outcome measures salient to providers (e.g., patient gains and reduced caseload). Because the benefits and costs do not always accrue to the same stakeholders, conflicts may arise. Many approaches to resolving such conflicts have been proposed (30–32) and their application to stakeholder engagement in economic analyses should be explored in future work.

Equity in economic analyses of behavioral health interventions

We use the term "health equity" broadly, referring to a range of ethical concerns from opportunity to achieve full health to respect for individual autonomy (33). As a US federal agency, VA must now incorporate equity into its program evaluations (34). As more program evaluations incorporate an economic component, equity concerns and economic analyses are increasingly intersecting. Fortunately, several innovations in economic analysis methods, along with insights from other fields, support addressing these concerns.

Evaluators can incorporate equity into the economic analysis process as early as goal specification. For example, decision-makers may request having patient financial costs included alongside a traditional budget impact analysis if they want to identify which of similar interventions is the least burdensome to patients. The emerging field of distributional cost-effectiveness analysis provides a framework for explicitly considering questions about how an intervention changes the distribution of health, health service access, cost, and protection from financial risk (35). Approaches borrowed from other fields, such as community-based participatory research, further increase equitable practices when those affected by the decision are given a greater voice in how it is framed.

Scoping costs and benefits is another potential point of intersection between economic analysis and equity. Economists necessarily specify the initial model structure; their choices about what costs and benefits are important will be shaped by their lived experience. However, stakeholders along the full continuum of the behavioral healthcare pathway may prioritize costs and benefits differently from those scoped into economic analyses with the standard decision-maker perspective. Patients likely have different perceptions of value (36), especially those like individuals living with serious mental illness who experience marginalization or are part of minoritized populations. Teams can engage in reflexivity practices throughout the process by explicitly considering differences in perceived value to enhance a focus on ethical, equitable use of economic evaluation (37, 38). Frequent interaction should be continued until the economic analysis is complete, with the team asking questions, sharing updates, and listening to and incorporating stakeholder feedback (26).

Once the scope is determined, equity can also inform decisions about processes for measuring costs and benefits. For example, many US federal agencies prohibit the use of cost-per-QALY measures to make decisions about whether an intervention will be available to patients, though some permit comparisons to choose between treatments for the same cohort of patients (39). The primary factor leading to US restrictions on QALYs was concern about discriminatory effects for people with chronic health conditions (39). Adopting more sensitive measures of health status change and better assessment of the value of that change could mitigate potential discrimination. Organizations, like VA, that conduct trials and have access to large populations of patients with behavioral health conditions are ideally positioned to further develop instruments that capture changes in quality of life salient to individuals with behavioral health conditions (40–42) and explore how decisions change when of using quality of life valuations from those with experience of a behavioral health condition instead of standard societal valuations (43). Research should also explore the result of incorporating spillover quality of life effects in analyses of behavioral health interventions and their implementation (44).

Economic analysis and equity also intersect when the decision is made to implement an intervention. Distributional

cost-effectiveness analysis provides economists with a framework for ensuring decision-makers have sufficient information to balance total health and equity tradeoffs. Another approach is the individualized comparative effectiveness framework that focuses on the relative benefit of different interventions for subgroups of patients (45). This information can be used by clinicians, patients, and caregivers to assume a greater role in the decision-making process.

As an example of where these principles apply, consider integration of mental health services into primary care, which reduces depression symptoms in patients and costly utilization (46, 47). A traditional economic analysis may be a cost-effectiveness analysis for policymakers considering endorsing the change or budget impact analysis for clinic owners. An equitable analysis would identify any pre-existing inequities that may be exacerbated. For example, decision-makers may want to know if the intervention disproportionately benefits patients who already have good access to primary care services. To be relevant to patients, the analysis also may include multifaceted "success" metrics that address symptom reduction and quality of life. Finally, integration may affect job satisfaction and burnout for primary care clinicians (48). Implementation scientists could use the results of such an analysis to design approaches to mitigate clinician burnout effects.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

RR wrote and edited the manuscript, conceptualized framing with respect to economic analysis challenges and new developments, and provided references to key literature. EW wrote and edited the manuscript, conceptualized framing in relation to clinical audience and in context of implementation, and provided references to key literature. JP wrote and edited the manuscript and provided references to key literature. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Center for Mental Healthcare and Outcomes Research Number CIN 13-411 from the United States Department of Veterans Affairs Health Services Research and Development Center of Innovation (RR and JP), the Evidence Policy, and Implementation Center Number EBP 22-104 from the United States Department of Veterans Affairs (RR and JP), Behavioral Health Quality Enhancement Research Initiative (JP), and Career Development Award Number CDA 18-192 from the United States Department of Veterans Affairs Health Services Research and Development (HSRD) Service (EW).

Acknowledgments

The authors thank Mary Kate Bartnik, MA, for assistance with literature search and manuscript preparation. The authors also thank reviewer and the guest editor for valuable feedback.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Author disclaimer

The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References

1. Whiteford H, Ferrari A, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PLoS One.* (2015) 10:e0116820.

2. Proudman D, Greenberg P, Nellesen D. The growing burden of major depressive disorders (MDD): implications for researchers and policy makers. *PharmacoEconomics.* (2021) 39:619–25. doi: 10.1007/s40273-021-01040-7

3. Shearer J, McCrone P, Romeo R. Economic evaluation of mental health interventions: a guide to costing approaches. *Pharmacoeconomics.* (2016) 34:651–64. doi: 10.1007/s40273-016-0390-3

4. Wagner T, Yoon J, Jacobs J, So A, Kilbourne A, Yu W, et al. Estimating costs of an implementation intervention. *Med Decis Making.* (2020) 40:959–67. doi: 10.1177/0272989X20960455

5. Duncan B, Miller S. Treatment Manuals Do Not Improve Outcomes. In: Norcross JC, Beutler LE, Levant RF editors. *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. Washington D.C: American Psychological Association Press (2005).

6. Waltz T, Powell B, Matthieu M, Damschroder L, Chinman M, Smith J, et al. Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert Recommendations for Implementing Change (ERIC) study. *Implement Sci.* (2015) 10:1–8. doi: 10.1186/s13012-015-0295-0

7. Liu C, Hedrick S, Chaney E, Heagerty P, Felker B, Hasenberg N, et al. Cost-effectiveness of collaborative care for depression in a primary care veteran population. *Psychiatr Serv.* (2003) 54:698–704. doi: 10.1176/appi.ps.54.5.698

8. Pyne J, Fortney J, Tripathi S, Maciejewski M, Edlund M, Williams D. Cost-effectiveness analysis of a rural telemedicine collaborative care intervention for depression. *Arch Gen Psychiatry.* (2010) 67:812–21. doi: 10.1001/archgenpsychiatry.2010.82

9. Luyten J, Naci H, Knapp M. Economic evaluation of mental health interventions: an introduction to cost-utility analysis. *Evid Based Ment Health.* (2016) 19:49–53. doi: 10.1136/eb-2016-102354

10. Knapp M, Wong G. Economics and mental health: the current scenario. *World Psychiatry.* (2020) 19:3–14. doi: 10.1002/wps.20692

11. Eisman A, Hutton D, Prosser L, Smith S, Kilbourne A. Cost-effectiveness of the Adaptive Implementation of Effective Programs Trial (ADEPT): approaches to adopting implementation strategies. *Implement Sci.* (2020) 15:1–3. doi: 10.1186/s13012-020-01069-w

12. Dopp A, Hanson R, Saunders B, Dismuke C, Moreland A. Community-based implementation of trauma-focused interventions for youth: Economic impact of the learning collaborative model. *Psychol Serv.* (2017) 14:57. doi: 10.1037/ser0000131

13. Sanders G, Neumann P, Basu A, Brock D, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA.* (2016) 316:1093–103. doi: 10.1001/jama.2016.12195

14. Eisman A, Kilbourne A, Dopp A, Saldana L, Eisenberg D. Economic evaluation in implementation science: Making the business case for implementation strategies. *Psychiatry Res.* (2020) 283:30752–8. doi: 10.1016/j.psychres.2019.06.008

15. Cidav Z, Mandell D, Pyne J, Beidas R, Curran G, Marcus S. A pragmatic method for costing implementation strategies using time-driven activity-based costing. *Implement Sci.* (2020) 15:28. doi: 10.1186/s13012-020-00993-1

16. Sullivan S, Mauskopf J, Augustovski F, Caro J, Lee K, Minchin M, et al. Budget impact analysis—principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. *Value Health.* (2014) 17:5–14. doi: 10.1016/j.jval.2013.08.2291

17. Drost R, van der Putten I, Ruwaard D, Evers S, Paulus A. Conceptualizations of the societal perspective within economic evaluations: a systematic review. *Int J Technol Assess Health Care.* (2017) 33:251–60. doi: 10.1017/S0266462317000526

18. Fairley M, Humphreys K, Joyce V, Bounthavong M, Trafton J, Combs A, et al. Cost-effectiveness of treatments for opioid use disorder. *JAMA Psychiatry.* (2021) 78:767–77. doi: 10.1001/jamapsychiatry.2021.0247

19. Goodman M, Sanders Thompson V. The science of stakeholder engagement in research: classification, implementation, and evaluation. *Transl Behav Med.* (2017) 7:486–91. doi: 10.1007/s13142-017-0495-z

20. Wallerstein N, Duran B. Community-Based participatory research contributions to intervention research: The intersection of science and practice to improve health equity. *Am J Public Health.* (2010) 100(Suppl 1):S40–6. doi: 10.2105/AJPH.2009.184036

21. Kirchner J, Kearney L, Ritchie M, Dollar K, Swensen A, Schohn M. Research & services partnerships: Lessons learned through a national partnership between clinical leaders and researchers. *Psychiatr Serv.* (2014) 65:577–9. doi: 10.1176/appi.ps.201400054

22. Xie R, Malik E, Linthicum M, Bright J. Putting stakeholder engagement at the center of health economic modeling for health technology assessment in the United States. *Pharmacoeconomics.* (2021) 39:631–8. doi: 10.1007/s40273-021-01036-3

23. Maignen F, Osipenko L, Gajraj E, Chivers R. Trends in early engagement between industry and HTA: analysis of scientific advice service provided by Nice since 2009. *Value Health.* (2014) 17:A441. doi: 10.1016/j.jval.2014.08.1159

24. Saldana L, Ritzwoller D, Campbell M, Block E. Using economic evaluations in implementation science to increase transparency in costs and outcomes for organizational decision-makers. *Implement Sci Commun.* (2022) 3:40. doi: 10.1186/s43058-022-00295-1

25. Eisman A, Quanbeck A, Bounthavong M, Panattoni L, Glasgow R. Implementation science issues in understanding, collecting, and using cost estimates: a multi-stakeholder perspective. *Implement Sci.* (2021) 16:75. doi: 10.1186/s13012-021-01143-x

26. Sharek P, Mullican C, Lavanderos A, Palmer C, Snow V, Kmetik K, et al. Best practice implementation: lessons learned from 20 partnerships. *Jt Comm J Qual Patient Saf.* (2007) 33:16–26. doi: 10.1016/S1553-7250(07)33120-6

27. Sculpher M, Basu A, Kuntz K, Meltzer D. Reflecting uncertainty in cost-effectiveness analysis. Second ed. In: Neumann P, Sanders G, Russell L, Siegel J, Ganiats T editors. *Cost Effectiveness in Health and Medicine*. Oxford: Oxford University Press (2017). p. 289–318. doi: 10.1093/acprof:oso/9780190492939.003.0011

28. Delafield R, Hermosura A, Ing C, Hughes C, Palakiko D, Dillard A, et al. A community-based participatory research guided model for dissemination of evidence-based interventions. *Prog Commun Health Partnersh.* (2016) 10:585. doi: 10.1353/cpr.2016.0067

29. Bauer M, Miller C, Kim B, Lew R, Weaver K, Coldwell C, et al. Partnering with health system operations leadership to develop a controlled implementation trial. *Implement Sci.* (2015) 11:22. doi: 10.1186/s13012-016-0385-7

30. Israel B, Schulz A, Parker E, Becker A. *Critical Issues in Developing and Following CBPR Principles. Community-Based Participatory Research for Health*. 3rd ed. San Francisco, CA: Jossey-Bass (2018). p. 31–46.

31. Harvey N, Holmes C. Nominal group technique: an effective method for obtaining group consensus. *Int J Nurs Pract.* (2012) 18:188–94. doi: 10.1111/j.1440-172X.2012.02017.x

32. Simoens S. Using the Delphi technique in economic evaluation: time to revisit the oracle? *J Clin Pharm Ther.* (2006) 31:519–22. doi: 10.1111/j.1365-2710.2006.00780.x

33. Cookson R, Culyer A, Norheim O. Principles of health equity. In: Cookson R, Griffin S, Culyer A, Norheim O editors. *Distributional Cost-Effectiveness Analysis: Quantifying Health Equity Impacts and Trade-Offs*. Oxford: Oxford University Press (2021). p. 18–43. doi: 10.1093/med/9780198838197.003.0002

34. The White House. *Executive Order on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government*. Washington, D.C: The White House (2021).

35. Cookson R, Griffin S, Culyer A, Norheim O. Designing a distributional cost-effectiveness analysis. In: Cookson R, Griffin S, Culyer A, Norheim O editors. *Distributional Cost-Effectiveness Analysis: Quantifying Health Equity Impacts and Trade-Offs*. Oxford: Oxford University Press (2021). p. 44–68. doi: 10.1093/med/9780198838197.003.0003

36. Watkins, J. Understanding value: the patients' perspective. *Value Outcomes Spotlight.* (2022) 8:26–28.

37. Jamieson M, Govaart G, Pownall M. Reflexivity in quantitative research: a rationale and beginner's guide. *PsyArXiv*[Preprint]. (2022): doi: 10.31234/osf.io/xvrhm

38. Gunn C, Bertelsen N, Regeer B, Schuitmaker-Warnaar T. Valuing patient engagement: Reflexive learning in evidence generation practices for health technology assessment. *Soc Sci Med.* (2021) 280:114048. doi: 10.1016/j.socscimed.2021.114048

39. National Council on Disability. *Quality-Adjusted Life Years and the Devaluation of Life with Disability: Part of the Bioethics and Disability Series*. (2019). Available online at: https://ncd.gov/sites/default/files/NCD_Quality_Adjusted_Life_Report_508.pdf (accessed May 10, 2022).

40. Pyne J, Tripathi S, French M, McCollister K, Rapp R, Booth B. Longitudinal association of preference-weighted health-related quality of life measures and substance use disorder outcomes. *Addiction.* (2011) 106:507–15. doi: 10.1111/j.1360-0443.2010.03299.x

41. Papaioannou D, Brazier J, Parry G. How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in schizophrenia? A systematic review. *Value Health.* (2011) 14:907–20. doi: 10.1016/j.jval.2011.04.006

42. Mulhern B, Mukuria C, Barkham M, Knapp M, Byford S, Brazier J. Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. *Br J Psychiatry.* (2014) 205:236–43. doi: 10.1192/bjp.bp.112.122283

43. Neil A, Carr V, Mackinnon A, Foley D, Morgan V. Health-related quality of life in people living with psychotic illness and factors associated with its variation. *Value Health.* (2018) 21:1002–9. doi: 10.1016/j.jval.2018.02.012

44. Basu A, Meltzer D. Implications of spillover effects within the family for medical cost-effectiveness analysis. *J Health Econ.* (2005) 24:751–73. doi: 10.1016/j.jhealeco.2004.12.002

45. Basu A. Economics of individualization in comparative effectiveness research and a basis for a patient-centered health care. *J Health Econ.* (2011) 30:549–59. doi: 10.1016/j.jhealeco.2011.03.004

46. Patel U, Blackmore M, Stein D, Carleton K, Chung H. Costs and utilization for low income minority patients with depression in a collaborative care model implemented in a community-based academic health system. *Health Serv Res.* (2020) 55:106–7. doi: 10.1111/1475-6773.13482

47. Leung L, Rubenstein L, Post E, Trivedi R, Hamilton A, Yoon J, et al. Association of veterans affairs primary care mental health integration with care access among men and women veterans. *JAMA Netw Open.* (2020) 3:e2020955. doi: 10.1001/jamanetworkopen.2020.20955

48. Leung L, Rose D, Rubenstein L, Guo R, Dresselhaus T, Stockdale S. Does mental health care integration affect primary care clinician burnout? Results from a longitudinal veterans affairs survey. *J Gen Intern Med.* (2020) 35:3620–6. doi: 10.1007/s11606-020-06203-4

frontiers | Frontiers in Psychiatry

TYPE Systematic Review
PUBLISHED 16 February 2023
DOI 10.3389/fpsyt.2023.1056210

OPEN ACCESS

EDITED BY
Ali Jalali,
Joan & Sanford I. Weill Medical College of Cornell University, United States

REVIEWED BY
Yee Ming Mok,
Institute of Mental Health, Singapore
Francesco Monaco,
Azienda Sanitaria Locale Salerno, Italy

*CORRESPONDENCE
Rory A. Cameron
✉ rory.cameron@uea.ac.uk

SPECIALTY SECTION
This article was submitted to Public Mental Health, a section of the journal Frontiers in Psychiatry

RECEIVED 28 September 2022
ACCEPTED 31 January 2023
PUBLISHED 16 February 2023

CITATION
Hannah LA, Walsh CM, Jopling L, Perez J, Cardinal RN and Cameron RA (2023) Economic evaluation of interventions for treatment-resistant depression: A systematic review.
Front. Psychiatry 14:1056210.
doi: 10.3389/fpsyt.2023.1056210

Economic evaluation of interventions for treatment-resistant depression: A systematic review

Laura A. Hannah [1,2], Cathy M. Walsh[1,2], Louise Jopling[3], Jesus Perez [1,2,4,5,6], Rudolf N. Cardinal [2,4] and Rory A. Cameron [1,5]*

[1]Applied Research Collaboration East of England, National Institute for Health and Care Research, Cambridge, United Kingdom, [2]Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, United Kingdom, [3]Eastern Academic Health Science Network, Cambridge, United Kingdom, [4]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom, [5]Norwich Medical School, University of East Anglia, Norwich, United Kingdom, [6]Psychiatry Unit, Department of Medicine, Institute of Biomedical Research of Salamanca, University of Salamanca, Salamanca, Spain

Background: The extraordinarily high prevalence of treatment-resistant depression (TRD), coupled with its high economic burden to both healthcare systems and society, underscore how critical it is that resources are managed optimally to address the significant challenge it presents.

Objective: To review the literature on economic evaluation in TRD systematically, with the aim of informing future studies by identifying key challenges specific to the area, and highlighting good practices.

Methods: A systematic literature search across seven electronic databases was conducted to identify both within-trial and model-based economic evaluations in TRD. Quality of reporting and study design was assessed using the Consensus Health Economic Criteria (CHEC). A narrative synthesis was conducted.

Results: We identified 31 evaluations, including 11 conducted alongside a clinical trial and 20 model-based evaluations. There was considerable heterogeneity in the definition of treatment-resistant depression, although with a trend for more recent studies to use a definition of inadequate response to two or more antidepressive treatments. A broad range of interventions were considered, including non-pharmacological neuromodulation, pharmacological, psychological, and service-level interventions. Study quality as assessed by CHEC was generally high. Frequently poorly reported items related to discussion of ethical and distributional issues, and model validation. Most evaluations considered comparable core clinical outcomes – encompassing remission, response, and relapse. There was good agreement on the definitions and thresholds for these outcomes, and a relatively small pool of outcome measures were used. Resource criteria used to inform the estimation of direct costs, were reasonably uniform. Predominantly, however, there was a high level of heterogeneity in terms of evaluation design and sophistication, quality of evidence used (particularly health state utility data), time horizon, population considered, and cost perspective.

Conclusion: Economic evidence for interventions in TRD is underdeveloped, particularly so for service-level interventions. Where evidence does exist, it is hampered by inconsistency in study design, methodological quality, and availability

of high quality long-term outcomes evidence. This review identifies a number of key considerations and challenges for the design of future economic evaluations. Recommendations for research and suggestions for good practice are made.

Systematic review registration: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=259848&VersionID=1542096, identifier CRD42021259848.

KEYWORDS

economic evaluation, health economics, treatment-resistant depression, persistent depression, values based commissioning

1. Introduction

Major depressive disorder (MDD) affects approximately 5% of the global population and continues to be a major contributor to the overall global burden of disease (1). There is strong evidence that the prevalence of MDD is increasing (2), with the COVID-19 pandemic driving prevalence rates yet higher. Response to the global health crisis and strategies used to prevent the spread of the virus, constructed an environment whereby factors contributing to MDD onset and reoccurrence were exacerbated; contributing to a 28% rise in global prevalence rates (3). Since many of these factors persist (including, but not restricted to: constrained healthcare resources; widened socioeconomic inequality; social isolation; neuropsychiatric sequelae), this trend is not expected to retreat in the near-term (4, 5).

Response to treatment of MDD varies, with many patients requiring more than one treatment step (6). A third of patients do not report improved symptoms despite multiple interventions, resulting in a persistent form of depression commonly described as "treatment-resistant depression" (TRD) (7). Defining TRD is problematic, since failure to respond to treatment "exists on a continuum" (8). A recent review found that while the most widely used definition for TRD was a failure to respond to two or more treatments at an adequate dose and duration, only 19% of recent interventional TRD studies were consistent with that definition (9).

Reflecting this heterogeneity in classification of TRD (10), and indeed in the patient population (11), no single treatment pathway exists, although a stepped-care approach is recommended. Such a model aims to address scarce treatment resources by ensuring that the most effective, least restrictive treatments (in terms of both healthcare resources, and patient convenience), are delivered first, with patients "stepped up" to more intensive treatments as needed (12). Recent UK National Institute for Health and Care Excellence (NICE) guidelines (13) advocate starting treatment for moderate to severe MDD with psychological interventions, such as cognitive–behavioural therapy (CBT), combined with an antidepressant. Where symptoms persist after 4–6 weeks, additional treatments and referral to secondary/specialist mental health services should be considered. Further treatments may include increasing the antidepressant dose, switching to another antidepressant medication of the same or different class, switching to another psychological therapy, adding a second-generation antipsychotic or lithium, or augmenting with electroconvulsive therapy (ECT), lamotrigine, or triiodothyronine. Other treatment options include repetitive transcranial magnetic stimulation (rTMS) and implanted vagus nerve stimulation.

Despite this diverse armamentarium, there remains a high unmet need for new and cost-effective interventions (14, 15). Unfortunately, the condition is highly recurrent—80% of TRD patients experience relapse within a year of remission and the probability of sustained remission over 10 years is just 40% (16). A well-established body of evidence has demonstrated that increasing treatment resistance is associated with poorer health-related quality of life (HRQoL) (8), increased direct medical costs (8, 17, 18), and indirect costs to society attributed to impairments in work productivity and activity (15, 19), and social care demands (20).

Against a background of increasingly constrained healthcare budgets, it is important that decision makers consider not only clinical effectiveness, but the economic evidence for interventions, in order to identify and prioritize those that make the best use of available resources (21). Previous systematic reviews of economic evaluations of interventions for MDD have reported considerable uncertainty in their findings due to inconsistent methodological quality and results (22), and highlighted a lack of evidence and good quality data in TRD (23, 24). Johnston et al. (8) reviewed the literature on the economic burden of TRD, and found significant methodological and population disparities, highlighting heterogeneity in defining TRD, the outcomes measured, and the health state utility values reported.

The aim of this review is to appraise the existing evidence and methods used in economic evaluations of interventions for TRD, and to make best-practice recommendations to inform the development of future evaluations. Promoting consistency in evaluation methodology will improve confidence when making resource allocation decisions, and increase the likelihood that promising interventions receive appropriate funding or support.

2. Concepts in health economic evaluation

2.1. Type of economic evaluation

A "full" health economic evaluation compares both the costs and the consequences of alternative courses of actions (25). The output of the evaluation is (typically) an incremental cost-effectiveness ratio (ICER) (26). Depending on the outcome measure used, economic evaluations may be classified as: cost-effectiveness analyses (CEA), when a clinical outcome measure is used; cost benefit analyses (CBA), when outcomes are valued in monetary terms; cost utility analysis (CUA), when health outcomes are valued as health state utilities to derive quality adjusted life years; cost consequence analysis (CCA), where multiple outcomes not easily summarized in a single summary measure are presented in a disaggregated format; and cost

minimisation analysis (CMA), which assumes that the outcomes from the alternatives under consideration are equivalent (27).

2.2. Health state utilities

Health state utilities are used to represent the "value" of different health states, based on a surveyed population's strength preferences for those health states. Utilities are conventionally scaled between 0 and 1, with 1 representing the value of perfect health and 0 representing the valuation of death (28). Some systems allow a negative utility value, whereby very poor health states may be valued as less preferable than death. When measured over time, utilities may be used to derive the quality adjusted life years (QALYs) associated with living in a particular health state (29).

2.3. Perspective

The perspective of the evaluation refers to the breadth of costs and benefits that are to be considered in the evaluation. Most commonly, the perspective of the healthcare provider or payer is adopted; at the broadest, a "societal" perspective reflects a comprehensive range of social opportunity costs associated with the alternatives under consideration (30). Where significant opportunity costs exist outside the healthcare system, for example in public health interventions, a broad perspective is advised, and there is growing support for such a broad perspective to be used in mental health economic evaluation (21). The 2016 Second Panel on Cost Effectiveness in Health and Medicine recommends analysts adopt a comprehensive approach, reporting separately both healthcare sector and societal perspectives (31). The Panel further recommends the societal perspective report costs and consequences in a comprehensive "impact inventory," and where possible, that non-health consequences are quantified and valued (31). While methodological guidance on choice of perspective varies by jurisdiction, it is generally agreed that the choice should be explicitly stated and determined by the study sponsor (and any stakeholders identified by the sponsor) (32).

2.4. Time horizon

The time horizon refers to the period over which the costs and benefits of the evaluation are captured. Choice of time horizon is influenced by the nature of the condition and intervention under evaluation, and the framework and purpose of the analysis. Ideally, the time horizon for economic evaluations should be sufficiently long to capture relevant differences in costs and outcomes between the comparators; for many interventions, this requires a lifetime horizon (33, 34). Where extrapolated data are used, this is likely to require the analyst to make assumptions about the continued efficacy of the interventions (35).

2.5. Study design

Economic evaluations of health care interventions typically follow one of two study designs: "within-trial" evaluations, where the costs and benefits of alternative courses of action are collected alongside clinical data in interventional clinical studies; and those that use decision analytic models.

2.5.1. Within trial designs

Within-trial evaluations have the advantage that the costs and consequences of the interventions under investigation are measured directly, but are constrained by the follow-up period, frequently precluding assessment of long-term cost effectiveness (36). Extrapolation may be possible using survival analysis models, though this approach requires related long-term data on costs, benefits and complications of the interventions (37).

Sample size and power estimates for trials are most commonly based on the primary clinical outcome. Owing to the tendency of cost variables to have much greater variance than clinical outcomes, trial-based economic evaluations are often underpowered to detect statistically significant differences in cost (38). Accordingly, health economic evaluations assess the probability of cost effectiveness against a certain threshold of willingness-to-pay (WTP), rather than employing statistical hypothesis tests concerning cost effectiveness (37). Typically, probability of cost-effectiveness is assessed against a range of WTP values, and is represented in a cost effectiveness acceptability curve, representing from the joint distribution of incremental costs and effects (37, 39). Most commonly, this distribution is estimated using non-parametric bootstrapping to address sampling uncertainty (39).

Best practice guidelines encourage the use of robust methods to address missing data, since exclusion of cases with missing or censored data may introduce bias (33). While several approaches may be adopted for handling missing data, (including complete case analysis, single imputation and inverse probability weighting), the use of multiple imputation models are usually recommended (40), although this approach may be contested when evaluating data with a high degree of missingness (41).

Combining methods for addressing sampling uncertainty and those for addressing missing data, however, is non-trivial and presents challenges both practical challenges (e.g., computational intensivity), and statistical challenges (e.g., the artificial reduction of sampling uncertainty through imputation) (33, 42). There is a need for further research in this area, as currently no consensus exists for best practice approaches (41).

2.5.2. Decision analytic model designs

Decision analytic models may be used to extrapolate the findings of clinical trial over a longer "time horizon," or to a different population, or may be used to compare interventions for which no head-to-head trials have yet been conducted. Economic models are mathematical abstractions of the real world: analysts will work with subject-matter experts to conceptualize a specific structure, the contingent assumptions, and required input parameters (43). The models describe the probability of specific outcomes following an intervention, with the costs and benefits of each outcome having an associated value. The expected value of that intervention is expressed as the sum of values for each outcome, weighted by the probability of the outcome (43).

Three approaches are commonly used in decision analytic economic evaluation models. The decision tree is a simple but widely used approach used to evaluate short-term prognoses, represented by a series of pathways (44). Markov cohort models may be used to evaluate outcomes over a lifetime horizon, and typically model a homogeneous population transitioning through a series of "health

states." Transitions are modeled in a series of cycles (of a length defined by the analyst); a key property (and frequently a problematic assumption) of Markov models is that no "memory" of the events of previous cycles is retained through each transition (45). Individual-level microsimulation models, which may take the same form as a Markov model, facilitate modeling of a heterogeneous population, and the impact of past events (e.g., number of treatment failures, or adverse events), on prognosis (46).

Analogous to bootstrapping in within-trial evaluations, parametric methods (e.g., Monte Carlo simulation), are recommended to generate sampling distributions of joint mean cost and efficacy estimates (47).

3. Methods

This systematic review follows guidance provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) group (48). The study protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration CRD42021259848).

3.1. Eligibility criteria

Predefined inclusion criteria, defined by the Population, Intervention, Comparator, Outcome, Study type (PICOS) framework (Table 1) were used to determine study selection. Evaluations were included if the author defined the population as "persistent/treatment-resistant/treatment-refractory depression," within adult populations (i.e., individuals aged at least 18 years). Any intervention, across all treatment settings (primary, secondary, and/or community care), relating to the treatment or management of TRD were eligible. Evaluations were excluded if there was no comparator, where comparators could include placebo, an alternative to standard treatment, or treatment as usual.

Evaluation types included any "full" economic evaluation that considered incremental changes across both costs and consequences (CUA, CEA, CBA, CMA, and CCA).

Included evaluations were required to be full-length, peer-reviewed interventional, observational, or modeling reports in journal or Health Technology Authority (HTA) publications in the English language. No date restrictions were imposed. Additionally, bibliographies of systematic reviews were examined to identify further potentially relevant evaluations; however, such reviews themselves were excluded.

TABLE 1 Review inclusion criteria.

Criteria	Notes
Population	Adults with treatment-resistant depression
Intervention	Any intervention for the management of TRD
Comparator	Any intervention for the management of TRD
Outcome	Incremental changes in costs and consequences
Study types	Full economic evaluations: cost-utility analyses (CUA); cost-effectiveness analyses (CEA); cost-benefit analyses (CBA); cost-minimization analyses (CMA); and cost-consequence analyses (CCA). Model and trial-based studies included
Language	English
Time frame	Any
Exclusion	No comparator No consideration of incremental Δcost and Δconsequences

3.2. Information sources and search strategy

Searches across seven electronic databases (MEDLINE; Embase; Cochrane Database of Systematic Reviews; NHS Economic Evaluation Database; Health Technology Assessment database; CINAHL; and PsycINFO) were conducted from inception to 19th May 2021. Searches used two primary concepts (population AND study type), described by Medical Subject Heading (MeSH) and free text search terms. Search terms were refined using Boolean, truncation and adjacency operators. Full search strategies are available in **Supplementary Table 1**.

3.3. Study selection

Records identified in the search strategy were uploaded to the Rayyan platform,[1] for de-duplication and screening. All papers were examined against the PICOS inclusion and exclusion criteria independently by two reviewers (RC and LH) in a two-stage process; title and abstract followed by full-text screening. Reviewers discussed conflicts after each phase and a consensus was reached.

3.4. Data extraction and quality assessment

Key study information was extracted using a pre-defined spreadsheet in Microsoft Excel. Two reviewers (RC and LH) conducted data extraction with a 30% overlap in evaluations. Level of agreement between the overlapping extractions were compared and discussed. Disagreements regarding the content of the extraction fields were resolved through discussion. Data extraction fields included: evaluation details (publication type, setting, objectives); population; general evaluation characteristics (type of intervention and controls, perspective, type of evaluation used, study design, time horizon and reference year); resource use and costs (type of category and costs, data source, and methods used to calculate costs); outcomes (primary clinical outcomes, other clinical outcomes, economic outcomes, and data source for outcomes); economic evaluation results (incremental costs and effects, summary measure of benefits, cost effectiveness results, analyses of uncertainty, and author's conclusions); and model-based evaluation characteristics (model type, model structure and assumptions, rationale for model type and structure, consideration of population heterogeneity).

The Consensus on Health Economic Criteria (CHEC) was used for quality-of-reporting assessment (49). The 19-item CHEC is recommended for systematic reviews that incorporate both trial-based and model-based economic evaluations (50). Additional items related to model conceptualization were included in the assessment: rationale for model type; rationale for model structure; whether

1 https://www.rayyan.ai/

sufficient information was provided to reproduce the model. These items have not been validated, but were informed by items within the "Phillips checklist" for decision analytic models (51).

3.5. Analysis

Evaluation characteristics, design, key cost and outcome parameters, and results were synthesized in summary tables and a narrative synthesis approach was used to describe common features and key differences amongst identified economic evaluations.

4. Results

4.1. Search results and evaluation selection

The evaluation selection process is summarized in **Figure 1**. A total of 539 records were identified through the literature searches, and one more was found through screening reference lists (52). After removing 85 duplicates, 400 records clearly failed to meet the inclusion criteria, or met at least one exclusion criterion, leaving 52 for full-text screening. Of these, 31 satisfied the inclusion criteria and were selected for review (52–82).

4.2. Summary of included evaluations

Key characteristics of the included economic evaluations are provided in **Tables 2A–D**. The interventions considered are categorized into four groups:

(a) Non-pharmacological neuromodulation (hereafter referred to as "neuromodulation"), $n = 14$:

1. Repetitive transcranial magnetic stimulation (rTMS) versus Electroconvulsive therapy (ECT), $n = 10$.
2. ECT versus treatment as usual (TAU), $n = 2$.
3. rTMS versus TAU, $n = 2$.

(b) Pharmacological agents $n = 9$:

1. Adjunctive esketamine versus TAU or placebo and TAU, $n = 3$.
2. Adjunctive atypical antipsychotics versus lithium or hypothetical monotherapy, $n = 2$.
3. Mirtazapine versus TAU, $n = 1$.
4. Multiple alternative antidepressant therapies, $n = 3$.

(c) Psychological therapies $n = 6$:

1. Cognitive-behavioral therapy (CBT]) versus TAU, $n = 3$.
2. Radically open dialectical behavior therapy [RO-DBT] versus TAU, $n = 2$.
3. Intensive short-term dynamic psychotherapy [ISTDP] versus TAU, $n = 1$.

(d) Service-level interventions versus TAU, $n = 2$:

The 31 evaluations, relate to 29 unique studies, with multiple economic evaluations included for two studies: a trial comparing the cost effectiveness of rTMS and ECT; (58, 69) and a trial of CBT as an adjunct to pharmacotherapy (73, 82). Of the 31 evaluations included, 11 were trial-based (predominantly psychological [$n = 6$], or service-level [$n = 2$] interventions), and 20 were model-based (predominantly non-pharmacological neuromodulation [$n = 12$] or pharmacological [$n = 8$] interventions). Twenty-four of the evaluations adopted a cost-utility analysis (CUA) as their primary analytical approach, six used cost-effectiveness analysis (CEA), and one adopted a cost-consequence analysis (CCA) as the primary method of evaluation. Six evaluations used multiple analytical approaches. The median time horizon was 1 year, eight evaluations used a time horizon of less than a year, and only two evaluations considered a lifetime horizon. The primary analysis for most evaluations ($n = 28$) considered costs from a healthcare provider perspective, three evaluations considered a (partial) societal perspective, and seven presented both societal and healthcare provider perspectives. The evaluations came almost exclusively from high income countries (UK [$n = 11$]; US [$n = 10$]; Canada [$n = 5$]; Australia [$n = 2$]; Singapore [$n = 1$]; Spain [$n = 1$]), with a single evaluation from Iran) (57).

