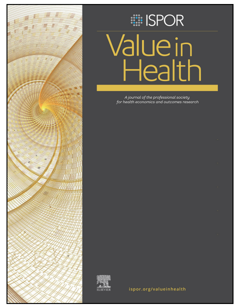
Article:

de Oliveira, Claire orcid.org/0000-0003-3961-6008, Matias, Maria Ana orcid.org/0000-0002-8782-3311 and Jacobs, Rowena orcid.org/0000-0001-5225-6321 (2024) Microsimulation models on mental health : a critical review of the literature. Value in Health. pp. 226-246. ISSN 1524-4733

https://doi.org/10.1016/j.jval.2023.10.015



Microsimulation models on mental health: a critical review of the literature

Claire de Oliveira, PhD, Maria Ana Matias, PhD, Rowena Jacobs, PhD

PII: S1098-3015(23)06162-4

DOI: https://doi.org/10.1016/j.jval.2023.10.015

Reference: JVAL 3914

To appear in: *Value in Health*

Received Date: 30 May 2023

Revised Date: 20 September 2023

Accepted Date: 26 October 2023

Target Journal: Value in Health

Title: Microsimulation models on mental health: a critical review of the literature

Authors: Claire de Oliveira PhD,*[1,2] Maria Ana Matias PhD,*[3] Rowena Jacobs[3] PhD

1. Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Canada
2. Institute of Health Policy, Management and Evaluation, University of Toronto, Canada
3. Centre for Health Economics, University of York, United Kingdom

* both authors are first authors

Corresponding author information:

Claire de Oliveira, PhD

Institute for Mental Health Policy Research

Centre for Addiction and Mental Health

Canada, Ontario, Toronto

Email: claire.deoliveira@camh.ca

Phone: 416-535-8501

Precis: This study sought to retrieve, synthesise, and appraise the existing literature on microsimulation models focused on mental health or that included a mental health component.

Word count: 4,108

Number of pages: 16

Number of figures: 1

Number of tables: 4

Supplementary materials:

Pages: 3

Figures: 0

Tables: 0

Author contributions:

Concept and design: de Oliveira, Jacobs

Acquisition of data: Matias

Analysis and interpretation of data: de Oliveira, Matias, Jacobs

Drafting of the manuscript: de Oliveira

Critical revision of the paper for important intellectual content: de Oliveira, Matias, Jacobs

Obtaining funding: de Oliveira, Jacobs

Supervision: de Oliveira

Funding: Funding to undertake this work was provided by The Closing the Gap network, which is funded by UK Research and Innovation (UKRI).

Acknowledgments: None.

Highlights

1. A cursory glance of the literature suggests there are few microsimulation models on mental health. Furthermore, there has only been one systematic review examining simulation modelling (e.g., microsimulation, discrete event simulation, Markov modeling) applied to mental health. However, this review is now outdated; moreover, it only synthesised the literature and did not critically appraise the models found.

2. This review focused only on studies that employed microsimulation models on mental health and found few microsimulation models on the topic. Moreover, this review critically appraised the models and found that few models were of high quality as many employed model inputs based on self-reported and/or cross-sectional data and small and/or non-representative samples; few undertook model validation.

3. The findings and recommendations from this review will be relevant to researchers and decision makers looking to build robust mental health-specific microsimulation models that can be used to inform policy development and guide health care delivery and service planning.

Microsimulation models on mental health: a critical review of the literature

Abstract

Objective: To retrieve and synthesise the literature on existing mental health-specific microsimulation models or generic microsimulation models used to examine mental health, and to critically appraise them.

Methods: All studies on microsimulation and mental health published in English in MEDLINE, Embase, PsycINFO, and EconLit between January 1, 2010, and September 30, 2022, were considered. Snowballing, Google searches and searches on specific journal websites were also undertaken. Data extraction was done on all studies retrieved and the reporting quality of each model was assessed using the Quality Assessment Reporting for Microsimulation Models checklist, a checklist developed by the research team. A narrative synthesis approach was used to synthesise the evidence.

Results: Among 227 potential hits, 19 studies were found to be relevant. Some studies covered existing economic-demographic models, which included a component on mental health and were used to answer mental health-related research questions. Other studies were focused solely on mental health and included models that were developed to examine the impact of specific policies and/or interventions on specific mental disorders. Most models examined were of medium quality. The main limitations included the use of model inputs based on self-reported and/or cross-sectional data, small and/or non-representative samples and simplifying assumptions, and lack of model validation.

Conclusions: This review found few high-quality microsimulation models on mental health. Microsimulation models developed specifically to examine mental health are important to guide health care delivery and service planning. Future research should focus on developing high-quality mental health-specific microsimulation models with wide applicability and multiple functionalities.

Keywords: mental health, mental illness, mental disorders, microsimulation, simulation, review

Introduction

In a time of limited health care resources, it is crucial to make informed decisions around resource allocation; to do so, decision makers require robust information to make sound investments. Microsimulation models are computer-based models that can be used to simulate the behaviour of micro-entities such as individuals or families.[1] They are commonly used to estimate the potential behavioural and economic effects of interventions and/or health policies, and to help guide decision-making.[1,2] Microsimulation models can also be used to test scenarios, which cannot be tested in the real world, such as trials.

According to Arias et al., 2022, in 2019 alone, 418 million disability-adjusted life-years (DALYs) could be attributable to mental disorders (i.e., 16% of global DALYs),[3] while the economic burden was estimated at roughly USD $5 trillion.[4] Given the large health and economic burdens of mental disorders, it is important to have robust microsimulation models on mental health to help decision makers make timely and cost-effective decisions to improve patient outcomes. Many existing microsimulation models, such as the Population Health Model (POHEM),[5] the United Kingdom Prospective Diabetes Study Outcomes Model,[6] and the Canadian Partnership Against Cancer's OncoSim model,[7] focus on physical health conditions, such as heart disease, osteoarthritis, diabetes, and cancer. It is not clear how many microsimulation models focus on mental health or on specific mental disorders and how these models have been developed.

A cursory glance at the literature suggests there are few microsimulation models on mental health.[8] One systematic review examined studies that employed simulation modelling (e.g., microsimulation, discrete event simulation, Markov modeling) in mental health.[9] This review found 10 papers (6.3% of all papers included) that used microsimulation to examine different mental disorders in several areas, such as medical decision making and treatment evaluation, prevention and screening, and health care design and planning. However, this review is now outdated, as the search only covers studies published until 2016; moreover, it only synthesised the literature and did not critically appraise the models found. This present review sought to address some of these limitations while focusing on studies/models that only used microsimulation to examine mental disorders. Although microsimulation models can be computationally intensive[10,11] and the size and complexity of a typical model can make it difficult to understand its properties intuitively (which may explain why microsimulation has not been widely adopted in the economics field),[12] microsimulation models have many advantages compared to other types of models. For example, microsimulation models represent hypothetical patients as unique individuals as opposed to average members of a representative cohort (cohort modeling) and can accommodate patient heterogeneity and interdependent health states, making them a more attractive alternative compared to other types of models.[10,13] Furthermore, from a technical point of view, microsimulation models are not subject to the restrictions that are common among other modeling approaches as they can handle any number of variables of any type. In addition to focusing on more recent literature on microsimulation models, this review also assessed the quality of these models. To encourage the uptake of microsimulation models by policy makers, it is important to ensure models are robust and produce valid outputs and thus are of high quality. Therefore, the objectives of this systematic literature review were to retrieve and synthesise the literature on existing microsimulation models focused on mental health or microsimulation models used to examine mental health, and to critically appraise them.

Methods

Study design

A systematic literature review was undertaken to identify the most significant papers on microsimulation models applied to mental health. To guide the analysis, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement[14] was followed.

Eligibility criteria

The population, intervention, control, outcomes, and study design (i.e., PICOS) criteria were used to guide the development of the search strategy and to inform the inclusion and exclusion criteria. The population of interest included all individuals (children, adolescents, and adults) diagnosed with mental illness, while the study design had to be a microsimulation model. There were no restrictions in terms of the type of outcomes examined (as long as these were mental health outcomes), while the control/comparison group criterion was not applicable within the context of this review. Only original studies were considered but reviews were examined, where available, to obtain studies that may not have been captured by the initial search.

Search strategy

All studies published in English (only) in MEDLINE (via PubMed), Embase (via OVID), PsycINFO (via OVID), and EconLit (via OVID) between January 1, 2010, and September 30, 2022, were considered. In addition to the database searches, additional searches were undertaken to increase the number of potential hits, namely Google searches (using Google and Google scholar), hand searches of relevant journals (e.g., International Journal of Microsimulation), hand searches of references of key papers and reviews (i.e., snowballing), and targeted searches on specific websites (e.g., https://www.microsimulation.ac.uk/). To guide the search, search terms/strings by concepts were developed; these can be found in **Table 1** (the full search strategies can be found in the supplementary materials).