4.3. Quality of reporting assessment

Quality of reporting of the evaluations was predominantly high; the range of fulfilled CHEC criteria across the evaluations fell between 47 and 100%, with an average of 83% of criteria fulfilled. Five evaluations met all criteria from the CHEC-list, and only two evaluations fulfilled fewer than 60% of the criteria (57, 61). The lowest-scoring items from the checklist were: discussion of ethical and distributional issues (45% of evaluations); reporting of structural assumptions and validation methods of models (55% of relevant evaluations); consideration of the generalizability of the results (61% of evaluations). Additional items used to evaluate reporting of conceptualization of model-based evaluations were less well reported: only 15% provided a rationale for choice of model type, and 55% provided a rationale for the model structure. Results of the quality assessment are presented in **Supplementary Table 2**.

4.4. TRD population

There was variation in patient populations considered by the included evaluations, reflecting a lack of consensus on the definition of TRD (83). Most commonly, treatment resistance was defined as a failure to achieve an adequate response to antidepressive treatment ($n = 24$), with half of these specifying a requirement for failure of at least two lines of therapy. Three evaluations used a definition based on the number of previous episodes, or duration of the current episode, and four evaluations did not clearly define treatment-resistant depression or the studied population. At baseline, the populations considered were typically severely depressed, however, severity was not well defined in most model-based studies and had to be intuited from the utility values reported.

4.5. Effects

4.5.1. Clinical outcomes

Trial-based evaluations tended to use either response ($n = 4$) or change in depressive symptoms ($n = 5$) as their primary clinical

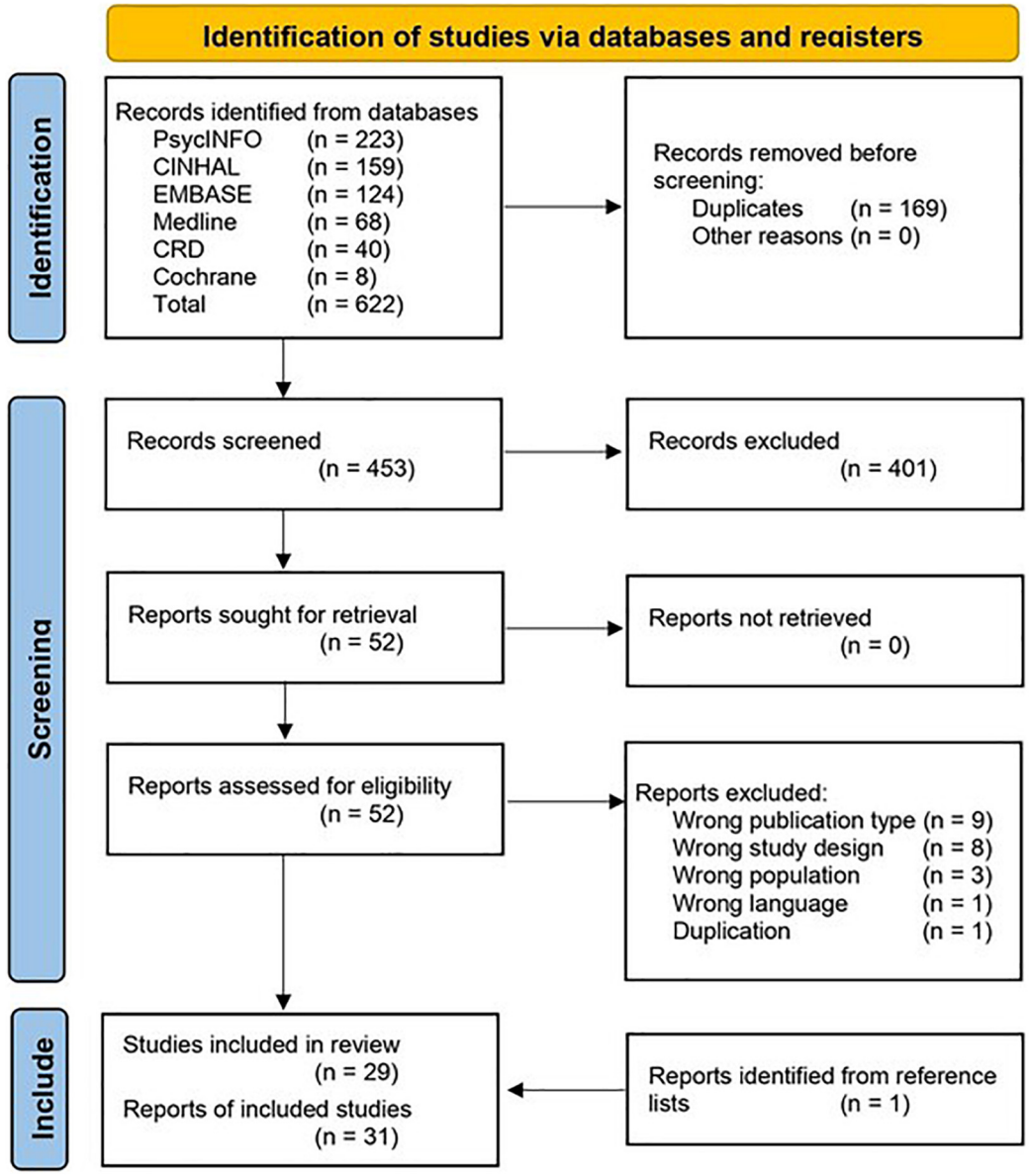


FIGURE 1
PRISMA flow diagram of study identification, adapted from (48).

outcome, with only one evaluation using remission (in addition to change in depressive symptoms). Other outcomes included relapse ($n = 2$), and depression-free days. Model-based evaluations tended to include both response and remission ($n = 13$), with five evaluations modeling remission only, one evaluation modeling response, and one modeling change in depressive symptoms.

In trial-based evaluations, the most common outcome measure was the Hamilton Depression Rating Scale (HAM-D, $n = 6$), followed by the Beck Depression Inventory II (BDI-II, $n = 3$). Other measures included the Montgomery–Åsberg Depression Rating Scale (MADRS, $n = 1$), Symptom Checklist–90 (SCL-90, $n = 1$), Beck Depression Inventory (BDI, $n = 1$), and the Global Assessment of Functioning scale (GAF, $n = 1$). Model-based evaluations typically synthesized outcomes from multiple sources, where outcomes may have been measured using several scales, though most frequently mentioned scales included the MADRS ($n = 10$) and the HAM-D ($n = 9$).

Response was typically defined as an improvement of ≥50% from baseline against the scales used, however, there was some variation in the scores used to define remission (HAM-D: <7 [$n = 1$]; ≤7 [$n = 6$]; ≤8 [$n = 3$]; MADRS:≤10 [$n = 7$]; <12 [$n = 1$]; ≤12 [$n = 2$]). In doing so, most evaluations diverged from broadly accepted cut-offs for defining remission: ≤7 for HAM-D (84), and <10 for MADRS (85, 86).

4.5.2. Health economic outcomes

All CUA evaluations used quality-adjusted life years (QALYs) as the primary economic outcome measure. Of these, for most ($n = 20$), utility values underpinning QALY estimates were derived from the EuroQol 5-Dimension 3-Level Health Scale (EQ-5D-3L), the most widely recommended measure of health-related quality of life by HTA authorities globally (87). Other measures included the EuroQol 5-Dimension 5-Level Health Scale (EQ-5D-5L, $n = 1$) (88), the Short-Form

Six-Dimension health index (SF-6D, n = 4) (73, 89), McSad (n = 2) (90), Health Utilities Index 3 (HUI3, n = 1) (91), and a vignette-based valuation of various levels of severity of MDD (n = 4) (92).

Of the nine evaluations that used a CEA approach (including four as secondary analyses), the most common economic outcome measures were cost per unit change in depression scale rating (n = 4), and cost per remitter (n = 3). Alternative outcomes included cost per relapse prevented (78), and cost per depression-free day (80).

The single CCA evaluation used maintenance of response and maintenance of relapse as outcome measures (72). All clinical and health economic outcome measures used are summarized in **Supplementary Table 3**.

4.6. Resource use and cost data

Generally, costs were well reported, although several evaluations only reported costs at an aggregate level (53, 57, 60, 61, 64, 65). Trial-based evaluations primarily used self-report questionnaires to collect resource use data (n = 8), but also relied on registry or hospital chart data (n = 4), and claims databases (n = 2). Most model-based evaluations drew data from the literature (n = 10), claims databases (n = 6), or registry or hospital chart data (n = 5).

Direct costs reported for all evaluations included treatment costs, with most also including outpatient (n = 27) and inpatient costs (n = 26); only three evaluations explicitly included costs for adverse events (AEs). Reported detail concerning assumptions and

TABLE 2A Characteristics of economic evaluations of non-pharmacological neuromodulation interventions.

References, country	TRD population/Definition	Comparators	Evaluation type[a]	Study design	Perspective(s)[a]	Time horizon
Nguyen and Gordon (62) Australia	Inadequate response to ≥2 AD treatments	rTMS vs TAU (pharmacotherapy)	CUA	Model	Health system	36 months
Simpson et al. (66) USA	Inadequate response to 1-4 AD treatment	TMS vs TAU (pharmacotherapy) vs Sham TMS	CUA	Model	Health system Societal	12 months
Ross et al. (64) USA	Inadequate response to ≥2 AD treatments	ECT at different therapy lines vs TAU (pharmacotherapy)	CUA	Model	Health system	48 months
McDonald et al. (72) USA	Geriatric TRD on maintenance treatment TRD not explicitly defined	ECT vs TAU (pharmacotherapy)	CCA	Within-trial (non-randomized)	Health system	12 months
Fitzgibbon et al. (56) Canada	Not explicitly defined	rTMS vs ECT vs rTMS + ECT stepped pathway	CUA	Model	Societal	Lifetime
Ghiasvand et al. (57) Iran	Not explicitly defined	rTMS vs ECT	CEA CUA	Model	Health system	7 months
Health Quality Ontario (58) Canada	Inadequate response to ≥ 2 AD treatments	rTMS vs ECT vs sham rTMS	CUA	Model	Health system	6 months
Unit University of Calgary (59) Canada	Inadequate response to ≥2 AD treatments	rTMS vs ECT	CUA	Model	Health system	1.5 months
Kozel et al. (60) USA	Not explicitly defined	rTMS vs ECT vs rTMS + ECT stepped pathway	CUA	Model	Societal	12 months
Vallejo-Torres et al. (67) Spain	Not explicitly defined	ECT vs rTMS vs rTMS + ECT stepped pathway	CUA	Model	Health system	12 months
Xie et al. (69) Canada	Inadequate response to ≥2 AD treatments	rTMS vs ECT vs sham rTMS + TAU	CUA	Model	Health system	6 months
Galletly et al. (52) Australia	Inadequate response to ≥2 AD treatments	rTMS vs ECT	CUA	Model	Health system	36 months
Zhao et al. (71) Singapore	Not explicitly defined	rTMS vs ECT	CUA	Model	Societal	12 months
Knapp et al. (75) UK	Not explicitly defined	rTMS vs ECT	CEA CUA	Within-trial	Health system Societal	6 months

[a]Primary analysis shown first.

TABLE 2B Characteristics of economic evaluations of pharmacological interventions.

References, country	TRD population/Definition	Comparators	Evaluation type[a]	Study design	Perspective(s)[a]	Time horizon
Atlas et al. (53) USA	Inadequate response to ≥2 AD treatments	Esketamine + Antidepressant vs Antidepressant	CUA CEA	Model	Health system Societal	Lifetime
Desai et al. (54) USA	Inadequate response to ≥2 AD treatments	Esketamine + Antidepressant vs Placebo + Antidepressant	CEA	Model	Health system (4 alternative payer perspectives)	12 months
Ross and Soeteman (65) USA	Inadequate response to ≥2 AD treatments	Esketamine + TAU (pharmacotherapy) vs TAU	CUA	Model	Health system Societal	60 months
Edwards et al. (55) UK	Inadequate response to ≥2 AD treatments	AAP + Antidepressant vs Lithium + Antidepressant	CUA	Model	Health system	12 months
Malone (61) USA	Inadequate response to single AD treatment	Antidepressants (escitalopram, paroxetine CR, sertraline, venlafaxine) Generic SSRIs	CEA	Model	Health system	6 months
Olgiati et al. (63) USA	Inadequate response to single AD treatment	Sequenced treatment (switch/augment following citalopram non-response) vs Continued citalopram	CUA	Model	Health system	6 months
Wang et al. (68) UK	Inadequate response to ≥2 AD treatments	Hypothetical monotherapy vs SSRI + AAP	CUA	Model	Health system	12 months
Young et al. (70) UK	Inadequate response to ≥2 AD treatments	Vortioxetine vs SSRIs vs SNRIs vs agomelatine	CUA	Model	Health system	24 months
Kessler et al. (74) UK	Inadequate response to single AD treatment	mirtazapine vs placebo + TAU (pharmacotherapy)	CUA CEA	Within-trial	Health system Societal	12 months

[a]Primary analysis shown first. SSRI, selective serotonin reuptake inhibitor; SNRI, serotonin/noradrenaline reuptake inhibitor; CR, controlled release; TAU, treatment as usual; AAP, atypical antipsychotic.

TABLE 2C Characteristics of economic evaluations of psychological interventions.

References, country	TRD population/Definition	Comparators	Evaluation type[a]	Study design	Perspective(s)[a]	Time horizon
Hollinghurst et al. (73) UK	Inadequate response to single AD treatment	CBT + TAU (usual clinical care for primary care TRD patients) vs TAU	CUA CCA	Within-trial	Health system Societal	12 months
Wiles et al. (82) UK	Inadequate response to ≥6 weeks AD treatment	CBT + TAU (usual clinical care for primary care TRD patients) vs TAU	CUA	Within-trial	Health system	Up to 46 months
Scott et al. (78) UK	Current residual symptoms of ≥8 weeks' duration following and MDD episode between last 2–18 months	CBT + TAU (usual clinical care) vs TAU	CEA	Within-trial	Health system	17 months
Town et al. (81) Canada	Inadequate response to ≥6 weeks AD treatment	ISTDP + TAU (usual clinical care for secondary care TRD patients) vs TAU	CUA CEA	Within-trial	Health system	18 months
Shearer et al. (79) UK	MDD lasting ≥ 2 years or ≥2 MDD episodes with inadequate response to ≥ 6 weeks AD treatment	RO-DBT + TAU (usual clinical care for secondary care TRD patients) vs TAU	CUA CEA	Within-trial	Health system	12 months
Lynch et al. (76) UK	Inadequate response to ≥2 AD treatments	RO-DBT + TAU (usual clinical care) vs TAU	CUA	Within-trial	Health system Societal	18 months

[a]Primary analysis shown first. AD, antidepressant (drug).

TABLE 2D Characteristics of economic evaluations of service-level interventions.

References, country	TRD population/Definition	Comparators	Evaluation type[a]	Study design	Perspective(s)[a]	Time horizon
Morriss et al. (77) UK	Inadequate response to ≥6 months secondary mental healthcare	Specialist depression service (SDS) vs TAU (usual clinical care for secondary care TRD patients)	CUA	Within-trial	Health system	18 months
Simon et al. (80) USA	≥2 depressive episodes	Collaborative care program vs TAU (usual clinical care for primary care TRD patients)	CEA	Within-trial	Health system	6 months.

[a]Primary analysis shown first.

methods for estimating attribution of capital equipment costs for neuromodulation interventions varied considerably. Indirect costs were considered by the ten evaluations that considered a broader cost perspective, but the scope of items collected varied considerably. Most (n = 9) considered productivity (in most cases measuring only absenteeism, although one also measured presenteeism) (76) others additionally considered out-of-pocket payments (n = 4), informal care (n = 4), formal societal or community care (n = 3), or transport (n = 3), but no two evaluations included the same set of indirect cost measures.

4.7. Modeling approaches and scope

The details of the 20 models appraised are given in **Tables 3A, B**. Six evaluations used a decision tree approach, the majority of which (n = 5) were evaluations of non-pharmacological neuromodulation interventions, while the sixth compared novel selective serotonin reuptake inhibitors (SSRIs)/serotonin and norepinephrine reuptake inhibitors (SNRIs) and generic SSRIs (61). In keeping with the associated restrictions of this analytical approach, all used a short time horizon, typically 6 months or less. The decision trees largely followed a similar structure, modeling three possible outcomes: remission, response (with no remission), and non-response. A representation of the generic decision tree structure is shown in **Figure 2**.

There were several notable variations from this structure. Both Kozel et al. (60) and Ghiasvand et al. (57) assumed that any response equated to full remission. This is a significant limitation as it does not allow for partial improvements in symptoms, and thereby is likely to overestimate the benefits of interventions.

Kozel's model also allowed for relapse. The model described by Malone et al. (61) compared the costs and consequences of various pharmaceutical treatment regimens, and augmented this generic structure with further steps that considered adverse events (AEs), and treatment changes. While four of the six evaluations described the conceptualization of the model structure, none described the rationale for selecting a decision tree approach, and only half described any structural assumptions or indicated that any validation assessment was undertaken.

Twelve evaluations used Markov cohort models, and three extended this approach with more sophisticated Markov microsimulation models (all for neuromodulation interventions). A key characteristic of this extension is that it enables the tracking of individual patient characteristics or event history through the model. Most Markov models had a minimum horizon of 12 months, but only two had a lifetime horizon (53, 56). A similar "base" generic structure, shown in shown in **Figure 3**, was used across the majority of models, with three key "health states": remission, response, and relapse (and/or non-response).

Several evaluations extend beyond this base structure, varying the levels of complexity and sophistication. Seven evaluations preceded the Markov model with a decision tree to represent a distinct acute phase of treatment. Other additional health states used (either as Markov health states or transition health states) included: death – particularly for models with a time horizon greater than 1 year (n = 7); treatment change (n = 7); severe depression (n = 6); discontinuation (n = 7); adverse events (n = 4); hospitalization (n = 2). Only one evaluation used an entirely different structure, modeling health states defined by four different levels of severity of depression (defined by MADRS score) (39).

The reporting of these models was generally good, with a majority describing a rationale for model structure (n = 9) and structural assumptions (n = 8). Nevertheless, some aspects of the health states included or omitted require some important limiting assumptions. Only seven models accounted for discontinuation of treatment, and none of those omitting discontinuation justified the omission. Discontinuation might feasibly be rolled into the "non-response" health state, however, this was not explicitly stated in any evaluation that omitted a "discontinued" health state; these may consequently overestimate treatment benefits by failing to account for discontinued patients. Of those evaluations that did include discontinuation, four either did not distinguish between discontinuation related to AEs or lack of efficacy, or assumed discontinuation due to AEs to be embedded in loss of treatment effect (53, 55, 67, 68). These four evaluations therefore considered AEs implicitly, but assumed no continued impact on quality of life beyond that of discontinuation due to lack of efficacy – an assumption that may not hold for severe or long-lasting AEs. AEs were considered explicitly in only five evaluations. Two considered both costs and utility decrements associated with AEs (52, 62), two considered only utility decrement (64), and one considered only costs (71). The majority did not model AEs and in most cases a rationale was not given, although it was suggested in two evaluations that the impact of AEs was expected to be limited, and similar between comparators (55, 59). While this assumption may be true of some comparators, it is an important structural assumption to validate, as omission will bias toward those interventions that have higher rates of AEs.

4.7.1. Utility data

There was considerable heterogeneity used in approaches to sourcing utility data for use in cost utility models: 11 different

sources, using six different methods of deriving utility (EQ-5D-3L [n = 13]; standard gamble [n = 4]; McSad [n = 2]; SF-6D [n = 1]; HUI-3 [n = 1]) were identified. The two main sources of utility values were studies by Sapin et al. (93) (n = 7) and Revicki et al. (92) (n = 4). The utility values derived from these two studies were based on MDD rather than TRD populations, and report only

TABLE 3A Model characteristics for economic evaluations of non-pharmacological neuromodulation.

References	Comparators	Study design	Model type	Horizon	Main health states modeled
Nguyen and Gordon (62)	rTMS TAU (pharmacotherapy)	CUA	Markov microsimulation	36 months	Acute treatment Full remission Partial remission No response/relapse Post treatment ECT Post-treatment lithium augmentation Acute episode hospitalization Death
Simpson et al. (66)	TMS Pharmacotherapy Sham TMS	CUA	Markov model	12 months	Well: MADRS 0-9 Mild: MADRS 10-17 Moderate: MADRS 18-27 Severe: MADRS > 28
Ross et al. (64)	ECT at different lines of therapy TAU (pharmacotherapy)	CUA	Markov model	48 months	Remission Response Relapse Non-response
Fitzgibbon et al. (56)	rTMS ECT combined rTMS + ECT stepped pathway	CUA	Markov microsimulation	Lifetime	Acute treatment Remission Maintenance treatment Severe depression Death
Ghiasvand et al. (57)	rTMS ECT	CEA CUA	Decision tree	7 months	Remission Relapse
Health Quality Ontario (58)	rTMS ECT sham rTMS	CUA	Decision tree	6 months	Non-response Response Remission
Unit University of Calgary (59)	rTMS ECT	CUA	Decision tree	6 weeks	Response Remission Relapse
Kozel et al. (60)	rTMS ECT combined rTMS + ECT stepped pathway	CUA	Decision tree	12 months	Non-response Response Continued response Relapse
Vallejo-Torres et al. (67)	ECT rTMS combined rTMS + ECT stepped pathway	CUA	Markov model	12 months	Acute treatment/relapse Continuation treatment Stable with/without treatment Moderate depression Severe depression Death
Xie et al. (69)	rTMS ECT sham rTMS + TAU	CUA	Decision tree	6 months	Non-response Response Remission
Galletly et al. (52)	rTMS ECT	CUA	Markov microsimulation	36 months	Acute treatment Full remission Partial remission No response/relapse Post treatment ECT Post-treatment lithium augmentation Acute episode hospitalization Death
Zhao et al. (71)	rTMS ECT	CUA	Markov model	12 months	Remission Non-remission Relapse Stable (remission) Severe depression Death (suicide or other causes)

TABLE 3B Model characteristics for economic evaluations of pharmacological interventions.

References	Comparators	Study design	Model type	Horizon	Main health states modeled
Atlas et al. (53)	Esketamine + Antidepressant Antidepressant	CUA	Markov model	Lifetime	Non-response Partial response Response Remission Treatment failure Relapse Discontinuation Death
Desai et al. (54)	Esketamine + Antidepressant Placebo + Antidepressant	CEA	Markov model	12 months	Response Remission Relapse
Ross and Soeteman (65)	Esketamine + TAU TAU	CUA	Markov model	60 months	Remission Response Relapse Non-response
Edwards et al. (55)	AAP + Antidepressant Lithium + Antidepressant	CUA	Markov model	12 months	Non-response Response (continue/discontinue) Remission (continue/discontinue) Relapse
Malone (61)	Antidepressants (escitalopram, paroxetine CR, sertraline, venlafaxine) Generic SSRIs	CEA	Decision tree	6 months	Non-response Response Remission
Olgiati et al. (63)	Sequenced treatment (either switch or augment following citalopram non-response) Continued citalopram	CUA	Markov model	26 weeks	Acute depression/non-remission/relapse Remission No treatment
Wang et al. (68)	Hypothetical monotherapy SSRI + AAP	CUA	Markov model	12 months	Full remission discontinued Full remission Partial remission discontinued Partial remission In episode discontinued Relapse discontinued
Young et al. (70)	Vortioxetine SSRIs, SNRIs, agomelatine	CUA	Markov model	24 months	Non-response Response Remission Recovery Long-term AEs

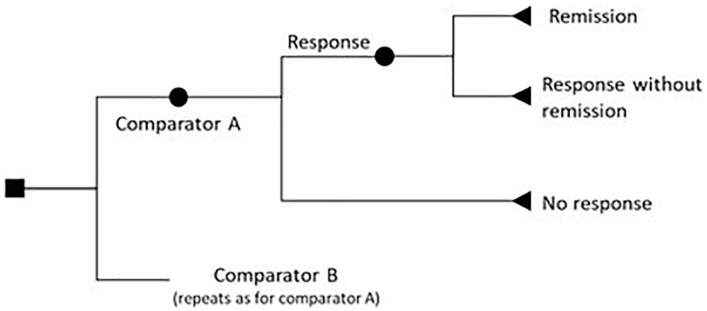


FIGURE 2
Generic decision tree structure for economic evaluations in TRD.

crude unadjusted values. Notably, only two evaluations used values derived from patients with TRD (53, 70). Use of values derived from the broader MDD population was driven by the scarcity of health-related quality of life data specific to patients with TRD. Ideally, values used should be population specific (94, 95) – the extent to which MDD values generalize to TRD is unknown. All evaluations that modeled the disutility of adverse events (n = 3) drew these values from a study of AEs associated with SSRIs in MDD (96). As a consequence of the heterogeneity in sources used (and concomitantly, the heterogeneity of the populations from which the source data were drawn), there was also considerable variability in the values used for common health states: baseline depression, 0.25–0.55; remission, 0.76–0.91; response 0.71–0.76; no response 0.52–0.58; relapse 0.30–0.63. Utility values used in the models, and their information sources are summarized in Supplementary Table 4.

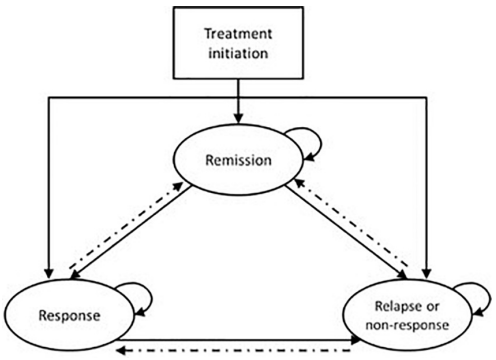


FIGURE 3
Generic Markov model structure for economic evaluations in TRD. Patients solid lines indicate pathways that are common amongst the majority of models; dashed lines indicate pathways that are included in a subset of models.

4.8. Results of evaluations

Table 4 summarizes evaluation results. Consistency of results varied across interventions. Four evaluations comparing neuromodulation (rTMS or ECT) to TAU consistently found the intervention to be cost-effective, and a dominant strategy (both more effective and less costly) in three evaluations. Direct comparisons of ECT and rTMS, however, were less consistent: six favored ECT and four favored rTMS. The source of these variations is not immediately clear; however, those that favored ECT tended to have a shorter (<12 month) time horizon, which may not have been long enough to capture benefits of maintenance treatment with rTMS. Those that adopted a societal perspective tended to favor rTMS, reflecting the higher indirect costs (care, time off work) of ECT. There is no clear indication that study or model design biased results in either direction. Notably, over half of these evaluations did not explicitly define the patient population in terms of severity or number of previously failed treatments. There was variation in the treatment protocol used for rTMS, which is likely to have a considerable impact on costs, as will the extent to which capital costs are attributed across different evaluations.

Only three pharmacotherapy evaluations considered the same comparators (esketamine vs TAU). Two CUAs found that despite improved outcomes esketamine was unlikely to be cost effective. The third evaluation, which was industry-sponsored, used a CEA approach, and found esketamine was likely to be cost efficient. In addition to differences in analytical approaches used, the two CUAs had much longer time horizons (5 years and lifetime), compared to the 12 month CEA. It is likely that the consideration of relapse over those longer horizons had a significant impact on cost-effectiveness.

The evaluations evaluating psychological interventions, which were all trial-based, were generally consistent in their findings: two CUAs comparing CBT to TAU and one comparing ISTDP to TAU found that these interventions were likely to be cost effective; two CUAs comparing RO-DBT found that the intervention was highly unlikely to be cost effective. The key driver of the cost inefficiency for RO-DBT were the costs of intensive treatment.

We reviewed two trial-based evaluations of service-level interventions which are not directly comparable. A US-based collaborative care program was found to be cost effective (80), while an evaluation of a specialist depression service in the UK found limited additional benefits associated with the service and concluded it was unlikely to be cost-effective (77).

All except one evaluation explored uncertainty in parameters and/or results (72). Bootstrapping or similar methods were used to account for sampling uncertainty in almost all (n = 9) trial-based evaluations, while probabilistic sensitivity analyses were used to account for the joint uncertainty of all key parameters in over half (n = 13) of the model-based evaluations. Although most (n = 16) model-based evaluations conducted some degree of one-way sensitivity analysis, fewer than half (n = 9) conducted a comprehensive sensitivity analysis, incorporating all important variables. Key drivers of uncertainty included the probability of response and remission, utility values used for acute/severe depression, and cost of intervention (particularly for rTMS where number of treatment courses varied).

5. Discussion

The aim of this review was to appraise the literature systematically to describe the methods used in the economic evaluation of interventions for the management of TRD, to inform design and development of future evaluations in this field. We identified 31 evaluations, including 11 trial-based and 20 model-based evaluations. A broad range of interventions and designs were considered by the included evaluations, but almost half evaluated the cost effectiveness of neuromodulation interventions (rTMS and/or ECT), enhancing our ability to consider consistency of evaluation design, and the factors that most strongly influence results.

There was a distinct paucity of evidence relating to the economic evaluation of service-level interventions, with only two studies identified in the literature search. In their evaluation of a dedicated specialist depression service for TRD, Morris et al. (77) noted significant loss to follow-up during the trial and indicated the evaluation may have been underpowered to detect statistical improvements in symptoms at follow-up. It has been argued that the objective of economic evaluation is estimation of expected value of an intervention, and that decision making should therefore be based upon the weight of evidence, rather than the application of statistical inference rules (38, 97). Lack of statistical significance may, however, suggest that there is value in obtaining further evidence (97).

Despite a growing interest in the application of digital technologies in the management and delivery of mental health care (98), no economic evaluations of such interventions were identified. Recent studies suggest the implementation of digital technologies (e.g., virtual reality, artificial intelligence) may improve diagnosis, intervention delivery, monitoring, access to care, and potentially reduce costs (98, 99). Economic evidence supporting digital technologies in healthcare generally is underdeveloped: there is a clear need for early-stage economic evaluations to support the development of these promising approaches (100).

The quality of reporting as assessed by the CHEC criteria was generally good, and some aspects were found to be relatively consistent across the evaluations. Most evaluations considered comparable clinical outcomes – encompassing remission, response/non-response to treatment, and relapse. There was good agreement on the definitions and associated threshold for these outcomes, and these were assessed by a relatively small pool of clinical outcome measures. The resource criteria used to inform

TABLE 4A Summary of authors' conclusions, and key drivers of uncertainty for economic evaluations of non-pharmacological neuromodulation interventions.

References	Comparators	Evaluation conclusion	Key drivers of uncertainty[1]
Nguyen and Gordon (62) Australia	rTMS TAU (pharmacotherapy)	rTMS dominates (cheaper and more effective) TAU	Probability of response, remission & relapse
Simpson et al. (66) USA	TMS Pharmacotherapy Sham TMS	rTMS is likely to be cost-effective compared to sham rTMS rTMS dominates pharmacotherapy	Cost of rTMS
Ross et al. (64) USA	ECT at different lines of therapy TAU (pharmacotherapy)	Offering ECT after 2 failed lines of pharmacotherapy/psychotherapy likely to be the most cost-effective algorithm	**Cost of ECT cost** **Utility value for non-response** **Probability of response**
McDonald et al. (72) USA	ECT TAU	ECT dominates TAU	None identified
Fitzgibbon et al. (56) Canada	rTMS ECT combined rTMS + ECT stepped pathway	rTMS dominates ECT in first line	**Frequency of rTMS administration**
Ghiasvand et al. (57) Iran	rTMS ECT	ECT more cost effective than rTMS	Costs of interventions
Health Quality Ontario (58) Canada	rTMS ECT sham rTMS	ECT likely to be cost effective compared to rTMS	**Probability of response to ECT and rTMS**
Unit University of Calgary (59) Canada	rTMS ECT	rTMS dominates ECT	**Probability of response & remission** **rTMS treatment cost**
Kozel et al. (60) USA	rTMS ECT combined rTMS + ECT stepped pathway	ECT unlikely to be more cost effective than rTMS Stepped rTMS-ECT pathway dominates	Probability of response & remission Intervention costs
Vallejo-Torres et al. (67) Spain	ECT rTMS combined rTMS + ECT stepped pathway	ECT dominates rTMS Stepped pathway unlikely to be cost effective	None identified
Xie et al. (69) Canada	rTMS ECT sham rTMS + TAU	Low probability of rTMS being cost effective using a non-inferiority framework (and a 75% preservation of effectiveness threshold)	None identified
Galletly et al. (52) Australia	rTMS ECT	rTMS unlikely to be cost effective compared to ECT	**Probability remission after treatment** **Probability remission after hospitalization** **Number of rTMS and ECT sessions per treatment course**
Zhao et al. (71) Singapore	rTMS ECT	rTMS likely to be cost effective compared to ECT	**Probability remission after** **Cost for hospitalization due to ECT**
Knapp et al. (75) UK	rTMS ECT	Very low probability that rTMS is cost effective compared to ECT	None identified

[1]Items in bold denote studies that conducted a comprehensive one-way uncertainty analysis.

the estimation of direct costs including inpatient stays, outpatient appointments, and pharmaceutical costs, were reasonably uniform. Predominantly, however, there was a high level of heterogeneity in terms of evaluation design and sophistication, quality of evidence used (particularly with respect to health state utility data), time horizon, population considered, and cost perspective adopted. The impact of these inconsistencies is highlighted by the fact that despite the inclusion of 10 evaluations comparing rTMS and ECT, there is still inconclusive evidence as to the cost effectiveness of rTMS vs ECT.

Our findings are in general agreement with the literature relating to economic evaluation of MDD, where reviews have found the evidence for multiple interventions to be inconclusive due to inconsistencies in evaluation design and methodological quality (21, 22), and that the paucity of evidence related to long-term outcomes in TRD restricts our ability to inform the long-term value of interventions in TRD (23, 24). In order to inform future economic evaluations in TRD, and promote greater consistency among them, a number of linked methodological considerations are identified and good practices suggested.

5.1. Evaluation population and incorporation of patient heterogeneity

There was considerable variation in the definition used to describe the TRD population under study, with a fifth of evaluations providing no explicit definition. The absence of a standardized definition of the population reduces the validity of comparison and data synthesis across evaluations (101). However, one must acknowledge that the population is highly heterogeneous, in terms of both degree of treatment resistance, and medical and psychiatric co-morbid conditions (102). Evaluations that restrict the their population to a narrow definition or TRD, or that model a homogeneous cohort will limit generalizability of the findings. Despite this, very few model-based evaluations in this review explored the impact of patient heterogeneity – and where

TABLE 4B Summary of authors' conclusions, and key drivers of uncertainty for economic evaluations of pharmacological interventions.

References	Comparators	Evaluation conclusion	Key drivers of uncertainty[2]
Atlas et al. (53) USA	Esketamine + Antidepressant Antidepressant	Esketamine unlikely to be cost effective	**Utilities for severe depression** **Probability of continued effect** **Probability of discontinuing therapy if effective**
Desai et al. (54) USA	Esketamine + Antidepressant Placebo + Antidepressant	Esketamine cost per remitter ($14-39k) is cost-effective	Probability of relapse free remission
Ross and Soeteman (65) USA	Esketamine + TAU TAU	Esketamine unlikely to be cost effective	**Probability of response and remission**
Edwards et al. (55) UK	AAP + Antidepressant Lithium + Antidepressant	Lithium dominates AAP (though subject to considerable uncertainty)	**Probability of acute efficacy** **Probability of discontinuation**
Malone (61) USA	Antidepressants (escitalopram, paroxetine CR, sertraline, venlafaxine) Generic SSRIs	Cost per remitter lowest for venlafaxine	None identified
Olgiati et al. (63) USA	Sequenced treatment (switch/augment following citalopram non-response) Continued citalopram	Sequenced treatment likely to be cost effective compared to remaining on citalopram	Utility values for acute depression and remitted depression
Wang et al. (68) UK	Hypothetical monotherapy SSRI + AAP	Hypothetical monotherapy dominates SSRI + AAP	**Probability of response and remission**
Young et al. (70) UK	Vortioxetine SSRIs, SNRIs, agomelatine	Vortioxetine in the third line likely to be cost effective compared to SSRIs	Secondary care costs
Kessler et al. (74) UK	mirtazapine placebo + TAU	No strong evidence that mirtazapine is cost-effective	None identified

[2]Items in bold denote studies that conducted a comprehensive one-way uncertainty analysis.

TABLE 4C Summary of authors' conclusions, and key drivers of uncertainty for economic evaluations of psychological interventions.