Study selection

Once all relevant studies were retrieved and duplicates were removed, two reviewers (MAM and CdO) screened all titles and abstracts and an additional reviewer (RJ) was brought in for discussion, if/where necessary (this happened in no instances). Studies were excluded either because they did not examine mental health and/or did not include a microsimulation model. Studies that made use of microsimulation techniques in their analysis but did not include a microsimulation model, such as studies of economic evaluations, were excluded as the focus of the review was on studies that described and/or included stand-alone microsimulation models that could be used for multiple purposes. Subsequently, all relevant full text articles were retrieved and screened by one reviewer (MAM) to confirm final eligibility and additional reviewers (CdO and RJ) were brought in, if/where necessary (this happened in a few instances).

Data extraction

The data extraction form, based on the Cochrane good practice data extraction form,[15] was developed by the research team and included the following elements: study information (author(s), year of publication, country); aim of the model; data source(s) and study population (entire population, children, adolescents, adults); mental disorder(s) and outcomes examined; type of microsimulation model (static, dynamic, spatial) and description of model and methods;

validation (i.e., internal validation, where simulated outcomes are compared to actual outcomes, and external validation, where model forecasts are compared to other forecasts), robustness checks, and model adjustments, such as calibration (i.e., parameter adjustments to ensure the model is able to simulate the distributions of key variables), where applicable; and limitations of the model. One reviewer (MAM) undertook the data extraction for each study and an additional reviewer (CdO) was assigned to review each entry for accuracy and consistency.

Quality assessment

Assessing the quality of microsimulation models is important to motivate their use by both academics and policy makers,[16] who are typically looking for robust models to guide decision making. There are many potentially relevant dimensions of quality; however, it can be challenging to discuss quality in abstract or general terms.[16] The existing quality reporting guidelines in health economics are mainly focused on economic evaluations. For example, the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist is typically used to evaluate the quality of reporting of economic evaluations,[17] some of which may use microsimulation techniques; however, issues like study perspective and measurement of effectiveness and costs may not be relevant when assessing microsimulation models. Other reporting checklists, such as the 2014 International Society for Pharmacoeconomics and Outcomes Research (ISPOR)[18] and the 2016 Assessment of the Validation Status of Health-Economic Decision Models (AdViSHE) checklists[19] were mainly developed to examine the credibility and assess the quality of economic models, respectively, and thus not all elements of these checklists are relevant to microsimulation models. Finally, there is a reporting quality checklist developed for discrete event simulations in health care.[20] This checklist is specific to discrete event simulations only and, while relevant, it is not directly applicable to other types of simulation models. Therefore, a reporting quality assessment checklist was developed specifically for this study – the Quality Assessment Reporting for Microsimulation Models (QARMM) checklist. The six main groups of the QARMM checklist were largely based on work done by Sutherland (2018), described as important dimensions to account for when considering the quality of microsimulation models,[16] as well as the reporting quality checklist developed for discrete event simulations in health care.[20] The six criteria are as follows: purpose of the model, data, transparency, uncertainty, validation, and generalisability, where the first three relate to the model development and structure and last three relate to the validity and scope of the results produced by the model; each item was determined to be worth one point (half points were given in cases where the criterion was not fully met). See **Table 2** for the proposed reporting quality checklist for microsimulation models. One reviewer (MAM) undertook the quality assessment of each model, and an additional reviewer (CdO) was assigned to review the quality assessment and resolve any disagreements, if/where necessary; any additional disagreements were resolved by a third reviewer (RJ) (this occurred in no instances). Models with a score greater than five (> 5/6) were considered high quality models (i.e., cases where >= 90% of criteria were met), while models with a score of three (=< 3/6) or less were considered low quality (i.e., cases where =< 50% of criteria were met); all models with scores in between were deemed medium quality models. It was decided that only models that met all (or almost all) criteria would be classified as high-quality models.

Data synthesis

Given the heterogeneity of studies and outcomes examined, undertaking a meta-analysis was not feasible. Therefore, a narrative synthesis approach was used to synthesise the evidence informed by the quality assessment.[21] In particular, the evidence was synthesised by study type (studies on generic models with built-in mental health components vs. studies on mental health-specific models), model type (static vs. dynamic; non-spatial vs. spatial), and mental disorder.

Results

Study selection

After all citations were merged across all databases (n = 294) and duplicates were removed (n = 67), the search produced 227 unique records. After titles and abstracts were reviewed, 36 full texts were assessed. Among these, 17 studies were ultimately included in the final review along with two other articles obtained from additional sources (i.e., Google search),[22-40] for a total of 19 studies (see **Figure 1**). As a result, there was an 8.4% (19/227) of relevant hits. The studies retrieved could be grouped into two groups: studies on existing economic-demographic models that included a component on mental health,[22-32] such as the Health&WealthMOD[22-27] and the Future Americans Model,[29] used to answer research questions on mental health (n = 11), and studies focused solely on mental health models developed to examine the impact of specific policies and/or interventions on mental disorders[33-40] (e.g., to understand the impact of social media on post-traumatic stress disorder (PTSD), examine the impact of smoking cessation initiatives among individuals with severe mental illness, or determine disease prevalence) (n = 8).

Study characterisation

Most models were from either the USA or Australia. For the first set of studies (i.e., studies on existing economic-demographic models that included a component on mental health), 7 models were based on data from Australia,[22-28] and one from the USA,[29] Ireland,[30] Scotland,[31] and England[32] each. For the second set of studies (studies focused solely on mental health models developed to examine the impact of specific policies and/or interventions on mental disorders), five models were based on data from the USA,[33,36-39] one from England,[40] and two used data that were not country-specific[34,35] as their aims were to create hypothetical patient populations (see **Table 3**). The aims of the models were diverse, from estimating long-term costs of lost productivity due to mental health to examining the long-term impact of policy changes on mental health-related outcomes. The main type of microsimulation models examined were static microsimulation models (n = 17), of which three were static spatial microsimulation models;[30,31,40] only two were dynamic microsimulation models.[29,32] The data sources employed in the models included survey data,[22-33,36,37,40] administrative/claims data,[33,36] and population statistics/Census data.[31-33,40] Some models also used data from existing models and published work (e.g., peer-reviewed studies, grey literature, and clinical trials).[33-35,37-39] Regarding the study population, most models (n = 14) focused on the general adolescent and adult populations (14 years and older). Among the other models, three models did not specify age/age range – one model focused on specifically on soldiers deployed to Operation Iraqi Freedom or Operation Enduring Freedom[33] and the other two examined hypothetical populations.[34,35] Only one examined individuals of all ages,[31] while the Population Ageing and Care Simulation (PACSim) model[32] focused solely on older individuals only (i.e., 65 years and older). A large range of mental disorders were examined. The studies using the Health&WealthMOD, a microsimulation model of health and disability and associated impacts on labour force participation, personal income, savings, and government revenue and expenditure, examined depression/mood affective disorders and other mental and behavioural disorders (n = 7). Apart from these studies, the most common mental disorder examined was schizophrenia/psychosis (n = 3), followed by severe mental illness (defined as psychosis, bipolar disorder, and major depression) (n = 2), and psychological/mental distress (n = 2). The other studies examined depression (n = 1), PTSD, major depression and comorbid PTSD, and major depression (n = 1), PTSD (n = 1), eating disorders (n = 1), and several mental disorders (n = 1). Some studies also examined other disorders in addition to mental health disorders such as the PACSim model,[32] a dynamic microsimulation model that simulates sociodemographic factors, health behaviours, chronic diseases and geriatric conditions of individuals, which examined the prevalence of multimorbidity (chronic heart disease, stroke, hypertension, diabetes, arthritis, cancer, respiratory disease, dementia, hearing impairment, vision impairment and cognitive impairment as well as depression), diabetes among individuals with schizophrenia,[36] and heavy alcohol consumption alongside psychological distress.[40] The outcomes examined in the models were diverse and included disease prevalence,[22-28,30,32,37,39,40] health-related outcomes (e.g., admissions[34,35] and life expectancy[29,32]), economic outcomes (e.g., labour force participation/employment[23,27,34,35]), and costs of mental illness.[32,33,36]

Only three models[29,31,32] undertook validation (however, among these, one undertook validation for all outcomes, except depression[32]). For example, the Future Americans Model, a dynamic microsimulation, which projects health, medical spending, social service use, and economic outcomes over time, was validated both internally and externally,[29] while the outputs of the SimAlba Model, a spatial microsimulation model used to estimate geographically sensitive health variables for Glasgow, were extensively compared against known Census totals.[31] However, three models[32,39,40] undertook calibration of model variables using empirical data (e.g., survey data, Census data) and several models undertook other robustness checks, such as sensitivity analyses.[34-36] The microsimulation models examined had many strengths, such as filling in knowledge gaps,[22,26-30,33] or providing relevant outputs for policy makers.[22-24,39] The main limitations of these models were the use of model inputs based on self-reported information, cross-sectional data, and small samples,[22-29,32,40] lack of generalisability (or generalisability not discussed),[34,37] and/or use of simplifying assumptions (e.g., one model employed a single state variable to represent the patient's willingness to quit smoking;[38] another model did not take into account other risk factors associated with eating disorders, such as other mental disorders, substance use, family history, or sexual orientation[39]).