References	Comparators	Evaluation conclusion	Key drivers of uncertainty
Hollinghurst et al. (73) UK	CBT + TAU TAU	CBT + TAU is likely to be cost effective compared to TAU	QoL measure used (more cost-effective with EQ-5d-3L *cf.* SF-6D)
Wiles et al. (82) UK	CBT + TAU TAU	CBT + TAU is likely to be cost effective compared to TAU	None identified
Scott et al. (78) UK	CBT + TAU TAU	£12.50 per relapse-free day (conclusion depends on willingness to pay for a relapse free day)	None identified
Town et al. (81) Canada	ISTDP + TAU TAU	ISTDP likely to be cost effective compared to TAU	None identified
Shearer et al. (79) UK	RO-DBT + TAU TAU	Highly unlikely that RO-DBT is cost effective compared with TAU	None identified
Lynch et al. (76) UK	RO-DBT + TAU TAU	RO-DBT unlikely to be cost effective compared with TAU	None identified

TABLE 4D Summary of authors' conclusions, and key drivers of uncertainty for economic evaluations of service-level interventions.

References	Comparators	Evaluation conclusion	Key drivers of uncertainty
Morriss et al. (77) UK	Specialist depression service (SDS) TAU	SDS unlikely to be cost effective compared to TAU	None identified
Simon et al. (80) USA	Collaborative care program TAU	$21 per depression-free day – likely to be comparable ROI to other widely accepted medical interventions	None identified

heterogeneity was considered, only a narrow range of aspects of heterogeneity were considered (age, gender, number of previous treatments). Equally, the under-reporting of severity at baseline is problematic when comparing economic evaluations, since this is likely to significantly impact outcomes (103).

To improve consistency across economic evaluations, we suggest that the widely used TRD definition of "failure to respond to two or more treatments at an adequate dose and duration" (9) be used as the base case for evaluation. Reflecting the concept that various degrees of resistance exist (102), more sophisticated evaluations might consider staging (for example by number of previous treatments), or at least characterizing the study population in this manner. Good practice guidelines for health economic models already highlight the importance of consideration of heterogeneity (47, 104). Cohort

models can achieve this through sensitivity testing of results with alternative patient cohorts; more sophisticated patient-level models incorporate the facility to directly model heterogeneity.

5.2. Time horizon

The persistent and highly recurrent nature of TRD is not well reflected in many of the evaluations: the time horizon for most models was only 12 months, and the average for trials was 18 months. Only two evaluations used a lifetime horizon, extrapolating outcomes from clinical evaluations with follow-up periods of 12 months or less (53, 56). A key driver for the use of models in economic evaluation is to extrapolate the results of clinical trials to a longer-term horizon (47). In the context of TRD, a short time horizon may underestimate the cost effectiveness of an intervention by failing to account for smaller incremental improvements in mental health (accruing substantially with a longer horizon), or the improvements that persist beyond the evaluation horizon – for example, MDD patients receiving cognitive therapy have been found to exhibit reduced relapse rates for up to 6 years (78). Conversely, bearing in mind the highly recurrent nature of TRD over periods of up to 36 months (105, 106), cost effectiveness might be overestimated through censoring of relapse or recurrence events. Extrapolation implicitly introduces additional uncertainty into the model, but one must balance the impact of that additional uncertainty on results against the benefits of decision support that reflects the longer-term costs and consequences of the intervention in question.

5.3. Analytical framework

Most evaluations included in this study used a CUA design, typically estimating incremental QALY changes associated with each alternative, with only five (mostly older evaluations) using only a CEA or CCA design. While the CEA approach has advantages – the results can be more intuitive for decision makers, and uncertainty is reduced since conversion of outcome measures to utility scores is not required – the results are of lesser value than those of a CUA for informing resource allocation decisions. Firstly, there is no immediately obvious decision rule: at what threshold of cost should a depression-free day be considered cost effective, for example? Perhaps more important, though, is the facility enabled by CUA to evaluate the cost effectiveness of an intervention within the whole healthcare sector. Mental healthcare provision is underfunded globally (107), and budgets for provision of mental healthcare are typically not ringfenced, but must compete with other healthcare priorities. To justify support for novel interventions, commissioners must be able to appraise the value of those interventions within the context of these competing priorities – e.g., mental health vs cardiovascular disease.

5.4. Summary measures of benefit

The most common economic outcome measure was the QALY, in most cases estimated using the EQ-5D-3L measure. Model-based evaluations predominantly used low-quality evidence to inform this parameter: sources were typically outdated, used unsophisticated valuation methods, and were usually drawn from the broader MDD population, rather than TRD specific. There is good evidence that an increased number of treatment failures within an episode is associated with both increased depression severity and decreased HRQoL (8). This would indicate that HRQoL in TRD follows a somewhat distinct profile from the broader MDD population, and highlights the importance of using values specific to the population under study. Generic preference-based HRQoL measures are increasingly deployed in interventional evaluations (including eight described in this review): synthesis of contemporary data specific to the TRD population should therefore considered for future economic evaluations.

Generic measures are typically recommended over condition-specific measures, since they facilitate comparable outcome collection across the healthcare spectrum, and (due to their brevity) are easy to collect (93). Despite their widespread use, however, there is a growing consensus amongst health economists working in mental health that generic measures such as the EQ-5D are not sufficiently sensitive to capture important changes in symptoms, functioning, or wellbeing in mental health conditions (108). While there is evidence that these issues may be valid in depression, concordance between generic HRQoL measures and clinical measures has been shown to reduce with severity (109). Partly in response to these concerns, there has been increased focus on measurement of wellbeing and quality of life in mental health (110), but to date, there exists no mental health domain-specific preference-based measure that has been sufficiently validated that it can be recommended as an alternative to the EQ-5D or the SF-6D. In the absence of such a measure, the quality of the evidence used to inform EQ-5D generated utility data is of particular importance, and extensive sensitivity testing of utility values is imperative. It should be noted that increasingly, the updated EQ-5D-5L (rather than the -3L) measure is used in interventional studies, owing to is superior psychometric properties (111). The value of supplementing a CUA with a secondary CEA or CCA analysis (for example incorporating mental-health specific outcomes, or patient preferences), in order to increase confidence in results, may additionally be considered.

Where a CEA approach was adopted, various outcomes were used (cost per remitter, cost per depression-free day, cost per relapse prevented, or simply incremental change in outcome). Cost per remission is arguably a more intuitive measure to present to decision makers, and conversion of the cost per unit change to this measure should be relatively straightforward, providing adequate availability of information.

5.5. Patient preference and priorities

Recent years have seen increasing interest in the adoption of a "values-based" framework for delivery of mental health care, explicitly incorporating the preferences, priorities, and values of mental health service users (112, 113). The incorporation of patient preferences in decisions related to resource allocation is justifiable on grounds of both ethics (since patients have agency in the decisions that affect their health), and on improving outcomes (patients are more likely to engage with interventions that match their preferences) (114). Despite this, none of the evaluations described incorporated patient values, preferences, or priorities in the presentation of their analysis. The HTA report by Atlas et al. (53) incorporated feedback from patient advocates, importantly highlighting concerns that the

clinical outcome measures typically used do not reflect the full burden of TRD, and calling for the incorporation of measures of impact on work, productivity, disability, and family or caregiver wellbeing. Elsewhere, patients have argued that remission is more accurately described by the presence of positive mental health features (optimism, vigor, and self-confidence) than the absence of symptoms (115). Although currently not a pre-requisite for HTA submissions, or best practice guidance, the growing recognition of the importance of the perspective of the patient in resource allocation decisions warrants serious consideration of how this might be incorporated explicitly in future economic evaluations. Longer-term objectives might consider the co-development of outcome measures that better reflect patient priorities; more immediately, methods such as discrete-choice experiments may be used to directly elicit and value both health and non-health impacts of interventions, facilitating direct incorporation of patient preferences in economic evaluations (116).

5.6. Reporting of resource use and cost data

Resource use in economic evaluation is highly context-specific – owing to the breadth of interventions, jurisdictions and cost perspectives considered by the evaluations in this study, a granular critical evaluation and comparison of resource use is unlikely to be informative. Focusing instead on broader resource item considerations, we found a reasonable level of consistency for direct costs across the evaluations. A third of the evaluations reviewed included indirect non-healthcare costs, although with considerable variation in the items included. In many cases this simply including productivity gains or losses which, when measured over relatively short time horizons, had a relatively small impact on results compared to the healthcare perspective. A minority considered a more comprehensive set of indirect costs. Variability in indirect costs that contribute to the broader "societal" perspectives is in part a reflection of the different contexts in which these evaluations were conducted: out-of-pocket costs, reliance on informal care, or transport costs may vary significantly between jurisdictions and in some cases may be so negligible that they are not considered for inclusion.

Good practice guidance relating to selection of costs for inclusion in economic evaluations recommends that either all relevant costs should be included, or (for more pragmatic studies) those costs that are most likely to meaningfully differ between comparators and thereby impact the result of the evaluation (47).

5.7. Perspective

The choice of cost perspective should be informed by the intended audience of the economic evaluation (47). Most commonly, the audience for economic evaluations is the payer; in the UK, NICE (whose remit is to determine if interventions should be funded by the NHS), requires that the perspective for economic evaluations should be that of the health service (104). Effective management of depression though, has been shown to have significantly greater impacts on productivity costs alone than on health care costs (21). When considering the global costs to society of poor mental health, choosing a narrow perspective that disregards those costs (or benefits) may be problematic, or even misleading.

Since mental health care is typically funded through public health care budgets, a health system perspective will be a pre-requisite for most decision makers, but we would reiterate the call from Knapp and Wong (21) that by providing a societal perspective in parallel, the broader societal impacts can also be taken into account. This broader perspective, however, is somewhat juxtaposed with our earlier recommendation that the primary analysis should use a CUA design. An immediate approach therefore might consider a secondary CCA analysis, adopting a societal perspective and reporting the non-health costs and benefits of alternatives.

5.8. Conceptualization and validation of model-based evaluations

None of the evaluations reviewed explicitly reported a formal conceptualization process, few presented a rationale for choice of model or model structure, and very few reported any robust validation of the model. The key health states described in most of the evaluations were consistent with established treatment goals of trials in MDD/TRD, including response, remission, and relapse (117). Sensitivity analyses of model-based evaluations frequently showed that it was these outcome parameters that were most likely to affect the results of the evaluations. Beyond these key endpoints, there was considerable variation in the structural complexity of model-based evaluations. Adverse events were rarely considered explicitly, although a minority of evaluations indicated that they had been considered and dismissed as having a negligible impact. Similarly, discontinuation was rarely considered, and where it was the reasons for discontinuation were poorly described.

Good practice guidance recommends an explicit process of conceptual modeling prior to implementation, to arrive at an appropriate scope for perspective, time horizon, choice of model type and structure, and which outcomes and costs to consider (118). The requirement to explicitly detail model conceptualization in reports has recently been added to the NICE HTA manual section 4.6.3 (104).

5.9. Limitations

This review restricted search criteria to English language only evaluations; by excluding foreign language records, our review may have limited consideration of aspects of economic evaluation that are prioritized differently in non-English speaking jurisdictions.

The review was deliberately designed with a "broad-brush" approach. Our aim was to develop a resource to inform the design of future economic evaluations in TRD agnostic of intervention, setting, or perspective. The review consequently incorporated all intervention types and all study design types; however, this introduces heterogeneity into the review, and limits the detail with which differences between evaluations may be explored. In keeping with the broad-brush approach, evaluation appraisal and recommendations are necessarily made at a generic level, and are not specific to context. Comparative evaluation of the results of included studies was conducted at a superficial level to illustrate how different evaluation design considerations may influence study conclusions. Where comparison of results is undertaken to inform resource allocation

decisions, it is critical that context is accounted for. Key factors that should be considered in further detail in such comparisons include severity; number of previously failed treatments; treatment setting; and jurisdictional variations in resource costs and cost-effectiveness thresholds.

6. Conclusion

Consistent with reviews of economic evaluations in MDD (23), our review found that the economic evidence for interventions in TRD is underdeveloped, particularly so for service-level interventions. Where evidence does exist, it is hampered by inconsistency in study design, methodological quality, and availability of high quality long-term outcomes evidence. Consequently there is limited data available to reassure policy makers involved in commissioning interventions and services in TRD of their cost effectiveness.

To strengthen the evidence base, this review identifies a number of key considerations and challenges for the design of future economic evaluations. While some considerations may be addressed immediately (e.g., appropriately defining the evaluation population, and selection of appropriate time-horizon and perspective), we also identify longer term challenges related to methodology development and building consensus in the research community to promote consistency in study design. The lack of long-term outcomes data limits the value of current economic evaluations. In particular we identified a need for more robust health-state utility data specific to TRD; consensus for a core outcome set that incorporates the measures from which these are derived would be a significant step forward.

Reflecting the growing recognition of the importance of incorporating the values of the patient in resource allocation decisions, we also suggest there is a need to develop methods to incorporate those values in economic evaluation frameworks systematically.

Data availability statement

The original contributions presented in this study are included in this article/**Supplementary material**; further inquiries can be directed to the corresponding author.

Author contributions

LJ and CW designed the research programme this review belongs to. LH, JP, and RAC developed the protocol and study design. LH and RAC conducted the literature search, screening of reports, data extraction, analysis, and drafted the manuscript. RNC contributed to the interpretation of literature review and critical revision of the manuscript. All authors reviewed and contributed to the final draft of this manuscript.

Funding

This study was supported by a Medical Education Grant from Janssen Pharmaceutical Companies of Johnson & Johnson [Medical Educational Goods and Services (MEGS) Agreement reference EM-31940], by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration East of England (NIHR ARC EoE) at Cambridgeshire and Peterborough NHS Foundation Trust, and by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). RNC's research was supported by the UK Medical Research Council (MR/W014386/1). The funders had no role in the design, review, interpretation, writing of the report, or the decision to submit it for publication.

Acknowledgments

We would like to thank Professor Jennifer Whitty, Honorary Professorial Fellow at Norwich Medical School, and Senior Research Scientist at Evidera for her role in designing the research programme this review belongs to, and guiding development of the original protocol.

Conflict of interest

The authors declare that this study received funding from Janssen Pharmaceutical Companies of Johnson & Johnson. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Publisher's note

Author disclaimer

This is a summary of research funded/supported by the National Institute for Health Research (NIHR) Applied Research Collaboration East of England. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1056210/full#supplementary-material

References

1. World Health Organization. *Depression Factsheet*. Geneva: World Health Organization (2021).

2. Proudman D, Greenberg P, Nellesen D. The growing burden of major depressive disorders (MDD): implications for researchers and policy makers. *Pharmacoeconomics*. (2021) 39:619–25. doi: 10.1007/s40273-021-01040-7

3. Santomauro D, Herrera A, Shadid J, Zheng P, Ashbaugh C, Pigott D, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet*. (2021) 398:1700–12. doi: 10.1016/S0140-6736(21)02143-7

4. Esterwood E, Saeed S. Past epidemics, natural disasters, COVID19, and mental health: learning from history as we deal with the present and prepare for the future. *Psychiatr Q*. (2020) 91:1121–33. doi: 10.1007/s11126-020-09808-4

5. Kathirvel N. Post COVID-19 pandemic mental health challenges. *Asian J Psychiatr*. (2020) 53:102430. doi: 10.1016/j.ajp.2020.102430

6. Trivedi M, Fava M, Wisniewski S, Thase M, Quitkin F, Warden D, et al. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. (2006) 354:1243–52. doi: 10.1056/NEJMoa052964

7. Rush A, Trivedi M, Wisniewski S, Nierenberg A, Stewart J, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. (2006) 163:1905–17. doi: 10.1176/ajp.2006.163.11.1905

8. Johnston K, Powell L, Anderson I, Szabo S, Cline S. The burden of treatment-resistant depression: a systematic review of the economic and quality of life literature. *J Affect Disord*. (2019) 242:195–210. doi: 10.1016/j.jad.2018.06.045

9. Gaynes B, Lux L, Gartlehner G, Asher G, Forman-Hoffman V, Green J, et al. Defining treatment-resistant depression. *Depress Anxiety*. (2020) 37:134–45. doi: 10.1002/da.22968

10. Demyttenaere K, Van Duppen Z. The impact of (the Concept of) treatment-resistant depression: an opinion review. *Int J Neuropsychopharmacol*. (2019) 22:85–92. doi: 10.1093/ijnp/pyy052

11. Akil H, Gordon J, Hen R, Javitch J, Mayberg H, McEwen B, et al. Treatment resistant depression: a multi-scale, systems biology approach. *Neurosci Biobehav Rev*. (2018) 84:272–88. doi: 10.1016/j.neubiorev.2017.08.019

12. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency: narrative literature review. *Br J Psychiatry*. (2005) 186:11–7. doi: 10.1192/bjp.186.1.11

13. National Institute for Health and Care Excellence. *Depression in Adults: Treatment and Management*. London: National Institute for Health and Care Excellence (2022).

14. McIntyre R, Millson B, Power G. Burden of treatment resistant depression (TRD) in patients with major depressive disorder in Ontario using institute for clinical evaluative sciences (ICES) databases: economic burden and healthcare resource utilization. *J Affect Disord*. (2020) 277:30–8. doi: 10.1016/j.jad.2020.07.045

15. Jaffe D, Rive B, Denee T. The humanistic and economic burden of treatment-resistant depression in Europe: a cross-sectional study. *BMC Psychiatry*. (2019) 19:247. doi: 10.1186/s12888-019-2222-4

16. Fekadu A, Wooderson S, Markopoulo K, Donaldson C, Papadopoulos A, Cleare A. What happens to patients with treatment-resistant depression? A systematic review of medium to long term outcome studies. *J Affect Disord*. (2009) 116:4–11. doi: 10.1016/j.jad.2008.10.014

17. Olchanski N, McInnis Myers M, Halseth M, Cyr P, Bockstedt L, Goss T, et al. The economic burden of treatment-resistant depression. *Clin Ther*. (2013) 35:512–22. doi: 10.1016/j.clinthera.2012.09.001

18. Sussman M, O'Sullivan AK, Shah A, Olfson M, Menzin J. Economic burden of treatment-resistant depression on the U.S. health care system. *J Manag Care Spec Pharm*. (2019) 25:823–35. doi: 10.18553/jmcp.2019.25.7.823

19. Mrazek D, Hornberger J, Altar C, Degtiar I. A review of the clinical, economic, and societal burden of treatment-resistant depression: 1996-2013. *Psychiatr Serv*. (2014) 65:977–87. doi: 10.1176/appi.ps.201300059

20. Perez-Sola V, Roca M, Alonso J, Gabilondo A, Hernando T, Sicras-Mainar A, et al. Economic impact of treatment-resistant depression: a retrospective observational study. *J Affect Disord*. (2021) 295:578–86. doi: 10.1016/j.jad.2021.08.036

21. Knapp M, Wong G. Economics and mental health: the current scenario. *World Psychiatry*. (2020) 19:3–14. doi: 10.1002/wps.20692

22. Grochtdreis T, Brettschneider C, Wegener A, Watzke B, Riedel-Heller S, Harter M, et al. Cost-effectiveness of collaborative care for the treatment of depressive disorders in primary care: a systematic review. *PLoS One*. (2015) 10:e0123078. doi: 10.1371/journal.pone.0123078

23. Zimovetz E, Wolowacz S, Classi P, Birt J. Methodologies used in cost-effectiveness models for evaluating treatments in major depressive disorder: a systematic review. *Cost Eff Resour Alloc*. (2012) 10:1. doi: 10.1186/1478-7547-10-1

24. Altunkaya J, Lee J, Tsiachristas A, Waite F, Freeman D, Leal J. Appraisal of patient-level health economic models of severe mental illness: systematic review. *Br J Psychiatry*. (2022) 220:2. doi: 10.1192/bjp.2021.121

25. York Health Economics Consortium. *Economic Evaluation [online] York*. Heslington: York Health Economics Consortium (2016).

26. York Health Economics Consortium. *Incremental Cost-Effectiveness Ratio (ICER) [online] York*. Heslington: York Health Economics Consortium (2016).

27. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Introduction to Economic Evaluation. Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press (2015).

28. York Health Economics Consortium. *Utility [online]*. Heslington: York Health Economics Consortium (2016).

29. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Introduction to the Measurement and Valuation of Health. Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press (2017). doi: 10.1093/med/9780198725923.001.0001

30. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Making Decisions in Health Care. Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press (2015).

31. Sanders G, Neumann P, Basu A, Brock D, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA*. (2016) 316:1093–103. doi: 10.1001/jama.2016.12195

32. Culyer A, Chalkidou K, Teerawattananon Y, Santatiwongchai B. Rival perspectives in health technology assessment and other economic evaluations for investing in global and national health. Who decides? Who pays? *F1000Res*. (2018) 7:72. doi: 10.12688/f1000research.13284.1

33. Ramsey S, Willke R, Glick H, Reed S, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II—an ISPOR good research practices task force report. *Value Health*. (2015) 18:161–72. doi: 10.1016/j.jval.2015.02.001

34. Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford: Oup Oxford (2006).

35. York Health Economics Consortium. *Time Horizon [online] York*. Heslington: York Health Economics Consortium (2016).

36. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Using Clinical Studies as Vehicles for Economic Evaluation. Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press (2015).

37. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ*. (2011) 342:d1548. doi: 10.1136/bmj.d1548

38. Briggs A. Economic evaluation and clinical trials: size matters. *BMJ*. (2000) 321:1362. doi: 10.1136/bmj.321.7273.1362

39. Fenwick E, Byford S. A guide to cost-effectiveness acceptability curves. *Br J Psychiatry*. (2005) 187:106–8. doi: 10.1192/bjp.187.2.106

40. Michalowsky B, Hoffmann W, Kennedy K, Xie F. Is the whole larger than the sum of its parts? Impact of missing data imputation in economic evaluation conducted alongside randomized controlled trials. *Eur J Health Econ*. (2020) 21:717–28. doi: 10.1007/s10198-020-01166-z

41. Jalali A, Tamimi R, McPherson S, Murphy S. Econometric issues in prospective economic evaluations alongside clinical trials: combining the nonparametric bootstrap with methods that address missing data. *Epidemiol Rev*. (2022) 44:67–77. doi: 10.1093/epirev/mxac006

42. Brand J, van Buuren S, le Cessie S, van den Hout W. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. *Stat Med*. (2019) 38:210–20. doi: 10.1002/sim.7956

43. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Economic Evaluation Using Decision-Analytic Modelling. Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press (2015).

44. York Health Economics Consortium. *Decision Tree [online]*. Heslington: York Health Economics Consortium (2016).

45. York Health Economics Consortium. *Markov Model [online]*. Heslington: York Health Economics Consortium (2016).

46. York Health Economics Consortium. *Micro-Simulation [online]*. Heslington: York Health Economics Consortium (2016).

47. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press (2015).

48. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. (2021) 372:n71. doi: 10.1136/bmj.n71

49. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. *Int J Technol Assess Health Care*. (2005) 21:240–5. doi: 10.1017/S0266462305050324

50. van Mastrigt G, Hiligsmann M, Arts J, Broos P, Kleijnen J, Evers S, et al. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: a five-step approach (part 1/3). *Expert Rev Pharmacoecon Outcomes Res*. (2016) 16:689–704. doi: 10.1080/14737167.2016.1246960

51. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess.* (2004) 8:ii–iv, ix–xi, 1–158. doi: 10.3310/hta8360

52. Galletly C, Clarke P, Fitzgerald P, Gill S, Loo C, Lyndon B, et al. *Repetitive Transcranial Magnetic Stimulation (rTMS), MSAC Application 1196*. Canberra, ACT: Commonwealth of Australia (2013).

53. Atlas S, Agboola F, Fazioli K, Kumar V, Adair E, Rind D, et al. *Esketamine for the Treatment of Treatment-Resistant Depression: Effectiveness and Value. Final Evidence Report*. Boston, MA: Institute for Clinical and Economic Review (2019).

54. Desai U, Kirson N, Guglielmo A, Le H, Spittle T, Tseng-Tham J, et al. Cost-per-remitter with esketamine nasal spray versus standard of care for treatment-resistant depression. *J Comp Eff Res.* (2021) 10:393–407. doi: 10.2217/cer-2020-0276

55. Edwards S, Hamilton V, Nherera L, Trevor N. Lithium or an atypical antipsychotic drug in the management of treatment-resistant depression: a systematic review and economic evaluation. *Health Technol Assess.* (2013) 17:1–190. doi: 10.3310/hta17540

56. Fitzgibbon K, Plett D, Chan B, Hancock-Howard R, Coyte P, Blumberger D. Cost-utility analysis of electroconvulsive therapy and repetitive transcranial magnetic stimulation for treatment-resistant depression in Ontario. *Can J Psychiatry.* (2020) 65:164–73. doi: 10.1177/0706743719890167

57. Ghiasvand H, Moradi-Joo M, Abolhassani N, Ravaghi H, Raygani S, Mohabbat-Bahar S. Economic evaluation of resistant major depressive disorder treatment in Iranian population: a comparison between repetitive transcranial magnetic stimulation with electroconvulsive. *Med J Islam Repub Iran.* (2016) 30:330.

58. Health Quality Ontario. Repetitive transcranial magnetic stimulation for treatment-resistant depression: an economic analysis. *Ont Health Technol Assess Ser.* (2016) 16:1–51.

59. Unit University of Calgary. *Repetitive Transcranial Magnetic Stimulation for Treatment-resistant DEPRESSION*. Calgary, AB: University of Calgary (2014). p. 2014

60. Kozel F, George M, Simpson K. Decision analysis of the cost-effectiveness of repetitive transcranial magnetic stimulation versus electroconvulsive therapy for treatment of nonpsychotic severe depression. *CNS Spectr.* (2004) 9:476–82.

61. Malone D. A budget-impact and cost-effectiveness model for second-line treatment of major depression. *J Manag Care Pharm.* (2007) 13:S8–18. doi: 10.18553/jmcp.2007.13.s6-a.8

62. Nguyen K, Gordon L. Cost-effectiveness of repetitive transcranial magnetic stimulation versus antidepressant therapy for treatment-resistant depression. *Value Health.* (2015) 18:597–604. doi: 10.1016/j.jval.2015.04.004

63. Olgiati P, Bajo E, Bigelli M, Montgomery S, Serretti A, C.E.A.P (Cost-Effectiveness Analysis on Psychiatric Disorders)-Group. Challenging sequential approach to treatment resistant depression: cost-utility analysis based on the sequenced treatment alternatives to relieve depression (STAR()D) trial. *Eur Neuropsychopharmacol.* (2013) 23:1739–46. doi: 10.1016/j.euroneuro.2013.08.008

64. Ross E, Zivin K, Maixner D. Cost-effectiveness of electroconvulsive therapy VS pharmacotherapy/psychotherapy for treatment-resistant depression in the united states. *JAMA Psychiatry.* (2018) 75:713–22. doi: 10.1001/jamapsychiatry.2018.0768

65. Ross E, Soeteman D. Cost-effectiveness of esketamine nasal spray for patients with treatment-resistant depression in the United States. *Psychiatr Serv.* (2020) 71:988–97. doi: 10.1176/appi.ps.201900625

66. Simpson K, Welch M, Kozel F, Demitrack M, Nahas Z. Cost-effectiveness of transcranial magnetic stimulation in the treatment of major depression: a health economics analysis. *Adv Ther.* (2009) 26:346–68. doi: 10.1007/s12325-009-0013-x

67. Vallejo-Torres L, Castilla I, Gonzalez N, Hunter R, Serrano-Perez P, Perestelo-Perez L. Cost-effectiveness of electroconvulsive therapy compared to repetitive transcranial magnetic stimulation for treatment-resistant severe depression: a decision model. *Psychol Med.* (2015) 45:1459–70. doi: 10.1017/S0033291714002554

68. Wang S, Anderson I, Mitchell D, Johnson S, Shiozawa A. Cost-effectiveness model for a hypothetical monotherapy vs standard of care in adult patients with treatment-resistant depression. *Clinicoecon Outcomes Res.* (2019) 11:257–70. doi: 10.2147/CEOR.S181718

69. Xie X, Falk L, Brophy J, Tu H, Guo J, Gajic-Veljanoski O, et al. A non-inferiority framework for cost-effectiveness analysis. *Int J Technol Assess Health Care.* (2019) 35:291–7. doi: 10.1017/S0266462319000576

70. Young A, Evitt L, Brignone M, Diamand F, Atsou K, Campbell R, et al. Cost-utility evaluation of vortioxetine in patients with major depressive disorder experiencing inadequate response to alternative antidepressants in the United Kingdom. *J Affect Disord.* (2017) 218:291–8. doi: 10.1016/j.jad.2017.04.019

71. Zhao Y, Tor P, Khoo A, Teng M, Lim B, Mok Y. Cost-effectiveness modeling of repetitive transcranial magnetic stimulation compared to electroconvulsive therapy for treatment-resistant depression in Singapore. *Neuromodulation.* (2018) 21:376–82. doi: 10.1111/ner.12723

72. McDonald W, Phillips V, Figiel G, Marsteller F, Simpson C, Bailey M. Cost-effective maintenance treatment of resistant geriatric depression. *Psychiatr Ann.* (1998) 28:47–52. doi: 10.3928/0048-5713-19980101-11

73. Hollinghurst S, Carroll F, Abel A, Campbell J, Garland A, Jerrom B, et al. Cost-effectiveness of cognitive-behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: economic evaluation of the CoBalT Trial. *Br J Psychiatry.* (2014) 204:69–76. doi: 10.1192/bjp.bp.112.125286

74. Kessler D, Burns A, Tallon D, Lewis G, MacNeill S, Round J, et al. Combining mirtazapine with SSRIs or SNRIs for treatment-resistant depression: the MIR RCT. *Health Technol Assess.* (2018) 22:1–136. doi: 10.3310/hta22630

75. Knapp M, Romeo R, Mogg A, Eranti S, Pluck G, Purvis R, et al. Cost-effectiveness of transcranial magnetic stimulation vs. electroconvulsive therapy for severe depression: a multi-centre randomised controlled trial. *J Affect Disord.* (2008) 109:273–85. doi: 10.1016/j.jad.2008.01.001

76. Lynch T, Hempel R, Whalley B, Byford S, Chamba R, Clarke P, et al. *Radically Open Dialectical Behaviour Therapy for Refractory Depression: The RefraMED RCT*. Southampton: NIHR Journals Library (2018). p. 12 doi: 10.3310/eme05070

77. Morriss R, Garland A, Nixon N, Guo B, James M, Kaylor-Hughes C, et al. Efficacy and cost-effectiveness of a specialist depression service versus usual specialist mental health care to manage persistent depression: a randomised controlled trial. *Lancet Psychiatry.* (2016) 3:821–31. doi: 10.1016/S2215-0366(16)30143-2

78. Scott J, Palmer S, Paykel E, Teasdale J, Hayhurst H. Use of cognitive therapy for relapse prevention in chronic depression. Cost-effectiveness study. *Br J Psychiatry.* (2003) 182:221–7. doi: 10.1192/bjp.182.3.221

79. Shearer J, Lynch T, Chamba R, Clarke S, Hempel R, Kingdon D, et al. Refractory depression–cost-effectiveness of radically open dialectical behaviour therapy: findings of economic evaluation of RefraMED trial. *BJPsych Open.* (2019) 5:e64. doi: 10.1192/bjo.2019.57

80. Simon G, Katon W, VonKorff M, Unutzer J, Lin E, Walker E, et al. Cost-effectiveness of a collaborative care program for primary care patients with persistent depression. *Am J Psychiatry.* (2001) 158:1638–44. doi: 10.1176/appi.ajp.158.10.1638

81. Town J, Abbass A, Stride C, Nunes A, Bernier D, Berrigan P. Efficacy and cost-effectiveness of intensive short-term dynamic psychotherapy for treatment resistant depression: 18-month follow-up of the Halifax depression trial. *J Affect Disord.* (2020) 273:194–202. doi: 10.1016/j.jad.2020.04.035

82. Wiles N, Thomas L, Turner N, Garfield K, Kounali D, Campbell J, et al. Long-term effectiveness and cost-effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: follow-up of the CoBalT randomised controlled trial. *Lancet Psychiatry.* (2016) 3:137–44. doi: 10.1016/S2215-0366(15)00495-2

83. Brown S, Rittenbach K, Cheung S, McKean G, MacMaster F, Clement F. Current and common definitions of treatment-resistant depression: findings from a systematic review and qualitative interviews. *Can J Psychiatry.* (2019) 64:380–7. doi: 10.1177/0706743719828965

84. Zimmerman M, Posternak M, Chelminski I. Is the cutoff to define remission on the Hamilton rating scale for depression too high? *J Nerv Ment Dis.* (2005) 193:170–5. doi: 10.1097/01.nmd.0000154840.63529.5d

85. Hawley C, Gale T, Sivakumaran T. Hertfordshire neuroscience research g. defining remission by cut off score on the MADRS: selecting the optimal value. *J Affect Disord.* (2002) 72:177–84. doi: 10.1016/S0165-0327(01)00451-7

86. Zimmerman M, Posternak M, Chelminski I. Defining remission on the montgomery-Asberg depression rating scale. *J Clin Psychiatry.* (2004) 65:163–8. doi: 10.4088/JCP.v65n0204

87. Kennedy-Martin M, Slaap B, Herdman M, van Reenen M, Kennedy-Martin T, Greiner W, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ.* (2020) 21:1245–57. doi: 10.1007/s10198-020-01195-8

88. Janssen M, Pickard A, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res.* (2013) 22:1717–27. doi: 10.1007/s11136-012-0322-4

89. McLoughlin D, Mogg A, Eranti S, Pluck G, Purvis R, Edwards D, et al. The clinical effectiveness and cost of repetitive transcranial magnetic stimulation versus electroconvulsive therapy in severe depression: a multicentre pragmatic randomised controlled trial and economic analysis. *Health Technol Assess.* (2007) 11:1–54. doi: 10.3310/hta11240

90. Bennett K, Torrance G, Boyle M, Guscott R. Cost-utility analysis in depression: the McSad utility measure for depression health states. *Psychiatr Serv.* (2000) 51:1171–6. doi: 10.1176/appi.ps.51.9.1171

91. Feeny D, Furlong W, Torrance G, Goldsmith C, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care.* (2002) 40:113–28. doi: 10.1097/00005650-200202000-00006

92. Revicki D, Wood M. Patient-assigned health state utilities for depression-related outcomes: differences by depression severity and antidepressant medications. *J Affect Disord.* (1998) 48:25–36. doi: 10.1016/S0165-0327(97)00117-1

93. Sapin C, Fantino B, Nowicki M, Kind P. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. *Health Qual Life Outcomes.* (2004) 2:20. doi: 10.1186/1477-7525-2-20

94. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Design and Analysis of Health State Valuation Data for Model-Based Economic Evaluations and for Economic Evaluations Alongside Clinical Trials. Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press (2017). doi: 10.1093/med/9780198725923.003.0009

95. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Identifying, Synthesizing, and Analysing Evidence for Economic Evaluation. Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press (2015).

96. Sullivan P, Valuck R, Saseen J, MacFall H. A comparison of the direct costs and cost effectiveness of serotonin reuptake inhibitors and associated adverse drug reactions. *CNS Drugs.* (2004) 18:911–32. doi: 10.2165/00023210-200418130-00006

97. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ.* (1999) 18:341–64. doi: 10.1016/S0167-6296(98)00039-3

98. Sin J, Galeazzi G, McGregor E, Collom J, Taylor A, Barrett B, et al. Digital interventions for screening and treating common mental disorders or symptoms of common mental illness in adults: systematic review and meta-analysis. *J Med Internet Res.* (2020) 22:e20581. doi: 10.2196/20581

99. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Health.* (2018) 5:e64. doi: 10.2196/mental.9782

100. Milne-Ives M, de Cock C, Lim E, Shehadeh M, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. *J Med Internet Res.* (2020) 22:e20346. doi: 10.2196/20346

101. Voineskos D, Daskalakis Z, Blumberger D. Management of treatment-resistant depression: challenges and strategies. *Neuropsychiatr Dis Treat.* (2020) 16:221–34. doi: 10.2147/NDT.S198774

102. Fava M. Diagnosis and definition of treatment-resistant depression. *Biol Psychiatry.* (2003) 53:649–59. doi: 10.1016/S0006-3223(03)00231-2

103. Russell J, Hawkins K, Ozminkowski R, Orsini L, Crown W, Kennedy S, et al. The cost consequences of treatment-resistant depression. *J Clin Psychiatry.* (2004) 65:341–7. doi: 10.4088/JCP.v65n0309

104. National Institute for Health and Care Excellence. *NICE Health Technology Evaluations: The Manual*. London: National Institute for Health and Care Excellence (2022).

105. Gillain B, Degraeve G, Dreesen T, De Bruecker G, Buntinx E, Beke D, et al. Real-world treatment patterns, outcomes, resource utilization and costs in treatment-resistant major depressive disorder: PATTERN, a retrospective cohort study in Belgium. *Pharmacoecon Open.* (2022) 6:293–302. doi: 10.1007/s41669-021-00306-2

106. Lex H, Nevers S, Jensen E, Ginsburg Y, Maixner D, Mickey B. Long-term quality of life in treatment-resistant depression after electroconvulsive therapy. *J Affect Disord.* (2021) 291:135–9. doi: 10.1016/j.jad.2021.05.012

107. World Health Organization. *WHO Report Highlights Global Shortfall in Investment in Mental Health*. Geneva: World Health Organization (2021).

108. Knapp M, Mangalore R. The trouble with QALYs. *Epidemiol Psichiatr Soc.* (2007) 16:289–93. doi: 10.1017/S1121189X00002451

109. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess.* (2014) 18:vii–viii, xiii–xxv, 1–188. doi: 10.3310/hta18340

110. van Krugten F, Feskens K, Busschbach J, Hakkaart-van Roijen L, Brouwer W. Instruments to assess quality of life in people with mental health problems: a systematic review and dimension analysis of generic, domain- and disease-specific instruments. *Health Qual Life Outcomes.* (2021) 19:249. doi: 10.1186/s12955-021-01883-w

111. Buchholz I, Janssen M, Kohlmann T, Feng Y. A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *PharmacoEconomics.* (2018) 36:645–61. doi: 10.1007/s40273-018-0642-5

112. Woodbridge-Dodd K. Values-based practice in mental health and psychiatry. *Curr Opin Psychiatry.* (2012) 25:508–12. doi: 10.1097/YCO.0b013e328359051c

113. Baggaley M. Value-based healthcare in mental health services. *BJPsych Adv.* (2020) 26:198–204. doi: 10.1192/bja.2019.82

114. Swift J, Mullins R, Penix E, Roth K, Trusty W. The importance of listening to patient preferences when making mental health care decisions. *World Psychiatry.* (2021) 20:316–7. doi: 10.1002/wps.20912

115. Zimmerman M, McGlinchey J, Posternak M, Friedman M, Attiullah N, Boerescu D. How should remission from depression be defined? The depressed patient's perspective. *Am J Psychiatry.* (2006) 163:148–50. doi: 10.1176/appi.ajp.163.1.148

116. Wildman J, Wildman J. Combining health and outcomes beyond health in complex evaluations of complex interventions: suggestions for economic evaluation. *Value Health.* (2019) 22:511–7. doi: 10.1016/j.jval.2019.01.002

117. Rush A, Thase M, Dube S. Research issues in the study of difficult-to-treat depression. *Biol Psychiatry.* (2003) 53:743–53. doi: 10.1016/S0006-3223(03)00088-X

118. Roberts M, Russell L, Paltiel A, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force-2. *Med Decis Making.* (2012) 32:678–89. doi: 10.1177/0272989X12454941

frontiers | Frontiers in Psychiatry

TYPE Methods
PUBLISHED 21 February 2023
DOI 10.3389/fpsyt.2023.1030407

OPEN ACCESS

EDITED BY
Ben Wijnen,
Trimbos Institute, Netherlands

REVIEWED BY
Ellie Brown,
The University of Melbourne, Australia
Eleanor Grieve,
University of Glasgow, United Kingdom

*CORRESPONDENCE
Jai L. Shah
jai.shah@mcgill.ca

SPECIALTY SECTION
This article was submitted to Public Mental Health, a section of the journal Frontiers in Psychiatry

RECEIVED 28 August 2022
ACCEPTED 03 January 2023
PUBLISHED 21 February 2023

CITATION
Shah JL, Moinfar Z, Anderson KK, Gould H, Hutt-Macleod D, Jacobs P, Mitchell S, Nguyen T, Rodrigues R, Reaume-Zimmer P, Rudderham H, Rudderham S, Smyth R, Surood S, Urichuk L, Malla AK, Iyer SN and Latimer E (2023) Return on investment from service transformation for young people experiencing mental health problems: Approach to economic evaluations in ACCESS Open Minds (Esprits ouverts), a multi-site pan-Canadian youth mental health project. *Front. Psychiatry* 14:1030407. doi: 10.3389/fpsyt.2023.1030407

Return on investment from service transformation for young people experiencing mental health problems: Approach to economic evaluations in ACCESS Open Minds (Esprits ouverts), a multi-site pan-Canadian youth mental health project

Jai L. Shah[1,2]*, Zeinab Moinfar[1], Kelly K. Anderson[3,4], Hayley Gould[5], Daphne Hutt-Macleod[5], Philip Jacobs[6], Stephen Mitchell[7], Thanh Nguyen[6], Rebecca Rodrigues[3], Paula Reaume-Zimmer[7,8], Heather Rudderham[5], Sharon Rudderham[5], Rebecca Smyth[7], Shireen Surood[9], Liana Urichuk[9], Ashok K. Malla[1,2], Srividya N. Iyer[1,2] and Eric Latimer[1,2]

[1]ACCESS Open Minds, Douglas Research Centre, Montreal, QC, Canada, [2]Department of Psychiatry, McGill University, Montreal, QC, Canada, [3]Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada, [4]Institute of Clinical and Evaluative Sciences, Toronto, ON, Canada, [5]Eskasoni First Nation, Eskasoni, NS, Canada, [6]Faculty of Medicine, University of Alberta, Edmonton, AB, Canada, [7]Canadian Mental Health Association Lambton Kent, Chatham-Kent, ON, Canada, [8]Bluewater Health, Chatham-Kent, ON, Canada, [9]Alberta Health Services, Edmonton, AB, Canada

Introduction: Mental health problems are common globally, and typically have their onset in adolescence and early adulthood—making youth (aged 11–25) an optimal target for prevention and early intervention efforts. While increasing numbers of youth mental health (YMH) initiatives are now underway, thus far few have been subject to economic evaluations. Here we describe an approach to determining the return on investment of YMH service transformation *via* the pan-Canadian ACCESS Open Minds (AOM) project, for which a key focus is on improving access to mental health care and reducing unmet need in community settings.

Approach: As a complex intervention package, it is hoped that the AOM transformation will: (i) enable early intervention through accessible, community-based services; (ii) shift care away toward these primary/community settings and away from acute hospital and emergency services; and (iii) offset at least some of the increased costs of primary care/community-based mental health services with reductions in the volume of more resource-intensive acute, emergency, hospital or specialist services utilized. Co-designed with three diverse sites that represent different Canadian contexts, a return on investment analysis will (separately at each site) compare the costs generated by the intervention, including volumes and expenditures associated with the AOM service transformation and any contemporaneous changes in acute, emergency, hospital or service utilization (vs.

historical or parallel comparators). Available data from health system partners are being mobilized to assess these hypotheses.

Anticipated results: Across urban, semi-urban and Indigenous sites, the additional costs of the AOM transformation and its implementation in community settings are expected to be at least partially offset by a reduction in the need for acute, emergency, hospital or specialist care.

Discussion: Complex interventions such as AOM aim to shift care "upstream": away from acute, emergency, hospital and specialist services and toward community-based programming which is more easily accessible, often more appropriate for early-stage presentations, and more resource-efficient. Carrying out economic evaluations of such interventions is challenging given the constraints of available data and health system organization. Nonetheless, such analyses can advance knowledge, strengthen stakeholder engagement, and further implementation of this public health priority.

KEYWORDS

economic evaluation, youth mental health, service transformation, return on investment, service utilization

Introduction

Following an era of relative neglect, mental health—and particularly the mental health of young people–is now seen to be of essential importance. Mental health problems usually begin before the age of 25, and can evolve or persist to adversely impact social, vocational and other trajectories (1–3). Mental health and substance use disorders are common worldwide and major contributors to the global disease burden, surpassing both cardiovascular disease and cancer (4, 5). In Canada, one in five young people are affected by them, making early identification and intervention during the critical period of age 12–25 central to reducing suffering and ensuring prompt, high-quality care (6). Particularly when these elements are absent, mental health problems are likely to have a substantial impact at both individual and population levels, and an associated economic cost.

From the perspective of youth mental health (YMH) services, better access to care in community settings should help to identify and provide services earlier on in the course of illness, with a corresponding improvement in population-level outcomes. Simultaneously, the provision of evidence-informed care should improve outcomes at the individual level. In response to a 2013 call for a pan-Canadian network in YMH service transformation, the ACCESS Open Minds (AOM)/Esprits ouverts project was conceived to implement related innovations for youth aged 11–25 years at 14 different sites across Canada (7, 8). Evaluation of data around five operational objectives of the project (**Box 1**) is now underway to determine the extent to which the project has increased youth referrals and help-seeking, sped up response times to requests for assessment, provided access to appropriate services, eliminated age-based transitions, and engaged youth and families (9).

The project was designed to harness multiple methods including a minimum evaluation protocol, a pragmatic trial, and qualitative approaches, which are or will be described in separate publications (8, 9). Here we articulate the protocol for a linked project—AOM's economic evaluations, taking place in three specific sites—in which we examine whether the AOM transformation was able to shift care toward community-based services, whose cost is at least partially offset by a reduction in acute, emergency, hospital or specialist care. It is hoped that a return on investment analysis will provide additional rationale for the effectiveness of broad, principles-based YMH service transformation, ultimately serving to inform policy-makers of sustainable solutions for mental health services for young people.

Methods

Rationale and selection of outcomes

Economic evaluations of complex intervention packages such as AOM are widely recognized to be worthwhile (10, 11), and yet are relatively rare in mental healthcare and healthcare in general (12, 13) as compared to the more common economic evaluations of specific health technologies (14). In part this may be because (i) such intervention packages are difficult to standardize; (ii) capturing key elements of local context and variation in implementation can be elusive; and (iii) the multiple links between intervention and outcome are complex (15).

In light of these challenges, the AOM economic evaluations are being designed to inform decision-makers about the extent to which a novel and complex intervention can achieve its projected impact of increasing access to and shifting provision of care away from acute/emergency or hospital/specialist services and toward community-based settings. To do this, we developed a conceptual model in which the AOM intervention was likely to have multiple effects (*via* rapid assessment, loose entry criteria, youth-friendly services, efficient triaging, etc.) with feed-back and feed-forward loops as in any complex mental health system. Through a process of stakeholder engagement with site-level partners, we determined that reducing unmet mental health needs in young people was a

BOX 1 At each site, the ACCESS Open Minds "intervention" transforms services to provide the following for youth aged 12–25.

1.Early case identification: targeted outreach, community awareness campaigns, etc., such that more youth self-refer or are referred sooner [16].

2.Rapid access that is engaging, including a offer of initial evaluation within 72 h in a non-emergency, community-based environment. A trained "ACCESS Clinician" will be deployed to conduct first evaluations; include family members in the process; and connect youth with services tailored to their needs and preferences. There are multiple portals of access; the elimination of referral or administrative requirements; and appropriate use of helplines, social media, etc.

3.Appropriate evidence-informed, illness-appropriate interventions offered within 30 days of initial evaluation (per Canadian Psychiatric Association benchmarks) (17). Treatment planning is guided less by symptoms (which can be non-specific and overlapping) or diagnoses and more by self-reported distress and functioning, and clinicians' impressions of problems and their severity. Care is focused on youth-defined goals, and provided in friendly, non-stigmatizing, and recovery-oriented settings. Where appropriate treatments are not available on site, youth will be connected to external services/specialists.

4.Continuity of care is prioritized to ensure that youth receive appropriate care for as long as needed. There is an emphasis on collaboration across services, stakeholders, sectors, and disciplines to reduce eliminate barriers, such as age-based transitions or transitions between other needed services, e.g., from primary to specialized care.

5.Engagement and involvement of youth and family/carers. Youth and families will be part of network- and site-level service design, oversight, and hiring committees; their input will be sought in designing youth spaces; intervention menus will be individualized, appointment times and venues will be flexible where possible; and clinician training will prioritize strengths-affirming and youth-friendly approaches. Transformation plans at all sites include core strategies such as deploying an ACCESS Clinician, responding to help-seeking/referrals within 72 h, designing and creating a physical space that is youth-friendly, and incorporating relevant evidence and local conditions.

meaningful population-level distal outcome, with multiple benefits for individuals as well as communities and health systems (7). Proximal to this outcome, however, is a YMH system in which patients are seen in primary care/community rather than acute or specialist care settings, due to key aspects of the AOM intervention such as sustained outreach and early case identification activities. If such efforts encourage youth to seek care at earlier stages of illness, then appropriate interventions reduce the need for later, more invasive or resource-intensive treatments and services.

The conceptual linkage between intervention and outcome is illustrated in **Figure 1**: traditional systems pose substantial barriers to accessing care, leading to individuals whose needs go unmet during early (and presumably less acute/severe) stages of illness. A lack of accessible treatments results in a proportion of these cases developing later-stage mental health problems that have a subsequent need for higher intensity care (including emergency or specialist services) (**Figure 1A**). In contrast, an AOM-transformed system of care is hypothesized to have reduced barriers to accessing services; this along with tailored outreach activities and youth-friendly services could encourage young people to access care at earlier stages of need, in lower-intensity primary/community care rather than high-intensity settings (**Figure 1B**). In at least some of these youth, obtaining treatment earlier would prevent or reduce the need for higher intensity care.

Objectives

Based on the principle that youth-friendly and stage-appropriate mental health services delivered in the community are preferable to and deliver an improved experience of care than in acute/hospital-based settings, the main AOM project seeks (among other things) to determine whether AOM's model significantly increases the number of youth receiving mental health-related services (9).

In AOM's economic evaluation, we will augment this at three study sites representing diverse Canadian settings to understand if:

- *Hypothesis 1:* There will be a significant increase in the average number of mental health-related primary care or community visits per person during the post-AOM period compared with the pre-AOM period.
- *Hypothesis 2:* There will be a significant reduction in the average number of mental health-related acute, emergency, hospital or specialist visits per person during the post-AOM period compared with the pre-AOM period.
- *Hypothesis 3:* The increase in the cost of mental health-related community/primary care visits in the post- compared with the pre-AOM period will be offset at least in part by a reduction in the cost of acute, emergency, hospital and specialist visits.
- *Exploratory Objective:* Where possible, we will attempt to examine non-mental health-related service use.

Setting/Sites

Overall, AOM examines how and to what extent the transformations identify youth in need (defined as any mental health problem), improve their access to high-quality mental healthcare, and the ways in which transformations are beneficial with respect to both individual- and service-level outcomes. Its 14 sites represent Canada's diverse geography, culture, resources, and population density. In recognition of this breadth, the AOM economic evaluation will take place at three sites representing different facets of the Canadian landscape:

- A remote Indigenous community, Eskasoni First Nation in the province of Nova Scotia (18). Indigenous communities tend to have relatively large youth populations, and some of them have experienced high rates of suicidality, vocational disengagement, involvement with youth protection and justice systems, as well as addiction and violence – much of which has been linked to colonial policies and the ensuing intergenerational trauma and cultural fragmentation.
- A semi-urban and rural community, Chatham-Kent in the province of Ontario (19). Prior to AOM, Chatham-Kent was an example of a siloed mental health system with resulting overlaps, lack of coordination and uncertainty regarding where individuals should access care.
- A large urban center, Edmonton in the province of Alberta (20). In Canada, cities are pluralistic and multicultural, including youth with particular vulnerabilities (ethnic minorities, homeless youth, post-secondary students, immigrants, refugees, etc.).

Beyond their sociodemographic contexts, these sites are located in different parts of Canada and therefore situated in different health systems. In keeping with the Canadian Institutes of Health Research–Strategy for Patient-Oriented Research stream under which AOM was funded, a high degree of site engagement and involvement

of local communities was needed when designing the economic evaluations (7, 8). This co-design has enabled alignment with local priorities, ensured access to needed data, and is consistent with values articulated by Indigenous and patient-oriented research advocates.

Study designs

AOM's multi-pronged programme of work includes a minimum evaluation protocol, qualitative methods, mapping exercises, stakeholder consultations, and other facets (9). The transformations are being studied through a multi-stakeholder led Research Advisory Group that includes individuals from all sites, amidst a broader governance structure (7).

Participants

Following a principles-based site-specific transformation of services, young people either self-refer, directly access (e.g., *via* walk-in sessions), or are referred by others to the AOM service. The referral process is open, meaning that referrals can be made by anyone—including but not limited to health providers. Youth are either seen initially, followed at the same site and/or during subsequent referral to an appropriate local service that is also affiliated with the overall transformation. While individuals could provide informed consent for the main AOM research project in the context of inclusion and exclusion criteria,[1] they can obtain services from the site even without consenting to the main project.

Unlike the main project, however, the economic evaluations will rely on secondary use of service data routinely collected by the surrounding health system during the course of care–regardless of individuals' involvement in the main AOM study. This means that the economic evaluations require no opt-in or opt-out consent; they instead utilize administrative data regarding all youth within the age range who received services at the site [or its comparator setting(s)].

Data

As costs will be estimated using administrative data, the perspective of each economic evaluation is that of the healthcare system.

The AOM economic evaluation integrates data collected at the site level with data collected *via* the "host" provincial health system. Because this system of care varies a great deal from site to site, the three economic evaluations are independent of each other: they will separately assess relevant service utilization alongside costs in those attending their AOM-transformed service, relative to the pre-AOM period and, where possible, a comparison site.

1 For the main AOM project (9), site clinicians identified potential study participants to research staff who would then explain the project and seek, in a youth-friendly fashion, written informed consent either at intake or at a later/more appropriate point. For minors or those with reduced capacity, consent was sought from a parent or legal representative but with assent from the youth [following applicable provincial/institutional regulations around age; see details in Iyer et al. (9)]. Most AOM projects are nested within the local service that aimed to provide care to a broad sample of youth within the 11–25 age range and their families/carers who were seeking help for mental health or substance-related problems. Exclusion criteria for research purposes included individuals younger than 11 or older than 25 (with the exception of family members/carers), those with a diagnosed intellectual disability, a history of organic brain damage, those unable to provide informed consent, or those who had received mental health or addiction services within the 6 months prior to site transformation (9).

Costs

Provincially-held administrative data will be used at all sites to estimate costs of acute, emergency, hospital and specialist services received outside of the AOM site both before and after its transformation. Hospitalization costs will be estimated using the Canadian Institute of Health Information case-mix group plus (CMG+) or cost per weighted case (CPWC) methodology (21).

Eskasoni

For Eskasoni, we will assess changes in local (community) and provincial (emergency, hospital and outpatient physician billing) service utilization over time in the population of youth aged 11–25. We will compare these as well as associated costs before versus after the advent of the AOM service. Including both those who did or did not use local services (See **Table 1** for details) will permit inclusion of the entire Eskasoni youth population, along with potential changes in case-mix and their impact on outcomes and costs. Site-level data will be encrypted then linked with provincial administrative data by Health Data Nova Scotia (HDNS).

Chatham-Kent

As in Eskasoni, population-level clinical and service-related outcomes (and their associated costs) for youth aged 11–25 years under AOM in Chatham-Kent will be compared with those prior to AOM. In addition, however, the difference between the two will be compared with the equivalent time period in a neighboring community, Sarnia, which did not have an AOM site. Services in Sarnia have not changed during the observation window; at no point did they correspond to those under AOM. Relevant service outcomes include emergency visits, hospitalizations, outpatient psychiatric and non-psychiatric visits assembled using physician billing and other provincial databases available *via* Institute of Clinical and Evaluative Sciences (formerly the Ontario ICES; see **Table 1**). Encrypted site-level data will be linked with provincial administrative data within secure ICES holdings where it will be analyzed.

Edmonton

In Edmonton, youth aged 15–25 attending provincial community mental health clinics that did not implement AOM will be compared with those receiving the AOM clinic's intervention package during the same observation period, and in the post-AOM versus the pre-AOM period. Outcomes of interest include the health services provided and their costs, for: hospitalizations, emergency department visits, outpatient clinic visits, specialist and family physician visits, prescription drug usage, community mental health clinic visits, and residential stays. These are available through provincially collected datasets (**Table 1**).

AOM implementation costs include infrastructure and setup, staff salaries, outreach/support, and overhead. These costs will be estimated using data collected at the site level.

Data linkage, encryption and transfer

At each site, service usage data will be sent securely to the responsible provincial department which will encrypt identifiers, including for those with unique health card numbers (*via* deterministic data linkage) and those without (*via* probabilistic data linkage). The data will then be linked to the respective health administrative data records. Once linked, the AOM and site team (with the support of an administrative data analyst if needed)

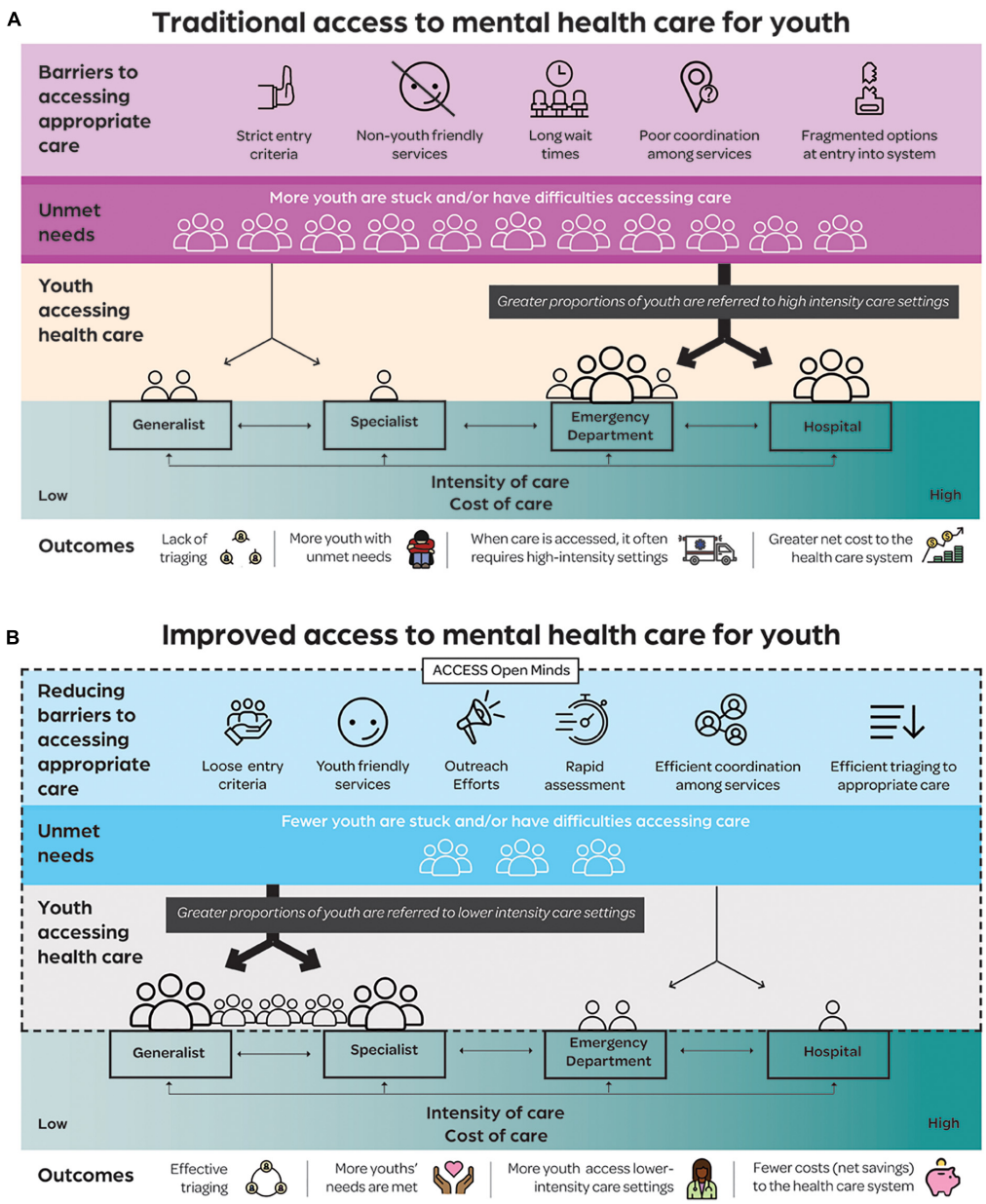


FIGURE 1
(A) Traditional service systems are characterized by being unfriendly to youth, having strict entry criteria, long wait times, and fragmented, poorly coordinated services. Youth are less likely to attempt to access such services when their needs are in early (less intense) stage, resulting in more individuals with unmet needs and difficulties accessing care when these needs grow. Care therefore becomes acute and intense in nature, including *via* emergency departments and other more costly services. **(B)** ACCESS Open Minds (AOM's) service transformations result in more welcoming and youth-friendly services, looser entry criteria, rapid assessments, improved triaging and better coordination. This, along with AOM's outreach programming, should result in more youth accessing services and reduced levels of unmet needs. When needs do arise, they are at earlier stages so can be met with lower intensity and less costly service settings.

will collaboratively conduct data analyses. All access to data will be *via* secure platform/systems; data and files will be destroyed upon termination of site-specific data sharing agreements. Figure 2 illustrates the data linkage process.

Study reporting

Each site's studies will be reported in separate manuscripts. The results will be developed in accordance with the Consolidated Health Economic Evaluation Reporting Standards (22), and a CHEERS 2022 checklist will accompany the site-specific reports.

Ethical considerations

The study has been approved centrally by the Douglas Research Centre's ethics review board, as well as by the designated ethics boards at each site. For Edmonton, ethics approval has been granted by the University of Alberta's research ethics board. In the case of Eskasoni and Chatham-Kent, since site data were to be linked with data held in provincial registries, ethics review took place both at the site level [Nova Scotia Mi'kmaw Client Linkage Registry (MCLR) Data Management Committee and Mi'kmaw Ethics

TABLE 1 Summary of relevant information for three sites undertaking economic evaluations for ACCESS Open Minds.

Exposed population and comparators	AOM intervention* start/end dates	Service utilization	Costs	Study design and key elements	Data sources (location)	Sensitivity analysis
Eskasoni First Nation, NS Exposed: Youth aged 11–25 years Historical Control: EMHS users from January 1, 2012 to July 20, 2016 Parallel control: non-EMHS users from January 1, 2012 to December 31, 2020	July 20, 2016 to December 31, 2020	• Number of referrals seen at site • Number of visits at site • Number of ER visits • Number of hospital admissions • Number of inpatient days • Number of outpatient psychiatry visits and services • Number of non-psychiatry visits	• Total cost of AOM implementation • Total cost of hospital admissions • Total cost of ER visits • Total cost of physicians visits	ROI (costs generated by the intervention will be compared to costs under control condition)	• Eskasoni Mental Health Services (local site) • Mi'kmaw Client Linkage Registry data (Medavie Blue Cross) • Health Data Nova Scotia linked datasets: DAD, MED^{+}, NARCS, MASTER^{++} (Provincial)	• Pre-post parallel trend assumption will be evaluated by examining the interaction between time and intervention • Time horizon over which the difference-in-differences are calculated will be varied • Analyses will be reconducted with inclusion of a washout period
Chatham-Kent, ON Exposed: Youth aged 11–25 years residing in Chatham-Kent from October 2016 to March 2020 Historical Control: Youth in Chatham-Kent catchment from October 1, 2012 to September 30, 2016 Parallel control: Youth in Sarnia catchment from October 1, 2012 to March 17, 2020	October 1, 2016 to March 17, 2020	• Number of referrals seen at site • Number of visits at site • Number of ER visits • Number of hospital admissions • Number of inpatient days • Number of outpatient psychiatry visits and services • Number of non-psychiatry visits covered under OHIP	• Total cost of AOM implementation and CMHA services • Total cost of hospital admissions • Total cost of ER visits • Total cost of physicians visits • Total cost of medications	ROI (costs generated by the intervention will be compared to costs under control condition) Time Horizon: no limit, repeated cross-sections of 6 months between October 1, 2012 and March 17, 2020 Washout period: 6 months before/after October 1, 2016	• Canadian Mental Health Association-Chatham-Kent (local site) • ICES linked datasets for cost analysis: ESTSOB, CCRS, HCD, DAD, NACRS, NRS, ODB, OHIP, OMHRS, SDS, ADP, CAPE, (provincial) • Additional ICES linked datasets for cohort description: CONTACT, RPDB, CPDB, IPDB, ONMARG, INST	• Pre-post parallel trend assumption will be evaluated by examining the interaction between time and intervention • Models will be re-run after excluding individuals with out-of-catchment service use • Analysis will be reconducted with removal of the washout period
Edmonton, AB AOM users, age 15–25 years Parallel control: Mental health service users from non-AOM community mental health clinics	April 6, 2017 to September 30, 2018	• Number of hospitalizations • Numbers of outpatient visits (ED, clinic, specialist, GP, CMHC) • Prescription drug usage • Residential admissions	• Total cost of AOM implementation • Total cost of hospital admissions • Total cost of ED, outpatient, specialist, GP, CMHC visits • Total cost of residential admissions • Total cost of physicians visits	ROI (costs generated by the intervention will be compared to costs under control condition) Time Horizon: Outcomes and costs were estimated for 1 year from the date of access to the AOM or control service, up to September 30, 2019	• Alberta Health Services (AHS) Mental Health and Addictions patient service data and associated costs • AHS community visit and residential stay data, and unit costs • Alberta Health (AH) hospital discharge data, outpatient visit data using CIHI case mix categories and associated costs • Alberta Health physician service data and Schedule of Medical Benefits • Alberta Health pharmaceutical data and unit costs	• Inclusion of all service types regardless of their statistical significance • Deterministic and probabilistic sensitivity analyses • Analysis will be reconducted with inclusion of a washout period

ADP, Assistive Devices Program; AHCIP, Alberta Health Care Insurance Plan; AHS, Alberta Health Services; AOM, ACCESS Open Minds; CAPE, Client Agency Program Enrolment; CCRS, Continuing Care Reporting System; CMHA LK, Canadian Mental Health Association Lambton Kent; CMHC, Community Mental Health Center; CONTACT, Yearly Health Services Contact; CPDB, Corporate Provider Database; DAD, Discharge Abstract Database; EMHS, Eskasoni Mental Health Services; ER, Emergency Room; ESTSOB, Estimated Schedule of Benefits; HCD, Home Care Database; HDNS, Health Data Nova Scotia; ICES, Institute for Clinical Evaluative Sciences; INST, Information about Ontario health care institutions funded by the Ministry of Health and Long-Term Care (MOHLTC); IPDB, ICES Physician Database; MCLR, Nova Scotia Mi'kmaw Client Linkage Registry; MHS, Mental Health Services; NACRS, National Ambulatory Care Reporting System; NRS, National Rehabilitation Reporting System; ODB, Ontario Drug Benefit Claims; OHIP, Ontario Health Insurance Plan Claims Database; OMHRS, Ontario Mental Health Reporting System; ONMARG, Ontario Marginalization Index; PIN, Pharmaceutical Information Network; ROI, Return on Investment; RPDB, Registered Persons Database; SDS, Same Day Surgery Database. *ACCESS Open Minds is the intervention in all three sites (reference)- the start and end dates reflect the economic evaluation, not necessarily the main AOM project. $^{+}$ MED, MSI Physician's Billings. $^{++}$ MASTER, Insured Patient Registry. Box 1 Study interventions.

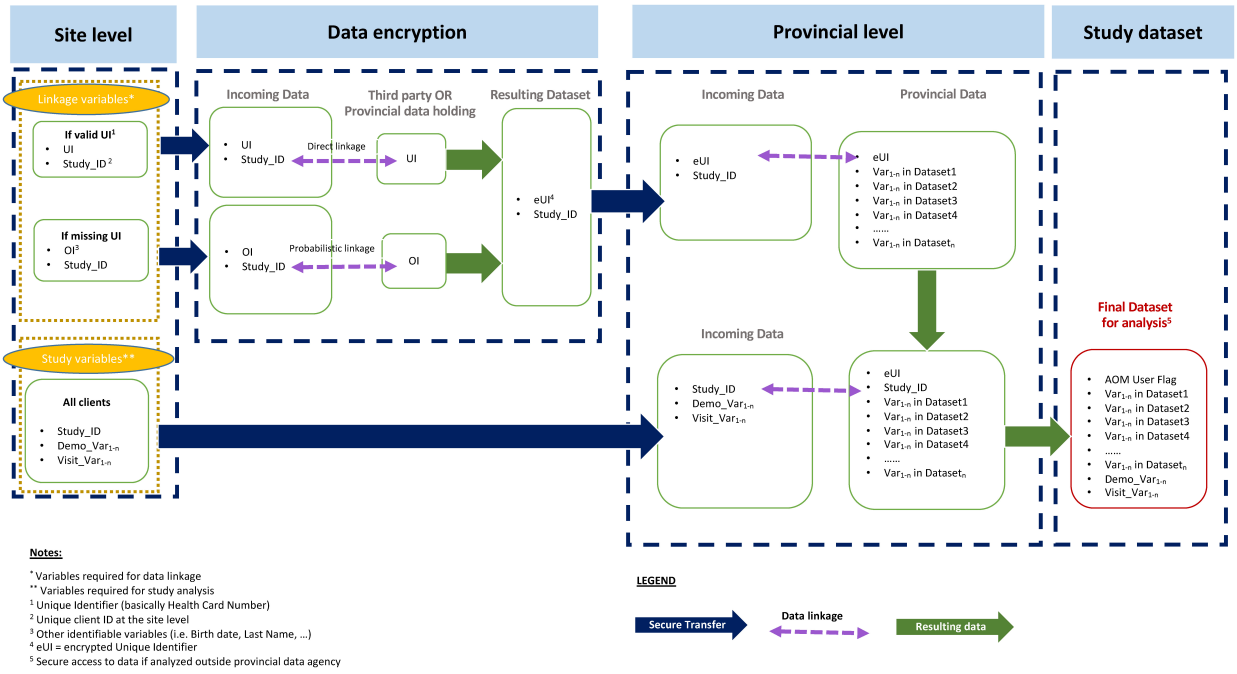


FIGURE 2
Data linkage, encryption, and transfer process. Site-level data will be sent to a responsible provincial department to encrypt identifiers via direct or probabilistic linkage. Following encryption, this will be linked to provincial health administrative data to create a final, combined dataset for analyses.

Watch and Chatham-Kent Research Ethics Board, respectively] as well as privacy assessments at the provincial level (HDNS and Ontario's ICES, respectively). The First Nations-led principles of data ownership, control, access and possession (OCAP) (23, 24, 25), as well as the Tri-Council Policy Statement regarding ethical conduct in research involving Indigenous peoples of Canada (26), have been acknowledged and privileged in partnership agreements between the Montréal-based central office and Indigenous sites and communities.

Analyses and anticipated results

Following confirmation that there is growth in the numbers of cases seen at each of the three sites, we will test hypotheses 1 and 2: namely, that there will be increases in the average number of mental health-related outpatient community visits per person, and decreases in the average number of mental health-related emergency and hospital visits per person, during the post-AOM period compared with the pre-AOM period. For hypothesis 3, a return on investment analysis will be conducted for which the costs generated by the intervention are compared in monetary terms to the costs in the absence of the intervention. At each site, net costs will be calculated separately for each outcome as the costs of services under AOM (including its implementation costs) minus the costs of services for the comparator intervention (whether a historical or parallel control).

Difference-in-differences (DID) approaches help researchers to control for unobserved biases or secular trends; any remaining difference between group-specific differences can be interpreted as likely to reflect (at least in part) the causal effect under investigation. In Eskasoni, changes in utilization of acute, emergency, hospital and specialist services and associated costs will be compared before and after AOM in two groups: those who did and those who did not receive local (AOM) mental health services. For sites where a parallel control exists, we will employ DID analyses that capture both the changes in costs and service utilization between the two periods, as well as the difference between those changes. For example, in Chatham-Kent, changes in acute, emergency, hospital, specialist as well as CMHA/AOM and primary care services will be examined for all youth in the region before and after AOM began, and compared with the same in Sarnia. For Edmonton, the costs and provision of similar services as well as prescription drug usage and residential admissions received by individuals before and after the AOM start date will be compared for two groups: those attending the AOM site and those attending the comparison community mental health clinics. The resulting data inputs into a return on investment calculation to ascertain the extent of net savings or expenditures due to the intervention.