Quality assessment

According to the quality assessment made using the QARMM checklist, most models (16/19) were classified as being of medium quality; two (2/19) models were considered to be of high quality,[29,31] while one (1/19) was deemed to be of low quality[38] (see **Table 4**). Generally, all studies performed well on all three criteria included under model development and structure (i.e., purpose of the model, data, and transparency). One model[38] used data from a randomised controlled trial but did not specify the data source and thus was only given 0.5 (however, it is worth noting that the goal of this study was not to develop a microsimulation model *per se* but rather to provide an illustration of a microsimulation application). Overall scores were much lower for the criteria under validity and scope of the results produced by the model, particularly for validation. Regarding generalisability, only four studies obtained full points,[29,34,36,37] while the remainder were given 0.5. This was typically the case for models where limitations were discussed but model generalisability was not (or at least not explicitly). Overall, most studies were attributed full points for uncertainty as model uncertainties were discussed and sensitivity analyses were performed and reported, where relevant (in some cases, some of the sensitivity analyses were undertaken in the original publication that described the model).

Evidence synthesis

Several studies[22-27] captured in this review employed an existing model, such as the Health&WealthMOD and the Care&WorkMOD, which were classified as medium quality models. The Health&WealthMOD models health and disability and associated impacts on labour force participation, personal income, savings, and government revenue and expenditure,[22-27] while the Care&WorkMOD is a microsimulation model designed to project the financial costs of reduced capacity to work due to provision of care.[28] Although validation was not performed for these specific models, the inputs that make up the models were obtained from validated models; if the validation of these other models had been considered, the Health&WealthMOD and the Care&WorkMOD models would likely have been classified as high-quality models in this review. The main strength of the Health&WealthMOD is its ability to produce a variety of policy relevant outputs, such as costs of mental illness, labour market outcomes (e.g., labour force participation), income/wealth, and disease prevalence. Future microsimulation models should strive to have multiple functionalities. Excluding these models, most studies on generic models with a built-in mental health component were considered to be of relatively high quality. The studies on mental health-specific models varied in terms of quality, with some being of higher quality[34,36,40] and some of lower quality.[33,38] While it is ideal to have models specifically built/focused on mental health, there are some existing models built for other purposes that can still be used to a satisfactory extent to examine research questions specifically targeted for mental health-related purposes. Regarding type of model, all static spatial models and one dynamic model were classified as being of higher quality. While the spatial models can only be used to produce outputs at the small area/regional level,[30,31,40] the Future Americans Model[29] can be used to examine many outcomes, such as life expectancy, quality of life, medical care spending, and economic outcomes, making it a very versatile model. These models were also among the few that undertook model validation. There was no pattern between the types of mental disorder studied/modelled and the quality of the models. Higher quality models examined a series of mental disorders, from schizophrenia to severe mental illness to mental/psychological distress and depression.

Discussion

Mental disorders are associated with large health and economic burdens, with many impacts across society. It is important to have robust microsimulation models on mental health to help decision makers make informed decisions regarding policy implementation and resource allocation. This review sought to retrieve and synthesise the literature on existing microsimulation models used to examine mental health or those focused on mental health, and to critically appraise them.

Overall, the review found few microsimulation models on mental health, particularly those that had been developed solely to examine mental health as most were generic economic-demographic models, which included a component on mental health. Only papers on models that included a mental health component and showed how that component was used were included. This was decided because it would not have been possible to fully evaluate a model without a mental health application. Moreover, microsimulation can be used in several different contexts. For example, there were several studies on economic evaluations of mental health interventions that employed microsimulation techniques within their analyses. These studies were ultimately not included, as the main objective of this review was to examine studies that covered or developed microsimulation models and not studies that made use of microsimulation techniques broadly speaking.

Most models examined in this review were of medium quality; only two models were considered high quality as they met all criteria required to be considered a robust microsimulation model. Based on the quality assessment reporting checklist, the main areas where models lacked quality/ robustness were validation, or rather lack thereof, followed by generalisability, namely the lack of discussion around generalisability. The main limitations of the models examined were the use of model inputs based on self-reported data (namely survey data),[22-30,32] which can be subject to reporting bias[41] due to stigma associated with some mental disorders; reliance on cross-sectional data,[24-27,30,31,40] which does not enable examining changes in individuals' behaviour over time; use of small samples and/or patient populations,[24,28,29] which are not representative of the entire population (e.g., exclusion of incarcerated, unhoused, or institutionalised individuals); simplifying assumptions,[23,28,33,37,39] which may lead to unrealistic model structures; and lack of validation,[22-28,33-39] which may question the robustness of the outputs produced by the model. Nonetheless, it is worth noting that some limitations many not necessarily be by choice but rather a feature of the available mental health data, especially for rarer mental illnesses, such as severe mental illness, where only survey data are available,[29] and for vulnerable populations, such as unhoused individuals with mental illness, where population-based samples may be challenging to obtain.[42] Thus, better data capture is likely needed first before models can be improved. In the absence of ideal data, modellers should use whatever data are available but always highlight any potential sample selection issues (e.g., inclusion of healthier individuals or individuals with less stigmatizing mental disorders). These limitations should be considered when developing robust microsimulation models in the future. Moreover, given that few models examined mental health disorders alongside other disorders, future microsimulation models should consider modelling the interplay between mental, physical and substance use disorders, where relevant, as these interactions can have important implications on several levels. For example, prior work has shown that individuals with chronic psychotic disorders can have a multitude of chronic physical health disorders, which impact how they interact with the health care system, and related health care expenditures, as well health outcomes.[43] Thus, a comprehensive microsimulation model should consider comorbidities when modelling lifetime outcomes, such as disease prevalence and resulting survival/life expectancy.

This review examined studies that included microsimulation models on mental health; given that most existing microsimulation models have mainly focused on physical health, this review sheds light on an important gap in the literature. This review expanded on previous literature[9] by assessing the quality of studies included and developed a quality assessment reporting checklist for microsimulation studies, based on several recommendations.[17,21] To the authors' knowledge, a checklist of this sort is currently lacking. However, the present review is not without limitations. This systematic literature review examined literature on microsimulation models from 2010 onwards only (though it is likely that few relevant microsimulation studies were published before then). Furthermore, this present review only examined literature published in English only in four databases and did not consider grey literature. Finally, the reporting quality assessment checklist developed assigned 1 point to each item (allowing for half-points, where applicable). While many checklists score each item equally (e.g., CHEERS), there may be some elements that are more important than others when developing microsimulation models (e.g., validation). However, creating different scores for each item required additional value judgments and the involvement of experts, which was beyond the main purpose of this study. Future work should seek to explore the use of different weighting criteria and validate the reporting quality assessment checklist for microsimulation models developed in this study.

The suggestions provided in this review will be relevant to researchers and decision makers looking to build mental health-specific microsimulation models to inform policy development and guide future health care delivery and service planning. Given that few microsimulation models have been developed for mental disorders, more work needs to be done in this space. In particular, future work should focus on developing mental health-specific microsimulation models with wide applicability and multiple functionalities, such as the possibility of examining the interdependence between mental and physical health.

Conclusion

A limited number of microsimulation models have been developed specifically for mental health disorders. Among the existing studies examining microsimulation models, few models have been developed solely to examine mental health as most were generic economic-demographic models, which included a component on mental health. Moreover, few microsimulation models were found to be of high quality; many used inputs based on self-reported data or cross-sectional data and the use of simplifying assumptions, and few did not undertake any type of model validation. Future research should focus on developing high-quality mental health-specific microsimulation models with wide applicability (i.e., applicable to all individuals living with a given mental disorder) and multiple functionalities (i.e., capable of modelling several policy relevant outcomes).

References

1. Abraham JM. Using microsimulation models to inform U.S. health policy making. *Health Serv Res.* 2013; 48(2 Pt 2):686-695.

2. Zucchelli E, Jones AM, Rice N. The evaluation of health policies through dynamic microsimulation methods. *International Journal of Microsimulation*. 2012;5(1):2-20.

3. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. *EClinicalMedicine*. 2022;54:101675.

4. The Lancet Global Health. Mental health matters. *Lancet Glob Health*. 2020;8(11):e1352.

5. Flanagan W. Overview of the Population Health Model (POHEM). Statistics Canada. 2008.

6. Clarke PM, Gray AM, Briggs A, et al.; UK Prospective Diabetes Study (UKDPS) Group. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia*. 2004;47(10):1747-1759.

7. Gauvreau CL, Fitzgerald NR, Memon S, et al. The OncoSim model: development and use for better decision-making in Canadian cancer control. *Curr Oncol*. 2017;24(6):401-440.

8. DJ Schofield, Zeppel MJB, Tan O, Lymer S, Cunich MM, Shrestha RN. A brief, global history of microsimulation models in health: Past applications, lessons learned and future directions. *International Journal of Microsimulation*. 2018;11(1):97-142.

9. Long KM, Meadows GN. Simulation modelling in mental health: A systematic review. *Journal of Simulation*. 2018;12(1):76-85.

10. 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 Mak*. 2012;32(5):678-689.

11. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force–7. *Value Health*. 2012;15(6):843-850.

12. Klevmarken A. Microsimulation. A Tool for Economic Analysis. *International Journal of Microsimulation*. 2022;15(1); 6-14.

13. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force–3. *Value Health*. 2012;15(6):812-820.

14. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PloS Med*. 2009;6(7):e1000097.

15. Cochrane Effective Practice and Organisation of Care (EPOC). Data collection form. EPOC Resources for review authors, 2017. Epoc.cochrane.org/resources/epoc-specific-resources-review-authors.

16. Sutherland H. Quality assessment of microsimulation models. The case of EUROMOD. *International Journal of Microsimulation*. 2018;11(1):198-223.

17. Husereau D, Drummond M, Augustovski F, et al.; CHEERS 2022 ISPOR Good Research Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated Reporting Guidance for Health Economic Evaluations. *Value Health*. 2022;25(1):3-9.

18. Caro J, Eddy DM, Kan H, et al.; ISPOR-AMCP-NPC Modeling CER Task Forces. Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health*. 2014;17(2):174-182.

19. Vemer P, Corro Ramos I, van Voorn GA, Al MJ, Feenstra TL. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics*. 2016;34:349-361.

20. Zhang X, Lhachimi SK, Rogowski WH. Reporting Quality of Discrete Event Simulations in Healthcare-Results From a Generic Reporting Checklist. *Value Health*. 2020;23(4):506-514.

21. Popay J, Roberts H, Sowden A, et al. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, 1(1), b92.

22. Schofield D, Cunich M, Shrestha R, et al. Indirect costs of depression and other mental and behavioural disorders for Australia from 2015 to 2030. *BJPsych Open*. 2019;5(3):e40.

23. Veerman JL, Shrestha RN, Mihalopoulos C, et al. Depression prevention, labour force participation and income of older working aged Australians: A microsimulation economic analysis. *Aust N Z J Psychiatry*. 2015;49(5):430-436.

24. Schofield DJ, Shrestha RN, Percival R, Passey ME, Callander EJ, Kelly SJ. The personal and national costs of mental health conditions: impacts on income, taxes, government support payments due to lost labour force participation. *BMC Psychiatry*. 2011;11:72.

25. Schofield DJ, Kelly SJ, Shrestha RN, Callander EJ, Percival R, Passey ME. How depression and other mental health problems can affect future living standards of those out of the labour force. *Aging Ment Health*. 2011;15(5):654-662.

26. Schofield DJ, Shrestha RN, Percival R, Kelly SJ, Passey ME, Callander EJ. Quantifying the effect of early retirement on the wealth of individuals with depression or other mental illness. *Br J Psychiatry*. 2011;198(2):123-128.

27. Schofield DJ, Callander EJ, Shrestha RN, Passey ME, Percival R, Kelly SJ. The indirect economic impacts of co-morbidities on people with depression. *J Psychiatr Res*. 2013;47(6):796-801.

28. Schofield D, Zeppel MJB, Tanton R, et al. Individual and national financial impacts of informal caring for people with mental illness in Australia, projected to 2030. *BJPsych Open*. 2022;8(4):e136.

29. Seabury SA, Axeen S, Pauley G, et al. Measuring The Lifetime Costs Of Serious Mental Illness And The Mitigating Effects Of Educational Attainment. *Health Aff (Millwood)*. 2019;38(4):652-659.

30. Morrissey K, Hynes S, Clarke G, O'Donoghue C. Examining the factors associated with depression at the small area level in Ireland using spatial microsimulation techniques. *Irish Geography*. 2010;43(1)1-22.

31. Campbell M, Ballas D. SimAlba: A Spatial Microsimulation Approach to the Analysis of Health Inequalities. *Front Public Health*. 2016;4:230.

32. Kingston A, Robinson L, Booth H, Knapp M, Jagger C; MODEM project. Projections of multi-morbidity in the older population in England to 2035: estimates from the Population Ageing and Care Simulation (PACSim) model. *Age Ageing*. 2018;47(3):374-380.

33. Kilmer B, Eibner C, Ringel JS, Pacula RL. Invisible wounds, visible savings? Using microsimulation to estimate the costs and savings associated with providing evidence-based treatment for PTSD and depression to veterans of Operation Enduring Freedom and Operation Iraqi Freedom. *Psychological Trauma: Theory, Research, Practice, and Policy*. 2011;3(2):201-211.

34. Horvitz-Lennon M, Predmore Z, Orr P, et al. Simulated long-term outcomes of early use of long-acting injectable antipsychotics in early schizophrenia. *Early Interv Psychiatry*. 2019;13(6):1357-1365.

35. Horvitz-Lennon M, Predmore Z, Orr P, et al. The Predicted Long-Term Benefits of Ensuring Timely Treatment and Medication Adherence in Early Schizophrenia. *Adm Policy Ment Health*. 2020;47(3):357-365.

36. Mulcahy AW, Normand SL, Newcomer JW, et al. Simulated Effects of Policies to Reduce Diabetes Risk Among Adults With Schizophrenia Receiving Antipsychotics. *Psychiatr Serv*. 2017;68(12):1280-1287.

37. Abdalla SM, Cohen GH, Tamrakar S, Koya SF, Galea S. Media Exposure and the Risk of Post-Traumatic Stress Disorder Following a Mass Traumatic Event: An *In-silico* Experiment. *Front Psychiatry*. 2021;12:674263.

38. Huang W, Chang CH, Stuart EA, et al. Agent-Based Modeling for Implementation Research: An Application to Tobacco Smoking Cessation for Persons with Serious Mental Illness. *Implement Res Pract*. 2021;2:10.1177/26334895211010664.

39. Ward ZJ, Rodriguez P, Wright DR, Austin SB, Long MW. Estimation of Eating Disorders Prevalence by Age and Associations With Mortality in a Simulated Nationally Representative US Cohort. *JAMA Netw Open*. 2019;2(10):e1912925.

40. Riva M, Smith DM. Generating small-area prevalence of psychological distress and alcohol consumption: validation of a spatial microsimulation method. *Soc Psychiatry Psychiatr Epidemiol*. 2012;47(5):745-55.

41. Mason J, Laporte A, McDonald T, Kurdyak P, de Oliveira C. Health Reporting from Different Data Sources: Does it Matter for Mental Health? The Journal of Mental Health Policy and Economics. 2023;25(1):33-57.

42. Richard L, Hwang SW, Forchuk C, et al. Validation study of health administrative data algorithms to identify individuals experiencing homelessness and estimate population prevalence of homelessness in Ontario, Canada. *BMJ Open*. 2019;9(10):e030221.

43. de Oliveira C, Iwajomo T, Kurdyak P. Health Care Expenditures Among Individuals With Chronic Psychotic Disorders in Ontario: An Analysis Over Time. *Front Health Serv*. 2022;2:848072.