To reduce bias due to differences in demographic and clinical characteristics between the intervention and comparator groups, we will apply adjusted regression models or propensity score matching techniques as needed. Where possible and appropriate, sensitivity analyses will be performed (see Table 1).

Dissemination plan

As mentioned, the project plan has been co-designed and executed in partnership with the sites themselves, ensuring that the knowledge generated will be meaningful and salient for local decision-makers and advocacy (27). It has already been disseminated to various stakeholder groups *via* the AOM website as well as through an extensive series of user-friendly graphics and reports, slide shows, and charts for youth, family, service providers, policy makers, and others. Similar accessible, engaging knowledge translation strategies will be employed once results are available and chosen in partnership

with each site and other key stakeholders in AOM such as its national youth, family and executive councils to ensure uptake and translatability of our findings. Once available, analyses will be added to these materials for scientific conferences and further dialogue with policy-makers.

Peer-reviewed journal publications will also be created for scientific audiences. In all cases, ownership, control, access and possession (OCAP) principles will take precedence in dissemination of findings involving Indigenous communities (23, 28). Project authorship guidelines (which prioritize inclusion of co-authors from the community and site) have been formulated by the AOM national publications committee, and are available upon request.

Discussion

Along with the main AOM study, its economic evaluations will manifest as three return on investment analyses. They will inform the extent to which YMH transformations that reduce barriers and improve access can also shift service provision away from relatively intensive and expensive care, and toward primary and community-based care that is also more resource-efficient. They will consider changes in health service utilization as well as associated costs, hypothesizing that the additional expenditures associated with implementation of new community-based care models (including the setup and operating cost of the transformed AOM site) will be at least partially offset by shifting care upstream with corresponding reductions in emergency, hospital, and other more resource-intensive services.

In our model, it is hoped that the service transformations will provide accessible and appropriate care at earlier stages of illness, thereby avoiding or reducing the risk of developing more severe conditions. Future work might complement the current studies by extrapolating longer-term consequences using economic modeling (which could provide disability- or quality-adjusted life year estimates if required) to complement our empirically measured service and cost metrics with cost-utility analyses. Of course, this will require conceptual advances, such as consensus around definitions and measurement of stage of illness (29); as well as substantial resourcing to scale up YMH services such that access is much improved across entire communities or regions, with data collected longitudinally over the course of routine clinical care. Finally, any reduction in development of late-stage mental illness due to care provision at earlier stages might yield additional benefits for education, justice, or social care. Capturing this would be greatly facilitated by the availability of linked datasets across jurisdictions and ministries.

Our approach to economic evaluations of YMH transformation is notable in its attempts to assess the effects of improved access to care and its desire to include a community-wide focus where possible. Previous economic evaluations of mental health interventions have often examined individual-level metrics such as quality- or disability-adjusted life years under a proposed intervention, compared with treatment as usual and often using a randomized design. While this would have been theoretically possible for AOM, it would be difficult to implement in practice for multiple reasons (30). First, the interventions integrated into AOM are consistent with existing best practices rather than experimental; a control condition in which some subjects were exposed to sub-standard care (or no formal services whatsoever) would not be ethically defensible. Second, a study in which individuals within a site were randomized to treatment arms would be unable to capture the community-level effect of improved access to care. The intensive, broad focus of the transformation means that its effects are unlikely to be specific or limited to AOM itself: the transformation has already been documented as having spillover effects on capacity and other outcomes (19, 31). Finally, the complexity of the main AOM study meant that additional data specifically for an economic evaluation (such as DALY- or QALY-based data) would be difficult to collect in a representative or comprehensive manner compared to secondary use of routine data collection. Instead, capturing changes in service provision (and the resulting costs) can be accomplished using a combination of site and administrative data.

Given recent and forthcoming investments and policy commitments to YMH both in Canada and globally, it is surprising that there are few if any economic evaluations of broad YMH service transformations, especially those that are inclusive of conditions that do not meet DSM/ICD threshold level criteria. In addition to this, AOM's economic evaluations will yield data across diverse contexts, including both urban as well as rural/remote and–critically–Indigenous communities whose youth have generally been neglected in service reform efforts. Together, the breadth of these contexts along with their tailored outcomes and data collection protocols should strengthen the generalizability of our findings, enabling sites to better advocate for sustainability and substantiating the benefits of the AOM transformation and network. Our project will also generate valuable insights on how to co-design, implement and disseminate economic evaluations with diverse stakeholders and community involvement.

The nature of the described economic evaluation does have its limitations. For example, the fact that we will evaluate the site model in its entirety with respect to changes in service utilization and associated costs means that it will be challenging to draw conclusions about which specific aspect(s) of the intervention are driving any observed shifts in care or cost. That said, our inclusion of comparison groups in each of the Eskasoni, Edmonton, and Chatham-Kent studies can (in different ways) account at least in part for unobserved biases and secular trends. Second, the timespan for the return on investment analysis is not the same as the timespan for the AOM intervention: while each site's transformation was assigned a discrete start date at which point the "AOM phase" of the economic evaluation also began, the momentum for transformation started before this and continued to evolve beyond the economic evaluation's end date. Thus, the long-term economic implications of these transformations cannot be depicted or understood within the scope of the current project. Indeed, beginning in March 2020 the COVID-19 pandemic wrought dramatic changes in service delivery and context which cannot be fully captured here.

Conclusion

With growing recognition of the large burden of unmet need in YMH, evaluation and implementation studies have increasingly considered shifts in care provision as a core metric of success. The AOM economic evaluations are designed to integrate an analysis of service utilization with an assessment of costs and the return on investment, furthering community-oriented research in YMH across a range of Indigenous, semi-urban, and urban settings across Canada. In doing so, the project's outcomes will be well poised to inform practice and to support decision-making around the future structure and function of YMH service transformations.

Data availability statement

The original contributions presented in this study are included in this article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JLS was involved in conceptualizing and coordinating the economic evaluations of ACCESS Open Minds, drafting the manuscript, and subsequent edits. ZM provided coordination during the design of the economic evaluations and commented on initial drafts of the manuscript. KA, PJ, and TN supported the design of the economic evaluation, will be involved in data analysis, and reviewed early drafts of the current manuscript. HG, SM, PR-Z, HR, RS, and SS have supported the design by identifying and developing relevant datasets and reviewed drafts of the current manuscript. AM and SI have led the overall AOM project and reviewed early drafts of the current manuscript. EL was involved in conceptualizing and coordinating the economic evaluations of ACCESS Open Minds, and reviewed drafts of the current manuscript. All authors contributed to the article and approved the submitted version.

Funding

ACCESS Open Minds is funded as a Strategy for Patient-Oriented Research (SPOR) of the Canadian Institutes of Health Research (CIHR), with joint funding from CIHR and the Graham Boeckh Foundation. The funders play no direct role in the design of the study or the collection, analysis, and interpretation of the data.

Acknowledgments

We thank Jason Madan for valuable suggestions regarding previous drafts of this manuscript, and Catherine Lau for assistance with **Figure 1**. We thank the following individuals for their contributions to AOM sites and their economic evaluations: Aaron Horvath and Alex Luby (Chatham-Kent); Adam Abba-Aji, Katherine Hay, Jill Kelland, and Christine Mummery (Edmonton); Karlee Francis, Stacey Lewis, and Jenika Macdonald (Eskasoni First Nation). In addition, we acknowledge the valuable contributions of the many additional ACCESS Open Minds Network members, not listed as authors on this manuscript, who have collaborated on this project and whose extensive efforts have contributed to its quality and spirit. The network includes a vibrant National Youth Council, a Family and Carers Council and an Indigenous Council. The corresponding author can be contacted for a list of all network members.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Kessler R, Berglund P, Demler O, Jin R, Merikangas K, Walters E. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Arch Gen Psychiatry.* (2005) 62:593–602. doi: 10.1001/archpsyc.62.6.593

2. Copeland W, Wolke D, Shanahan L, Costello E. Adult functional outcomes of common childhood psychiatric problems: a prospective, longitudinal study. *JAMA Psychiatry.* (2015) 72:892–9. doi: 10.1001/jamapsychiatry.2015.0730

3. Shah JL, Jones N, van Os J, McGorry PD, Gülöksüz S. Early intervention service systems for youth mental health: integrating pluripotentiality, clinical staging, and transdiagnostic lessons from early psychosis. *Lancet Psychiatry.* (2022) 9:413–22. doi: 10.1016/S2215-0366(21)00467-3

4. Bloom D, Cafiero E, Jané-Llopis E, Abrahams-Gessel S, Bloom L, Fathima S, et al. *The Global Economic Burden of Noncommunicable Diseases.* PGDA Working Papers. (2012). Available online at: https://ideas.repec.org//p/gdm/wpaper/8712.html (accessed August 24, 2022).

5. Erskine H, Moffitt T, Copeland W, Costello E, Ferrari A, Patton G, et al. A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. *Psychol Med.* (2015) 45:1551–63. doi: 10.1017/S0033291714002888

6. Malla A, Shah J, Iyer S, Boksa P, Joober R, Andersson N, et al. Youth mental health should be a top priority for health care in Canada. *Can J Psychiatry.* (2018) 63:216–22. doi: 10.1177/0706743718758968

7. Malla A, Iyer S, Shah J, Joober R, Boksa P, Lal S, et al. Canadian response to need for transformation of youth mental health services: access open minds (Esprits ouverts). *Early Interv Psychiatry.* (2019) 13:697–706. doi: 10.1111/eip.12772

8. Iyer S, Boksa P, Lal S, Shah J, Marandola G, Jordan G, et al. Transforming youth mental health: a Canadian perspective. *Ir J Psychol Med.* (2015) 32:51–60. doi: 10.1017/ipm.2014.89

9. Iyer S, Shah J, Boksa P, Lal S, Joober R, Andersson N, et al. A minimum evaluation protocol and stepped-wedge cluster randomized trial of ACCESS open minds, a large Canadian youth mental health services transformation project. *BMC Psychiatry.* (2019) 19:273. doi: 10.1186/s12888-019-2232-2

10. Tsiachristas A, Stein K, Evers S, Rutten-van Mölken M. Performing economic evaluation of integrated care: highway to hell or stairway to heaven? *Int J Integr Care.* (2016) 16:3. doi: 10.5334/ijic.2472

11. Meacock R. Methods for the economic evaluation of changes to the organisation and delivery of health services: principal challenges and recommendations. *Health Econ Policy Law.* (2019) 14:119–34. doi: 10.1017/S1744133118000063

12. Madan J, Bruce Kumar M, Taegtmeyer M, Barasa E, Singh SP. SEEP-CI: a structured economic evaluation process for complex health system interventions. *Int J Environ Res Public Health.* (2020) 17:6780. doi: 10.3390/ijerph17186780

13. Sutton M, Garfield-Birkbeck S, Martin G, Meacock R, Morris S, Sculpher M, et al. *Economic Analysis of Service and Delivery Interventions in Health Care.* Southampton: NIHR Journals Library (2018).

14. Campbell M, Fitzpatrick R, Haines A, Kinmonth A, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ.* (2000) 321:694–6.

15. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new medical research council guidance. *Int J Nurs Stud.* (2013) 50:587–92. doi: 10.1016/j.ijnurstu.2012.09.010

16. Access Open Minds. *Early Identification: An Evidence-Based Practical Guide. (English and French)*. Montreal, QC: ACCESS Open Minds (2017).

17. Canadian Psychiatric Association. *Wait Time Benchmarks for Patients with Serious Psychiatric Illnesses*. Ottawa, ON: Canadian Psychiatric Association (2006).

18. Hutt-MacLeod D, Rudderham H, Sylliboy A, Sylliboy-Denny M, Liebenberg L, Denny J, et al. Eskasoni first nation's transformation of youth mental healthcare: partnership between a Mi'kmaq community and the ACCESS open minds research project in implementing innovative practice and service evaluation. *Early Interv Psychiatry.* (2019) 13(Suppl. 1):42–7. doi: 10.1111/eip.12817

19. Reaume-Zimmer P, Chandrasena R, Malla A, Joober R, Boksa P, Shah J, et al. Transforming youth mental health care in a semi-urban and rural region of Canada: a service description of ACCESS open minds Chatham-Kent. *Early Interv Psychiatry.* (2019) 13(Suppl. 1):48–55. doi: 10.1111/eip.12818

20. Abba-Aji A, Hay K, Kelland J, Mummery C, Urichuk L, Gerdes C, et al. Transforming youth mental health services in a large urban centre: access open minds Edmonton. *Early Interv Psychiatry.* (2019) 13(Suppl. 1):14–9. doi: 10.1111/eip.12813

21. Canadian Institute for Health Information. *CMG+*. (2022). Available online at: https://www.cihi.ca/en/cmg (accessed August 27, 2022).

22. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs A, Carswell C, et al. Consolidated health economic evaluation reporting standards (CHEERS) 2022 explanation and elaboration: a report of the ISPOR CHEERS II good practices task force. *Value Health.* (2022) 25:10–31. doi: 10.1016/j.jval.2021.10.008

23. Mecredy, G, Sutherland R, Jones C. First Nations data governance, privacy, and the importance of the OCAP® principles. *Int J Popul Data Sci.* (2018) 3. doi: 10.23889/ijpds.v3i4.911

24. National Aboriginal Health Organization [NAHO], First Nations Centre. *OCAP: Ownership, Control, Access and Possession: Sanctioned by the First Nations Information Governance Committee*. Ottawa, ON: National Aboriginal Health Organization [NAHO] (2007).

25. The First Nations Information Governance Centre. *The First Nations Principles of OCAP®*. (2020). Available: https://fnigc.ca/ocap-training/ (accessed August 24, 2022).

26. Hyett S, Marjerrison S, Gabel C. Improving health research among indigenous peoples in Canada. *CMAJ.* (2018) 190:E616–21. doi: 10.1503/cmaj.171538

27. Iyer S, Boksa P, Joober R. Editorial: how youth mental healthcare is being transformed in diverse settings across Canada: reflections on the experience of the ACCESS open minds network. *Early Interv Psychiatry.* (2019) 13(Suppl. 1):8–11. doi: 10.1111/eip.12811

28. Bharadwaj L. A framework for building research partnerships with first nations communities. *Environ Health Insights.* (2014) 8:15–25.

29. Shah J, Scott J, McGorry P, Cross S, Keshavan M, Nelson B, et al. Transdiagnostic clinical staging in youth mental health: a first international consensus statement. *World Psychiatry.* (2020) 19:233–42. doi: 10.1002/wps.20745

30. Ratnasingham S, Cairney J, Manson H, Rehm J, Lin E, Kurdyak P. The burden of mental illness and addiction in Ontario. *Can J Psychiatry.* (2013) 58:529–37.

31. Vallianatos H, Friese K, Perez J, Slessor J, Thind R, Dunn J, et al. ACCESS open minds at the university of Alberta: transforming student mental health services in a large Canadian post-secondary educational institution. *Early Interv Psychiatry.* (2019) 13(Suppl. 1):56–64. doi: 10.1111/eip.12819

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 02 June 2023
DOI 10.3389/fpsyt.2023.1031159

OPEN ACCESS

EDITED BY
Huajie Jin,
King's College London, United Kingdom

REVIEWED BY
Yuhua Bao,
Weill Cornell Medicine, United States
Min Peng,
King's College London, United Kingdom

*CORRESPONDENCE
Marie Le Novere
✉ m.novere@ucl.ac.uk

RECEIVED 29 August 2022
ACCEPTED 27 April 2023
PUBLISHED 02 June 2023

CITATION
Le Novere M, Johnson S, Lloyd-Evans B, Marston L, Ambler G, Clarke CS, Osborn D, Lamb D and Hunter RM (2023) Cost-effectiveness of peer-supported self-management for people discharged from a mental health crisis team: methodological challenges and recommendations. *Front. Psychiatry* 14:1031159. doi: 10.3389/fpsyt.2023.1031159

Cost-effectiveness of peer-supported self-management for people discharged from a mental health crisis team: methodological challenges and recommendations

Marie Le Novere[1]*, Sonia Johnson[2], Brynmor Lloyd-Evans[2], Louise Marston[1], Gareth Ambler[3], Caroline S. Clarke[1], David Osborn[2], Danielle Lamb[4] and Rachael Maree Hunter[4]

[1]Research Department of Primary Care and Population Health, Faculty of Population Health Sciences, University College London, London, United Kingdom, [2]Divison of Psychiatry, Faculty of Brain Sciences, University College London, London, United Kingdom, [3]Department of Statistical Science, Faculty of Mathematical & Physical Sciences, University College London, London, United Kingdom, [4]Department of Applied Health Research, Faculty of Population Health Sciences, University College London, London, United Kingdom

Background: Mental health acute crisis episodes are associated with high inpatient costs. Self-management interventions may reduce readmission by enabling individuals to manage their condition. Delivery of such interventions by Peer Support Workers (PSWs) may be cost-effective. CORE, a randomized control trial of a PSW self-management intervention compared to usual care, found a significant reduction in admissions to acute mental healthcare for participants receiving the intervention. This paper aims to evaluate the cost-effectiveness of the intervention over 12 months from a mental health service perspective. Analysis methods of increasing complexity were used to account for data missingness and distribution.

Methods: Participants were recruited from six crisis resolution teams in England from 12 March 2014 to 3 July 2015 (trial registration ISRCTN: 01027104). Resource use was collected from patient records at baseline and 12 months. The EQ-5D-3L was collected at baseline and 4 and 18 months, and linear interpolation was used to calculate 12-month values for quality-adjusted life-years (QALYs). The primary analysis of adjusted mean incremental costs and QALYs for complete cases are calculated separately using OLS regression. Secondly, a complete-case non-parametric two-stage bootstrap (TSB) was performed. The impacts of missing data and skewed cost data were explored using multiple imputation using chained equations and general linear models, respectively.

Results: Four hundred and forty-one participants were recruited to CORE; 221 randomized to the PSW intervention and 220 to usual care plus workbook. The probability that the PSW intervention was cost-effective compared with the workbook plus usual care control at 12 months varied with the method used, and ranged from 57% to 96% at a cost-effectiveness threshold of £20,000 per QALY gained.

Discussion: There was a minimum 57% chance that the intervention was cost-effective compared to the control using 12-month costs and QALYs. The

10.3389/fpsyt.2023.1031159

probability varied by 40% when methods were employed to account for the relationship between costs and QALYs, but which restricted the sample to those who provided both complete cost and utility data. Caution should therefore be applied when selecting methods for the evaluation of healthcare interventions that aim to increase precision but may introduce bias if missing data are heavily unbalanced between costs and outcomes.

KEYWORDS

crisis resolution teams, peer-support, cost-effectiveness, quality adjusted life years, economic evaluation, EQ-5D, mental health

1. Introduction

Between 1998 and 2012, the number of psychiatric beds in England fell by 39%, shifting activity away from acute services and toward care focused on recovery and self-management for those going through an acute crisis episode (1, 2). Crisis Resolution Teams (CRTs) were introduced in England with the aim of encouraging early discharge from hospital or providing intensive home treatment when possible (3). Evidence suggested that these have been successful in reducing hospital admissions and in turn reducing health service costs (4–6). However, more recent evidence has found that CRTs' service delivery and organization varies and model fidelity is not high, both in the UK (7) and internationally (8). Naturalistic studies suggest they may not consistently have an impact on hospital admissions to an acute mental health ward (9, 10). This may be related to high relapse rates given around 50% of patients are readmitted to acute care within 1 year of contact with a CRT (11). Self-management interventions, which aim to educate and empower individuals to control or reduce the impact of their condition (12), may be useful in reducing readmission to acute care by enabling individuals to keep the severity of their condition in check following discharge from a CRT. There is evidence to suggest that the delivery of such interventions by Peer Support Workers (PSWs) may be cost-effective (13). PSWs are individuals who have shared experiences with the patients, facilitating their ability to provide support and mentorship to those receiving the intervention (14–16). Studies have found that the benefits of employing PSWs, such as reduction in hospital admission to an acute mental health ward and improvement in other aspects of patients' lives such as social functioning (17–19), outweighed the costs of employing PSWs (13, 18). PSWs are increasingly commonly employed within the English National Health Service (NHS) mental health services and internationally and are advocated in the mental health implementation guidance for the NHS Long term Plan (20–22). The findings from the CORE trial (23) provide evidence to support this approach, demonstrating significant reduction in admissions to acute mental healthcare for participants receiving the CORE peer-supported self-management intervention compared with the control. To our knowledge, the cost-effectiveness of such an intervention following a mental health crisis has not previously been tested. We therefore carried out an economic evaluation alongside the clinical trial with the aim to calculate the probability that the CORE peer-provided self-management intervention was cost-effective compared to control. The control was Treatment as Usual (TAU) accompanied by a self-management workbook without guidance on how to use it.

Recruitment, retention and follow-up are known issues in clinical trials; loss to follow up may occur if the participant's state of health, particularly mental health worsens, and they are no longer able to engage with the trial. These issues can be more pronounced in mental health trials, especially those involving complex interventions, where the participant commonly knows if they have been randomized to the intervention or control. Participants randomized to control may lose interest once they know they will not immediately receive the trial intervention (24).

Economic evaluations alongside clinical trials often face a high level of missing cost data due to their reliance on self-reported measures such as the Client Service Receipt Inventory (CSRI) to collect resource use information (25), that ask participants or carers to recall what appointments and other treatments they have had. Trial participants who are missing this type of outcome data may be systematically different from participants with complete data, so to simply ignore the missing data potentially introduces bias. As a result, different methods have been explored in order to minimize missing cost data in economic evaluations (26), including using electronic healthcare records to supplement or replace self-completed questionnaires, and in this study resource use information is collected using medical records from mental health Trusts. While the use of electronic healthcare records has some shortfalls in terms of scope, it reduces the risk of missing data caused by illness, disengagement with the trial, patient recall and questionnaire design (27). Instead, there may now be more missing data on the self-reported health-related quality of life side of the equation, which may affect the interpretation of the results in a different way to missing data on the cost side.

The aim of this paper is to report the 12- month cost-effectiveness of CORE, a peer-provided self-management intervention, compared with the control, where data were collected over 12 months for resource use and 18 months for health-related quality of life. The analysis used data from patient medical records for healthcare resource use in addition to self-completed questionnaires for health-related quality of life to calculate utilities and quality adjusted life years (QALYs). Medical records are considered to be relatively complete, whereas self-completed questionnaires are subject to a larger quantity of missing data. This imbalance in data completeness between costs and outcomes leads to methodological challenges which must be addressed in order to achieve our aim. As a result, in this paper we explore the differential

impact of economic evaluation methods of increasing complexity to account for missing data. We also explore the impact of accounting for resource use skew, which, although always present in economic evaluations, is particularly marked in acute crisis care due to the high use of expensive inpatient care.

2. Methods

2.1. Sample

Participants were identified from caseloads from CRTs in six NHS Trusts in London, South East and South West England from 12 March 2014 to 3 July 2015. Participants were recruited after they were discharged from the CRT and were eligible if they had been on the caseload for at least a week because of a crisis. More detail on the eligibility and exclusion criteria is available elsewhere (23). The study included an internal pilot in which 40 participants were recruited (23).

2.2. Treatment offered

Participants and care providers were not blinded but neither were they informed of the participants' allocation until after they had been discharged from the CRT, to minimize any impact on discharge planning from trial participation. Those in the treatment group were given a personal recovery workbook and offered up to 10 sessions with a PSW, aimed to be completed within 4 months, to support them in the completion of the workbook in addition to usual care. For a more detailed description of the intervention components please see Johnson et al. (23). Those in the control group received usual care and the workbook by post only, without additional guidance.

2.3. Measures

2.3.1. EQ-5D-3L

The EQ-5D-3L (28) was collected at baseline and 4 months initially. During the trial, additional funding was received to add a follow-up point for the self-completed questionnaire at 18 months, so EQ-5D-3L was also collected at this point. The formula developed by Dolan (29) and the area under the curve method were used to calculate QALYs for each group from baseline to 4 and 18 months (30). For participants who died during the trial their utility was assessed as 0 at the date of death and a straight line was assumed from their last completed EQ-5D-3L to the time of death. To calculate the mean difference in QALYs and 95% confidence intervals between the intervention group and control, a regression with 5,000 bootstrapped replications was used controlling for group, baseline EQ-5D-3L utility score and clustering by peer support worker (30). For the 18-month analysis, a discount rate of 3.5% was used to discount QALYs from 12 to 18 months in line with NICE guidance (31).

To match the QALY follow-up duration with resource use data collected from clinical records, QALYs were calculated over 12 months using linear interpolation, a straight line between the 4- and 18-month follow-up points, with the value on that line at 12 months assumed to be the utility value that would have occurred at 12 months.

2.4. Service utilization and costs

2.4.1. Cost of mental health service use

Acute and community mental health service use for both groups was collected at baseline and 12 months for the previous 12 months from electronic patient records held by mental health Trusts. Unit costs were gathered from published sources including the Personal Social Services Research Unit (PSSRU) (32) and NHS reference costs (33) to be applied to mental health service use over 12 months. The cost of mental health clusters was estimated based on diagnosis. Mental health clustering is used in the UK to allow patients to be grouped together by severity while still allowing a degree of variation in the combination and severity of needs.

2.4.2. Cost of intervention

The cost of training PSWs and supervision by clinical staff was included in the intervention cost. The hourly cost of an 'Agenda for change' Band 3 staff member (pay bands used by the NHS, example role: emergency care assistant, occupational therapy support worker) from the PSSRU (32) was used for the hourly cost of a PSW. Costing for supervision was varied by grade and frequency (see Supplementary material), with clinical supervision predominately being provided by Band 8a. The cost of the intervention also included PSWs time providing support based on the number of appointments participants had and the duration of appointments. The cost of the workbook is not included in the intervention costs as both groups received it.

A linear regression with 5,000 bootstrap replications, controlling for baseline service use, and clustered by peer support worker, was used to calculate the mean difference in costs between the intervention and control group and 95% confidence intervals.

As costs were reported for baseline and 12 months only, there was no discounting of costs. All costs reported are in 2015/2016 British Pounds.

2.5. Data analysis

The planned primary analysis was a complete-case analysis calculating the incremental cost per QALY gained by dividing the mean difference in costs between the two groups by the mean difference in QALYs found using the linear interpolation for 12-month utility. To account for any potential relationship between costs and QALYs, cost-effectiveness analyses commonly use seemingly unrelated regression (SUR; Stata command SUREG), which account for the relationship through correlated error terms, to calculate mean incremental costs and QALYs (34). This method does not allow for clustering by PSW. Ignoring clustering in randomized trials can lead to biased and incorrect conclusions (35, 36). In the case where a non-pharmaceutical intervention is delivered by multiple health professionals, those participants who are treated by the same health professional may have similarities or be clustered due to differences in the healthcare professionals. This violates the assumption of independence and appropriate statistical methods are needed to account for this (37, 38). As a result, for our original primary analysis, we calculated the mean incremental costs and QALYs using complete-case linear regression controlling for baseline service use and

including clustering for PSW with 5,000 bootstrap replications. Regression analyses for costs and QALYs were run separately.

Other methods for use in cost-effectiveness analyses (CEA) of cluster randomized trials include non-parametric two-stage bootstrap (TSB) (35) which accounts for the relationship between the costs and outcomes by sampling the costs and effects in pairs which maintains the relationship between the two in the bootstrapped results (34). The different methods and their benefits and pitfalls were explored in the context of this analysis considering the high levels of missing data present for QALYs. They are laid out in this paper as follows:

i. The original separate primary regression analyses
ii. Sensitivity analyses including joint analysis of costs and QALYs using TSB
iii. Missing data analyses
iv. Sensitivity analysis around resource use skew

2.5.1. Original separate primary analyses

2.5.1.1. Incremental cost-effectiveness ratio

The planned primary analysis was a complete-case analysis calculating the incremental cost per QALY gained by dividing the mean difference in costs between the two groups by the mean difference in QALYs found using the assumed 12-month utility. 12 month was chosen as the more conservative option given we have costs at this timepoint and utility before and after. This was considered more robust than extrapolating costs to 18 months (cost data only being available up to 12 months). The analysis was also designed to be aligned with the main statistical analysis which was comparing readmission within 1 year between the two groups using a logistic regression (23).

2.5.1.2. Cost-effectiveness plane and cost-effectiveness acceptability curve

A CEP is used to report the bootstrapped 12-month QALYs and 12-month costs. These results are also reported on a CEAC to show the probability that the intervention was cost-effective compared with the control for a range of cost-effectiveness threshold values from £0 to £100,000, with probabilities reported for a £20,000 cost-effectiveness threshold. We also report the probability that the intervention was cost-effective compared with control for this range for:

i. 12-month costs and 18 months QALYs
ii. 12-month costs and 4 months QALYs

The primary 12-month costs and QALYs analysis was repeated using the non-parametric TSB.

2.5.2. Sensitivity analysis

Uncertainty around the following aspects of the analysis were explored in sensitivity analysis using the TSB method:

i. The cost of the intervention (Supplementary material)

The analysis was repeated using supervision and training costs provided by mental health Trusts to calculate the cost of PSWs as well as exploring how the results might change if supervision was weekly rather than fortnightly.

ii. Calculating 12-month utility

In the primary analyses, we assume that trial participants' utility changes in a linear way between timepoints. To test the impact of this assumption, the last values were carried forward using utility at 4 months to impute utility at 12 months and recalculating QALYs at 12 months. We then did the same again but with next value carried backward, i.e., using utility at 18 months to impute utility at 12 months. We present these results on a CEAC alongside the estimated 12-month QALY results.

2.5.3. Missing data analysis

Only 52% of participants have complete data for all time points of the EQ-5D-3L. Given high proportions of missing data can lead to misleading results if not dealt with appropriately, we have followed the process laid out by Faria et al. (39), on how to deal with missing data in within-trial CEAs. The process is broken down into 3 stages: descriptive statistics to inform assumptions on the missing data mechanism, choosing an appropriate method to deal with the missing data for the base-case analysis using these assumptions, and finally, sensitivity analysis to explore how the results change with the asssumptions made. The first stage is to explore the data in order to inform whether the data are likely to be missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR). The classifications of missing data are explained further in Faria et al. (39).

For the data to be MCAR, missing data must be independent of both observed and unobserved characteristics, although covariate dependent missingness (CD-MCAR) occurs when the probability of missingness is dependent on baseline covariates but is independent of the missing and observed outcome. Data can be MAR if missingness can be accounted for using the observed data and the probability of missingness is independent of unobserved characteristics. MNAR occurs when missingness is dependent on unobserved factors, and this may introduce bias if for example, individuals are more likely to have missing data depending on if they have good or bad outcomes.

To determine the type of missing data present, we used logistic regressions to investigate the relationship between observed variables and missingness. Predictors of missingness in 4- and 18-month EQ-5D-3L data included whether participants were in employment and their level of educational attainment. Being in employment and higher levels of educational attainment were associated with lower levels of missing data. This analysis included the main trial only as the wording of questions changed between the pilot and main trial.

We used logistic regression to test if there was a relationship between missingness and previously observed outcomes and found no association between utility score at 4 months and missing utility data at 18 months. This suggests that there was no association with having a worse or better observed outcome at 4 months and likelihood of missing outcome data at 18 months.

When using linear interpolation to calculate 12-month QALYs, there was 48% missing data for QALYs. Multiple imputation using chained equations (MICE) and predictive mean matching was therefore used to impute 4- and 18-month utility data for 48 imputations, stratified by group. The imputed utility scores were then used to calculate imputed 12-month QALYs using linear interpolation (40).

While the descriptive analysis suggested the data can be described as MAR as missingness can be accounted for using the observed data (employment and level of educational attainment), this is never

10.3389/fpsyt.2023.1031159

certain given we cannot observe which unobserved factors we may be missing. As such, to evaluate the uncertainty around this assumption and avoid bias, it is best practice to explore how the results may change if we assume the data are MNAR. Leurent et al. (40) recommend conducting scenario analysis around the imputed values, and as such we apply a utility decrement of varying severity based on whether the participant has been readmitted to acute care. The multiple imputation process was repeated but with a utility decrement weighting applied to the imputed utilities so that the imputed utility was multiplied by 0.9 if the participant had been readmitted to acute care in scenario 2, 0.8 in scenario 3 and 0.7 in scenario 4. Scenario 1 is the MAR scenario where no utility decrement is applied.

2.5.4. Sensitivity analysis around resource use skew

The costs associated with healthcare resource use are often skewed, with a high number of participants accumulating at very low or zero values, and is certainly the case here due to the high costs associated with readmission. Therefore using TSB, we estimate a generalized linear model (GLM) using a gamma distribution to evaluate how accounting for this pattern in resource use costs may impact the cost-effectiveness results using the MICE data set.

All analyses were conducted in Stata 16.

3. Results

3.1. Baseline characteristics

Baseline characteristics are shown in Table 1. Participants are split into those with complete utility data and those missing utility data at one or more time points, to begin investigating whether there are any significant differences between these groups and if this varies between the intervention and control group. There is no evidence to suggest that there are any significant differences between the four groups at baseline.

TABLE 1 Comparison of sample characteristics at baseline.

Characteristic	Complete utility data (*N*=223)		Missing utility at one or more timepoints (*N*=218)	
	Intervention (*N*=107)	Control (*N*=116)	Intervention (*N*=114)	Control (*N*=104)
Male sex: *n* (%)	43 (40)	46 (40)	47 (40)	42 (40)
Age: mean years (SD)	46 (13)	46 (12)	46 (14)	46 (13)
Ethnicity: *n* (%)				
White (UK and non-UK)	63 (59)	77 (66)	80 (70)	62 (60)
Black (UK, African, Caribbean, mixed, and other)	24 (22)	23 (20)	21 (18)	20 (19)
Asian (UK, South Asian, Chinese, mixed, and Other)	8 (7)	7 (6)	7 (6)	6 (6)
Other	12 (11)	7 (6)	6 (5)	13 (13)
UK born	79 (74)	89 (76)	97 (85)	75 (72)
Marital status: *n* (%)				
Single	62 (58)	74 (64)	79 (70)	71 (68)
Married or cohabiting	27 (25)	31 (27)	19 (16)	21 (20)
Separated or divorced	16 (15)	11 (9)	11 (12)	12 (12)
Widowed	2 (2)	0 (0)	5 (4)	0 (0)
Lifetime admissions to psychiatric hospital: *n* (%)				
Never	67 (63)	72 (62)	67 (59)	60 (58)
1	15 (14)	20 (17)	12 (11)	18 (18)
2–5	18 (17)	19 (16)	21 (18)	21 (20)
> 5	7 (7)	5 (4)	14 (12)	5 (5)
Periods of support from crisis resolution teams				
1	58 (54)	54 (47)	53 (46)	48 (46)
2	20 (19)	23 (20)	23 (20)	20 (19)
3–5	20 (19)	24 (21)	27 (24)	26 (25)
6–10	6 (6)	7 (6)	6 (5)	5 (5)
>10	3 (3)	8 (7)	5 (4)	6 (6)

SD: Standard deviation.

3.2. Costs and effects

3.2.1. Cost of the intervention

PSWs are costed at £25 per hour (32). PSW supervision varied in frequency and grade of clinical staff providing the supervision. Supplementary Table 1 shows a comparison of the costs depending on whether supervision was weekly or fortnightly. The most common structure was a fortnightly session with a grade 8 supervisor. Therefore, to calculate the cost per PSW, sessions were assumed to be fortnightly, and the cost was weighted for supervisor seniority. Including overheads, the cost of training and supervision per PSW was £2,548. On average, each PSW was allocated 6.5 participants, which equated to a cost per participant in the intervention group of £392.

Participants on average had 5.8 (95% CI 5.3–6.3) appointments with their PSW. According to the intervention manual, each appointment was scheduled to last an hour, at a cost of £25 per hour of PSW time, the average cost of appointments per patient was £145 (95% CI £131 to £159). The total mean cost per participant of the intervention including training and supervision was £537 (95% CI £523 to £551). The cost of the workbook was not included given both groups received it.

3.2.2. Cost of 12-month mental health service use

Table 2 reports the mean cost of mental health service use at baseline and 12 months for both the intervention group and the control group. The total cost of mental health services at 12 months, adjusting for baseline differences was £6,586 (95% CI: £4,922–£8,249) for the intervention group and £6,605 (95% CI: £4,951–£8,259) for the control group. Including the cost of the intervention and adjusting for baseline, the complete-case mean incremental cost of the intervention group compared with the control group at 12 months was -£261 (95% CI: £2,450–£1928).