Table 1. Concepts and search terms used to identify relevant studies

Concept	Search terms
Population	children, youth, teen, adolescents, adults, seniors
Intervention/exposure	mental disorder, mental health, mental illness, mental wellbeing, mental hygiene, severe mental illness, severe mental disorder, serious mental illness, serious mental disorder, schizophrenia, schizoaffective disorder, psychotic, psychosis, bipolar disorder, bipolar disease, mania, depression, major depression, unipolar depression, persistent depressive disorder, mood disorder, mood, dysthymia, anxiety, stress, phobia, panic disorder, neurosis, anorexia nervosa, bulimia nervosa, binge eating disorder, eating disorder, personality disorder, obsessive compulsive disorder, suicide, suicide ideation, self-harm, ADHD
Outcome	n/a
Study design	microsimulation

Note: All studies published in English (only) in MEDLINE, Embase, PsycINFO, and EconLit between January 1, 2010, and September 30, 2022, were considered.

Legend: ADHD – attention deficit hyperactivity disorder

Table 2. The Quality Assessment Reporting for Microsimulation Models (QARMM) checklist

Items	**Points**
Model development and structure	
Purpose of the model • Are the objectives/goals of the model well defined? • Is the target population (i.e., individuals with a given mental health condition/disorder) described appropriately?	1
Data • Are the data used in the model development representative of the population examined? • Are the data sources informing parameter estimations provided? • Are the parameters used to populate model frameworks specified?	1
Transparency • Is the model structure well described (e.g., are assumptions clear, are there choices for the user to make)? • Is the time horizon of the model provided? • Are all simulated strategies/scenarios specified and/or explained clearly?	1
Validity and scope	
Uncertainty • Are model uncertainties discussed? • Are sensitivity analyses performed and reported? • Is model calibration performed and reported, where required?	1
Validation • Is internal and/or external validation performed and reported? • Is predictive validation performed and/or attempted?	1
Generalisability • Is model generalisability discussed? • Are model limitations discussed?	1

Table 3. Details of included studies

Author, year, country	Aim of the model	Data source and study population	Mental disorder and outcomes examined	Type of microsimulation model and description of model and methods	Validation*	Limitations
Schofield et al., 2019; Australia	To quantify long-term costs of lost productive life-years due to mental disorders of Australians aged 45-64 years old	Outputs from the Health&WealthMOD2030. Databases used to build Health&WealthMOD2030: Australian Bureau of Statistics Survey of Disability, Ageing and Carers 2003 and 2009; Static Incomes Model 2013; Australian Population and Policy Simulation Model 2010-2030; the Treasury 2013-14 (population and labour force projections); Australian Burden of Disease and Injury Study 2003 Individuals aged 45-64 years old	Depression/mood affective disorders (excluding postpartum depression); other mental and behavioural disorders Costs of mental illness projected over 2015–2030	Static microsimulation model The Health&WealthMOD2030 dataset provides prevalence of diseases, socio-demographic, and economic characteristics of Australians aged 45-64 years old every 5 years from 2010 to 2030. Static ageing techniques were used to capture changes in population structure. To generate estimates of more detailed economic variables (e.g., income, income tax), synthetic matching was used to link the Survey of Disability, Ageing and Carers with the Australian Population and Policy Simulation Model. This dataset was used to estimate the indirect costs of depression and other mental and behavioural disorders. A counterfactual simulation using Monte Carlo methods was used to estimate differences in costs.	No validation was performed. However, the Australian Population and Policy Simulation Model and Static Incomes Model, used as inputs in the model, were validated.	Findings based on self-reported labour market behaviour and mental illness
Veerman et al., 2015; Australia	To quantify the potential economic impact of 5-yearly screening for sub-syndromal depression in general practice of Australians aged 45-64 years old	Outputs from the Health&WealthMOD2030. Databases used to build Health&WealthMOD2030: Australian Bureau of Statistics Survey of Disability, Ageing and Carers 2003 and 2009; Static Incomes Model 2013; Australian Population and Policy Simulation Model 2010-2030; the Treasury 2013-14 (population and labour force projections); Australian Burden of Disease and Injury Study 2003 Individuals aged 45-64 years old	Depression/mood affective disorders (excluding postpartum depression); other mental and behavioural disorders Labour force participation, personal income, tax paid, and transfer income	Static microsimulation model The Health&WealthMOD2030 dataset provides prevalence of diseases, socio-demographic and economic characteristics of Australians aged 45-64 years old every 5 years from 2010 to 2030. Static ageing techniques were used to capture changes in population structure. To generate estimates of more detailed economic variables (e.g., income, income tax), synthetic matching was used to link the Survey of Disability, Ageing and Carers with the Australian Population and Policy Simulation Model. This dataset was used to estimate the prevalence of depression and labour force participation rates and the impact of depression prevention interventions on personal income, savings, taxation revenue, and welfare expenditure. An epidemiological Markov model was used to estimate reductions in the prevalence of depression if interventions had been in place.	No validation was performed. However, the Australian Population and Policy Simulation Model and Static Incomes Model, used as inputs in the model, were validated.	i) Assumed that the reduction in self-reported depression due to an intervention would be similar to a reduction in diagnosed depression; ii) Assumption of uniform reductions in depression prevalence across all levels of severity

Schofield et al., 2011a; Australia	To quantify the personal cost of lost income and the cost to the state as a result of early retirement due to mental health conditions in Australians aged 45-64 years old in 2009	Outputs from the Health&WealthMOD. Databases used to build Health&WealthMOD: Australian Bureau of Statistics Survey of Disability, Ageing and Carers 2003; Static Incomes Model 2003 Individuals aged 45-64 years old	Depression/mood affective disorders (excluding postpartum depression); other mental and behavioural disorders Weekly income, weekly transfer income received by individuals, and weekly tax liability paid by individuals	Static microsimulation model The Health&WealthMOD provides the economic impacts of illness on retirement and projects retirement due to illness to 2020. To generate estimates of income and wealth, synthetic matching was used to link the Survey of Disability, Ageing and Carers with the Static Incomes Model. This dataset was used to quantify the personal cost of lost income and the cost to the state from lost income taxation, increased benefits payments, and lost GDP as a result of early retirement due to mental health conditions. Multiple linear regression model was used to analyse differences between weekly incomes of people in the labour force with no health condition and people not in the labour force due to depression/other mental health conditions.	No validation was performed. However, the Static Incomes Model, used as an input in the model, was validated.	i) Results based on a relatively small sample size of individuals who were not in the labour force due to depression/other mental disorders; ii) Results based on cross-sectional data and self-reported data
Schofield et al., 2011b; Australia	To determine the impact of early retirement due to depression and other mental health problems on future retirement savings of Australians aged 45-64 years old in 2009	Outputs from the Health&WealthMOD. Databases used to build Health&WealthMOD: Australian Bureau of Statistics Survey of Disability, Ageing and Carers 2003; Static Incomes Model 2003 Individuals aged 45-64 years old	Depression/mood affective disorders (excluding postpartum depression); other mental and behavioural disorders Value of savings at age 65	Static microsimulation model The Health&WealthMOD provides the economic impacts of illness on retirement and projects retirement due to illness to 2020. To generate estimates of income and wealth, synthetic matching was used to link the Survey of Disability, Ageing and Carers with Static Incomes Model. Accumulated savings were determined by the model. Multiple linear regression model was used to analyse the differences between savings and annuity of people working full-time with no chronic condition, persons working part-time with no chronic condition, and people not in the labour force due to depression and other mental health problems.	No validation was performed. However, the Static Incomes Model, used as an input in the model, was validated.	Results based on cross-sectional data (not possible to know individuals' economic status before onset of disease); ii) Results based on self-reported health status