3.2.3. QALYs

Mean unadjusted utility scores generated from participant-completed EQ-5D-3L are reported in Table 3. The four participants who died during the trial are included; these were all in the control group. The mean QALYs at 12 months, for which the utility value was taken by drawing a straight line between 4 and 18 months (shown in Supplementary Figure 1), were 0.651 (95% CI 0.612 to 0.689) for the intervention group and 0.640 (95% CI 0.600 to 0.679) for the control group, a mean difference of 0.011 (95% CI: −0.043 to 0.065). The mean QALYs at 18 months, adjusted for baseline and discounted at 3.5% per year after 12 months, were 0.991 (95% CI: 0.931–1.051) for the intervention group and 0.968 (95% CI: 0.907–1.03) for the control group. The mean difference between the two groups was 0.023 (95% CI: −0.062 to 0.107).

3.3. Cost-effectiveness—original primary analysis

The intervention dominates the control group as it results in more QALYs and lower costs, although the differences were not significant. Figure 1 shows the CEP using the 12-month QALYs and 12-month costs from the original analysis. The CEAC in Figure 2 reports the probability of cost-effectiveness at different thresholds using 12-month costs with 4 and 18-month QALYs and 12 months calculated as a linear change between 4-and 18-month QALYs.

At a cost-effectiveness threshold of £20,000 per QALY gained, the probability that the intervention was cost-effective compared to the control was 65% based on 12-month QALYs calculated using linear interpolation. The probability of the intervention being cost-effective compared to control increases as the duration of follow-up increases (see Figure 2). This occurs from a combination of the maximum QALYs achievable increasing with a longer follow-up duration and the difference in utility between the two groups appearing to persist through time. This is in addition to the costs remaining constant as we do not have any costs past 12 months.

3.4. Nonparametric two-stage bootstrap

The results of the TSB are shown in Figures 3, 4, showing the results on a CEP and CEAC, respectively. Using 12-month QALYs calculated using linearly interpolated utility at 12 months, the intervention is 96% cost-effective at a threshold value of £20,000 per QALY. Comparing the results from the CEP in Figure 3 to those in Figure 1, the CEP for the separate regressions, illustrates that this is because, for the TSB, the majority of bootstrap iterations lie in the bottom two quadrants (cost-saving). Despite the apparent advantage provided by the TSB of accounting for the relationship between costs and outcomes by sampling costs and QALYs at the same time, the analysis is potentially biased as it only includes costs for trial participants who have complete utility data ($N = 223/441$), hence missing many individuals. Table 4 shows how costs differ between those with complete and incomplete utility data across the two groups. Those with missing utility data have significantly higher acute care costs at 12 months than those with complete utility data [£5,855 (95% CI: £3,888–£7,822) vs. £1885 (95% CI: £1,045–£2,725); $p < 0.001$].

TABLE 2 Mean costs and 95% CIs for mental healthcare resource use.

		Intervention		Control	
		Baseline	12months	Baseline	12months
Acute care costs	Mean	£6,008	£3,673	£5,351	£4,023
	95% CI	£4,631 to £7,385	£2,156 to £5,220	£3,846 to £6,855	£2,525 to £5,522
Community costs	Mean	£1,740	£2,390	£1,941	£2,581
	95% CI	£1,362 to £2,119	£1,954 to £2,825	£1,478 to £2,405	£2,076 to £3,086
Total	Mean	£7,748	£6,586	£7,292	£6,605
	95% CI	£6,328 to £9,260	£4,923 to £8,949	£5,614 to £8,970	£4,951 to £8,259

3.5. Uncertainty in 12-month estimated QALYs (using TSB)

When the analysis was replicated using utility at 4 months to calculate 12-month QALYs using last value carried forward, the probability that the intervention was cost-effective compared to TAU fell to 85% at a cost-effectiveness threshold of £20,000/QALY gained. The analysis using utility at 18 months to calculate 12-month QALYs using next value carried backward had very similar results to the analysis using linearly interpolated utility at 12 months. This suggests that the results are driven by an improvement in recorded utility at 18 months (Table 3) rather than simply having more QALYs available and hence a larger potential incremental benefit. The CEAC is shown in Figure 5.

TABLE 3 Mean utility scores generated from the EQ-5D-3L and unadjusted 12- and 18-month QALYs. 3.5% discounting for utility scores over 12months.

		Intervention	Control
Baseline	*N*	217	220
	Mean (SD)	0.613 (0.323)	0.595 (0.331)
	N missing (%)	4 (2)	0 (0)
4 months	*N*	173	169
	Mean (SD)	0.670 (0.310)	0.658 (0.328)
	N missing (%)	48 (22)	51 (23)
18 months	*N*	122	124
	Mean (SD)	0.698 (0.331)	0.675 (0.322)
	N missing (%)	99 (45)	96 (44)
12 months QALYs	*N*	107	116
	Mean (SD)	0.664 (0.271)	0.627 (0.308)
	N missing (%)	114 (52)	104 (47)
18 months QALYs	*N*	107	116
	Mean (SD)	1.011 (0.403)	0.950 (0.450)
	N missing (%)	114 (52)	104 (47)

3.6. Missing data analysis

3.6.1. MAR analysis

Following multiple imputation, the mean difference in QALYs between the intervention and control group at 12 months was 0.012 (95% CI: −0.033 to 0.057). The CEP and CEAC were constructed using the TSB following multiple imputation, and are shown in Figures 6, 7. The probability that the intervention was cost-effective compared to the control was 66% at a cost-effectiveness threshold of £20,000/QALY gained.

3.6.2. MNAR sensitivity analysis

The example provided by Leurent et al. (40) experiments with applying different weights to the imputed utility in different scenarios based on the assumption that those with missing utility data may be systematically worse off. This is likely to be the case here given those missing utility data have significantly higher acute care costs indicating they are in worse health than those with complete utility data. However, unlike the example in which they apply different weights to the treatment and control group, we apply a different weight based on whether the participant has been readmitted to acute care. Logistic regression showed that

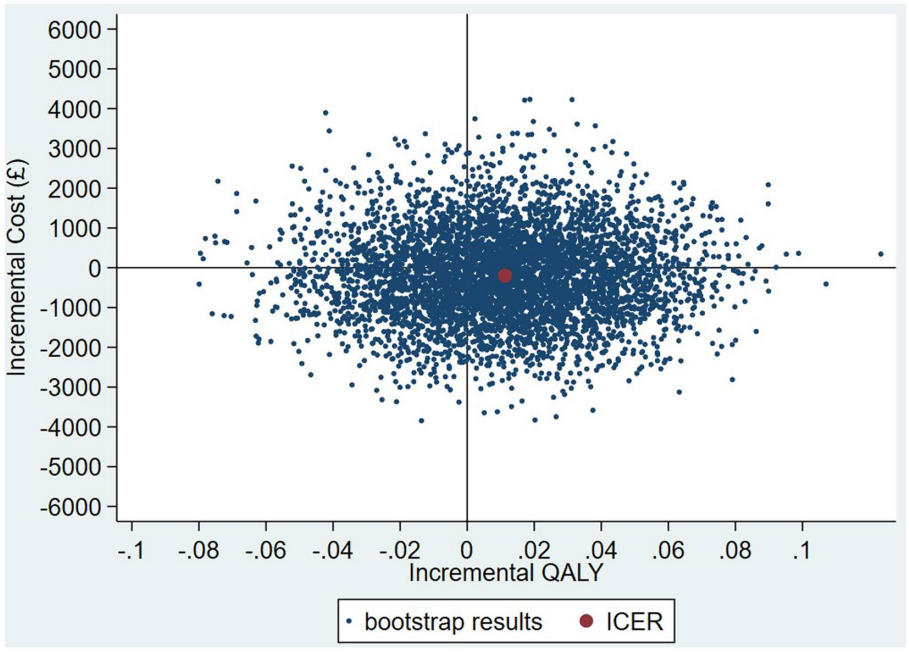


FIGURE 1
Cost-effectiveness plane (CEP) for 12-month QALYs and 12-month costs based on running separate bootstrap regressions for costs and QALYs (N_{costs}=441/441, N_{QALYs}=223/441).

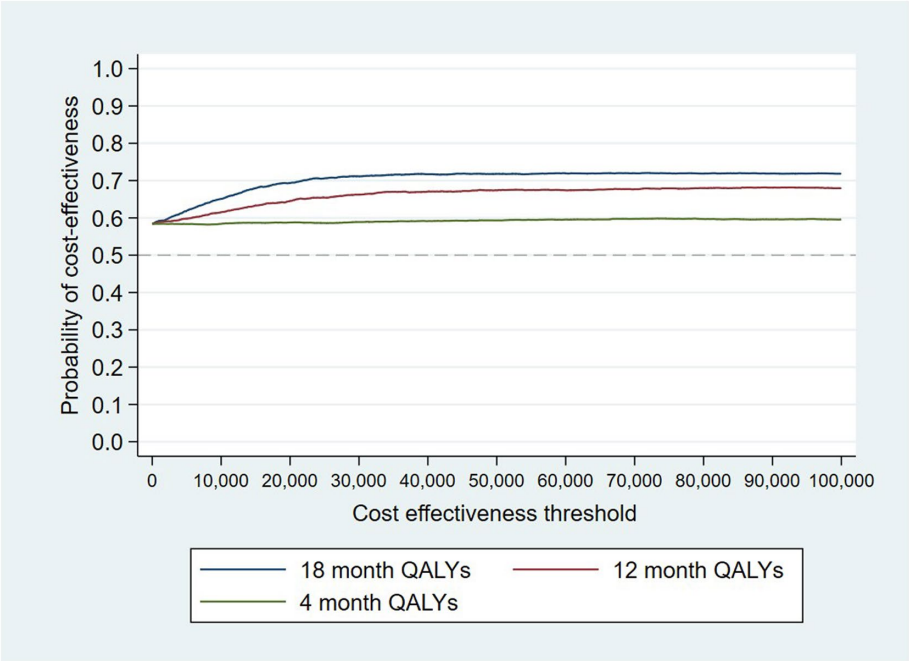


FIGURE 2
Cost-effectiveness acceptability curves (CEACs) for 4-, 12-, and 18-month QALYs based on running separate bootstrap regressions for costs and QALYs (N_{costs}=441/441, N_{QALYs}=223/441).

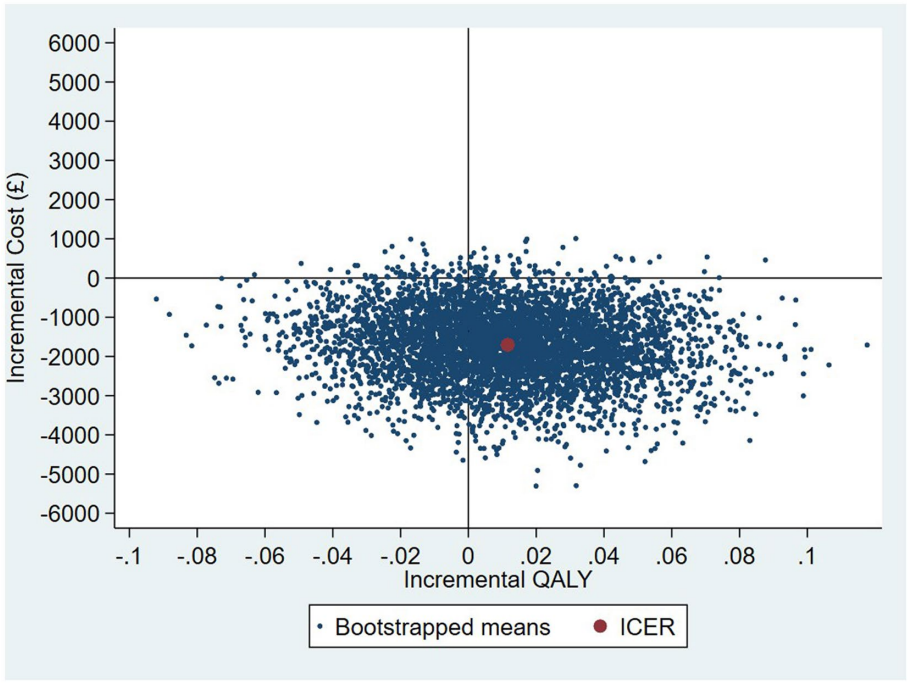


FIGURE 3
CEP for 12-month QALYs (using TSB method, *N*=223/441).

those who were readmitted to acute care within 1 year were 20% more likely to have missing utility data at 4 months and 7% more likely to be missing utility data at 18 months compared with those who were not readmitted. It is plausible that those who were missing utility data and had been readmitted to acute care had a lower health-related quality of life.

The probability of the intervention being cost-effective compared with TAU increased as the utility decrement weighting increased. All

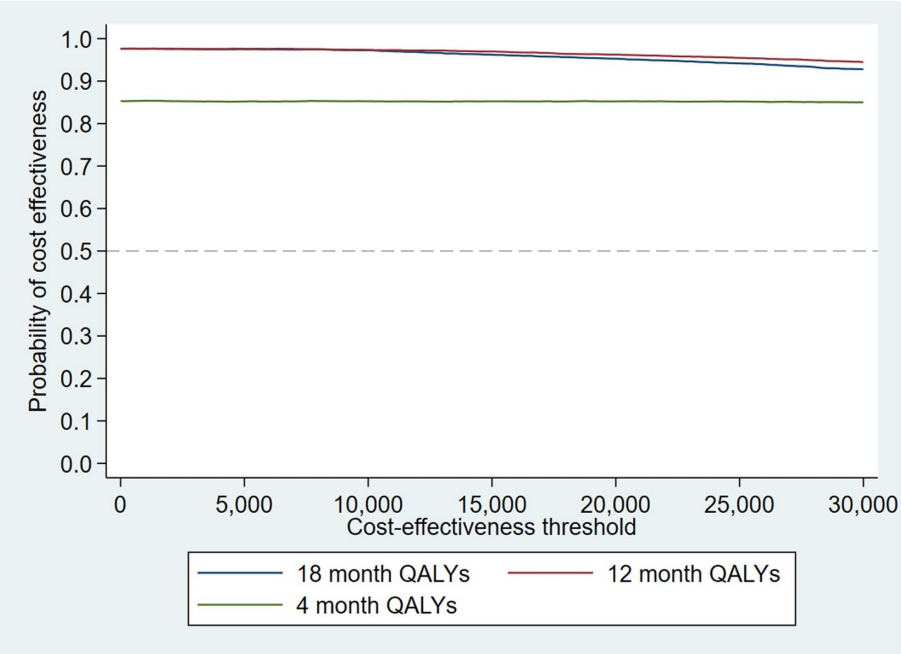


FIGURE 4
CEAC using TSB method (*N*=223/441).

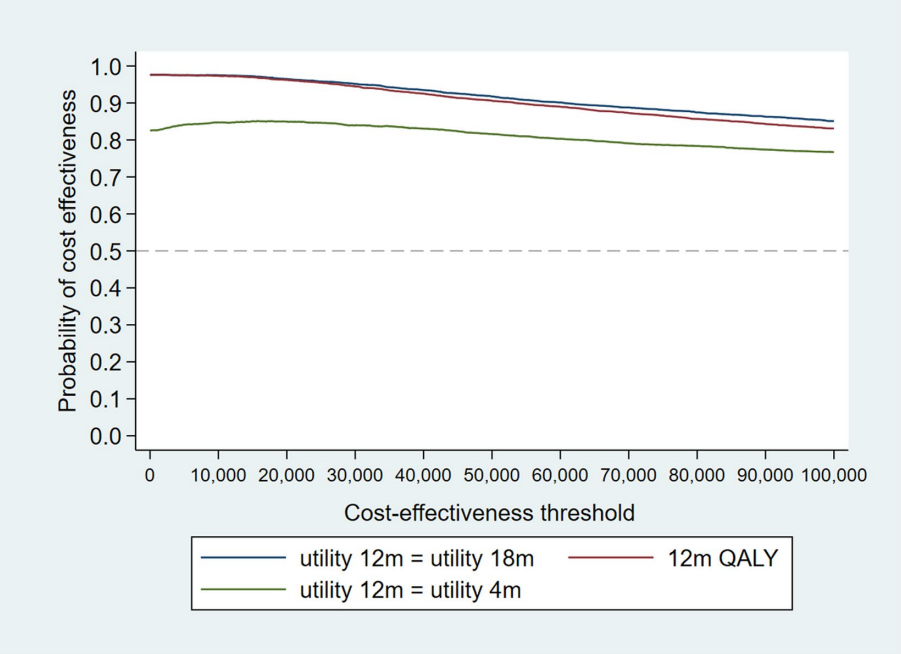


FIGURE 5
CEAC for sensitivity analysis of 12-month QALYs (using TSB method, *N*=223/441).

four scenarios are presented on a CEP in Supplementary Figure 2. The results were very close to those found in the MAR analysis with the probability of cost-effectiveness ranging from 64.7% (MAR) to 66.4% (imputed utility multiplied by 0.7 if the participant has been readmitted) at a threshold value of £20,000/QALY gained, shown on a CEAC in Supplementary Figure 3.

3.6.3. Accounting for resource use skew

Accounting for the skew in the resource use cost data by using a GLM model, the probability that the intervention was cost-effective compared with control is 57% at a cost-effectiveness threshold of £20,000 per QALY gained (see Figures 8 and 9 for the bootstrapped results illustrated on a CEP and CEAC). The mean cost difference is

TABLE 4 Mean cost of mental healthcare resource use for those with complete and incomplete utility data.

		Complete utility data (*N*=223)		Missing utility data at one or more time points (*N*=218)	
		Intervention (*N* =107)	Control (*N* =116)	Intervention (*N* =114)	Control (*N* =104)
Acute care costs baseline	Mean (SD)	£5,639	£3,980	£6,356	£6,879
	95% CI	£3,588 to £7,689	£2,328 to £5,631	£4,476 to £8,235	£4,284 to £9,474
Community costs baseline	Mean	£1,314	£1,507	£2,141	£2,427
	95% CI	£928 to £1,699	£1,079 to £1,936	£1,505 to £2,777	£1,571 to £3,282
Total acute care costs 12 months	Mean	£1,122	£2,589	£6,067	£5,624
	95% CI	(£440 to £1,805)	(£1,101 to £4,077)	(£3,184 to £8,949)	(£2,926 to £8,322)
Total community care costs at 12 months	Mean	£1,888	£2,212	£2,861	£2,993
	95% CI	(£1,393 to £2,382)	(£1,628 to £2,796)	(£2,159 to £3,563)	(£2,144 to £3,843)

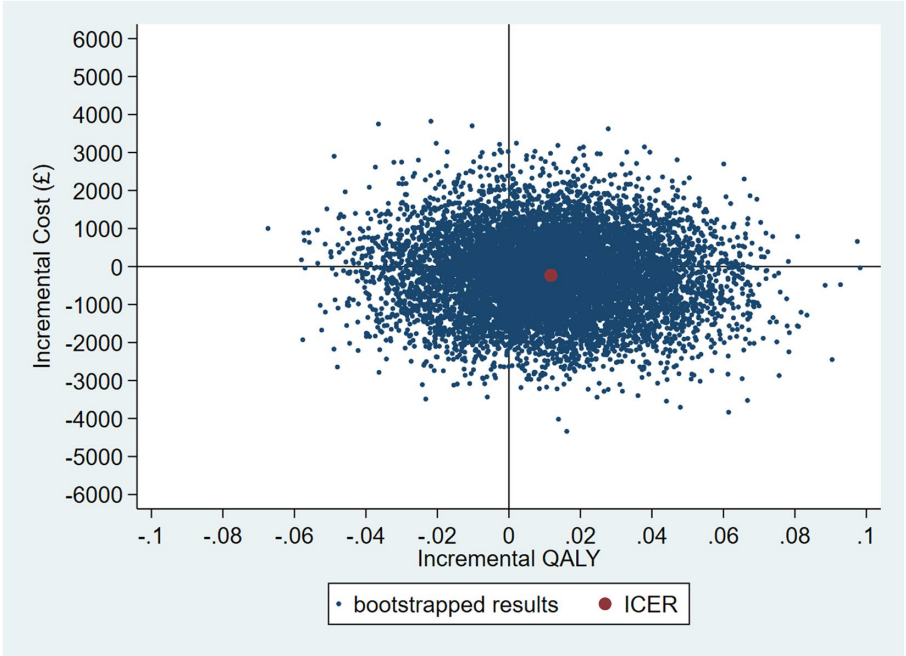


FIGURE 6
CEP MAR using 12-month QALYs and TSB method (using MICE data).

–£427, with 90% of iterations from the bootstrap falling between -£9,186 and £8,522.

Table 5 summarizes the probability of cost-effectiveness for each analysis for ease of comparison.

4. Discussion

The aim of this study was to determine whether a peer-supported self-management intervention delivered by PSWs was cost-effective compared with a self-management workbook plus TAU control. As we had complete data for resource use at 12 months and baseline, and self-report data for utilities at baseline, 4 and 18 months with a large proportion of missing data, we conducted a range of analyses to evaluate the impact of conducting more complex analyses on the results. The intervention dominated the control, as it cost less and yielded more QALYs, although this difference was not significant and had wide confidence intervals. Both the complete-case linear regression and MAR multiple imputation analysis had a probability of 65% that the intervention was cost-effective compared to control at a £20,000/QALY cost-effectiveness threshold over 12 months. This increased to 69% if 18-month utility data and 12-month costs were used as the intervention had a sustained health-related quality-of-life increase.

Resource use came from mental health service use only, and as this was collected from patient records the analysis benefitted from a high level of follow-up for resource use (intervention = 218/221, control = 216/220). This meant that the cost perspective of the

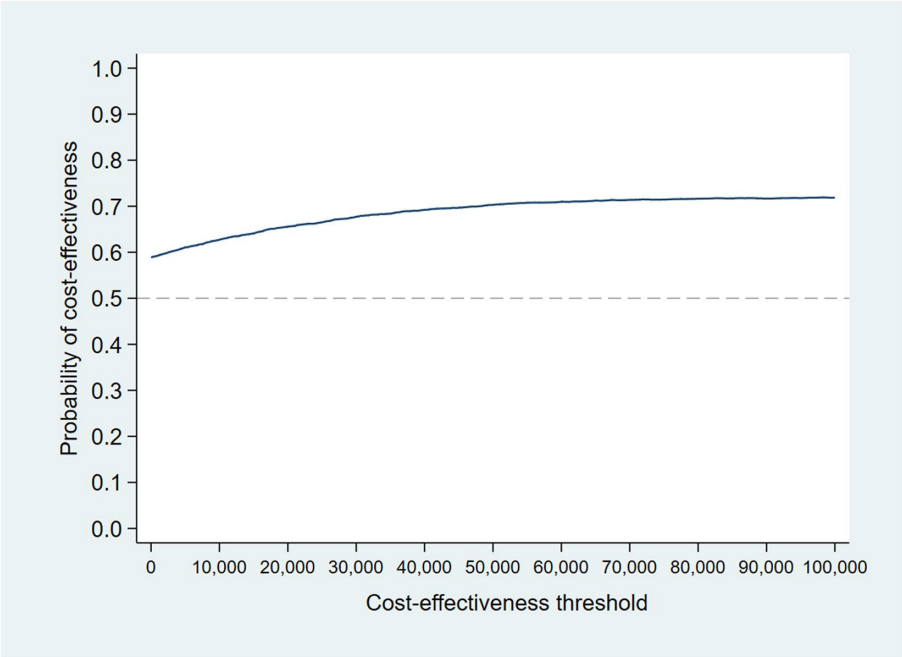


FIGURE 7
CEAC MAR using 12-month QALYs and TSB method (using MICE data).

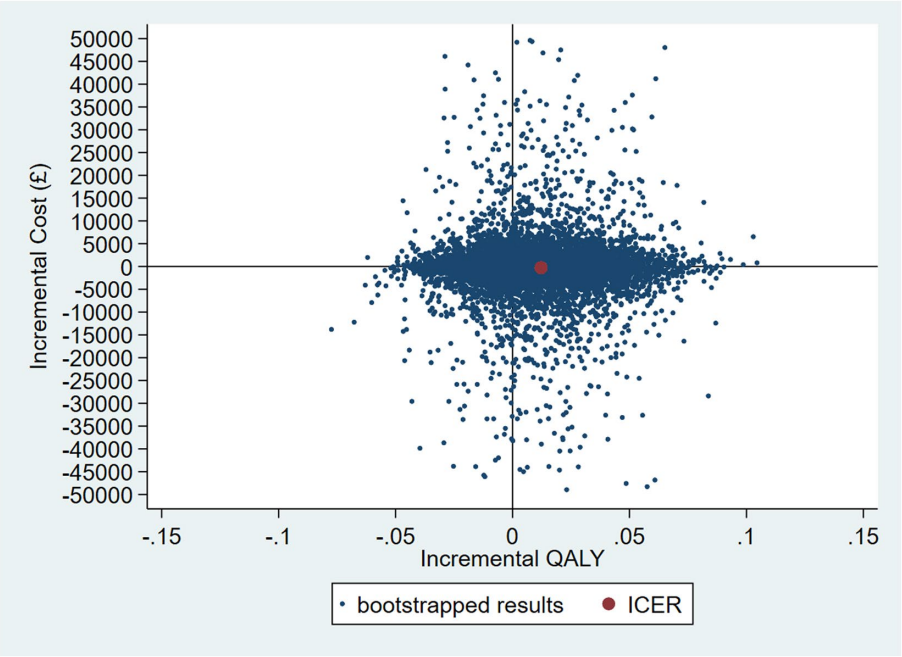


FIGURE 8
CEP for GLM model accounting for resource use skew using 12-month QALYs and TSB method applying a gamma distribution (using MICE data).

analysis was limited to mental health costs only. As the probability that the intervention was cost-effective increased with increasing follow-up periods from 4 to 18 months, this suggests that the benefit of the intervention may be maintained over time, potentially increasing the probability that the intervention is cost-effective through increased QALYs and cost-savings. Given the different follow-up duration for costs and QALYs these results should be interpreted with caution.

This analysis brings into perspective the importance of parsimony when choosing evaluation methods. Given that

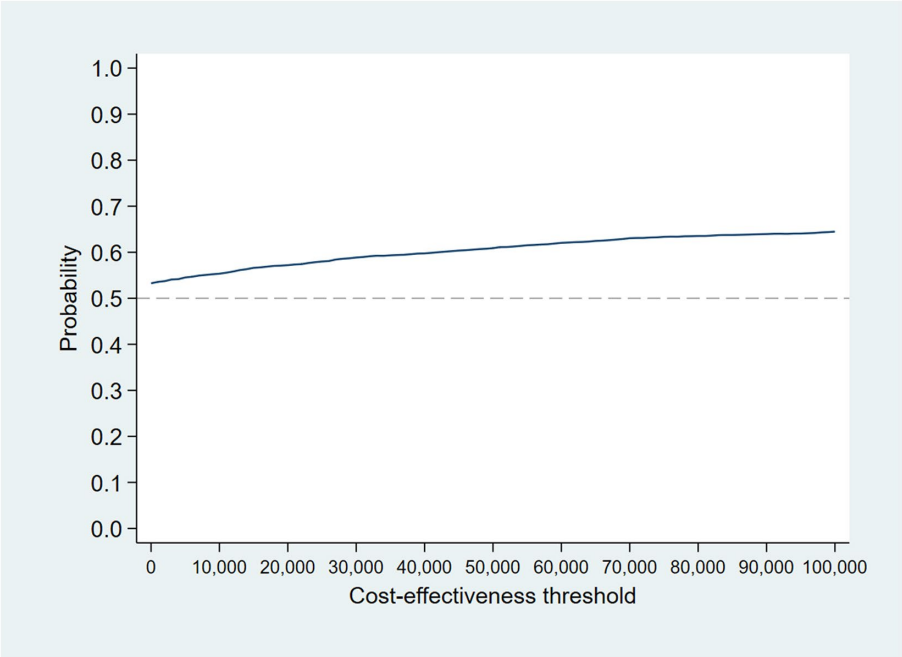


FIGURE 9
CEAC for GLM model accounting for resource use skew using 12-month QALYs and TSB method applying a gamma distribution (using MICE data).

TABLE 5 Summary of results from each analysis.

Analysis	Probability at £20,000	Probability at £30,000
Original analysis	65%	66%
Non-parametric TSB	96%	94%
MI MAR	65%	69%
MI MNAR Utility = imputed utility × 0.8 for those missing & readmitted	66%	69%
GLM MAR	57%	59%

healthcare costs and health-related quality of life are intrinsically linked, it is sensible that we should seek to use methods which take this relationship into account when assessing the cost-effectiveness of a treatment. This, however, requires that both resource use and preference-based health-related quality of life information are present to calculate costs and QALYs, respectively. The results of the complete-case TSB provide evidence of the possible bias that can be introduced when, in this case, information for the denominator of the ICER (ICER = difference in costs/difference in outcome) is missing, restricting the number of cases available for the numerator. Here, analyses using complete case and multiple imputation of utility values are consistent in suggesting the intervention is cost-effective compared to control at 12 months, with a 65% probability that the intervention is cost-effective at a cost-effectiveness threshold of £20,000 per QALY gained. This decreased to 57% when the distribution of the data was taken to account. It is clear in this example that the complete-case TSB leads to an over-estimate of cost-effectiveness and if used incorrectly in other similar analyses, it could lead to an intervention which is not cost-effective being recommended for use, or to not recommending an intervention due to underestimating the cost-effectiveness resulting in patients not receiving the best care available. The results of this analysis show that, when the level of missing data is heavily unbalanced between costs and outcomes, multiple imputation can allow us to implement the preferred method while avoiding introducing bias into the results.

4.1. Strengths and limitations

This analysis was based on data from a randomized control trial in mental healthcare Trusts in England, and provides a robust estimate of the cost-effectiveness of the intervention in this setting. We had relatively complete follow-up for mental health service use data, although the choice of statistical methods for the cost-effectiveness analysis could potentially introduce bias into the analysis when incorporating QALYs, something we have explored in this paper. The cost perspective was limited to specialist mental health services given that this was all that we could obtain from patient files and asking patients to complete questionnaires regarding resource use was considered an onerous addition. Consequently, we are unable to say anything about impact on wider healthcare service resource use or employment and productivity as a result of the trial.

A complete analysis at 18 months was not possible, as although we had EQ-5D-3L data for participants for the calculation of QALYs, we had no resource use information

beyond 12 months. A 4 month cost analysis was also not possible because of the way data was collected from clinical records, giving the number of attendances over 12-months, not when they occurred. Given improvements in utility continue through to 18 months, there may be further QALY gains and cost-savings to be made beyond 18 months, potentially further extending cost-effectiveness if these improvements are related to lower admissions and therefore lower costs. As a result the 18-month cost-effectiveness analysis is potentially a conservative one, if one that should be interpreted with caution given the different time horizon for costs and QALYs.

The EQ-5D is potentially not the best outcome measure to have used as it is not as sensitive in serious mental illness (41). Since the trial, a tariff for calculating utility scores from the Recovering Quality of Life (ReQoL) questionnaire has been developed (42). The measure was designed to assess the quality of life of people with different mental health conditions and may be more suitable in future studies of this patient population.

4.2. Conclusion

There is a high probability that PSW plus workbook is cost-effective compared to usual care plus workbook for a range of cost-effectiveness thresholds. This is likely to be driven by reduced readmissions (23). The probability of cost-effectiveness though is highly dependent on the statistical methods used for the analysis. As a result, it is important that analysts take into account the potential bias from missing data as part of trials in serious mental illness. We would recommend ensuring that resource use is collected as best as possible from patient files. This needs to be complemented though with methods to ensure minimum loss to follow-up for preference-based measures of health-related quality of life for calculating QALYs to reduce the potential bias in the analysis.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by London Camden and Islington Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

ML and RH designed, conducted, and interpreted the analyses. SJ and BL led the study. LM and GA provided advice on the analyses and designed, conducted, and interpreted the statistical analyses for the clinical paper. DO and DL helped to design the study. All authors contributed to the paper and approved the final version, and all took responsibility for its content.

Funding

The paper reports work undertaken as part of the CORE study which was funded by the National Institute for Health Research under its Programme Grants for Applied Research program (reference RP-PG-0109-10078). The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1031159/full#supplementary-material

References

1. The King's Fund (2015). Mental health under pressure.

2. Crisp N, Smith G, Nicholson K. *Old problems, new solutions—improving acute psychiatric care for adults in England* The Commission on Acute Adult Psychiatric Care (2016).

3. Johnson S. Crisis resolution and home treatment teams: an evolving model In: . *Advances in psychiatric treatment*, vol. *19*. Cambridge, United Kingdom: Cambridge University Press (2013). 115–23. doi: 10.1192/apt.bp.107.004192

4. Jethwa K, Galappathie N, Hewson P. Effects of a crisis resolution and home treatment team on in-patient admissions. *Psychiatr Bull*. (2007) 31:170–2. doi: 10.1192/pb.bp.106.010389

5. Johnson S, Nolan F, Pilling S, Sandor A, Hoult J, McKenzie N, et al. Randomised controlled trial of acute mental health care by a crisis resolution team: the North Islington crisis study. *BMJ*. (2005) 331:599. doi: 10.1136/bmj.38519.678148.8F

6. McCrone P, Johnson S, Nolan F, Pilling S, Sandor A, Hoult J, et al. Economic evaluation of a crisis resolution service: a randomised controlled trial. *Epidemiol Psichiatr Soc*. (2009) 18:54–8. doi: 10.1017/S1121189X00001469

7. Lamb D, Lloyd-Evans B, Fullarton K, Kelly K, Goater N, Mason O, et al. Crisis resolution and home treatment in the UK: a survey of model fidelity using a novel review methodology. *Int J Ment Health Nurs*. (2020) 29:187–201. doi: 10.1111/inm.12658

8. Hasselberg N, Holgersen KH, Uverud GM, Siqveland J, Lloyd-Evans B, Johnson S, et al. Fidelity to an evidence-based model for crisis resolution teams: a cross-sectional multicentre study in Norway. *BMC Psychiatry*. (2021) 21:231. doi: 10.1186/s12888-021-03237-8

9. Jacobs R, Barrenho E. Impact of crisis resolution and home treatment teams on psychiatric admissions in England. *Br J Psychiatry*. (2011) 199:71–6. doi: 10.1192/bjp.bp.110.079830

10. Wheeler C, Lloyd-Evans B, Churchard A, Fitzgerald C, Fullarton K, Mosse L, et al. Implementation of the crisis resolution team model in adult mental health settings: a systematic review. *BMC Psychiatry*. (2015) 15:74. doi: 10.1186/s12888-015-0441-x

11. Werbeloff N, Chang CK, Broadbent M, Hayes JF, Stewart R, Osborn DPJ. Admission to acute mental health services after contact with crisis resolution and home treatment teams: an investigation in two large mental health-care providers. *Lancet Psychiatry*. (2017) 4:49–56. doi: 10.1016/S2215-0366(16)30416-3

12. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. *Patient Educ Couns*. (2002) 48:177–87. doi: 10.1016/S0738-3991(02)00032-0

13. Trachtenberg M, Parsonage M, Shepherd G, Boardman J. (2013). *Peer support in mental health care: is it good value for the money?* London: School of Economics and Political Science, LSE Library;1–16.

14. Davidson L, Chinman M, Sells D, Rowe M. Peer support among adults with serious mental illness: a report from the field. *Schizophr Bull*. (2005) 32:443–50. doi: 10.1093/schbul/sbj043

15. Watson E. The mechanisms underpinning peer support: a literature review. *J Ment Health*. (2019) 28:677–88. doi: 10.1080/09638237.2017.1417559

16. Gillard S, Foster R, Gibson S, Goldsmith L, Marks J, White S. Describing a principles-based approach to developing and evaluating peer worker roles as peer support moves into mainstream mental health services In: . *Mental health and social inclusion*, vol. *21*. Bingley, United Kingdom: Emerald Group Publishing Ltd. (2017). 133–43. doi: 10.1108/MHSI-03-2017-0016

17. Repper J, Carter T. A review of the literature on peer support in mental health services. *J Ment Health*. (2011) 20:392–411. doi: 10.3109/09638237.2011.583947

18. Stant AD, Castelein S, Bruggeman R, Van Busschbach JT, Van Der Gaag M, Knegtering H, et al. Economic aspects of peer support groups for psychosis. *Commun Ment Health J*. (2011) 47:99–105. doi: 10.1007/s10597-009-9193-8

19. Mahlke CI, Priebe S, Heumann K, Daubmann A, Wegscheider K, Bock T. Effectiveness of one-to-one peer support for patients with severe mental illness—a randomised controlled trial. *Eur Psychiatry*. (2017) 42:103–10. doi: 10.1016/j.eurpsy.2016.12.007

20. NHS Long Term Plan (n.d.). *NHS mental health implementation plan 2019/20–2023/24*. [Internet]. Available at: https://www.longtermplan.nhs.uk/publication/nhs-mental-health-implementation-plan-2019-20-2023-24/ (Accessed July 23, 2019).