Schofield et al., 2011c; Australia	To quantify costs of lost savings and wealth to Australians aged 45-64 years old who retire early due to depression or other mental illness	Outputs from the Health&WealthMOD. Databases used to build Health&WealthMOD: Australian Bureau of Statistics Survey of Disability, Ageing and Carers 2003; Static Incomes Model 2003 Individuals aged 45-64 years old	Depression/mood affective disorders (excluding postpartum depression); other mental and behavioural disorders Wealth for individuals employed full- and part-time, and those not in labour force owing to depression/other mental illness	Static microsimulation model The Health&WealthMOD provides the economic impacts of illness on retirement and projects retirement due to illness to 2020. To generate estimates of income and wealth, synthetic matching was used to link the Survey of Disability, Ageing and Carers with Static Incomes Model. Logistic regression model was used to compare the odds of owning wealth by those who reported being out of the labour force due to depression/other mental illness with those who were in full-time work and had no chronic condition. Multiple linear regression model was used to analyse the differences between the wealth of people working full time with no chronic condition, people working part time with no chronic condition and people not in the labour force because of depression and other mental illness.	No validation was performed. However, the Static Incomes Model, used as an input in the model, was validated.	i) Not clear how long individuals in this study had been out of the labour force; ii) Results based on cross-sectional data and self-reported health status
Schofield et al., 2013; Australia	To quantify the association between co-morbid health conditions and labour force status and economic circumstances of Australians aged 45-64 years old with depression	Outputs from the Health&WealthMOD. Databases used to build Health&WealthMOD: Australian Bureau of Statistics Survey of Disability, Ageing and Carers 2003; Static Incomes Model 2003 Individuals aged 45-64 years old	Depression/mood affective disorders (excluding postpartum depression) Number/proportion of individuals with and without chronic health condition and/or depression, proportion of individuals in/not in the labour force, weekly private income, transfer payments and tax liability	Static microsimulation model The Health&WealthMOD provides the economic impacts of illness on retirement and projects retirement due to illness to 2020. To generate estimates of income and wealth, synthetic matching was used to link the Survey of Disability, Ageing and Carers with Static Incomes Model. Multiple linear regression model used to analyse differences between weekly private incomes/transfer income/tax liability.	No validation was performed. However, the Static Incomes Model, used as an input in the model, was validated.	i) Not clear how long individuals in this study had been out of the labour force; ii) Results based on cross-sectional data and self-reported health status
Schofield et al., 2022; Australia	To estimate costs of lost labour force participation among primary carers due to the provision of informal care for people with mental illness in Australia from 2015 to 2030	Outputs from the Care&WorkMOD. Databases used to build Care&WorkMOD: Australian Bureau of Statistics Survey of Disability, Ageing and Carers 2003, 2009 and 2012; Intergenerational Report 2015; Australian Population and Policy Simulation Model; and Static Incomes Model 2015 Primary carers of individuals with mental illness aged 15-64 years old	Depression/mood affective disorders, dementia, schizophrenia, nervous tension/stress, phobic and anxiety disorders and other mental and behavioural disorders Weekly income, weekly welfare payments and weekly tax payments	Static microsimulation model The Care&WorkMOD provides the economic costs of early exit from the labour force from 2015 to 2030. To generate estimates of economic variables, synthetic matching was used to link the Survey of Disability, Ageing and Carers with Static Incomes Model. Differences in financial costs for those not in the labour force (lost productive life-years) owing to care provision compared with those for people in the labour force who were not providing care were estimated from counterfactuals using Monte Carlo methods. Similar analysis undertaken to estimate differences in financial outcomes of carers	No validation was performed. However, the Australian Population and Policy Simulation Model and Static Incomes Model, used as inputs in the model, were validated.	i) Results based on self-reported data; ii) Study only considered carers aged 15-64 years caring for people in the same household; iii) Results based on 88 survey records of informal carers who were out of the labour force due to caring for someone with mental illness; iv) Study only focused on main reason for being out of the labour force; v) Results did not account for possibility of informal carers working part-time or working in a lower-paid full-time position to support their caregiving needs;

				not in the labour force and people employed who were not carers.		vi) Due to small number of informal carers caring for specific mental illnesses, broader grouping of all informal carers caring for ‘mental and behavioural disorders’ was used; vii) Study data collected prior to the COVID-19 pandemic and thus does not include the mental health effects of COVID-19/strategies to contain the disease.
Seabury et al., 2019; USA	To project the impact of increased educational attainment on health and economic outcomes by age 25 among people diagnosed with serious mental illness	Outputs from the Future Americans Model (FAM). Databases used to build the FAM: Panel Study of Income Dynamics 1999-2015; Health and Retirement Study 1998-2012; Medical Expenditure Panel Survey 2001/03 and 2007/10; Medicare Current Beneficiary Survey 2007-2010 Individuals aged 25 years and older	Psychosis, bipolar disorder, or depression Life expectancy, quality of life, medical care spending, and economic outcomes	Dynamic microsimulation model The FAM calculates transition probabilities for specified health states (e.g., chronic disease incidence, functional status, body mass index, and mortality), which are modeled as first-order Markov processes, with probabilities based on predicted values from probit regressions. Chronic health conditions are treated as absorbing states. Serious mental illness is measured using the Kessler Psychological Distress (K6) Scale. Individuals' health care spending and economic outcomes are projected based on health transitions, functional status, body mass index, K6 score, and demographics. Economic outcomes are estimated by regressing spending on risk factors, health, and functional status. The resulting dataset was used to conduct cohort simulations based on two scenarios: i) status quo, the current lifetime burden of SMI across a wide range of outcomes; ii) improved education scenario - the effect of extending the RAISE-ETP trial to all patients with onset of SMI by age 25, where the RAISE-ETP trial was a randomised controlled trial focusing on improving educational outcomes for patients with first episode psychosis.	Both internal and external validation was performed. Demographic, health, and economic outcomes were compared between the simulated and actual Panel Study of Income Dynamics populations. FAM underestimates the prevalence of activities of daily living and claiming of federal disability, and overpredicts Social Security retirement claiming. Supplemental security and working for pay were not statistically different between the FAM and the observed population. Furthermore, the FAM forecasts were compared to Census forecasts of the US population, which showed that FAM forecasts remained within 2% of Census forecasts.	i) Study relies on multiple datasets and estimation techniques subject to potential error; ii) Intervention modelled based on findings of an RCT focused on first-episode psychosis (may not be generalisable); iii) Results based on self-reported health status; iv) Study may have missed cases of SMI that were so severe early on that it prevented people from ever forming a household (PSID only samples people if they form a household); v) Results do not capture incarceration, homelessness, or institutionalisation associated with SMI.
Morrissey et al., 2010; Ireland	To examine the spatial prevalence of depression in Ireland	Outputs from the SMILE. Databases used to build the SMILE: Weighted Living in Ireland Survey 2000; Irish Small Area Population Statistics 2002 Individuals aged 16 years and older	Depression Distribution of depression, access to acute psychiatric hospitals, access to community-based psychiatric services	Static spatial microsimulation model The SMILE produces a micro-level synthetic dataset for Ireland using the simulated annealing technique to match the weighted Living in Ireland dataset to the Irish Small Area Population Statistics. The type of household/individuals suffering from depression were cloned in the microsimulation process at the electoral division level. A stochastic process was	No validation was performed. However, the authors examined the determinants of the spatial distribution of depression at the electoral division level.	i) Depression based on self-reported data; ii) Estimates of the distance were not examined to check whether estimates were proportional to the actual number of patients treated at each facility.

				incorporated into the alignment process using logistic regressions, which were used to assign a probability of suffering from depression to each individual in the dataset. The resulting dataset was used to investigate health status and health service utilisation patterns at the small area level. Logistic regression model was used to examine determinants of depression at national level. Stepwise regression was used to determine variables with the most significant relationship. The Spatial Interaction Model was used to measure: i) access scores from each electoral division to the nearest psychiatric inpatient facility at the national level; ii) access to community-based psychiatric services at the sub-national level. ArcGIS used to calculate distance from each electoral division to each psychiatric unit.		
Campbell and Ballas, 2016; Scotland	To identify which Census output areas have the greatest proportions of "unhappy" people	Outputs from the SimAlba. Databases used to build the SimAlba: Scottish Health Survey 2003; Census data for Scotland 2001 Individuals of all ages	Mental distress/unhappiness (measured by GHQ-12) Mental distress	Static spatial microsimulation model The SimAlba produces estimates of geographically sensitive health variables (e.g., smoking, alcohol consumption, physical activity, general health) for Glasgow, Scotland. Data from the Scottish Health Survey was upscaled through deterministic reweighting to reflect the populations of census areas as closely as possible. This dataset was used to estimate geographical distribution of 'unhappy' people.	Both internal and external validation was performed. The small-area microdata set was found to be a reasonably robust estimate of the data within 5-10% of the actual census data. The outputs of SimAlba were extensively compared against known Census totals.	Difficult to verify outputs against real population data.
Kingston A et al., 2018; England	To examine how long-term conditions and multi-morbidity will evolve between 2015 and 2035 in the population aged 65 years old and over in England	Outputs from the PACSim. Databases used to build the PACSim: Understanding Society wave 1; English Longitudinal Study of Ageing wave 5; Cognitive Function and Ageing Study II Individuals aged 65 years and older	Multi-morbidity (chronic heart disease, stroke, hypertension, diabetes, arthritis, cancer, respiratory disease, dementia, hearing impairment, vision impairment and cognitive impairment) including depression Prevalence of individual diseases, impairments and multi-morbidity in 2015, 2025, 2035	Dynamic microsimulation model The PACSim produces estimates of future prevalence, incidence, and life and health expectancies. Individuals' characteristics were updated monthly over the full time period of the simulation. All characteristics (e.g., marital status, education) are stochastic apart from sex, education, and socio-economic status, which are fixed; age is deterministic. Transition models for stochastic characteristics were calculated by fitting binary, ordinal or generalised logistic regression models to the base and 2-year follow-up waves of the combined studies. Monthly survival probabilities were derived from the annual probabilities underlying the 2014-based principal population projection for England. This dataset was used to simulate individual characteristics between 2015-2035.	Validation was performed for most conditions but not depression. The level of agreement between simulated numbers in 5-year age groups for each year and Office for National Statistics 2014 projections for England was good. Age-sex-specific prevalence of stroke, diabetes, current smoking, overweight, and obesity from PACSim were compared with those from the Health Survey for England 2014 and there was good agreement, except for obesity.	i) Morbidities based on self-reported data; ii) Transition rates based on observations from only two consecutive waves of each survey; iii) Transitions between states of all characteristics independent of time; iv) Lack of confidence intervals for all outcomes that account for error in the transition rates; v) Authors did not validate model for depression.

Kilmer et al., 2011; USA	To estimate social costs of depression and PTSD to veterans of Operation Enduring Freedom and Operation Iraqi Freedom	Defense Manpower Data Center 2000, Armed Services Medical Surveillance Monthly Report 2007, Office of the Under Secretary of Defense for Personnel & Readiness 2007, Congressional Budget Office 2004, Defense Manpower Data Center 2000, peer-reviewed studies, reports from the Institute of Medicine 2007, published guidance, randomised controlled trials, TRICARE and Medicare reimbursement rates, published prices for drugs, Veterans Administration negotiated rates, Department of Defense pay tables, Current Population Survey 2006 from the Bureau of Labour Statistics Soldiers deployed to Operation Iraqi Freedom or Operation Enduring Freedom (age range not specified)	PTSD, major depression, and comorbid PTSD and major depression Social costs of PTSD and depression including lost productivity, mental health treatment, medical costs of suicide, cost of lives lost to suicide	Static microsimulation model Soldiers' depression and PTSD trajectories was modelled over a 2-year period after they returned home, accounting for mental health treatments received and events that may have occurred due to a mental health condition. Individuals were constrained from switching across conditions. All personal characteristics alongside mental health status could influence wages, labour force status, and the probability of suicide. An individual's initial assignment to a mental health state was based on prevalence data. Since mental health outcomes were assigned stochastically, realised rates of mental health conditions were variable. Modeled individuals with a mental health condition had a probability of receiving evidence-based treatment or usual care, and these treatments influenced the course of illness. The simulation assumed that individuals with active mental illness had a higher probability of leaving the Department of Defense service. It also accounted for the fact that individuals with active mental illness who are discharged have a lower probability of working in the civilian sector. For those with a mental health condition, the model reduced the probability of working and wages conditional on working. Only individuals with active mental illness could attempt suicide. The probability of dying due to a suicide attempt was higher for active duty individuals relative to discharged individuals. Several scenarios were modeled by varying assumptions of key parameters.	No validation was performed. However, all parameter estimates were vetted by a group of experts from the RAND, University of California-Los Angeles, and Uniformed Services University of the Health Sciences.	i) Assumptions and required data likely to change over time, therefore cost estimates are likely conservative as they only include costs related to treatment, lost productivity, and suicide; ii) Limited data to derive some parameters and cost estimates; iii) Study did not include all costs associated with mental health and cognitive conditions.

Horvitz-Lennon et al., 2019; NA	To estimate potential long-term effects of pro-adherence interventions in early schizophrenia	Peer-reviewed studies; grey literature Hypothetical patients with early stage-schizophrenia (age range not specified)	Schizophrenia/psychosis Schizophrenia-related admissions, competitive employment and independent or family living at end of chronic phase, receipt of disability benefits at start of chronic phase	Static microsimulation model The model was used to predict the trajectories of hypothetical patients over 10 years from the first onset of psychosis. Two components were included in the model: i) critical period (lasting 3 years) divided into time between the onset of psychosis and treatment entry (i.e., duration of untreated psychosis, DUP) and the calibration phase, during which treatment starts and treatment decisions are varied; ii) chronic phase (lasting 7 years), during which long-term outcomes are predicted. The model started at psychosis onset. Patients were randomly assigned a DUP and underwent antipsychotic treatment during a calibration phase that lasted until the end of the critical period. Patients were exposed to different antipsychotic treatment sequences that could vary based on whether/when long-acting injectable agents were used. Simulated patients entering treatment with a DUP of 3 years or more did not go through this calibration phase but were entered directly into the chronic phase. Patients differed with respect to DUP at treatment entry but were assumed to have comparable socio-demographic and clinical characteristics. The likelihood of symptom control (effectiveness) was simulated. Simulation ended when they either achieved symptom control or reached the end of the critical period. In the chronic phase, long-term outcomes were predicted as a function of symptom severity at the calibration phase. Treatment effects were assumed to be identical across simulated patients. Patients were assumed to relapse and thus needed re-treatment.	No validation was performed. However, probabilistic sensitivity analyses, which varied all parameters randomly and simultaneously, were undertaken. Results showed that varying parameters individually did not lead to disproportionate changes in outcome estimates. When all parameters were varied randomly by 10%, the interquartile range remained within ±10% of the median prediction (except for predicted receipt of disability benefit). Differences among treatment pathways were sensitive to the chosen parameter values for long-acting injectable vs. oral adherence.	i) Findings are suggestive and in need of confirmatory empirical research; ii) Results may not be generalisable to other health care systems as most studies used were from the USA.

Horvitz-Lennon et al., 2020; NA	To predict long-term benefits of interventions to reduce the time in psychosis during the early phase of the illness, with and without accompanying efforts to improve medication adherence	Peer-reviewed studies; grey literature Hypothetical patients with early stage-schizophrenia (age range not specified)	Schizophrenia/psychosis Schizophrenia-related admissions, competitive employment and independent or family living at end of chronic phase, receipt of disability benefits at start of chronic phase	Static microsimulation model The model was used to predict the trajectories of hypothetical patients over 10 years from the first onset of psychosis. Two components were included in the model: i) critical period (lasting 3 years) divided into the time between the onset of psychosis and treatment entry (i.e., duration of untreated psychosis, DUP) and the calibration phase, during which treatment starts and treatment decisions are varied; ii) chronic phase (lasting 7 years), during which long-term outcomes are predicted. The model started at psychosis onset. Patients were randomly assigned a DUP and underwent antipsychotic treatment during a calibration phase that lasted until the end of the critical period. Patients were exposed to different antipsychotic treatment sequences that could vary based on whether/when long-acting injectable agents were used. Simulated patients entering treatment with a DUP of 3 years or more did not go through this calibration phase but entered directly into the chronic phase. Patients differed with respect to DUP at treatment entry but were assumed to have comparable socio-demographic and clinical characteristics. The likelihood of symptom control (effectiveness) was simulated. Simulation ended when they either achieved symptom control or reached the end of the critical period. In the chronic phase, long-term outcomes were predicted as a function of symptom severity at the calibration phase. Treatment effects were assumed to be identical across simulated patients. Patients were assumed to relapse and thus needed re-treatment.	No validation was performed. However, one-way and probabilistic sensitivity analyses were undertaken to test the outcome effects of several model parameters. The authors claimed that the sensitivity analysis provided support for the soundness of the model.	Findings are suggestive and in need of confirmatory empirical research

Mulcahy et al., 2017; USA	To determine potential effects of policies aimed at increasing metabolic testing rates among beneficiaries with schizophrenia receiving antipsychotics	California Medicaid Analytic eXtract 2002-2009; National Health and Nutrition Examination Survey 2007-2010 Medicare beneficiaries aged 20-65 years old with schizophrenia.	Schizophrenia (and diabetes) Metabolic testing rates, rate of diabetes conditions, years with diabetes conditions, time to diagnosis, and short-term costs	Static microsimulation model The model was developed to examine prescriber decision making and disease progression over a 10-year horizon. In each iteration of the model, 10 annual periods were simulated starting in 2002 through 2012 for the simulation cohort considering five states (healthy, undiagnosed prediabetes, diagnosed prediabetes, undiagnosed diabetes, or diagnosed diabetes). The model allowed for individuals to transition to other undiagnosed health states but did not allow anyone to die or leave Medicaid. A set of assumptions was developed for model inputs with insufficient empirical evidence: i) medium- and high-risk antipsychotics increase the risk of transitioning to diabetes condition states; ii) testing rates do not depend on the antipsychotic prescribed; iii) test results reveal all diabetes conditions to the prescriber without error; iv) test results revealing diabetes conditions lead prescribers to switch patients on medium- and high-risk drugs to a lower-risk drug 100% of the time.	No validation was performed. However, a sensitivity analysis was undertaken by considering different switching rates, antipsychotic risk levels, and costs of policy implementation. Results were robust to different assumptions.	i) Study relied on imputed data for health states; ii) Results might not be generalisable since data from only a single USA state; iii) Benefits from screening might be underestimated since individuals assumed not to transition toward healthier states; iv) Total costs might be underestimated since study focuses only on incremental costs associated with diabetes diagnoses.
Abdalla et al., 2021; USA	To examine the role of exposure to TV and social media coverage of a mass traumatic event in shaping community PTSD prevalence	American Community Survey 2010-2015; Pew research center; peer-reviewed articles Artificial population of Parkland and Coral Springs, Florida, USA, aged 14 years and older	PTSD PTSD population prevalence based on exposure to TV coverage of Parkland shooting	Static microsimulation model The model was used to simulate the 2018 Stoneman Douglas High School (Parkland) shooting. Authors initialised a population of agents that was demographically comparable to the population of Parkland and Coral Springs, Florida, where the shootings took place. An Iterative Proportional Updating approach, which computes the selection probabilities of different household types, was used. An internal social network was created to connect those directly affected with agents assigned as family or close friends. Multiple scenarios were implemented regarding the total number of hours of using TV in the population to estimate the potential association of changing media exposure on PTSD prevalence in the population.	No validation was performed.	i) Limited generalisability since findings are based on several epidemiological studies; ii) Results only apply to those who use TV as their preferred source of news; iii) Authors do not offer insight about how different types of exposure could affect PTSD prevalence.

Huang et al., 2021; USA	To determine the characteristics related to willingness to quit in smokers with serious mental illness	Data from the IDEAL trial (a randomised controlled trial of a comprehensive cardiovascular risk reduction program for persons with serious mental illness) Individuals aged 18 years or older with serious mental illness and at least one cardiovascular risk factor	Schizophrenia, schizoaffective disorder, bipolar disorder, and depression Willingness to quit smoking, abstinence status	Static microsimulation model The model was developed with explicit time dependence of the internal state and smoking status of each smoker. A single-state variable was created to represent patient's willingness to quit smoking (low, medium, and high) and assigned to each one using longitudinal data. Ordinal logistic regression was used to calculate likelihood of state variable at baseline and the transition probabilities. Markov analysis was used to propagate the probabilities of the states of willingness to quit at baseline to the probabilities of the states at 6 and 18 months. Lasso regression was used to select and reduce the number of regression variables in the final predictive model of quit status at 18 months.	No validation was performed.	Need for detailed data to calibrate model parameters.
Ward et al., 2019; USA	To model individual-level disease dynamics of eating disorders from birth to age 40 years old and to estimate the association of increased treatment coverage with eating disorders-related mortality	Global Burden of Disease Study 2017; peer-reviewed studies Individuals from birth until 40 years old	Eating disorders Age-specific 12-month and lifetime eating disorder prevalence and number of deaths per 100,000 general population individuals by age 40	Static microsimulation model The model included a Markov state transition model with 6 states (healthy, anorexia nervosa, bulimia nervosa, binge eating disorder, other specified feeding and eating disorders, and deceased). Transitions among all states were allowed, except for deceased. An annual model cycle was used to model transitions and to follow a simulated cohort of 100,000 individuals from birth until age 40. 1,000 simulations of 100,000 individuals were run to explore eating disorder dynamics using a parameter set from the best 100 sets identified in calibration in each simulation sampling. The approach considered individual-level, stochastic, and parameter uncertainty around the estimates. Counterfactual scenarios were run to estimate eating disorder-associated excess mortality.	No validation was performed. However, calibration was performed, which involved comparing the model simulated estimates with empirical data to find parameter sets that achieved a good fit. The model was calibrated using empirical data on the prevalence of eating disorders at different ages using nationally representative surveys. During calibration, if a sampled set of parameters was invalid, the lower bounds of the relevant search bounds were iteratively lowered until a valid parameter set was sampled. 87% of the simulated estimates fell within the target confidence intervals. The model-estimated prevalence of anorexia nervosa and bulimia nervosa tended to be higher and lower, respectively, than the 2017 Global Burden of Disease Study at older ages.	Results do not take into account other risk factors associated with eating disorders; ii) Results are affected by transportability of standardised mortality ratios and eating disorders crossover estimates due to comparability issues; iii) Estimates of association of treatment with mortality may be conservative since it is assumed that treatment is associated with remission probabilities only; iv) By using a cohort, rather than an open population model, authors were unable to explore potential secular trends in eating disorders that might occur over time; v) Avoidant restrictive food intake disorder not included because not established as a diagnosis until DSM-5.

Riva and Smith, 2012; England	To generate small area estimates of psychological distress and alcohol consumption	Health Survey for England 2004-2006; Census 2001; Index of Multiple Deprivation 2007 Individuals aged 18 years and older	Psychological distress (and alcohol consumption) Prevalence estimates of psychological distress (and heavy alcohol consumption)	Static spatial microsimulation model The model consisted of a derivation of a deterministic re-weighting methodology used for spatial microsimulation of populations. The method employed produced prevalence estimates for health outcomes through a process of matching individuals from a large, population-representative dataset to known local populations based on similar socio-demographic characteristics using indirect standardisation. The model iterates through the variables identified to predict the health outcome. Ordered logistic regression was used to model the proportion of psychological distress at Lower Super Output Areas and heavy alcohol drinking.	No validation was performed. However, the 'goodness of fit' and the 'construct' and 'convergent' validity of the microsimulated Lower Super Output Areas (LSOAs) estimates of psychological distress (and alcohol consumption levels) were established through several analyses. Results showed no more than 10% errors in at least 90% of districts for mental health problems (and moderate and heavy drinking). The 'construct validity' results suggested that the method for generating synthetic area estimates of psychological distress performs well against known mental health needs indicator. The 'convergent validity' was supported by strong correlation with indicators of years of potential life lost, illness and disability ratio, acute morbidity, and 'mood and anxiety disorders'.	i) Results based on self-reported health status; ii) Microsimulated prevalence might be underestimated in some localities since national level datasets used; iii) Discrepancies between the years of data used in the model and the data used to validate the microsimulated estimates; iv) Microsimulated estimates could not be validated directly to 'real-world' Lower Super Output Areas prevalence data.

* including robustness checks and model adjustments

Legend: NA – not applicable

Table 4. Quality assessment scores of included studies

Study	Model development and structure			Validity and scope of results			Total
	Purpose	Data	Transparency	Uncertainty	Validation*	Generalisability	
Studies on existing economic-demographic models that included a component on mental health							
Seabury et al., 2019	1	1	1	1	1	1	**6.0**
Campbell and Ballas, 2016	1	1	1	1	1	0.5	**5.5**
Morrissey et al., 2010	1	1	1	1	0.5	0.5	**5.0**
Kingston et al., 2018	1	1	1	0.5	1**	0.5	**5.0**
Schofield et al., 2019	1	1	1	1	0	0.5	**4.5**
Veerman et al., 2015	1	1	1	1	0	0.5	**4.5**
Schofield et al., 2011a	1	1	1	1	0	0.5	**4.5**
Schofield et al., 2011b	1	1	1	1	0	0.5	**4.5**
Schofield et al., 2011c	1	1	1	1	0	0.5	**4.5**
Schofield et al., 2013	1	1	1	1	0	0.5	**4.5**
Schofield et al., 2022	1	1	1	1	0	0.5	**4.5**
Studies focused solely on mental health models developed to examine the impact of specific policies/interventions on mental disorders							
Riva and Smith, 2012	1	1	1	1	0.5	0.5	**5.0**
Mulcahy et al., 2017	1	1	1	1	0	1	**5.0**
Horvitz-Lennon et al., 2019	1	1	1	1	0	1	**5.0**
Horvitz-Lennon et al., 2020	1	1	1	1	0	0.5	**4.5**
Abdalla et al., 2021	1	1	1	0.5	0	1	**4.5**
Ward et al., 2019	1	1	1	1	0	0.5	**4.5**
Kilmer et al., 2011	1	1	1	0.5	0	0.5	**4.0**

Huang et al., 2021	1	0.5	0.5	0.5	0	0.5	**3.0**

* In addition to validation, robustness checks and model adjustments were considered.
** Although this study validated the model for physical health outcomes, it did not do this for depression

Figures

Figure 1. PRISMA flow diagram