21. Ibrahim N, Thompson D, Nixdorf R, Kalha J, Mpango R, Moran G, et al. A systematic review of influences on implementation of peer support work for adults with mental health problems. *Soc Psychiatry Psychiatr Epidemiol [Internet]*. (2020) 55:285–93. doi: 10.1007/s00127-019-01739-1

22. Mahlke CI, Krämer UM, Becker T, Bock T. Peer support in mental health services. *Curr Opin Psychiatry [Internet]*. (2014) 27:276–81. doi: 10.1097/YCO.0000000000000074

23. Johnson S, Lamb D, Marston L, Osborn D, Mason O, Henderson C, et al. Peer-supported self-management for people discharged from a mental health crisis team: a randomised controlled trial. *Lancet [Internet]*. (2018) 392:409–18. doi: 10.1016/S0140-6736(18)31470-3

24. Liu Y, Pencheon E, Hunter RM, Moncrieff J, Freemantle N. Recruitment and retention strategies in mental health trials – a systematic review. *PLoS One*. (2018) 13:e0203127. doi: 10.1371/journal.pone.0203127

25. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. *Med Care Res Rev*. (2006) 63:217–35. doi: 10.1177/1077558705285298

26. Franklin M, Lomas J, Walker S, Young T. An educational review about using cost data for the purpose of cost-effectiveness analysis. *PharmacoEconomics*. (2019) 37:631–43. doi: 10.1007/s40273-019-00771-y

27. Franklin M, Thorn J. Self-reported and routinely collected electronic healthcare resource-use data for trial-based economic evaluations: the current state of play in England and considerations for the future. *BMC Med Res Methodol*. (2019) 19:8. doi: 10.1186/s12874-018-0649-9

28. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy (New York)*. (1990) 16:199–208. doi: 10.1016/0168-8510(90)90421-9

29. Dolan P. Modelling valuations for EuroQol health states. *Med Care*. (1997) 35:1095–108. doi: 10.1097/00005650-199711000-00002

30. Hunter RM, Baio G, Butt T, Morris S, Round J, Freemantle N. An educational review of the statistical issues in Analysing utility data for cost-utility analysis. *Pharmacoeconomics [Internet]*. (2015) 33:355–66. doi: 10.1007/s40273-014-0247-6

31. National Institute for Health and Care Excellence. *Guide to the methods of technology appr appraisal 2013*. London: National Institute for Health and Care Excellence (NICE) (2013).

32. Curtis LA, Burns A. (2017). *Unit costs of health and social care 2017*. Personal Social Services Research Unit.

33. DH PbR Finance and Costing Team (2017). *National Schedule of reference costs 2015–16*. Department of Health.

34. Mutubuki EN, el Alili M, Bosmans JE, Oosterhuis T, Snoek JF, RWJG O, et al. The statistical approach in trial-based economic evaluations matters: get your statistics together! *BMC Health Serv Res*. (2021) 21:475. doi: 10.1186/s12913-021-06513-1

35. Gomes M, Ng ES, Grieve R, Carpenter J, Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Med Decis Making*. (2012) 32:350–61. doi: 10.1177/0272989X11418372

36. El Alili M, van Dongen JM, Goldfeld KS, Heymans MW, van Tulder MW, Bosmans JE. Taking the analysis of trial-based economic evaluations to the next level: the importance of accounting for clustering. *PharmacoEconomics*. (2020) 38:1247–61. doi: 10.1007/s40273-020-00946-y

37. Walters SJ. Therapist effects in randomised controlled trials: what to do about them. *J Clin Nurs [Internet]*. (2010) 19:1102–12. doi: 10.1111/j.1365-2702.2009.03067.x

38. Lee KJ, Thompson SG. Clustering by health professional in individually randomised trials. *BMJ [Internet]*. (2005) 330:142–4. doi: 10.1136/bmj.330.7483.142

39. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *PharmacoEconomics*. (2014) 32:1157–70. doi: 10.1007/s40273-014-0193-3

40. Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *PharmacoEconomics*. (2018) 36:889–901. doi: 10.1007/s40273-018-0650-5

41. Brazier J. Is the EQ–5D fit for purpose in mental health? *Br J Psychiatry*. (2010) 197:348–9. doi: 10.1192/bjp.bp.110.082453

42. Keetharuth A, Brazier J, Connell J, Carlton J, Buck ET, Ricketts T, et al. Development and validation of the RECOVERING quality of LIFE (REQOL). *Outcome Measures [Internet]*. (2017)

frontiers | Frontiers in Psychiatry

TYPE Systematic Review
PUBLISHED 21 June 2023
DOI 10.3389/fpsyt.2023.1031037

OPEN ACCESS

EDITED BY
Huajie Jin,
King's College London, United Kingdom

REVIEWED BY
Filipa Sampaio,
Uppsala University, Sweden
Cláudia Camilo,
Universidade de Lisboa, Portugal

*CORRESPONDENCE
Tom Kugener
✉ tom.kugener@outlook.com

RECEIVED 29 August 2022
ACCEPTED 24 April 2023
PUBLISHED 21 June 2023

CITATION
Kugener T, Wiethoff I, van Mastrigt G, van den Berg B and Evers SMAA (2023) Economic evaluations of interventions focusing on child abuse and neglect in high-income countries: a systematic review.
Front. Psychiatry 14:1031037.
doi: 10.3389/fpsyt.2023.1031037

Economic evaluations of interventions focusing on child abuse and neglect in high-income countries: a systematic review

Tom Kugener[1]*, Isabell Wiethoff[1], Ghislaine van Mastrigt[1], Bram van den Berg[2] and Silvia M. A. A. Evers[1,3]

[1]Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, Netherlands, [2]Department of Quality, Policy and Monitoring, Nederlands Jeugdinstituut, Utrecht, Netherlands, [3]Centre for Economic Evaluation and Machine Learning, Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands

Introduction: Child abuse and neglect are together considered to be an important public health problem with a high individual and societal burden. Different interventions have been developed to prevent, diagnose, or treat maltreatment. While their effectiveness has been synthesized in prior reviews, the analysis of their cost-effectiveness is less common. The aim of this study is to synthesize and analyse economic evaluations of interventions focusing on child abuse and neglect in high-income countries.

Methods: A systematic literature review was performed using MEDLINE, EMBASE, EconLit, PsycInfo and NHS EED. This study follows the PRISMA guidelines and double scoring was performed. The review includes trial- and model-based economic evaluations of preventive, diagnostic, and treatment related interventions in children up to 18 years or their caregivers. Risk of bias was assessed using the CHEC-extended checklist. The results are presented in a cost-effectiveness plane.

Results: Of 5,865 search results, the full texts of 81 were analyzed, resulting in the inclusion of 11 economic evaluations. Eight of the included studies focus on prevention of child abuse and neglect, one study on diagnosis, and two on treatment. The heterogeneity between studies did not allow for the quantitative pooling of results. Most interventions were cost-effective, with the exception of one preventive and one diagnostic intervention.

Conclusion: This study was subject to some limitations, as no gray literature was included, and the selection of studies may have been arbitrary due to varying terminologies and methodologies in the field. However, the quality of studies was high, and several interventions showed promising results.

Systematic review registration: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021248485, identifier: CRD42021248485.

KEYWORDS

child abuse and neglect, maltreatment, economic evaluation, cost-effectiveness, review-systematic

Introduction

Child abuse and neglect is highly prevalent in high-income countries, having a great impact on the child and their surroundings, consequently leading to a high burden on society. According to a review by Gilbert et al. (1) 4–16% of children in high-income countries experience physical abuse yearly, and 10% experience neglect or psychological

abuse. The cumulative prevalence for sexual abuse of children ranges between 5% and 30% (1). Current estimates by the Centers for Disease Control and Prevention (CDC) show similar results. The CDC (2) estimates that in 2020, one in seven children in the US experienced child abuse or neglect. Maltreatment prevalence rates are expected to be similar for various high-income countries, such as the US, Canada, and European countries (3). A vast number of studies have established an association between maltreatment and different adverse outcomes, including an increased risk for several physical and mental health conditions, emotional and functional impairment, lower wellbeing, and higher risk of delinquent behavior (3–5).

Besides the individual burden, child maltreatment represents a global public health issue with high economic and societal costs (3, 4). Based on several studies, the United Nations estimates that the global burden of violence against children ranges between 2 and 10% of the global Gross Domestic Product (GDP) (6). According to the European Commission, in European countries the annual economic burden of maltreatment represents 4% of the GDP (3). These estimates include costs related to child welfare services, educational services, criminal justice services and productivity losses, in addition to healthcare costs (3).

The current article focuses on economic evaluations of interventions in child maltreatment or child abuse and neglect. The latter terms, i.e., child maltreatment or child abuse and neglect, will be used interchangeably throughout the article. Child maltreatment or child abuse and neglect are defined in this study as "all forms of physical and/or emotional ill-treatment, sexual abuse, neglect or negligent treatment or commercial or other exploitation, resulting in actual or potential harm to the child's health, survival, development or dignity in the context of a relationship of responsibility, trust or power" (7). Four types of child abuse or neglect are commonly distinguished in the literature, namely physical abuse, emotional abuse, sexual abuse, and neglect (7, 8). The definitions of the different types of abuse and neglect can be found in the World Health Organization's (WHO) 1999 Report of the consultation on child abuse prevention (7).

The high burden of child abuse and neglect has led to the development of different interventions to prevent the maltreatment of children or adolescents and, in case of prior maltreatment, provide them with adequate help and treatment. Even though the effectiveness of child abuse and neglect related interventions has been extensively covered in prior reviews, the economic evaluation of these interventions has been given less consideration (9–11). Economic evaluations can be defined as "the comparative analysis of alternative courses of action in terms of both their costs and their consequences" (12). There are four types of full economic evaluations, namely cost-minimization analysis (CMA), cost-effectiveness analysis (CEA), cost-utility analysis (CUA) and cost-benefit analysis (CBA) (12, 13). In each of these types of economic evaluations, the costs and effects of two or more interventions are analyzed and compared. While the costs are analyzed in the relevant monetary unit or currency, the types of outcomes vary across the types of economic evaluations, ranging from monetary outcomes to clinical effectiveness or quality of life (12, 14). In a CMA, the outcomes for both interventions are the same and their costs are compared. In a CEA, clinical outcomes or assessments from validated tools, as well as their costs are compared. A CUA includes quality adjusted life years as outcome and in a CBA benefits are measured in monetary units (12).

The current literature reveals that several interventions to either prevent, treat, or diagnose child maltreatment have been developed. Some existing reviews have analyzed the cost-effectiveness of such interventions (9, 15, 16). According to Dalziel and Segal (9) the cost-effectiveness of home-based interventions varies strongly, and is highest for more complex interventions targeting high-risk populations through professionals from different disciplines. El-Banna et al. (15) conclude that most social care interventions seem cost-effective but highlight the lack of standardized procedures or methods for analyzing the cost-effectiveness of such interventions. Peterson and Kearns (16) came to a similar conclusion, analyzing violence prevention interventions.

Existing reviews, however, mainly differ from the current review in three aspects. First, they may be outdated, as they include relatively old studies conducted prior to 2010 (9). Second, they focus on a broader range of interventions, related to, e.g., general violence prevention and social care interventions (15, 16). Finally, the systematic assessment of the quality of individual economic evaluations is given less attention or is not reported.

The preliminary review of the literature consequently shows the necessity of an up-to-date review, focusing on economic evaluations of child maltreatment interventions in high-income countries conducted after 2010. Accordingly, the aim of this study is to analyze the current evidence on the cost-effectiveness of various interventions focusing on the prevention of child abuse and neglect or services aimed at children and adolescents who have experienced abuse or neglect.

Methods

To analyze the current evidence on economic evaluations of relevant interventions in high- income countries, a multipurpose systematic literature review of model- and trial-based economic evaluation was performed. A systematic review was considered appropriate to ensure a systematic and reproducible collection, analysis, and synthesis of relevant primary studies. A multipurpose

Abbreviations: AAPI-2, Adult Adolescent Parenting Inventory; AHT, Abusive Head Trauma; BCAP, Brief Child Abuse Potential Inventory; BCR, Benefit-Cost Ratio; CADTH, Canadian Agency for Drugs and Technologies in Health; CAN, Child Abuse and Neglect; CAP, Child Abuse Potential Inventory; CBA, Cost-Benefit Analysis; CCEMG, Campbell and Cochrane Economics Methods Group; CEA, Cost-Effectiveness Analysis; CDC, Centers for Disease Control and Prevention; CHEC, Consensus Health Economic Criteria; CMA, Cost-Minimization Analysis; CUA, Cost-Utility Analysis; EE, Economic Evaluation; EPPI, Evidence for Policy and Practice Information and Coordinating; GDP, Gross Domestic Product; ICER, Incremental Cost-Effectiveness Ratio; ICUR, Incremental Cost-Utility Ratio; IMF, International Monetary Fund; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; QALY, Quality-Adjusted Life Year; SEEK, Safe Environment for Every Kid; SSRI, Selective Serotonin Reuptake Inhibitors; TF-CBT, Trauma Focused Cognitive Behavioral Therapy; WHO, World Health Organization; WSIPP, Washington State Institute for Public Policy.

systematic review was chosen, as the primary goal of the study is to identify knowledge gaps and inform policy decisions (17). The pre-specified methods follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (18) and the 5-step approach proposed by Van Mastrigt et al. (17). Furthermore, a protocol of this review (CRD42021248485) has been published in the International Prospective Register of Systematic Reviews (PROSPERO) and can be found in the Supplementary material (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021248485). The study selection as well as the data extraction and quality appraisal were assessed by two researchers independently. Initial disagreements were discussed between the researchers and a third researcher was consulted to reach consensus if needed.

Data collection/literature search

A literature search was performed using MEDLINE, EMBASE, EconLit, PsycInfo, and NHS EED. As NHS EED, a database focusing on economic evaluations in health care, is no longer publishing, only publications up to March 2015 could be included from this source. Furthermore, the references from included studies were analyzed and a citation search was performed. The search strategy was constructed based on the following keywords related to the research aim: "Youth," "Economic Evaluations," "cost-of-illness" and "child abuse and neglect (interventions)." The search strategy includes "cost-of-illness" studies, as this article is part of a larger project by Maastricht University, in collaboration with the Dutch Youth Institute, looking at the economic impact of child abuse and neglect. Furthermore, the Pediatric Economic Database Evaluation (PEDE) was checked for additional, relevant studies.

As the development of a valid new search strategy is time-consuming, existing verified search filters were retrieved through the InterTASC Information Specialists' Sub-Group Search Filter Resource. In general, search filters with high sensitivity are most desirable for a systematic review of economic evaluations (19). The chosen Youth-related keywords were based on two search filters from the Canadian Health Libraries Association for children and adolescents. For "economic evaluations," a search filter from the Canadian Agency for Drugs and Technologies in Health (CADTH) was included. For "Child Abuse & Neglect (interventions)," the search strategy from El-Banna et al. (15) was adapted to the aim and search strategy of this review. Finally, a conceptual approach was applied to establish a search filter to retrieve cost-of-illness studies.

Synonyms for one concept were combined through the Boolean operator "OR," while different concepts were combined through the Boolean operator "AND." The search strategy was adapted individually for each database, as the transferability of database-specific search filters is often limited (19). The literature search for all databases was performed on 4 May 2021. To manage and analyze the search results, EndNote (version X9.3.3) was used as reference software. Further details on the final search strategy for the different databases can be found in the Supplementary material. Reference checking and citation tracking was performed to identify studies which may have been missed through the search strategy.

Inclusion/exclusion criteria

The eligibility of retrieved studies was assessed based on the following inclusion and exclusion criteria. To be included, studies had to include a full economic evaluation (EE). Full economic evaluations are defined as a comparison between two or more programs or interventions in terms of both costs and effects or benefits (17). To include all the relevant literature, model-based and trial-based EEs were included. Second, to be included, studies had to focus on interventions for children and adolescents aged 0 to 18, according to the definition of child abuse and neglect (CAN) by the WHO, and/or interventions for their caregivers. Third, relevant studies had to include an intervention specifically focusing the prevention of CAN or services provided to children or their family after abuse or neglect occurred. To be included, interventions had to focus on either the prevention, diagnosis or treatment of children at risk of or experiencing abuse and neglect. Studies focusing on interventions not specifically addressing CAN or focusing on mental health conditions in parents or children were excluded. Only studies performed in a high-income country (based on the World Bank Atlas Method Classification in 2020) were included, as the rates of CAN and the interventions used in these countries are expected to be comparable. To ensure the inclusion of the most recent evidence, studies published prior to 2010 were excluded, as well as studies written in languages other than English, German, and French. Conference abstracts, editorials and letters were excluded. Although systematic reviews were excluded, their references were analyzed for further results.

Data extraction and quality assessment

An extraction sheet for systematic reviews of economic evaluations was constructed based on the 35 items described in Wijnen et al. (20). To ensure a systematic application of the extraction sheet, a picklist was constructed in Excel (version 2101). The results of the data extraction were summarized in two tables: one focusing on the study characteristics and the second focusing on the study results.

As the review includes model- and trial-based EEs, the extended Consensus Health Economic Criteria (CHEC) checklist has been applied to critically assess eligible EEs, as it is recommended by Cochrane (21). Furthermore, it is the only consensus based quality assessment tool (20). The CHEC-extended checklist includes 20 items or questions which can be answered by Yes, No or Suboptimal and scored to assess the methodological quality of full EEs. To adequately implement the checklist and increase transparency, the assessment instructions provided by Maastricht University were applied (https://www.maastrichtuniversity.nl/research/caphri/our-research/creating-value-based-health-care/chec-list-consensus-health-economic). The results of the risk of bias assessment were summarized in a table enabling the ranking of the studies based on their quality. The CHEC-extended checklist can be found in the Supplementary material.

All monetary units were adapted to a single currency and reference year using a tool provided by the Campbell and Cochrane Economics Methods Group (CCEMG) and the

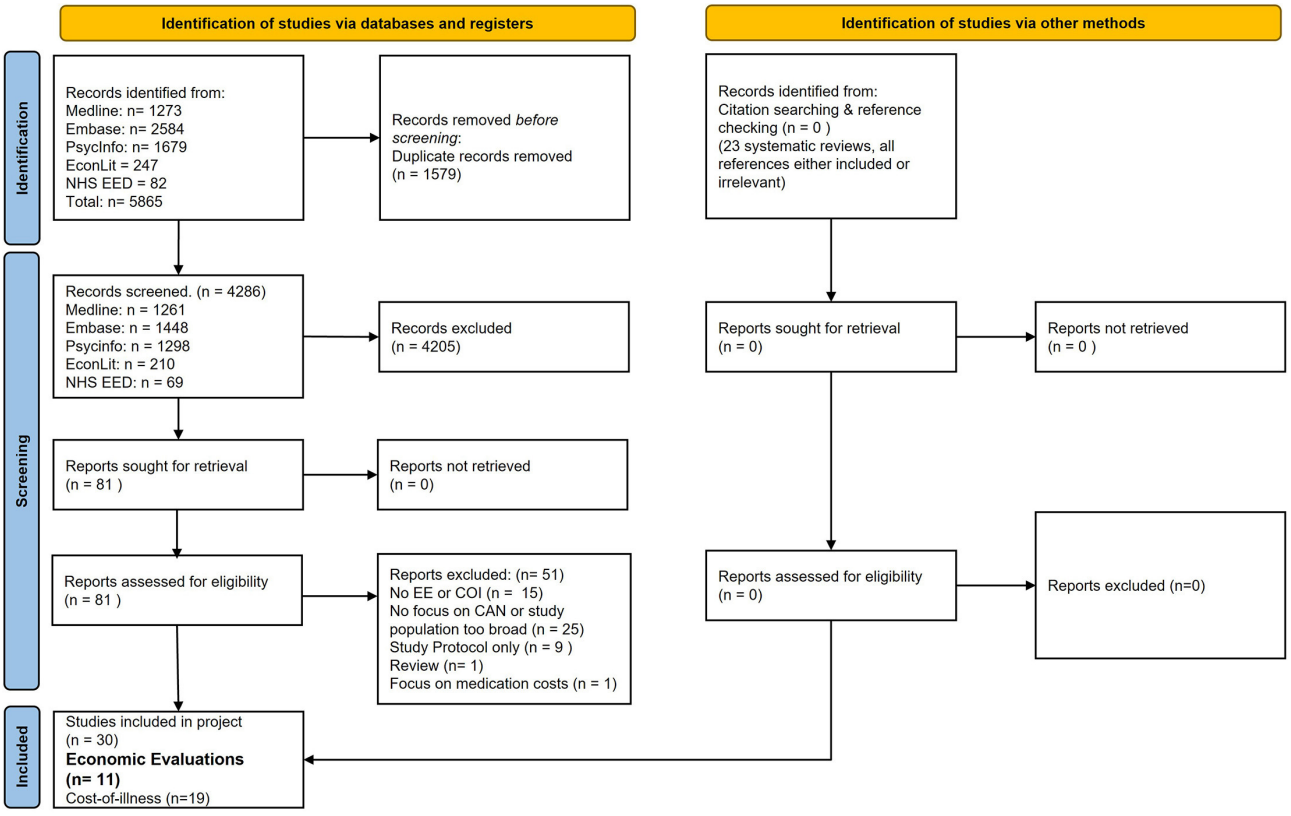


FIGURE 1
PRISMA flowchart (18).

Evidence for Policy and Practice Information and Coordinating Center (EPPI-Center) (https://eppi.ioe.ac.uk/costconversion/default.aspx). Consequently, the values for benefit-cost estimates expressed as a ratio between benefits and costs do not change. The analysis of included studies was clustered according to the type of intervention. The results of all included studies were visualized in a cost-effectiveness plane with a fixed threshold for the willingness to pay per gain of quality-adjusted life-year (QALY). A threshold of 20,000 Euros per QALY was chosen based on the most conservative cost-effectiveness threshold value in the Netherlands (22).

Results

As can be seen in Figure 1 a total of 5,865 studies were retrieved through the previously established search strategy. After deduplication ($n = 1{,}579$), 4,286 studies remained eligible for initial screening. After scanning the title and abstract of these studies, 4,205 studies were excluded, and 81 studies remained eligible for full-text screening. The main reasons for exclusion in abstract scanning were the focus on lower income countries or on the prevention or treatment of specific mental health conditions in children and their caregivers. The full-text analysis of the remaining studies resulted in the inclusion of 11 economic evaluations and 19 cost-of-illness studies. In this article, only the retrieved economic evaluations are analyzed, while the cost-of-illness studies will be covered in a second article. The main reasons for exclusion in the full-text analysis were a study design which included neither a cost-of-illness nor an economic evaluation ($n = 15$), a focus on other things than maltreatment, including conditions or events such as depression, anxiety, self-harm, conduct disorder or behavioral disorders ($n = 25$). While similar interventions were identified in the studies focusing on the mentioned conditions or events, the studies focused on other outcomes not related to abuse and neglect. Nine articles were excluded as they included only the study protocol for economic evaluations. Finally, one study focused on medication costs only and one study included a review. The abstracts of identified systematic reviews were analyzed. Twenty-three systematic reviews were considered potentially relevant based on their abstract, and their references were screened for relevant economic evaluations. The analysis of references to relevant systematic reviews did not lead to the inclusion of any further studies. All checked references were either included already or did not fit the inclusion criteria. Other reviews did not include any relevant studies and were not further analyzed after scanning their abstracts. No additional relevant studies were identified through the PEDE database, as they were either already included, or did not meet all the inclusion criteria (conducted prior to 2010).

General characteristics

Of the 11 studies deemed eligible, 4 studies include a trial-based EE (23–26), while 7 studies focus on model-based EEs (27–33).

Regarding the type of economic analyses, 4 studies produced a cost-utility analysis (23, 24, 30, 31), 5 studies a cost-benefit analysis (26, 27, 29, 32, 33), and 6 studies a cost-effectiveness analysis (23–26, 28, 33). Two studies included both a cost-effectiveness and a cost-utility analyses (23, 24) and two studies included a cost-effectiveness and cost-benefit analysis (26, 33). However, QALYs in children were assessed in only two studies (30, 31). This may be due to the complications associated with determining QALY scores in young children. Five of the cost-effectiveness studies include prevented maltreatment cases as outcome (23–26, 33), while one study focuses on additional convictions (28). The CEA focusing on avoided cases of maltreatment and the CBA were the most applied types of economic evaluations. CBA may be beneficial, as the broad range of consequences of maltreatment and possible effects of interventions can be summarized in a single monetary value.

Five studies were performed in the United States (26, 28, 29, 32, 33), two in Australia (25, 31), two in the United Kingdom (23, 24), one in New Zealand (30) and one in Canada (27). Two studies applied a UK NHS and personal social services perspective (23, 24) and four studies a societal perspective (23, 25, 27, 33), with some studies additionally including more narrow perspectives. Finally, two studies applied a participant, taxpayer and society perspective based on the Washington State Institute for Public Policy (WSIPP) model for cost-benefit analyses (29, 32). For three studies the perspective had to be assumed, as it was not specifically mentioned (26, 28, 30). Eight studies focused on the prevention of maltreatment before or after first contact with child protection services (23–27, 30, 32, 33), one on diagnosis (28), and two studies on the treatment of CAN (29, 31). The age of included children ranged from birth to 17 years of age.

The analysis of the results of individual studies is divided according to the type of intervention. Preventive interventions, which account for most of the included studies, are analyzed first, followed by EEs of diagnostic- and treatment-related interventions. The general characteristics of all included studies can be seen in Table 1.

Quality of the identified studies

Applying the CHEC-extended checklist for the quality of economic evaluations, the average quality of included studies was 89.21%. The lowest score was 58.3% (28). The highest score of 100% was achieved by three studies (24, 29, 31). The other studies all had scores ranging between 82.5% and 97.4%. No direct trends or associations between study characteristics or outcomes and quality scores were observed.

As can be seen in Table 2, most points were deducted for "Q7", "Q8" and "Q15". A complete list of the questions can be found in the Supplementary material. "Q7" is related to the chosen perspective. Points were deducted as some studies did not specifically mention the applied perspective, which consequently had to be assumed (26, 28, 30). "Q8" focuses on the inclusion and reporting of relevant costs for both the intervention of interest and the comparator. The deduction of points was caused by a lack of transparency in reporting costs for both alternatives or for missing costs that should be included, considering the chosen perspective (23, 25–28, 30). "Q15" asks whether costs and outcomes are discounted properly. In some studies, the discount rate was not reported or not applied to all relevant costs and outcomes (26, 28, 33).

Results of the included study

Preventive interventions

Parents under pressure

Barlow et al. (23) analyze the cost-effectiveness of the *Parents Under Pressure* program compared to usual care from a UK NHS and personal social services and from a societal perspective. *Parents under Pressure* is a mainly home-based intervention based on "attachment theory, behavioral parenting skills, and adult psychopathology" (23). The aim is to improve emotional regulation in caregivers to decrease the risk of child maltreatment (23). Study participants were parents in substance abuse treatment with children aged 2.5 years or younger. The trial-based EE had a time frame of 12 months and analyzed QALY gains in parents, as well as the risk of child abuse through the Brief Child Abuse Potential Inventory (BCAP) (23). Even though quality of life gains in parents are valuable outcomes, they are less relevant regarding child abuse and the wellbeing of the children. Consequently, the analysis of the results focuses on the costs for improvements in BCAP scores. The study resulted in €1,234.8 per BCAP score improvement from the personal social services perspective and €2,037.9 per BCAP score improvement from the societal perspective (23).

Dalziel et al. (25) conducted a trial-based CEA to analyse over six months the *Parents Under Pressure* program in Australia, for methadone-receiving parents of children between 2 and 8 years of age. The authors focus on the Child Abuse Potential Inventory (CAP) to distinguish between abusive and non-abusive parents. The results suggest that the program could be cost-effective in preventing maltreatment with an ICER of €26,527.4 per case of maltreatment avoided (25). Dalziel et al. (25) furthermore report net cost savings ranging from €1.5 million to €6.2 million for 100 families profiting from the *Parents Under Pressure* program.

Both studies seem to be of high quality (>90%) and show results in favor of *Parents Under Pressure* compared to usual care. However, they differ in regard to the age of included participants, the time frame and the reported outcome measure. Barlow et al. (23) did not report individual costs, which affects the transparency of the study. Dalziel et al. (25) had a relatively short follow-up period to assess persisting intervention effects. The conclusions of both studies, however, seem justified considering the given data. *Parents Under Pressure* could be a cost-effective solution to reducing the risk of maltreatment in complex situations involving substance-abusing caregivers.

Group family nurse partnership

Barnes et al. (24) provide a trial-based CUA and CEA of the *Group Family Nurse Partnership* program compared to care as usual. As the name indicates, the intervention is conducted by family nurses in a group setting, following young mothers from pregnancy until their children are 1 year old (24). The

TABLE 1 Characteristics of included studies.

Authors (year)	Participants	Perspective	Type of economic Evaluation and Intervention	Country	Trial- or model-based	Comparator	Outcome measure	Conclusion
Barlow et al. (23)	Children under 2.5 years and parents in substance abuse treatment	a) NHS and Personal social services b) Societal	CEA: *Parents under Pressure* (Prevention)*	UK	Trial	Treatment as usual	QALY & child abuse potential	Cost-effective
Barnes et al. (24)	Expectant mothers	NHS and Personal social services	CEA: *Group Family Nurse Partnership* (Prevention)*	UK	Trial	Usual Care	QALY, child abuse potential & maternal sensitivity	Not cost-effective
Beaulieu et al. (27)	Based on children aged 0-24 months	a) Societal b) Health services perspective	CBA: *PURPLE* program (Prevention)	Canada	Model	No program/Period before intervention	Monetary outcomes	Dominant
Block et al. (28)	Possibly abused children	Societal assumed	CEA: *Multiple Interviews* (Diagnosis)	US	Model	Single interviews	Additional convictions	Unclear
Dalziel et al. (25)	Children aged 2-8 years and parents in methadone treatment	Societal	CEA: *Parents Under Pressure* (Prevention)	Australia	Trial	Usual care and brief intervention	Prevented cases of maltreatment	Cost-effective
Dopp et al. (29)	Children aged 10–17 (with determination by CPS that CAN occurred)	Participant, taxpayer and society (WSIPP model)	CBA: *Multisystemic Therapy* for child abuse & Neglect (Treatment/ Prevention after reported abuse)	US	Model	Enhanced outpatient treatment	Monetary values	Cost-effective
Friedman et al. (30)	National Births	Societal assumed	CUA: Shaken Baby Prevention program (Prevention)	New Zealand	Model	No treatment comparator	QALY	Dominant
Gospodarevskaya and Segal (31)	Based on 10-year-old baseline cohort with PTSD due to sexual abuse	Mental healthcare system a) 12-month timeframe b) 31 years' timeframe	CUA: *TF-CBT, TF-CBT and SSRI, and Non-Directive Supportive Counseling* (Treatment)	Australia	Model	No treatment comparator	QALY	Non-directive counseling: (Least) cost-effective TF-CBT (& SSRI): Cost-effective
Kuklinski et al. (32)	Children aged 10 to 24 months	Participant, Taxpayer and Society (WSIPP model)	CBA: *Promoting First Relationships* (Prevention after open CPS report)	US	Model	Resource and/or referral	Monetary outcome	Cost-effective
Lane et al. (26)	Children below 6 years old	Health care assumed	CEA: *Safe Environment for every Kid (SEEK)* (Prevention)**	US	Trial	Routine pediatric care	Prevented cases of maltreatment	Cost-effective
Peterson et al. (33)	Based on hypothetical cohort estimated for each US state	a) Government payer b) Societal	CEA: *Child Parent Centers (CPC)* and *Nurse-Family Partnership (NFP)***	US	Model	Control groups from prior studies	Prevented cases of maltreatment	CPC: Dominant (less than NFP) NFP: Dominant

*Studies additionally include outcomes of a cost-utility analysis.

**Studies additionally include outcomes of a cost-benefit analysis.

10.3389/fpsyt.2023.1031037

TABLE 2 Quality assessment based on the CHEC-extended checklist.

Author(s) and year	Q1 Study Population described	Q2 Alternatives described	Q3 Well-defined research question	Q4 Appropriate study design	Q5 Structural assumptions & validation	Q6 Appropriate time horizon	Q7 Appropriate perspective	Q8 Relevant costs for both alternatives	Q9 Costs measured appropriately	Q10 Costs valued appropriately	Q11 Relevant outcomes for both alternatives	Q12 Outcomes measured appropriately	Q13 Outcomes valued appropriately	Q14 Incremental analysis of costs and outcomes	Q15 Future costs and outcomes discounted	Q16 Sensitivity analysis performed	Q17 Conclusion supported by data	Q18 Generalizability discussed	Q19 Potential conflict of interest	Q20 Ethical issues discussed	Total* (%)
Barlow et al. (23)	●	●	▲	●	⊘	●	●	▲	●	●	●	●	●	●	⊘	●	●	▲	●	●	91.7
Barnes et al. (24)	●	●	●	●	⊘	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	100
Beaulieu et al. (27)	●	●	●	●	●	●	●	▲	▲	▲	●	●	●	●	●	▲	●	●	●	●	90.0
Block et al. (28)	▲	●	●	●	●	○	○	▲	○	○	●	●	⊘	▲	○	▲	●	●	▲	⊘	58.3
Dalziel et al. (25)	●	●	●	●	⊘	▲	●	▲	●	●	●	●	⊘	●	⊘	●	●	●	▲	⊘	90.6
Dopp et al. (29)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	100
Friedman et al. (30)	●	▲	●	●	●	●	○	▲	●	●	●	●	▲	●	●	▲	●	●	▲	●	82.5
Gospodarevskaya and Segal (31)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	⊘	100
Kuklinski et al. (32)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	▲	●	**97.4**
Lane et al. (26)	●	●	●	●	⊘	▲	▲	▲	●	●	●	●	●	●	○	●	●	●	▲	⊘	**83.3**
Peterson et al. (33)	●	●	●	●	▲	●	●	●	●	●	▲	●	▲	●	●	▲	▲	●	●	●	**87.5**
Total* (%)	95.5	95.5	95.5	100	92.9	81.8	77.3	72.7	86.4	81.8	100	95.5	94.4	95.5	72.2	81.8	100	90.9	81.8	100	

● YES – 1 point given, ○ NO – 0 points given, ▲ Suboptimal (SO) – 0.5 points given, ⊘ Not applicable.

* Total (%) $= \frac{Number\ of\ YES * 1 + Number\ of\ SO * 0.5}{(Number\ of\ YES + Number\ of\ SO + Number\ of\ NO)} * 100$.

10.3389/fpsyt.2023.1031037

principles of the intervention are adapted from the *Family Nurse Partnership* intervention. The aim of the program is to provide group sessions to mothers with similar characteristics and generally low educational levels to improve parenting skills and increase infant and maternal health (24). Similarly to the study by Barlow et al. (23), the QALY analysis focused on parents, and consequently the outcome of interest is the risk of child maltreatment. The Adult Adolescent Parenting Inventory (AAPI-2) and the CARE index for maternal sensitivity were used to distinguish between abusive and non-abusive parenting. The intervention showed a low probability for being cost-effective at a cost-effectiveness threshold of €20,000, with an ICER of €130,543 per AAPI-2 score improvement. The study fulfilled all the criteria in assessing the methodological quality, based on the CHEC-extended checklist (100%).

Abusive head trauma prevention

Beaulieu et al. (27) study the costs of abusive head trauma (AHT) and provide a CBA of the Period of PURPLE crying program, a prevention program for AHT. The Period of PURPLE crying intervention aims to educate parents about the normality of increased crying of their (healthy) baby in the first few months of their life, termed "period of PURPLE crying" (27, 34). The model-based study is based on the number of AHT cases reported in British Columbia (Canada) between 2002 and 2014 and estimates the lifetime costs of AHT. The cost-benefit estimates are based on a 35% prevention of AHT cases, on the average costs in the study population and the probability of AHT with or without the PURPLE program (27). The 35% are based on a study conducted in British Columbia, where a 35% decrease in AHT hospitalizations was observed after the implementation of the program (34). From a societal perspective, a one Euro investment would result in a savings of €54. From a healthcare perspective, a one Euro investment would result in a savings of €2.9. As the expected costs per child are lower in the intervention group compared to the group not receiving the program, and the effectiveness higher, the intervention is considered dominant.

Another study analyzing a prevention program for AHT has been conducted by Friedman et al. (30) in New Zealand. The intervention consists of the provision of information by maternity nurses on crying in babies and the risks and consequences of shaken baby syndrome, through a leaflet and a video (30). The authors conducted a model-based CUA of a national primary prevention program for AHT compared to no intervention, including lifetime costs. The costs were based on the review of a 5-year cohort and the incidence taken from a 3-year prospective study (30). QALYs were derived from the CHIP study conducted in the Netherlands. For an effectiveness of 5% and a cost of NZ$20 intervention (€12.5) the ICUR would be €4,436.5 cost per QALY saved. The study concludes that a higher effectiveness with reasonable costs would result in the intervention being dominant, saving money per QALY gained (30).

The quality of both studies was considered acceptable. Beaulieu et al. (27) did not describe the costs of the intervention in detail. Friedman et al. (30) did not mention the perspective, which had to be assumed. The quality of the study by Friedman et al. (30) was slightly lower as details on the perspective were missing, and not all outcomes were reported. However, based on the results from both studies, preventive interventions for abusive head trauma show a high likelihood of being dominant Consequently, both studies or interventions can be found in the SE quadrant in the cost-effectiveness plane (Figure 2).

Promoting first relationships

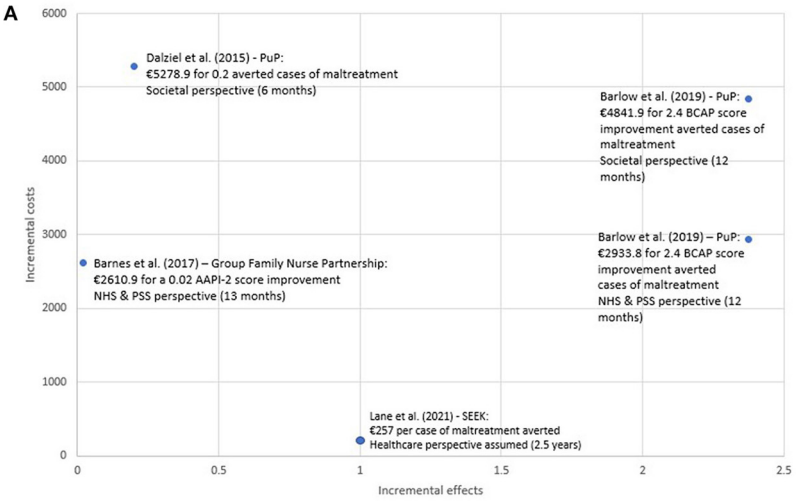
Kuklinski et al. (32) provide a CBA based on the WSIPP model to determine the cost-effectiveness of *Promoting First Relationships* in households with an open Child Protective Services report of possible abuse or neglect. *Promoting first Relationships* is a home-based intervention for children aged 0 to 5 years old and their caregivers (32). Out-of-home placements were used as a proxy to determine reductions in child abuse and neglect. As child abuse and neglect had to be deducted from out-of-home placements, different values were used for the effect size of monetizable child abuse and neglect benefits (32). Consequently, benefit-cost ratios for different effect sizes of abuse and neglect, as a percentage of the effect size of out-of-home placements, were calculated. For a 20% effect size of CAN compared to out-of-home placements, the authors estimated a benefit-cost ratio (BCR) of €4.13 in scenario 2, including systems and victims benefits, and a BCR of €5.19 in scenario 3, including quality-of-life related benefits. Besides the assumptions on the effect size, which may not be completely accurate, the study showed high quality (97.4%) and the intervention could be cost-effective

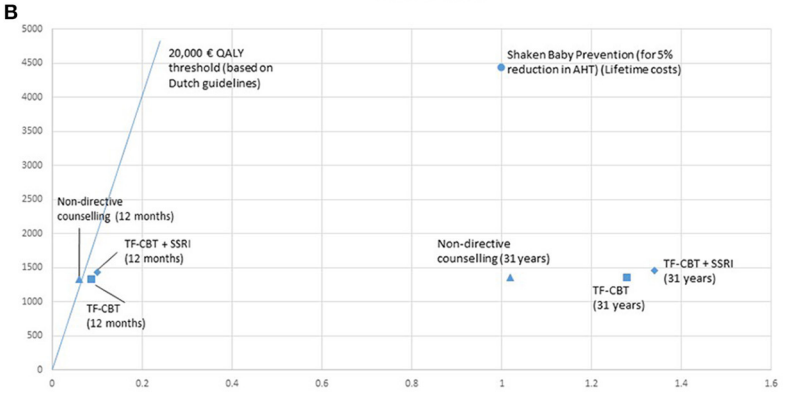
Safe environment for every kid

Lane et al. (26) provide an analysis of the cost-effectiveness of the *Safe Environment for every Kid* (*SEEK*) in comparison with routine pediatric care over 3 years in children below 6 years old. *SEEK* is embedded in pediatric primary care services and consists of a questionnaire filled out by the parents to assess psychosocial risk factors for child maltreatment. The identified risk factors are then addressed by the primary care provider (26, 35). Based on a trial conducted previously by Dubowitz et al. (36), the cost-effectiveness of *SEEK* was estimated for a population of 29,610 children (26). Even though the quality of the study was acceptable (83.3%), the perspective had to be assumed. Their results include an ICER of €257 per case of maltreatment averted, which is considered cost-effective (26).

Child parent centers and nurse-family partnership

Peterson et al. (33) provide estimates of the cost-effectiveness of *Child Parent Centers* (*CPC*) and the *Nurse-Family Partnership* in each US state. To simplify reporting of their results, the estimates of the net present value per avoided case of CAN for the total population were analyzed and adjusted. The *CPC* in preschool resulted in an ICER of only €49,627.1 per averted case of CAN (payer perspective) and savings of €84,211.5 per averted case of CAN (societal perspective). The *CPC* in preschool and school age resulted in an ICER of €45,600.2 per averted case of CAN (payer perspective) and savings of €88,336.6 per averted case of CAN (societal perspective). The *Nurse-Family Partnership* showed savings for both the payer and the societal perspectives, with €24,817.6 and €167,664.3 savings, respectively, per averted case of maltreatment. Furthermore, the BCR for the societal perspective was given, and equals to €1.73 per euro invested for *CPC* and €6.37





Study	Intervention	Incremental effect (QALY)	Incremental cost (€)
Friedman et al. (2012)	*Shaken Baby Prevention* (for 5% reduction in Abusive Head Trauma, 47% found in prior study by Dias et al. in 2005)	1.00	4436.50
Gospodarevskaya & Segal (2012)	*Non-directive counselling* (12 months)	0.06	1332.68
	TF-CBT (12 months)	0.09	1317.97
	TF-CBT + SSRI (12 months)	0.10	1430.55
	Non-directive counselling (31 years)	1.02	1364.30
	TF-CBT (31 years)	1.28	1346.63
	TF-CBT + SSRI (31 years)	1.34	1457.86

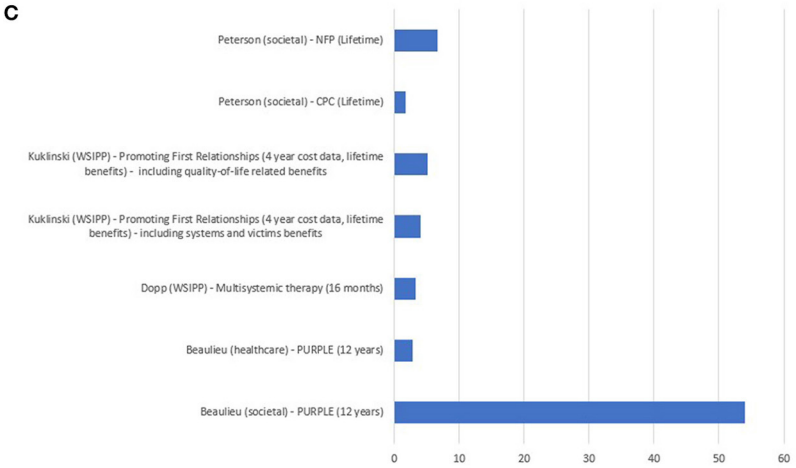


FIGURE 2
(A) Cost-effectiveness plane for studies including a cost-effectiveness analysis (CEA). Block et al. (28) not included in graph due to scale: Block et al. (28): €89,268 for one additional criminal conviction (societal perspective assumed). **(B)** Cost-effectiveness plane for studies including a cost-utility analysis (CUA). **(C)** Results of benefit-cost ratio studies (Cost-savings from a one Euro investment).

for the *Nurse-Family Partnership* (33). Both interventions—*CPC* and *Nurse-Family Partnership*—are considered dominant from a societal perspective as they result in savings per averted case of maltreatment.

Treatment and diagnosis

Multisystemic therapy

Dopp et al. (29) conduct a CBA of *Multisystemic Therapy* for child abuse and neglect. The community-based program is composed of different interventions involving the whole family and their surrounding to identify and address risk factors for maltreatment, treat consequences of maltreatment and prevent further abuse or neglect (29). The intervention was evaluated in families with recently diagnosed physical abuse in children aged 10 to 17. The cost-benefit ratio, based on trial data with a 16-month follow-up, shows that €3.3 could be saved for a 1 Euro investment (cost-effective). The authors reported all relevant aspects required by the CHEC-extended checklist.

Trauma-focused cognitive behavioral therapy and non-directive counseling

A model-based CUA was applied over a time frame of 12 months and 31 years, comparing different treatments from an Australian mental health system perspective in children aged 10 years at baseline (31). The CUA includes *Trauma-Focused Cognitive Behavioral Therapy (TF-CBT), TF-CBT in combination with selective serotonin reuptake inhibitors (SSRI),* and *Non-Directive Supportive Counseling. TF-CBT* and *Non-Directive Counseling* are both flexible treatment programs for post-traumatic stress disorder in children consisting of several sessions including the child or the child and caregivers (31). For a 12-month time frame the ICERs for the different interventions were equal to €22,211.3 per QALY gained (*Non-Directive Counseling*), €14,647 per QALY gained (*TF-CBT*) and €14,305.5 per QALY gained *(TF-CBT & SSRI)*. For a 31-year timeframe, they were equal to €1,337.55 per QALY gained (non-directive counseling), €1060.34 per QALY gained *(TF-CBT)* and €1,096.61 per QALY gained *(TF-CBT & SSRI)* (31). *TF-CBT* as well as *TF-CBT in combination with SSRI* were more likely to be cost-effective than *Non-Directive Counseling.*

Multiple interviews

Block et al. (28) conducted a model-based CEA of *Multiple Interviews* in the diagnosis of possible sexual abuse in children compared to the usual interviewing procedure. The outcome of interest was the number of additional convictions based on a 6.1% increase in the likelihood of criminal convictions (28). The ICER given in the study is equal to €89,268.6 per additional criminal conviction, which the authors consider to be acceptable regarding the high costs of CAN and the number of cases that may be prevented through one additional conviction (28). Based on the comparatively low study quality (58.3%) and the methodological difficulties of measuring additional criminal convictions, no conclusion can be drawn whether *Multiple Interviews* are cost-effective in comparison with single interviews.

Cost-effectiveness plane

In Figures 2A–C, the results are summarized in cost-effectiveness planes (Figures 2A, B) and one additional graph for the studies including a BCR (Figure 2C). Figure 2A focuses on studies including a CEA or clinical outcomes, while Figure 2B focuses on CUA studies using QALY as outcomes. The cost-effectiveness planes visually presents the results from different EEs, based on their incremental effects and incremental costs. Due to the heterogeneity of individual studies, several graphs were created. As the outcomes are varying, the position of different studies or interventions may be to some extent arbitrary, but the figures provides first insight into the cost-effectiveness of different interventions. As can be seen, most included studies are situated in the north-eastern (NE) quadrant. This means that they show beneficial effects for additional costs, which requires a decision to be taken based on a willingness-to-pay threshold for a certain outcome. Four interventions seem particularly likely to be cost-effective, or dominant, as they have high incremental effects for lower costs (27, 30, 33). Dominant interventions can be found in the south-eastern (SE) quadrant. On the other hand, two interventions seem unlikely to be cost-effective compared to other interventions, as they show low effects for high incremental costs (24, 28).

Discussion

The aim of this review was to assess the current evidence on economic evaluations of interventions aimed at the prevention or treatment of child abuse and neglect in high-income countries.

Only a small number of economic evaluations focusing on child abuse and neglect have been retrieved and fulfilled the inclusion criteria. This highlights the need for further studies analyzing the cost-effectiveness of interventions relating to child abuse and neglect. Even though a sensitive search strategy was applied, only 11 economic evaluations were eligible. Most studies are considered to be cost-effective, which may be partly due to publication bias. Due to the heterogeneity of the studies, the results were not pooled as they are incomparable to a large extent.

The quality of the economic evaluations was high, with an average of 89.21%. Only one study scored below 80% on the CHEC-extended checklist. The applied methodology varied considerably between studies in regard to the type of economic evaluations performed and the type of outcomes measured. Most quality concerns were related to the description of the chosen perspective, the included costs, and discounting procedures.

The results have shown that most of the included studies were cost-effective in tackling child abuse and neglect. The *Parents Under Pressure* program has shown evidence of improving outcomes at lower costs than the comparators in caregivers of children aged 2.5 years or younger in the UK and in children aged 2 to 8 years in Australia (23, 25). Two studies have shown that interventions focusing on the prevention of abusive head trauma or shaken

baby syndrome show a high likelihood of leading to beneficial outcomes at lower costs than the comparators. In other words, they show a high likelihood of being dominant. *The Period of PURPLE crying* implemented in Canada was compared to no program and showed an acceptable return on investment and cost-savings (27). Another basic shaken baby prevention program in New Zealand has shown low costs or even cost savings per QALY gain (30). *Child Parent Centers* and *Family-Nurse Partnership* also showed a high likelihood of being dominant, i.e., resulting in cost savings per case of maltreatment averted from a societal perspective (33).

Only one included study focused on an intervention implemented in primary care services. The implementation of *Safe Environment for Every Kid* in a population of around 30,000 children showed low costs per case of maltreatment averted (26). The *Group Family Nurse Partnership* did not show evidence of being cost-effective (24). The interventions focusing on the treatment of children who have experienced child abuse and neglect were also found to be cost-effective (29, 31). *Cognitive Behavioral Therapy* focused on trauma seems to be more cost-effective than non-directive counseling. The addition of selective serotonin reuptake inhibitors to *Trauma-Focused Cognitive Behavioral Therapy* may provide even lower costs per QALY gains (31). However, the use of selective serotonin reuptake inhibitors in children may also include other risks, not included in the economic evaluation. *Multisystemic Therapy* has shown an acceptable benefit-cost ratio (29). Based on the comparatively low study quality and the methodological difficulties of measuring additional criminal convictions, no definite conclusion can be drawn about whether multiple interviews are cost-effective in comparison with single interviews (28).

This review has several strengths. Following the PRISMA framework and guidelines on conducting a systematic review of economic evaluations is expected to ensure the methodological quality of this review. Furthermore, a PROSPERO protocol was developed before conducting the review. In addition, the data extraction and quality assessment was checked by two researchers independently.

The review is, however, subject to several limitations. Publication bias has not been estimated. No gray literature was included, and studies published in languages other than German, French or English were excluded. Due to the heterogeneity of retrieved studies and varying terminologies and methodologies, the selection of articles may have been arbitrary. While the selection of studies has been done by two researchers independently, no intercoder agreement score was determined. In addition, transferability was not assessed, as the review does not focus on the implementation of interventions in one particular country or setting.

Several shortcomings of individual economic evaluations in the field of child abuse and neglect were identified. There appears to be a need for more standardized reporting methods, as the results of the review depend strongly on methodological choices and the reporting quality of included studies. Dalziel et al. (25) and Barlow et al. (23), for example, include different methods for discriminating between abusive and non-abusive parents. Barlow et al. (23) report the cost per Brief Child Abuse Potential Inventory (BCAP) score improvement, while Dalziel et al. (25) report the cost per prevented case of maltreatment based on Child Abuse Potential Inventory (CAP) cut-off values and the respective risks of maltreatment. Preferably, studies applying either the CAP or the BCAP, which correlate strongly, should report the same outcome. The benefits and disadvantages of using BCAP or CAP scores, or cut-off points to estimate the number of prevented cases of maltreatment should be further analyzed. Standardized reporting of the outcome would allow further comparisons and pooling of results. Other methods used to assess the number of maltreatment cases or prevented maltreatment cases include the Adult Adolescent Parenting Inventory or the Conflict Tactics Scales: Parent-Child version (24, 26). These methods may yield different results and reflect prevented child abuse and neglect cases more or less accurately. The assessment of QALY in children is also subject to several limitations (37).

Possible influences could be the sample sizes, the handling of missing data or the models used, which may also have a large influence on the effectiveness and cost-effectiveness results. In addition, the included costs and time frames vary, highlighting the lack of a common methodology for analyzing cost-effectiveness of interventions for abuse and neglect. Therefore, the results of studies have not been pooled and comparisons between studies should be made with caution.

Furthermore, there is no common terminology applied in research on child abuse and neglect. A highly sensitive search strategy was applied to ensure the inclusion of relevant studies. However, as there are no strictly defined boundaries on what should be considered abuse and neglect, it is difficult to determine which studies to include. Furthermore, even if boundaries are well-defined, it is difficult to accurately measure the prevalence or number of maltreatment cases in a certain population.

The most recent identified reviews including economic evaluations of child abuse and neglect interventions reported similar limitations. Peterson and Kearns (16) mention the need for better reporting standards to increase comparability between economic evaluations of violence prevention interventions. El-Banna et al. (15) furthermore highlight the lack of standardized outcome measures and cost-effectiveness threshold in children's social care interventions. In addition, the time frame of the economic evaluations is often too short to include long-term costs and effects of the interventions (15). Based on the identified limitations, El-Banna et al. (15) developed ten recommendations for future systematic reviews of economic evaluations children's social care interventions.

Due to the mentioned limitations of the field of study and of the studies included, some arbitrary decisions had to be taken while developing and applying the inclusion and exclusion criteria. This study included articles focusing on the four main types of child abuse and neglect, including abusive head trauma. Intimate partner violence was not included as it does not necessarily lead to maltreatment. Furthermore, studies had to focus specifically on children at risk of abuse or neglect or children who experienced abuse or neglect. Economic evaluations of studies focusing on broader outcomes with possible effects on abuse and neglect did not meet the inclusion criteria. Children or caregivers with mental health disorders were not included. Other studies that did not

meet the inclusion criteria but may provide additional information include, among others: Aas et al. (38), Dijkstra et al. (39), Johnson-Motoyama et al. (40), Lynch et al. (41) and Reynolds et al. (42). These studies were excluded for different reasons. Aas et al. (38) focused on children who experienced a trauma which does not exclusively focus on traumas related to abuse and neglect. Johnson-Motoyama et al. (40) focused on out-of-home placements in substance-affected families, which is not necessarily linked on abuse and neglect. Reynolds et al. (42) also did not focus specifically on abuse and neglect related outcomes. Lynch et al. (41) included children in foster care and permanent placements as main outcome, which does not reflect abuse and neglect. The outcome measure in Dijkstra et al. (39) was considered insufficient to measure prevented cases of maltreatment.

Regarding the generalizability of findings, it should be kept in mind that a study that has been found to be cost-effective in a specific setting and population is not necessarily cost-effective in another setting and population. The generalizability of the results presented in this study are limited to high-income countries. Primary care and childcare services vary across countries. Transferability analysis should be performed to ensure that an intervention will remain cost-effective in a different setting. Therefore, one should have a clear idea about the structure of childcare services in the country of interest and the basic level of care. For example, abusive head trauma prevention interventions which have been found to be cost-effective may already be part of the basic care provided to parents in other countries. Otherwise, it might be a cost-effective prevention measure to reduce abusive head traumas in maternal care or primary, pediatric care. Treatment interventions might be integrated into existing childcare services. To determine the transferability of the economic evaluation, the Welte checklist may be used, including general checkout criteria, methodological characteristics, healthcare system characteristics and population characteristics (20). To further investigate the transferability, the PIET-t model in "Models of Child Health Appraised" may serve as a helpful tool for assessing similarities between childcare systems and identifying possible barriers to implementation (43).

Based on the results of the review and the identified limitations, several recommendations for policy and future research will be made. The presented results are expected to provide insight to policymakers in high-income countries on financially sustainable possibilities to tackle child abuse and neglect. Interventions to prevent abusive head trauma through simple educational means (e.g., fact sheets) have shown high cost-effectiveness. As they show considerable effects for low costs and efforts, they are expected to be cost-effective in varying settings. Furthermore, home-based, individualized interventions to prevent maltreatment may be of interest for policymakers. While the priority should be on preventing abuse and neglect, the treatment options have shown promising results for being cost-effective.

To overcome current shortcomings, expertise from different fields is required when conducting economic evaluations in the field of child abuse and neglect and should be integrated into the development and evaluation of interventions. A common methodology would strongly benefit future reviews and economic evaluations of relevant interventions. More research is needed to determine the most accurate and useful outcome measure in economic evaluations on child abuse and neglect. Developing a common methodology and outcome measure will allow further comparisons or pooling of data in a meta-analysis. Additionally, a common methodology and outcome measure will facilitate the development and application of strict inclusion and exclusion criteria for future systematic reviews. Spillover effects should be estimated in future research as child abuse and neglect interventions impact various dimensions of a caregiver's and/or children's life as well as the people around them. Neglecting spillover effects such as costs of informal care and benefits to family members, may result in an underestimation of the benefits of child abuse and neglect interventions (44).

Further research is required to determine which intervention shows the most favorable cost-effectiveness in different settings. Future researchers should adhere to the guidelines established by the Professional Society for Health Economics and Outcomes Research (ISPOR). Furthermore, there are several reporting guidelines for economic evaluations, such as the recently updated Consolidated Health Economic Evaluation Reporting Standards (CHEERS), to ensure all relevant study aspects are reported (45). Using recognized guidelines ensures that all relevant study aspects will be reported, facilitating future reviews and the development of replicable methodologies.

Conclusion

This study provides an overview of economic evaluations of preventive, diagnostic and treatment interventions related to child abuse and neglect in high-income countries. The results show that little research has been done in this field, but the evaluated interventions have a high potential for cost-effectiveness, especially individualized home- or community-based interventions and educational interventions. The transferability should, however, be assessed before implementing the interventions in a new setting. Future research could benefit from a more strictly defined terminology for child abuse and neglect, and from clear boundaries on which caregiver practices are considered to be abuse or neglect. Furthermore, a common methodology could increase the comparability of interventions focusing on child abuse and neglect.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

TK, IW, and SE were involved throughout the conception and the development of the design of the study. BB provided expert opinion on child abuse and neglect, methodological guidance (e.g., search strategy, definitions, inclusion/exclusion criteria), as well as important aspects to consider in the discussion and conclusion. GM and SE provided guidance on the design of systematic reviews

of economic evaluations and the reporting and interpretation of results. TK conducted the analysis and wrote the first draft of the article and integrated the feedback. IW independently conducted certain steps of the analysis. All authors approved the final article for publication, critically reviewed the manuscript, and provided feedback.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1031037/full#supplementary-material

SUPPLEMENTARY APPENDIX A
Prospero Protocol.

SUPPLEMENTARY APPENDIX B
Search strategy.

SUPPLEMENTARY APPENDIX C
CHEC-extended checklist.

References

1. Gilbert R, Widom CS, Browne K, Fergusson D, Webb E, Janson S. Burden and consequences of child maltreatment in high-income countries. *Lancet.* (2009) 373:68–81. doi: 10.1016/S0140-6736(08)61706-7

2. CDC. Preventing child abuse and neglect. (2021). Available online at: https://www.cdc.gov/violenceprevention/childabuseandneglect/fastfact.html (accessed August, 2022).

3. Ferrara P, Corsello G, Basile MC, Nigri L, Campanozzi A, Ehrich J, et al. The Economic burden of child maltreatment in high income countries. *J Pediat.* (2015) 167:1457–9. doi: 10.1016/j.jpeds.2015.09.044

4. Fang X, Brown DS, Florence CS, Mercy JA. The economic burden of child maltreatment in the United States and implications for prevention. *Child Abuse Negl.* (2012) 36:156–65. doi: 10.1016/j.chiabu.2011.10.006

5. Gardner MJ, Thomas HJ, Erskine HE. The association between five forms of child maltreatment and depressive and anxiety disorders: a systematic review and meta-analysis. *Child Abuse Neglect*. (2019) 96:4082. doi: 10.1016/j.chiabu.2019.104082

6. Pais MS. *The Economic Costs of Violence Against Children*. New York: United Nations (Special representative of the Secretary-General on Violence Against Children) (2015).

7. Consultation on Child Abuse Prevention WHO, Violence and Injury Prevention Team & Global Forum for Health Research. *Report of the Consultation on Child Abuse Prevention*. Geneva: World Health Organization. (1999).

8. Norman RE, Byambaa M, De R, Butchart A, Scott J, Vos T. The long-term health consequences of child physical abuse, emotional abuse, and neglect: a systematic review and meta-analysis. *PLoS Med Public Library Sci.* (2012) 9:e1001349. doi: 10.1371/journal.pmed.1001349

9. Dalziel K, Segal L. Home visiting programmes for the prevention of child maltreatment: cost-effectiveness of 33 programmes. *Arch Dis Child.* (2012) 97:787–98. doi: 10.1136/archdischild-2011-300795

10. Mikton C, Butchart A. Child maltreatment prevention: a systematic review of reviews. *Bull World Health Organ.* (2009) 87:353–61. doi: 10.2471/BLT.08.057075

11. van der Put CE, Assink M, Gubbels J, Boekhout van Solinge NF. Identifying effective components of child maltreatment interventions: a meta-analysis clinical child and family. *Psychol Rev.* (2018) 21:171–202. doi: 10.1007/s10567-017-0250-5

12. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press (2015).

13. Kernick DP. Introduction to health economics for the medical practitioner. *Postgrad Med J.* (2003) 79:147–50. doi: 10.1136/pmj.79.929.147

14. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. *BMC Med.* (2013) 11:80. doi: 10.1186/1741-7015-11-80

15. El-Banna A, Petrou S, Yiu HHE, Daher S, Forrester D, Scourfield J, et al. Systematic review of economic evaluations of children's social care interventions. *Child Youth Serv Rev*. (2021) 121:5864. doi: 10.1016/j.childyouth.2020.105864

16. Peterson C, Kearns MC. Systematic review of violence prevention economic evaluations, 2000–2019. *Am J Prev Med.* (2021) 60:552–62. doi: 10.1016/j.amepre.2020.11.002

17. van Mastrigt GA, Hiligsmann M, Arts JJ, Broos PH, Kleijnen J, Evers SM, et al. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: a five-step approach (part 1/3). *Expert Rev Pharmacoecon Outcomes Res.* (2016) 16:689–704. doi: 10.1080/14737167.2016.1246960

18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* (2009) 6:e1000097. doi: 10.1371/journal.pmed.1000097

19. Thielen FW, Van Mastrigt G, Burgers LT, Bramer WM, Majoie H, Evers S, et al. How to prepare a systematic review of economic evaluations for clinical practice guidelines: database selection and search strategy development (part 2/3). *Expert Rev Pharmacoecon Outcomes Res.* (2016) 16:705–21. doi: 10.1080/14737167.2016.1246962

20. Wijnen B, Van Mastrigt G, Redekop WK, Majoie H, De Kinderen R, Evers S. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: data extraction, risk of bias, and transferability (part 3/3). *Expert Rev Pharmacoecon Outcomes Res.* (2016) 16:723–32. doi: 10.1080/14737167.2016.1246961

21. Cochrane C, Higgins JPT, Thomas J. *Cochrane Handbook for Systematic Reviews of Interventions*. London: Cochrane Collaboration (2021).

22. Zwaap J, Knies S, van der Meijden C, Staal P, van der Heiden L. Zorginstituut Nederland. *Kosteneffectiviteit in de prakttijk*. Diemen: Zorginstituut Nederland. (2015).

23. *Barlow J, Sembi S, Parsons H, Kim S, Petrou S, Harnett P, et al. A randomized controlled trial and economic evaluation of the parents under pressure program for parents in substance abuse treatment. *Drug Alcohol Depend.* (2019) 194:184–94. doi: 10.1016/j.drugalcdep.2018.08.044

24. *Barnes J, Stuart J, Allen E, Petrou S, Sturgess J, Barlow J, et al. Randomized controlled trial and economic evaluation of nurse-led group support for young mothers during pregnancy and the first year postpartum versus usual care. *Trials.* (2017) 18:508. doi: 10.1186/s13063-017-2259-y

25. *Dalziel K, Dawe S, Harnett PH, Segal L. Cost-effectiveness analysis of the Parents under Pressure programme for methadone-maintained parents. *Child Abuse Rev.* (2015) 24:317–31. doi: 10.1002/car.2371

26. *Lane WG, Dubowitz H, Frick KD, Semiatin J, Magder L. Cost effectiveness of SEEK: A primary care-based child maltreatment prevention model. *Child Abuse Neglect.* (2021) 111:4809. doi: 10.1016/j.chiabu.2020.104809

27. *Beaulieu E, Rajabali F, Zheng A, Pike I. The lifetime costs of pediatric abusive head trauma and a cost-effectiveness analysis of the Period of Purple crying program in British Columbia, Canada. *Child Abuse Neglect.* (2019) 97:104133. doi: 10.1016/j.chiabu.2019.104133

28. *Block SD, Foster EM, Pierce MW, Berkoff MC, Runyan DK. Multiple forensic interviews during investigations of child sexual abuse: a cost-effectiveness analysis. *Appl Develop Sci.* (2013) 17:174–83. doi: 10.1080/10888691.2013.836033

29. *Dopp AR, Schaeffer CM, Swenson CC, Powell JS. Economic impact of multisystemic therapy for child abuse and neglect. *Admin Policy Mental Health.* (2018) 45:876–87. doi: 10.1007/s10488-018-0870-1

30. *Friedman J, Reed P, Sharplin P, Kelly P. Primary prevention of pediatric abusive head trauma: a cost audit and cost-utility analysis. *Child Abuse Neglect.* (2012) 36:760–70. doi: 10.1016/j.chiabu.2012.07.008

31. *Gospodarevskaya E, Segal L. Cost-utility analysis of different treatments for post-traumatic stress disorder in sexually abused children. *Child Adoles Psych Mental Health.* (2012) 6:15. doi: 10.1186/1753-2000-6-15

32. *Kuklinski MR, Oxford ML, Spieker SJ, Lohr MJ, Fleming CB. Benefit-cost analysis of promoting first relationships: implications of victim benefits assumptions for return on investment. *Child Abuse Neglect.* (2020) 106:4515. doi: 10.1016/j.chiabu.2020.104515

33. *Peterson C, Florence C, Thomas R, Klevens J. Cost-benefit analysis of two child abuse and neglect primary prevention programs for US States. *Prevent Sci.* (2018) 19:705–15. doi: 10.1007/s11121-017-0819-8

34. Barr M. What is the Period of PURPLE crying? n.d. Available online at: http://purplecrying.info/what-is-the-period-of-purple-crying.php (accessed August, 2022).

35. SEEK. The SEEK Model. (2020). Available onlin at: https://seekwellbeing.org/the-seek-model/ (accessed August, 2022).

36. Dubowitz H, Lane WG, Semiatin JN, Magder LS. The SEEK model of pediatric primary care: can child maltreatment be prevented in a low-risk population? *Acad Pediatr.* (2012) 12:259–68. doi: 10.1016/j.acap.2012.03.005

37. Petrou S. Methodological challenges surrounding QALY estimation for paediatric economic evaluation. *Cost Effect Res Allocation.* (2022) 20:10. doi: 10.1186/s12962-022-00345-4

38. Aas E, Iversen T, Holt T, Ormhaug SM, Jensen TK. Cost-effectiveness analysis of trauma-focused cognitive behavioral therapy: a randomized control trial among Norwegian youth. *J Clin Child Adoles Psychol.* (2019) 48:S298–311. doi: 10.1080/15374416.2018.1463535

39. Dijkstra S, Creemers HE, van Steensel FJA, Dekovic M, Stams G, Asscher JJ. Cost-effectiveness of family group conferencing in child welfare: a controlled study. *BMC Public Health.* (2018) 18:848. doi: 10.1186/s12889-018-5770-5

40. Johnson-Motoyama M, Brook J, Yan Y, McDonald TP. Cost analysis of the strengthening families program in reducing time to family reunification among substance-affected families. *Child Youth Serv Rev.* (2013) 35:244–52. doi: 10.1016/j.childyouth.2012.11.008

41. Lynch FL, Dickerson JF, Saldana L, Fisher PA. Incremental net benefit of early intervention for preschool-aged children with emotional and behavioral problems in foster care. *Child Youth Serv Rev.* (2014) 36:213–9. doi: 10.1016/j.childyouth.2013.11.025

42. Reynolds AJ, Temple JA, White BA, Ou SR, Robertson DL. Age 26 cost-benefit analysis of the child-parent center early education program. *Child Dev.* (2011) 82:379–404. doi: 10.1111/j.1467-8624.2010.01563.x

43. Kocken P, Vlasblom E, de Lijster G, Wells H, van Kesteren N, van Zoonen R, et al. The Transferability of Primary Child Healthcare Systems. In: Blair M, Rigby M, Alexander D, editors. *Issues and Opportunities in Primary Health Care for Children in Europe.* (Emerald Publishing Limited) 2019. p. 331-44. doi: 10.1108/978-1-78973-351-820191019

44. Grosse SD, Pike J, Soelaeman R, Tilford JM. Quantifying family spillover effects in economic evaluations: measurement and valuation of informal care time. *Pharmacoeconomics.* (2019) 37:461–73. doi: 10.1007/s40273-019-00782-9

45. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *BMC Med.* (2022) 20:23. doi: 10.1186/s12916-021-02204-0

*Studies included in the systematic review.

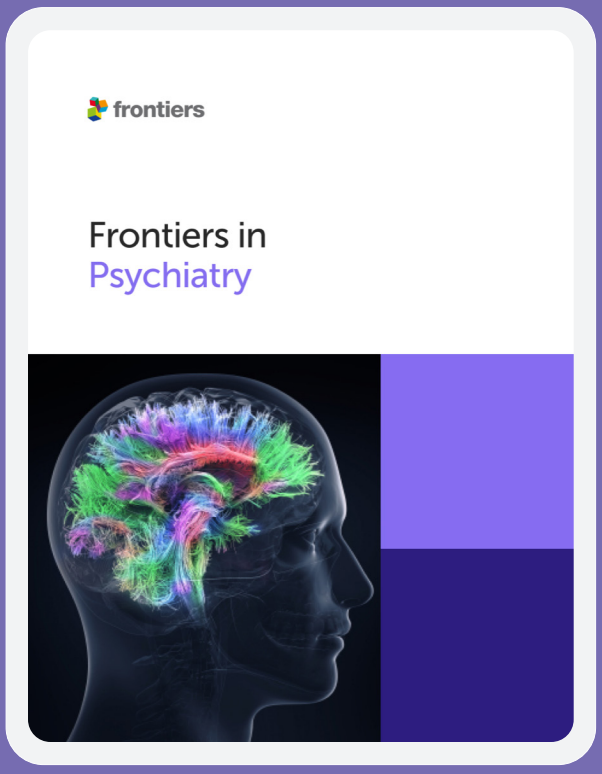
Frontiers in Psychiatry

Explores and communicates innovation in the field of psychiatry to improve patient outcomes

The third most-cited journal in its field, using translational approaches to improve therapeutic options for mental illness, communicate progress to clinicians and researchers, and consequently to improve patient treatment outcomes.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact