Effectiveness of the Nationwide Depression Screening: A Systematic Review and A Target Trial Emulation Study Using National Health Insurance Service Database

Hyejin Kim

The Graduate School
Yonsei University
Department of Public Health

Effectiveness of the Nationwide Depression Screening: A Systematic Review and A Target Trial Emulation Study Using National Health Insurance Service Database

A Doctoral Dissertation
Submitted to the Department of Public Health
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Public Health

Hyejin Kim

December 2024

This certifies that the Doctoral Dissertation of
Hyejin Kim is approved.

_____
Thesis Supervisor: Sun Jae Jung

_____
Thesis Committee Member #1: Chung Mo Nam

_____
Thesis Committee Member #2: Hyeon Chang Kim

_____
Thesis Committee Member #3: Jae Kwan Jun

_____
Thesis Committee Member #4: Jae-Won Kim

The Graduate School
Yonsei University
December 2024

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

GLOSSARY OF ABBREVIATIONS

ATC	Anatomical Therapeutic Chemical
CTFPHC	Canadian Task Force on Preventive Health Care
CCI	Charlson Comorbidity Index
CI	Confidence intervals
DALY	Disability-adjusted life year
ED	Emergency department
EPDS	Edinburgh Postnatal Depression Scale
HIRA	Health Insurance Review & Assessment Service
HR	Hazard ratio
ICD-10	International Classification of Diseases-10
ITT	Intent-to-treat
NHIS	Korea National Health Insurance Service
NRS	Non-randomized observational study
PHQ-9	Patient Health Questionnaire-9
PI	Predictive intervals
PICOS	Population/intervention/comparison/outcome/study design
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	Prospective Register of Systematic Reviews
RCT	Randomized controlled trial
RoB 2	Version 2 of the Cochrane Risk-of-Bias Tool for Randomized Trials
ROBINS-I	Risk of Bias in Non-Randomized Studies - of Interventions
SNRI	Serotonin–norepinephrine reuptake inhibitors
SSRI	Selective serotonin reuptake inhibitors
TCA	Tricyclic antidepressants
TTE	Target trial emulation
UKNSC	UK National Screening Committee
USPSTF	US Preventive Services Task Force

ABSTRACT IN ENGLISH

Effectiveness of the Nationwide Depression Screening: A Systematic Review and A Target Trial Emulation Study Using National Health Insurance Service Database

INTRODUCTION

Over the past two decades, previous studies about the effectiveness of organized depression screening at the primary care or national level have reached mixed conclusions. Despite its early age of onset and substantial disease burden, depression often remains undetected and untreated in the general population. Although screening for depression has been proposed to prevent the delay of treatment, the public health effectiveness of such screening remains inconclusive. Therefore, the first study used a systematic review and meta-analysis to comprehensively evaluate the effectiveness of depression screening based on both randomized controlled trials (RCTs) and non-randomized observational studies. Following this, a target trial emulation study analyzed the effectiveness of the 2019 nationwide depression screening program in South Korea, which used the Patient Health Questionnaire-9 (PHQ-9), leveraging data from the National Health Insurance Service (NHIS) database to support more informed decision-making.

METHODS

In the first study, a systematic review and meta-analysis was conducted to identify and synthesize evidence evaluating the effectiveness of depression screening published from June 1959 to August 2024. Databases including PubMed, EMBASE, PsychINFO, and Web of Science were searched, pooling 4,039 articles initially. Two independent reviewers performed the study search, data extraction, and study evaluation. Each included study was summarized based on the predefined population, intervention, comparison, outcome, and study design (PICOS) framework. The quality of the studies was assessed using Version 2 of the Cochrane Risk-of-Bias Tool for Randomized Trials and the Risk of Bias in Non-

Randomized Studies - of Interventions. The evaluation of the evidence was based on six criteria: (1) targeting specific population, (2) exclusion of prevalent cases of depression, (3) eligibility determined before follow-up, (4) inclusion of additional components besides screening, (5) use of a validated screening tool, and (6) analysis of outcomes using individual-level data. Random effects models were used to conduct the meta-analysis, applying restricted maximum likelihood estimation method. Heterogeneity across studies was assessed using the I^2 statistic and prediction intervals (PI). Publication bias was assessed using visual evaluation of Begg's funnel plot and Egger's test. Subgroup analyses were performed based on outcome definition, sex, and geographic region, while sensitivity analysis excluded studies with a high risk of bias.

In the second study, an emulation of a hypothesized target trial was conducted using the Korea NHIS database (2019 – 2022). The estimands were derived using the traditional intent-to-treat (ITT) and modified ITT approaches. As the depression screening for individuals in their 20s and 30s was introduced in 2019, the follow-up period started on January 1, 2019, and ended on December 31, 2022. The primary outcome was the incidence of hospitalization for mood disorders, with secondary outcomes including (1) initial use of antidepressants, (2) emergency department visits for mood disorders, and (3) suicide and suicidal behaviors. Individuals with a psychiatric diagnosis prior to 2019 or those who had participated in the depression screening before 2019 were excluded. Coarsened exact matching (1:1 ratio) based on age, sex, and subscriber types was used to match the screened and non-screened individuals. The multivariate cause-specific Cox regression models were used to estimate the hazard ratios (HRs), which adjusted for urbanicity, prior health check-ups, income, and the Charlson Comorbidity Index. Subgroup analyses were conducted based on sex, age groups (20–50s and 60–70s), and income levels. To address unmeasured confounding, a negative control outcome (hospitalization for cancer) and the E-value were applied. Exploratory analyses were also performed to examine the distribution of depression-related outcomes, the diagnostic performance of the PHQ-9, and the impact of post-screening interventions using a dataset from a cohort of screeners.

RESULTS

The systematic review and meta-analysis found significant but weak evidence that depression screening for the general adult population in primary care or at the national level is effective (pooled OR 0.74, 95% CI: 0.62–0.87), particularly among women and Asian populations. However, the included studies were mostly of poor methodological quality,

with significant bias and considerable heterogeneity across studies. None of the studies met all six criteria required to provide valid evidence for the effectiveness of depression screening in the general adult population.

The target trial emulation study aimed to estimate the effectiveness of depression screening based on the modified ITT approach (n = 53,688), and the traditional ITT (n = 276,214) as secondary. After adjusting for confounders, the screened group showed a 44% lower risk of hospitalization for mood disorders compared to the non-screened group (adjusted hazard ratio [aHR] for modified ITT 0.56, 95% confidence intervals [CI] 0.41–0.76), especially in women and in older adults. The initiation of antidepressant use was significantly increased in screened individuals than in non-screened (aHR for modified ITT 1.21, 95% CI 1.10–1.33). The hazards of emergency department visits were significantly lower in the screened group than in non-screened group (aHR for modified ITT 0.71, 95% CI 0.51–0.98). No significant difference was observed in suicide and suicidal behaviors between the screened and non-screened groups (aHR for modified ITT 0.62, 95% CI 0.32–1.21). Hospitalization for cancer, used as a negative control outcome, showed no difference between these groups.

CONCLUSION

The systematic review and meta-analysis found statistically significant but limited evidence supporting effectiveness of depression screening in general adult populations. The target trial emulation study using NHIS data in South Korea showed that the 2019 nationwide depression screening program reduced hospitalization rates for mood disorders, increased antidepressant use, and decreased emergency department visits, particularly among women and older adults. However, the observed differences between screened and non-screened individuals may result from non-specific effects of the screening program, highlighting the need for ongoing evaluation to support evidence-based decisions on nationwide depression screening.

Keywords: nationwide depression screening, secondary prevention, systematic review, meta-analysis, target trial emulation, Korean National Health Insurance Service Database

1. INTRODUCTION

1.1. Backgrounds

Depression, or major depressive disorder, is characterized by clinically depressive episodes marked by persistently sad, empty, or irritable moods that last for at least two weeks. [1] The diagnosis of depression requires the presence of at least one of the following symptoms: a depressed mood or a loss of interest or pleasure, along with at least five of the following symptoms: changes in appetite and weight, alterations in sleep patterns, psychomotor agitation or retardation, loss of energy, feelings of worthlessness or guilt, difficulty thinking or making decisions, and recurrent suicidal ideation. [1] The pooled lifetime prevalence of depression was estimated to be 10.8%, with a one-year prevalence of 7.2% reported across 30 countries from 1994 to 2014. [2] The prevalence of depression is higher among older adults, with women experiencing rates about twice as high as men; however, this gender difference in prevalence decreases with age. [3-6] Additionally, the socioeconomic status comprising income, education, occupation, social class, or wealth is known to be associated with depression in many studies. [7-9]

Depression is one of the leading cause of disability worldwide. According to the Global Burden Disease Study 2010, the disease burden of depression was reported to be 40% of the global burden of disease, and the greatest proportion of disability-adjusted life years (DALYs) associated with depression was observed in individuals within the age range of 10 to 29 years. [10] Moreover, depression is a key risk factor for suicide. [11] A systematic review found that the population attributable risks for suicide due to affective disorders, including depressive and anxiety disorders, were estimated to be 26% for men and 32% for women. [12] Additionally, depression is associated with difficulties in role functioning, including marital quality, work performance, and employment. [13]

The natural history of depression is not clear yet. The first onset of depression often occurs in early adulthood (in the 20s) and later in life (over the age of 60), and it is typically associated with greater impairment in quality of life and increased disease burden, particularly when it arises in early adulthood. [14,15] There is no consensus on the etiology of depression, as genetic, biological, psychological, and social factors are interconnected in the development of pathological depressive episodes. [16] The course of depression after onset varies widely among patients. The NIMH Collaborative Depression Study reported that while most patients recover within one year, about 20% remain depressed from their

baseline episode after two years, and 7% do not recover even after ten years. [17,18] Previous studies using data from the non-clinical general population have reported that the recurrence rate of depression is approximately 40%.[19] Also, the etiology of the first depressive episode is likely distinct from the factors associated with subsequent episodes. [20] Therefore, proper and timely treatment for depression is crucial, as early detection and intervention are key, with longer durations of untreated depression being associated with worse outcomes. [21-25]

Unfortunately, the underdiagnoses and undertreatment of depression is common. A previous study using data from three European countries found that only about 21% patients with depressive symptoms in primary care were diagnosed with major depression by their general practitioners. [26] A study using the 2012 Canadian Community Health Survey found that only half of Canadians aged 15 and older with depressive symptoms were diagnosed with a mood disorder. Untreated cases were generally younger, single, less educated, had lower incomes, and lacked physical comorbidities compared to those treated for depression. [27] Another study using Brazilian National Survey data reported a 64% underdiagnosis rate of depression, with underdiagnosis more common among men, older individuals, and those with lower income, education levels, and fewer chronic diseases. [28] The World Health Organization's World Mental Health Surveys from 15 countries found that the median delays among cases eventually making contact ranged from 1 to 14 years for mood disorders. [29] A recent psychological autopsy study in Spain revealed that the majority of individuals with depression who died by suicide had not received a diagnosis or treatment, even though they had visited health services in the weeks prior. [30] Individuals with undiagnosed depression may fail to seek psychiatric treatment because the somatic symptoms associated with depression can be mistakenly attributed to physical causes. [31,32] As a result, the diagnosis and treatment of psychiatric disorders are highly stigmatized, contributing to the underdiagnoses of depression. [33]

1.2. Theoretical Framework and Evidence on Depression Screening

To address delays in treating depressive episodes, screening for depression has been proposed as a means to identify undiagnosed cases. [34] This approach employs a different framework than screening for cancer or diabetes, as the stages of depression are less distinct, and diagnosis relies on the features, severity, and duration of symptoms rather than on pathophysiological markers. [35,36] Clinical depression is highly heterogeneous, and while its natural history remains incompletely understood, its progression can be outlined in six stages: no depression, onset of depressive episode, response to intervention, remission, recovery, and recurrence (Figure 1). [37] These trajectories fluctuate along the continuum of depressive symptom severity, ranging from asymptomatic to mild, moderate, and severe. [1] The optimal window for depression screening and further evaluation lies between mild and moderate symptom severity, as clinical benefits are unlikely during asymptomatic states or in cases of severe depression. A well-designed screening program would increase the detection and treatment of clinical depression at these optimal stages, reducing the progression to severe depression associated with hospitalization, significant daily impairment, or suicide.

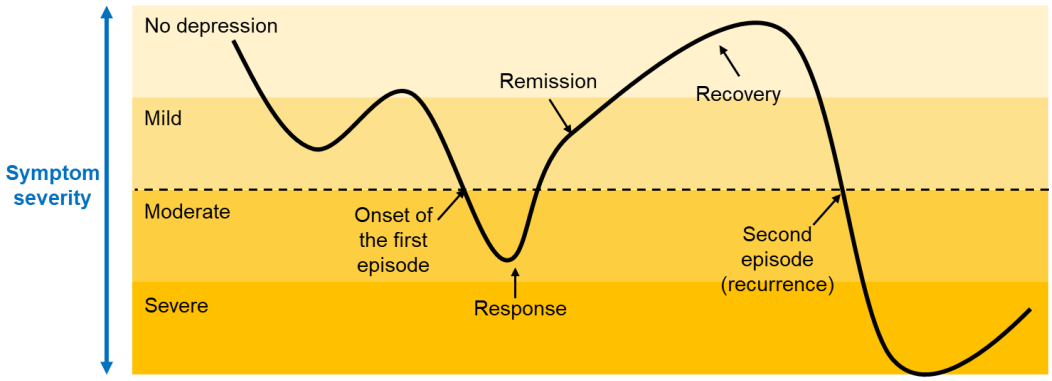


Figure 1. Hypothetical progression of depression across different levels of symptom severity

Notes. The gradation of yellow area indicates the severity of depressive symptoms. The screening can detect clinical depression cases regardless of the disease phase.

Previous reviews on the effectiveness of depression screening have been independently conducted by the US Preventive Services Task Force (USPSTF), the Canadian Task Force on Preventive Health Care (CTFPHC), and the UK National Screening Committee (UKNSC). [38-40] However, these organizations reached different conclusions, leading to varying national recommendations for depression screening in the US, Canada, and the UK over the past 20 years. Thombs and colleagues (2017) suggested that the discrepancies arose because the three organizations prioritized different aspects when selecting, evaluating, and interpreting the evidence. [41] They pointed out that the USPSTF's review included indirect evidence, such as the effectiveness of depression treatment, to support their recommendation for depression screening in primary care. However, only direct evidence comparing depression-related outcomes between screened and non-screened individuals should be used to inform health policy decisions regarding nationwide depression screening. In Australia, the Royal Australian College of General Practitioners, referencing findings from the USPSTF and CTFPHC, does not recommend depression screening for the general population due to insufficient evidence but, similar to UK guidelines, advises primary care providers to monitor for depressive symptoms, particularly in high-risk adolescents aged 12 to 18. [42] Other countries, such as Japan, Taiwan, and South Korea, have introduced nationwide depression or mental health screening programs in response to the increasing public health burden of mental health issues, despite the lack of evaluation of their clinical benefits. [43-45]

Thombs and Ziegelstein (2014) suggested the three criteria that should be considered when evaluating the evidence of the effectiveness of routinely depression screening for general population in primary care. [46] The first criterion is determining eligibility and randomizing participants before screening begins, to minimize selection bias resulting from post-randomization exclusions. The second criterion is the exclusion of patients who are already diagnosed with or receiving treatment for depression, as screening aims to detect asymptomatic or undiagnosed cases before they are identified or treated by a physician. The third criterion assesses whether the study provided similar treatment options to patients in both trial arms, aside from the screening intervention, to minimize the impact of co-interventions on the study findings. These criteria were used to evaluate the studies included in the CTFPHC and UKNSC reviews, but not in the USPSTF's review. [39,40]

Moreover, none of these reviews pooled the results from non-randomized observational studies, they included only randomized controlled trial (RCTs). RCTs are deemed as gold-standard when it comes to assess the effectiveness of intervention program. RCT has advantages to prevent confounding, length time bias, and lead time bias via random allocation of treatment conditions. [36] However, in some cases, the RCT design is

not feasible, unethical, or timely. [47] Moreover, both RCTs and non-randomized observational studies have strengths and limitations and can be complementary in assessing intervention effects. While RCTs excel in evaluating the efficacy of an intervention within a specific population under ideal conditions, non-randomized observational studies are better suited for assessing the effectiveness of the intervention in real-world settings. [48] Therefore, including the results from non-randomized observational studies in the evidence synthesis is important to comprehensively assess the effectiveness of depression screening to reflect real-world effect of the program in the population.

Meanwhile, a case study examining the inconsistency between two meta-analyses addressing the same research question on the effectiveness of depression screening noted that the reviews included different sets of studies and applied different weightings to the same studies. [49] The authors argued that confirmation bias within the author groups could influence the selection and interpretation of the included studies. Therefore, it is necessary to revisit the evidence on the effectiveness of depression screening by an independent author group.

1.3. Nationwide Depression Screening in South Korea

In South Korea, the prevalence of depression diagnosis is increasing from 2.8% in 2002 to 5.3% in 2013. [50] Nevertheless, South Korea has maintained the highest suicide rate among high-income countries over the past two decades (24.1 per 100,000 in 2020), while its depression prevalence remains below the global average, indicating continued underdiagnosis of clinical depression. [2,51] Additionally, the economic burden of depression in South Korea was estimated to be $4,049 million, accounting for approximately 0.74% of the total healthcare expenditure. [52] Therefore, preventing the severe prognosis of depression and suicide is one of the primary public health issues in South Korea.

Cultural obstacles unique to Korean society often pose challenges to the effective diagnosis and treatment of depression. Mental health problem is often viewed as a disorder caused by supernatural forces or a sign of character weakness in Korea. [53] Koreans also value *'chemyon*,' which refers to the concern about losing face in social contexts, and those with a heightened sensitivity to *chemyon* are more likely to develop stigma toward suicide. [54] A study using data from the Korean Community Health Survey found that only 27% of South Koreans with depressive symptoms sought help. [55] Those who were male, older, had lower education levels, and were covered by National Health Insurance were less likely to seek help for their depression. [55]

The South Korean government has launched several public health programs to address mental health issues, particularly those involving clinical depression and suicide, under the management of the Ministry of Health and Welfare and with assistance from the National Health Insurance Service (NHIS). The NHIS is a compulsory social insurance program, with premiums determined by income and wealth. While salaried employees have their premiums partially deducted from their wages, self-employed individuals are responsible for paying their premiums in full. For roughly 3% of the population unable to afford premiums, the government offers support via the Medical Aid Program. [56] The NHIS oversees payment administration, while the Health Insurance Review & Assessment Service (HIRA) is tasked with evaluating the quality of healthcare services. The majority of healthcare providers (over 90%) are private, and 92.6% of physicians are specialists. [57] Patients are not required to obtain referrals to visit specialty clinics.

Taking into account the public health significance of depression and the cultural nuances of mental health in Korea, there was a growing need for active case finding for clinical depression instead of relying solely on individuals' voluntary help-seeking behaviors. Thus, in 2018, the biannual National General Health Screening Program was launched, which includes depression screening. This program is funded by the NHIS and targets the entire population aged 40, 50, 60, and 70. [56] Recognizing the importance of equity and accessibility, health screenings were extended to people in their 20s and 30s, and due to the pressing issue of suicide among young adults, depression screening was initiated, although the evidence remains insufficient. [43] However, there are currently no studies on the effectiveness of depression screening in Korea.

1.4. Objectives of Study

The aims of this study are twofold: (1) to synthesize previous studies about the impact of depression screening on general adult population, and (2) to generate new evidence by examining the outcomes of the depression screening program implemented in South Korea in 2019 and its impact on the prognosis of depression.

In the first study, a systematic review and meta-analysis were conducted with the following objectives:

1. To qualitatively and quantitatively summarize research on the effectiveness of depression screening in primary care, non-mental health practice, or general population including both RCT and non-randomized observational study designs,

2. To describe the heterogeneity and key differences in the previous studies according to the study design and population to discuss the implications of subsequent study using NHIS database.

In the second study, a target trial emulation study using NHIS database was conducted with the following objectives:

1. To provide the descriptive statistics from 2019 nationwide depression screening,

2. To compare the incidence of hospitalization for mood disorders between screened and non-screened groups using traditional intent-to-treat (ITT) and modified ITT approaches,

3. To compare the incidence of antidepressant use, emergency department visits for mood disorders, and suicide or suicidal behaviors between screened and non-screened groups, using traditional ITT and modified ITT approaches,

4. To evaluate whether the effects of depression screening differed based on sex (men versus women), age group (20–50s versus 60–70s), and income level (lower versus higher income),

5. To exploratory assess the effects of post-screening intervention by comparing depression-related outcomes between individuals with positive and negative PHQ-9 results.

2. MATERIALS AND METHODS

2.1. Systematic Review and Meta-Analysis

The purpose of the systematic review and meta-analysis was to comprehensively evaluate the level of evidence regarding the effectiveness of depression screening for general adult population. This systematic review was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO; ID: CRD42024584471), and conducted in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines. [58] Following the search strategy, two reviewers independently searched relevant articles with same search strategy, assessed study eligibility, extracted data, and evaluated study quality and any discrepancies were resolved by consensus. Differences in the effectiveness of depression screening were examined in relation to the study design, comparing RCTs and non-randomized observational studies. Pre-specified framework for systematic review was based on the Population-Intervention-Comparison-Outcome-Study design (PICOS) tool. The PICOS table is presented in the Table 1.

Table 1. Detailed inclusion and exclusion criteria based on PICOS

PICOS	Descriptions
Population	Adult general population (aged ≥ 18 years)
Intervention	Screening program aimed to improve the prognosis of depression. This involves identifying individuals with untreated depression and providing them with appropriate post-screening interventions.
Comparison	Individuals who were not assigned to or did not undergo depression screening served as the comparison group. Studies that lacked such comparison groups were excluded.
Outcomes	Outcomes related to the prognosis of depression or the pathways through which depression screening exerts its effects were included to assess the effectiveness of depression screening. Outcomes unrelated to depression, such as the prognosis of other illnesses, were excluded. Outcomes aimed at improving the quality of the screening program, such as response rates, were excluded.
Study design	Randomized controlled trials, non-randomized controlled trials, and cohort studies were included. Reviews, cross-sectional studies, case reports, case series, protocols for clinical trials, and animal studies were excluded. Economic evaluation and validation study of the screening tools were also excluded.
Type of literature	Published peer-reviewed research articles. Commentary, editorials, letters, media, news stories, and conference abstracts were excluded.

Abbreviation: PICOS, Population-Intervention-Comparison-Outcome-Study design

2.1.1. Search strategy and eligibility criteria

A database search was conducted on August 14th, 2024, across PubMed, Web of Science, EMBASE, and PsycINFO. The search for studies was conducted without a time limit, covering publications from June 1959 to August 2024. The search was limited to publications in English and Korean. Predefined search terms included various combinations of the following keywords: (*screen** OR *casefinding* OR *case-finding* OR *case finding*) AND (*depress** OR *mood* OR *affective*) in the title. Studies obtained from the search were transferred into EndNote version 21.0.0 (Clarivate), and duplicates were removed. Two independent reviewers independently extract relevant data from each eligible article and enter the data directly into formatted Excel spreadsheets.

The screening of literature began with a review of titles and abstracts, followed by a full-text evaluation of selected studies. Studies were included if they assessed the effectiveness of depression screening by comparing mood disorder outcomes, including depression, between screened and non-screened groups. Only peer-reviewed articles, such as randomized controlled trials, non-randomized controlled trials, and cohort studies, were included. Excluded were pre-prints, commentaries, editorials, letters, media reports, news articles, conference abstracts, reviews, case reports, case series, clinical trial protocols, qualitative studies, animal studies, screening tool development or validation studies, general guidelines, economic evaluations, quality improvement studies, and studies targeting children or adolescents. To ensure no studies were missed, existing evidence syntheses and reviews that informed nationwide recommendations in the UK, Canada, and the USA were additionally included in the review. [38,40,59,60]

2.1.2. Data extraction

The summaries of included study were qualitatively organized to cover aspects of PICOS: title, country, study design, population, data source, eligibility, sample size, intervention, follow-up/study period, outcomes and main findings, and adjusted covariates. For the meta-analysis, effect sizes and the corresponding 95% confidence intervals or *p*-values related to depression screening outcomes were collected. Outcome measures were summarized in the form of odds ratios, with continuous variables being translated into log odds ratios according to Chinn's procedure. [61] All the extracted ORs were adjusted for the measured covariates.

2.1.3. Study assessment

The quality of study was assessed using Version 2 of the Cochrane Risk-of-Bias tool for randomized trials (RoB 2) and Risk of Bias in Non-Randomized Studies - of Interventions (ROBINS-I). [62,63] The risk of bias in the included RCT studies was assessed using RoB 2 as low, some concern, or high for each domain of the following: randomization process, deviation from intended interventions, missing outcome data, outcome measurement, and selection of the reported results. [62] If any of the items were rated as high, the overall bias of the study was concluded to be high risk of bias; if any of the items included some concerns, the overall bias was concluded to be some concerns; and if all areas were low risk, the overall bias was concluded to be low risk of bias. The risk of bias for the included observational studies was evaluated using the ROBINS-I tool. Each study was scored on the following items: confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, outcome measurement, and selection of reported results. The scoring categories were: low risk, moderate risk, serious risk, critical risk, or no information (NI). [63] NRSs can be categorized based on their risk of bias as follows: (1) Low risk of bias indicates that the study is comparable to a well-performed randomized trial; (2) Moderate risk of bias means that while the study provides sound evidence for a non-randomized study, it cannot be considered comparable to a well-performed randomized trial; (3) Serious risk of bias suggests that the study has some important problems; (4) Critical risk of bias denotes that the study is too problematic to provide any useful evidence and should not be included in any synthesis; and (5) No information on which to base a judgement about risk of bias signifies a lack of sufficient information to assess the study's risk of bias.

To assess the level of evidence for the effectiveness of depression screening, the six predetermined criteria were used: (1) targeting a specific population, (2) exclusion of prevalent cases of depression, (3) determination of eligibility prior to follow-up, (4) incorporation of additional components beyond screening, (5) utilization of a validated screening tool, and (6) analysis of outcomes using individual-level data. The evaluation of whether the study targeted the general population or specific groups—such as patients with particular illnesses, postpartum women, or specific age cohorts—was used to assess the generalizability of the study results to the broader population. The third, fourth, and fifth criteria was suggested by Thombs and Ziegelstein (2014), and these criteria were applied as eligibility criteria in reviews conducted in the UK and Canada. [39,40,46] Excluding prevalent cases is essential in studies evaluating the effectiveness of depression screening, as the primary aim of such screening is to identify previously undiagnosed cases of depression and enhance their prognosis. To avoid the risk of collider bias associated with

using postbaseline data during follow-up for participant selection, it is crucial to establish eligibility criteria beforehand. [64] The presence of co-interventions alongside the screening process could confound the findings of the included studies. [46,65] Utilizing screening tools that have been validated in prior studies is crucial for ensuring the effectiveness of screening programs and mitigating potential harms. [66] Finally, the criterion for whether the included studies used individual-level outcome data instead of aggregated data was used, as individual-level data typically yield more accurate estimates compared to aggregated data. [67] Moreover, studies using aggregated data are prone to ecological fallacy and Simpson's paradox, which can result in misleading interpretations of the data by masking individual-level relationships. [68]

2.1.4. Statistical methods for meta-analysis

Random effect models using restricted maximum likelihood method were used to account for the potential heterogeneity within studies. Heterogeneity across studies was assessed using the I^2 statistic and prediction intervals (PI). [69] Publication bias was assessed using visual evaluation of Begg's funnel plot and Egger's test. The pooled ORs and corresponding 95% CIs were presented seperately depending on the study design, and the overall estimates were also presented. Subgroup analyses were conducted across outcome definition (depressive symptoms, psychiatric diagnosis and treatment, or suicide), gender (men and women), and geographic region (Asia, America, and Europe). Sensitivity analysis was conducted excluding studies with high-risk of bias. Statistical analyses for meta-analysis were conducted using *metan* command of STATA software version 18.0 (StataCorp LLC, College Station, TX, USA). [70]

2.2. Target Trial Emulation using NHIS Database

The target trial emulation study was conducted to evaluate the depression-related outcomes of the nationwide screening program for depressive symptoms launched in 2018 for all adults in South Korea. The gold standard for quantifying the effectiveness of nationwide depression screening is the RCT design. In an ideal trial setting, eligible participants would be randomly assigned to either the screened or non-screened groups to ensure comparability between them. Therefore, randomization can prevent confounding effect related to prognostic factors and allow for causal inference regarding the effectiveness of an intervention by comparing outcomes between the groups. [71] Additionally, in a RCT, the timing for determining eligibility, assigning treatment, and initiating follow-up is clearly defined and consistently aligned. [72] When an RCT design is impractical due to ethical or logistical constraints, the non-randomized observational studies are used to evaluate intervention effectiveness, and the target trial emulation framework helps reduce biases and improve interpretation by hypothesizing an ideal trial design. [73] In the context of this study, the depression screening program, introduced in South Korea as part of health insurance services, has been offered at no cost to all adults, making it impractical to conduct an RCT. Additionally, the program spans multiple healthcare providers and services, such as antidepressant prescriptions and psychiatrist referrals, necessitating the establishment of external validity for assessing its effectiveness in real-world settings. To address these issues, an observational study based on the target trial emulation framework was conducted using Korean National Health Insurance Service database.

2.2.1. Target trial specification

The first stage of target trial emulation is to explicitly specify the protocol of a hypothetical randomized trial. [73] The second stage is emulation of the specified protocol using the observational data. The specification and emulation of a target trial of the nationwide depression screening is specified in Table 2. The hypothetical ideal trial started in 2019, aligning with the availability of nationwide depression screening program to individuals aged 20–70 years, encompassing the entire target age range.

Causal contrasts of interest

The ideal trial aimed to assess the causal contrasts of traditional ITT and modified ITT effects on depression screening outcomes. Traditional intent-to-treat estimates the impact of being assigned to the intervention, whereas modified intent-to-treat adjusts for protocol

violations. The as-treated estimates were also provided as a supplementary analysis to evaluate the effect of receiving the intervention. Due to low adherence in the national health check-up program, modified intent-to-treat estimates were chosen as the primary estimands to ensure more meaningful reflection of real-world effectiveness. [74] The modified intent-to-treat estimate was found to be comparable to the intent-to-treat estimate regarding attrition bias in meta-analyses of randomized trials. [75]

Eligibility criteria

The purpose of the nationwide depression screening is to identify individuals with depressive symptoms who might not otherwise be recognized or seek treatment on their own. [76] Therefore, the target trial would involve the general adult population alive in 2019, excluding those with a history of depression or those who participated in the screening prior to 2019.

Assignment procedure

In the ideal trial, randomization should be used to assign treatment arms among participants. Therefore, eligible individuals—adults without known depression or previous experience with the nationwide depression screening—would be randomly assigned to either the screening or non-screening group. The non-screening group continued to have access to standard healthcare services as usual.

Follow-up

For the intent-to-treat analysis, follow-up began on January 1, 2019, when eligibility and screening were determined. It ended upon the occurrence of the first of the following events: an outcome event, death, a competing event relevant to the analysis, or the conclusion of follow-up on December 31, 2022. The as-treated analysis, conducted as part of a secondary analysis, defined the start of follow-up as the date of screening.

Outcomes

The effectiveness of the depression screening should be estimated based on the natural history of depression. The nationwide depression screening acts as a form of secondary prevention, aiming to detect and treat depressive symptoms early on before they progress to a chronic or severe state. [76] In an ideal trial, depressive symptoms would be evaluated through clinical interviews following a period of follow-up to determine whether the screening intervention group had a lower incidence of severe depression compared to the control group.

Table 2. Specification of target trial emulation framework.

Protocol	Target trial	Emulation using observational data
Eligibility criteria	General adult population without prior diagnosis of psychiatric disorders and prior participation of the depression screening.	Same.
Treatment strategies	- Screening for depression with identification of risk groups, counseling, and psychiatric referral compared to the treatment-as-usual group. The screening program is performed according to the national guideline for general health checkup. - Tool for depression screening was the PHQ-9. The cut-off point for classifying risk group was 10. - Settings: Any healthcare provider certified to perform general health checkup. The general health checkup including depression screening is performed by licensed medical doctor who completed a specific curriculum. - The cost for the general health check-up including depression screening was fully covered by the Korean NHIS.	Same.
Assignment procedures	The individuals are randomly assigned to undergo depression screening or not.	All South Korean citizens who were 20/30/40/50/60/70 years old in 2019 were invited to the depression screening. To emulate randomization, coarsened exact matching was used to match the screening and non-screened groups by age, sex, and insurance status. Additional confounders, including income, urbanicity of residence, medical comorbidities, and the number of previous health check-ups, were adjusted for in the multivariate Cox regression modeling. The look-back period for identifying confounders was five years prior to the time zero.
Follow-up period	Time zero: time of randomization. Follow-up until the occurrence of outcomes, death, or the end of study period.	Time zero: 1/1/2019. Follow-up until the occurrence of outcomes, death, or the end of study period (12/31/2022).

Protocol	Target trial	Emulation using observational data
Outcome	Direct evaluation of depressive symptoms via clinical interview.	A proxy for severe depression or depression-related outcomes that can be identified within the NHIS database. The records from the Korean NHIS in the form of ICD-10 must be available for all outcome variables except suicide deaths. For suicide deaths, the cause of death must be recorded in ICD-10 format in the database provided by the Korean National Statistical Office. - Primary outcome: hospitalization for mood disorders - Secondary outcomes: use of antidepressants, emergency department visit for mood disorders, and suicide and suicidal behaviors.
Causal contrast of interest	Intention-to-treat (traditional and modified) effects.	Same.
Statistical Analysis plan	Both traditional and modified intention-to-treat analysis contain the adjustment for the baseline characteristics.	The analogue of the intent-to-treat analysis was feasible since the date of eligibility determination, treatment assignment, and the start of follow-up were the same, excluding cases with major protocol violations. Cause-specific Cox proportional hazard models were fitted to adjust for several confounders and control for competing risks.

Abbreviation: NHIS, National Health Insurance Service; PHQ-9, Patient Health Questionnaire-9

2.2.2. Overview of depression screening program in South Korea

The treatment under study was the nationwide depression screening for adults in the general population in settings with integrated feedback and treatment system in South Korea. Participants of the National General Health Screening program who were 20, 30, 40, 50, 60, and 70 years old in 2019 were eligible for the depression screening. Eligibility for depression screening was determined for administrative reasons, independent of prognosis of depression, to manage the number of individuals being screened within the healthcare system. NHIS invited all eligible South Korean citizens to participate in the general health screening and sent them a reminder message. The participation of health checkup with depression screening was determined by individuals' choice. The NHIS covered all costs associated with the National General Health Screening program, making the depression screening free of charge for all participants. The Medical Act requires that the examination and discussion of results of National General Health Screening program be conducted by a trained physician.

The nationwide depression screening in South Korean used Patient Health Questionnaire–9 (PHQ-9). PHQ-9 is a self-report questionnaire that is developed to identify the nine depressive symptom criteria: anhedonia, depressive mood, sleep disturbance, fatigue, appetite disturbance, guilt/worthlessness, trouble concentrating, feeling slowed down or restless, and suicidality/thoughts of death. [77] PHQ-9 evaluates depressive symptoms using a scale from 0 to 3, indicating the frequency of symptoms over the past two weeks: "not at all" (score 0), "several days" (1), "more than half the days" (2), and "nearly every day" (3) (Appendix 1). In a Korean validation study, the PHQ-9 demonstrated a Cronbach's α of 0.852, with a sensitivity of 81.8%, specificity of 89.9%, and a positive likelihood ratio of 8.10 at a cutoff score of 10. The area under the ROC curve was 0.944 ($P<0.05$), indicating excellent diagnostic performance. [78] The accuracy of the PHQ-9 in Korea was slightly lower in sensitivity (pooled estimate: 0.88, 95% CI: 0.83, 0.92) and slightly higher in specificity (pooled estimate: 0.85, 95% CI: 0.82, 0.88) compared to the results of a meta-analysis that synthesized the results of several studies. [79] For elderly or visually impaired individuals, the physician or nurse can read the questionnaire aloud and record the answers.

After completing the self-administration of the PHQ-9, the physician provides several brief interventions as outlined in the manual of in the National General Health Screening program. A score of 5-9 out of a total score of 27 is considered mild depression. In such cases, individuals receive information on understanding and managing depression. Additionally, the presences of certain risk signs are assessed, including: (1) current suicidal

thoughts or plans, (2) a history of suicide attempts, (3) current or past treatment by a mental health professional, (4) current alcohol or drug addiction, (5) episodes of feeling excessively happy or excited for more than four days that disrupt work or daily life, (6) episodes of extreme irritability lasting more than four days that disrupt work or daily life, (7) paranoia or the belief that others are watching you, and (8) hearing voices that others do not hear. If any of these eight risk factors are identified, the participant is referred to a psychiatrist. In the absence of high-risk signs, the physician provided empathy, brief counseling, and, if necessary, prescribes short-term medication to the individual.

If the PHQ-9 score is 10 or higher, indicating moderate or severe depression, if any of the eight risk signs are present, or if symptoms do not improve after an adequate period of antidepressant treatment, the individual should be referred to a local psychiatrist.

2.2.3. Emulation of the target trial

Causal contrasts of interest

This observational study, without randomization, enabled the estimation of ITT effect analogues rather than the true ITT effects. Screening invitations in 2019 were based on specific ages: 20, 30, 40, 50, 60, and 70 years. In the analogue of the ITT analysis, individuals eligible for screening in 2019 were compared to those ineligible. The modified ITT analogue further excluded eligible individuals with major protocol violations before making the same comparison.

Data source

The NHIS conducts the nationwide general health screenings, including depression screening, and stores participants' screening results along with their medical information. For this study, the NHIS database was used to emulate the target trial. Representing 98% of Korea's population's medical expenses, the NHIS database provides representative data, including medical institution types, visit dates, admission lengths, ICD-10 codes, and treatments such as procedures and medications. The Cause of Death Statistics data was linked to the NHIS database using personal identification numbers unique to every citizen.

Study population

The selection of the study population was based on three datasets corresponding to the estimands under analysis: traditional and modified ITT (Figure 2). A separate population of depression screening participants was also selected to provide descriptive data about the

program (Figure 3). From the NHIS database, a 4% random sample of South Koreans (n = 2,123,299) was selected. Exclusions included individuals diagnosed with psychiatric disorders before 2019 (n = 457,119) and those who had undergone depression screening before 2019 (n = 50,834). Among the remaining population, 138,107 individuals met the eligibility criteria for depression screening based on their age (20, 30, 40, 50, 60, or 70 in 2019). To address the specific age distribution from the eligibility criteria, coarsened exact matching was used to match screened and non-screened individuals by sex, age (±2 years), and insurance subscriber type, resulting in 138,107 matched pairs. This dataset, matched purely on screening eligibility, was used for traditional intent-to-treat analysis (Figure 2).

The dataset for the modified intent-to-treat analogue was constructed by excluding individuals with major protocol deviations. This encompassed 86,912 individuals who were invited but did not participate in the depression screening in 2019, as well as 21 individuals in the control group who underwent screening during the same year. An additional 58,993 individuals in the non-screened group, who became eligible for screening between 2020 and 2022 due to administrative factors unrelated to depression prognosis, were also excluded in the current 2019 trial emulation. Finally, to uphold the matched-cohort design, 76,597 individuals without a matched pair were removed. The resulting sample comprised 26,844 individuals in the screened group and an equal number in the non-screened group (Figure 2).

For the observational analogue of the as-treated effect, outcomes were assessed based on actual intervention received, regardless of initial assignments in January 1, 2019. Participants who underwent depression screening during 2019–2022 were compared with those who never participated in screening. To accommodate different follow-up start dates, a sequence of trials was emulated, enabling individuals not invited in 2019 to join trials in subsequent years (2020–2022) as they became eligible. Screeners from each year were matched to non-screeners who did not participate throughout the study period, and the matched non-screeners were assigned an index date corresponding to their matched screener's screening date (Appendix 5). The as-treated analysis followed participants from the date they completed their depression screening until the first occurrence of an outcome event, death, a competing event, or the end of follow-up on December 31, 2022.

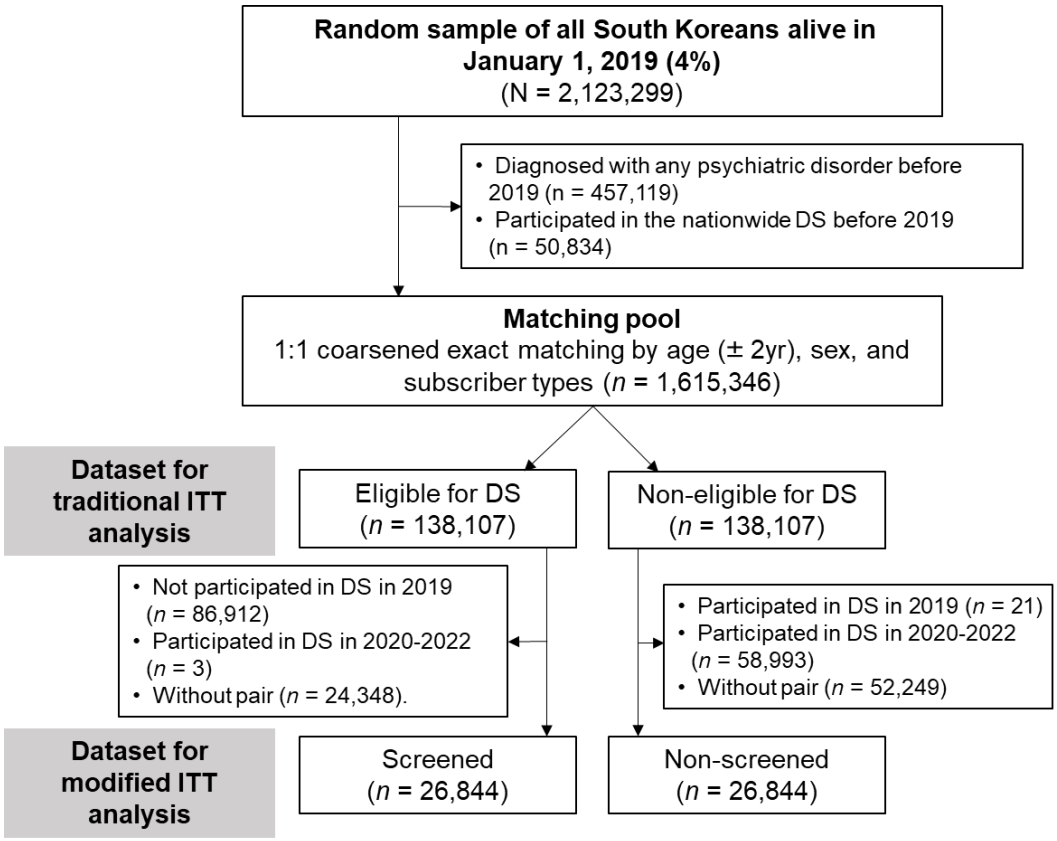


Figure 2. Flow chart of the study population (traditional ITT and modified ITT approach)

Note. The flow chart outlines the selection process for the study population to emulate the traditional ITT and modified ITT approaches in the 2019 trial. The traditional ITT dataset was used to evaluate the effect of assignment to depression screening. The modified ITT dataset assessed the effect of assignment to depression screening, excluding significant protocol deviations. Abbreviation: ITT, intent-to-treat; DS, depression screening; yr, year.

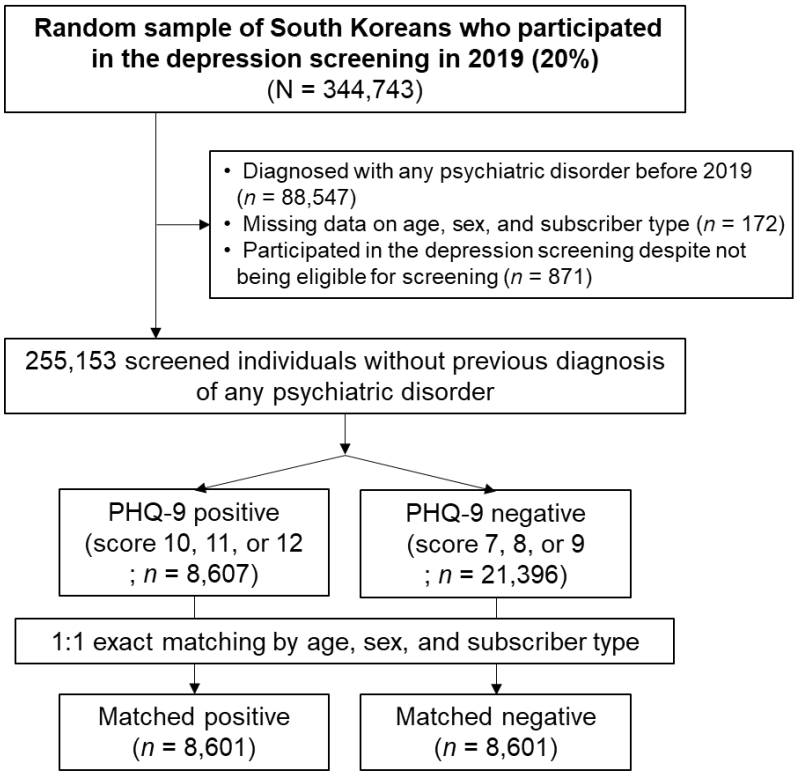


Figure 3. Flowchart of the study population comparing PHQ-9 positives and negatives in 2019 depression screening participants.

Abbreviation: PHQ, Patient Health Questionnaire-9.

To evaluate the post-screening intervention, an exploratory analysis was conducted to compare depression-related outcomes between individuals with positive and negative PHQ-9 results. The dataset included 344,743 individuals, representing 20% of the 2019 screened population. After excluding 88,547 participants with preexisting psychiatric diagnoses, 172 with missing demographic data, and 871 who were ineligible for the screening due to age, the final sample consisted of 255,153 individuals. These data were used to describe the diagnostic accuracy of the PHQ-9 and outcome distributions within participants of the depression screening. Outcomes were analyzed by categorizing participants into positive (PHQ-9 scores: 10–12) and negative (scores: 7–9) groups, using a regression discontinuity-inspired design to control for extraneous differences. [80] To

further balance the groups, positive and negative screeners were matched by age, sex, and subscriber type (Figure 3).

Follow-up

The study design for the follow-up period aligns with the protocol outlined for the target trial (Figure 4). As this study used the NHIS database, only individuals who had died or emigrated lost their NHIS subscription, therefore loss to follow-up. In the ITT analyses, the time zero (start of follow-up), the specification of eligibility criteria, and the treatment assignment were aligned and synchronized. This specification of time zero prevents the possibility of immortal time bias and selection bias when emulating the randomized assignment. [72]

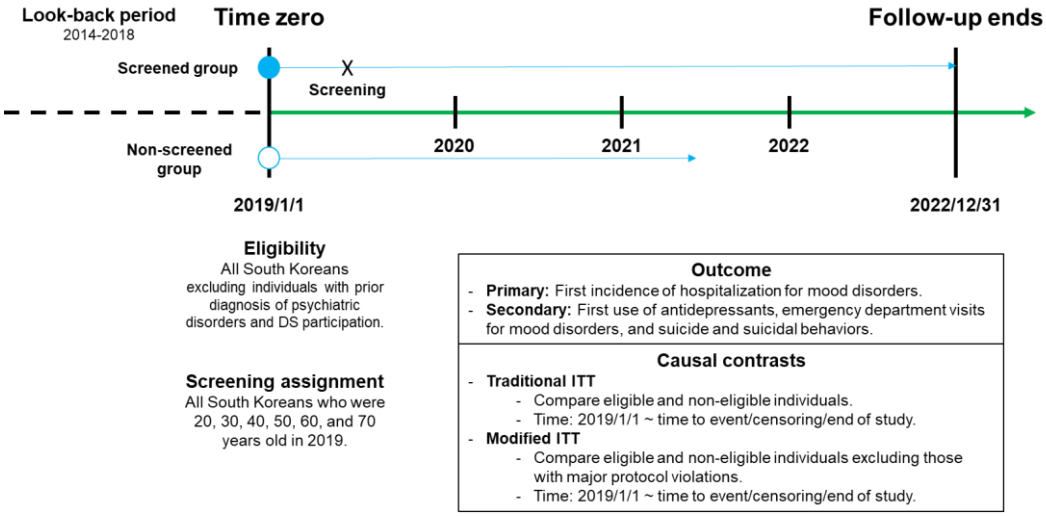


Figure 4. Illustration of the study timeframe

Abbreviation: DS, depression screening; ITT, intention-to-treat.

Outcomes

The target trial protocol outlined an improvement of depression-related prognosis, but this NHIS-based observational study lacks direct measures of depression severity. As a proxy, hospitalizations for mood disorders (ICD-10 codes F3) were chosen as the primary outcome, requiring at least a three-day stay. Hospitalizations occurring within 30 days of discharge were treated as one episode. The secondary outcomes included the initial use of antidepressants, suicide and suicidal behaviors, and emergency department visits for mood disorders. Antidepressants were identified based on clinical guidelines for depression treatment, textbook references, and review by a qualified psychiatrist. [81,82] The identified antidepressants included selective serotonin reuptake inhibitors (SSRIs), serotonin–norepinephrine reuptake inhibitors (SNRIs), and tricyclic antidepressants (TCAs) (Appendix 2). These drugs were classified using the Anatomical Therapeutic Chemical classification system. Suicide-related behaviors and deaths were determined using ICD-10 codes X60–X84 and Y10–Y34, derived from diagnostic claims or death records. Emergency department visits for mood disorders were defined as those associated with a diagnosis of mood disorder (ICD-10 codes F3) and involving hospitalization for fewer than two days.

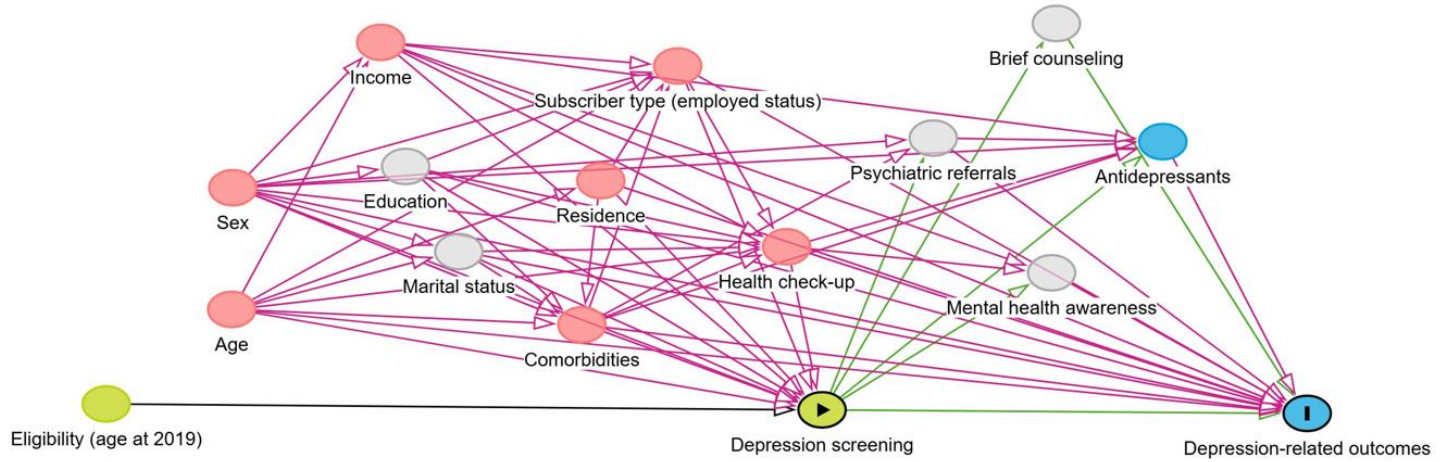


Figure 5. Directed acyclic graph adopted for this study

Note. The directed acyclic graph illustrates the causal pathway from eligibility for depression screening, determined by specific ages in 2019, through participation in screening, to depression-related outcomes. Red circles indicate observed confounders, gray circles indicate unobserved variables, and the blue circle indicates observed mediators. Red lines indicate biasing paths, while green lines indicate causal pathways.

Covariates

A five-year look-back period starting from the index dates was used to identify confounders for all analyses. The covariate set required to approximate randomization of conditions was determined using a directed acyclic graph which was created using *daggity* software (Figure 5). Covariate adjustments were performed through coarsened exact matching and multivariate regression models. To address the sparse age distribution and potential violations of positivity assumptions in causal inference, coarsened exact matching by age, sex, and insurance subscriber type was applied, considering the age-based eligibility for depression screening. This approach further improved the control of key confounders in large-scale datasets. The subscriber type was categorized into three groups based on the socio-economic status of the head of the household: medical aid recipients, employed individuals, and self-employed individuals. An individual who cannot afford coverage for NHIS on their own is eligible for health insurance as a dependent under the head of household's insurance status.

Next, adjustments were made in the multivariate Cox regression models to account for confounders such as urbanicity of residence, frequency of previous health check-ups, relative income level, and Charlson Comorbidity Index (CCI) scores. Residential regions were registered in the NHIS database and classified into metropolitan, urban, and rural areas. The frequency of prior health check-ups was considered a proxy measure for participants' engagement with their health. Relative income levels were divided into five categories—medical aid, Q1 representing the lowest, Q2, Q3, and Q4 representing the highest. The CCI was used to control for medical comorbidities, with CCI scores calculated from the appearance of specific ICD-10 codes for each comorbidity prior to the index date.[83] The CCI included acute myocardial infarction, congestive heart failure, peripheral vascular disease, cerebral vascular accident, dementia, pulmonary disease, connective tissue disorder, peptic ulcer, liver disease, diabetes, diabetes complications, paraplegia, renal disease, cancer, metastatic cancer, severe liver disease, and HIV (Appendix 3). For analyses that used hospitalization for cancer as a negative control outcome, the modified version of CCI that excluded cancer was used.

2.2.4. Statistical analysis

First, the response rates were analyzed and presented by subgroup, including sex, age, insurance type, income level, urbanicity of residence, and CCI scores among individuals who were eligible for the depression screening in 2019. Second, sensitivity, specificity, positive predictive value, negative predictive value, and the positive likelihood ratio were derived from the 2019 screening dataset to determine the diagnostic performance of the PHQ-9 for diagnosing new mood disorders within 3 months. Third, the distribution of outcomes, such as mood disorder-related hospitalizations, antidepressant type and usage, emergency department visits, and incidents of suicide or suicidal behavior, was illustrated by PHQ-9 score categories. Fourth, categorical variables describing participants' baseline characteristics were reported as numbers and percentages, while continuous variables were summarized as mean ± standard deviation or median (interquartile range). Since the study generated four datasets tailored to the estimands and research objectives, the baseline characteristic distributions were presented individually for each dataset.

The main statistical analysis was conducted using cause-specific Cox proportional hazards models, stratified on matched pairs, to estimate hazard ratios (HRs) and corresponding 95% confidence intervals (CIs) for the associations between nationwide depression screening and outcomes, accounting for death as a competing risk. The time to first admission due to mood disorders was compared between the screened and non-screened groups from the start of follow-up. In the intent-to-treat analysis, the index date was January 1, 2019. The proportional hazards assumption was checked using Schoenfeld residuals, nonlinearity was assessed by Martingale residuals, and influential observations were examined using deviance residuals. In the first model, the age-, sex-, and insurance status-matched screened group and non-screened group were compared without covariate adjustment. In the adjusted model, additional covariates including income, urbanicity of residence, medical comorbidities, and the number of previous health check-ups were included. The same analyses were also applied to examine secondary outcomes of the nationwide depression screening, such as the use of antidepressants, emergency department visits due to psychiatric symptoms, and instances of suicide and suicidal behaviors. Subgroup analyses were conducted based on sex, age group (over 60 years vs. younger), and income level (Q3 and Q4 vs. Medical aid, Q1, and Q2).

Sensitivity analyses of the modified ITT results examined the use of subdistributional hazard models instead of cause-specific hazard models the lag time for hospitalizations due to mood disorders, the duration of antidepressant use, and the types of antidepressants (SSRI, SNRI, TCA). Hospitalizations were assessed with lag periods of 180 days and 1 year for outcome identification. Antidepressant duration was analyzed based on continuous use for 30 and 90 days, while antidepressant types were categorized as SSRIs, SNRIs, and TCAs. Additionally, the E-values were calculated to evaluate the robustness of the identified significant associations in the presence of potential unmeasured confounders. [84]

All statistical analyses were conducted using SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA). The visual graphs were generated using *gglot2* and *ComplexUpset* from the *r* package (R Core Team , Vienna, Austria). [85]

2.2.5. Ethical statement

The study used data that had been anonymized and de-identified to eliminate the possibility of identifying individuals, allowing the waiver of informed consent. This study protocol was approved by the Institutional Review Board of Severance Hospital at Yonsei University Health System (4-2024-0242) and the Health Insurance Review and Assessment Service (NHIS-2024-1-464).

3. RESULTS

3.1. Systematic Review and Meta-Analysis

3.1.1. Included and excluded Studies

From PubMed, Web of Science, EMBASE, and PsychINFO, 12,978 publications were identified. After excluding 8,939 duplicates, two independent researchers reviewed the titles and abstracts of 4,039 publications. Of the 3,984 removed studies, the majority were development or validation studies for depression screening tools (n=1,420) or non-peer-reviewed articles (e.g., conference abstracts, commentaries) (n=905). Additionally, 725 studies were excluded for describing or applying the depression screening results to find associations between variables without comparing outcomes between screened and non-screened group, 336 studies were unrelated to depression screening effectiveness (i.e., association between depression and screen time), and 208 studies focused on children and adolescents instead of adults. Finally, 390 studies were excluded for evaluating outcomes unrelated to depression prognosis. During the full-text review, 37 articles were excluded. Finally, nine RCTs and nine non-randomized observational studies were included in this review. The flow diagram suggested by the PRISMA is presented in Figure 6.

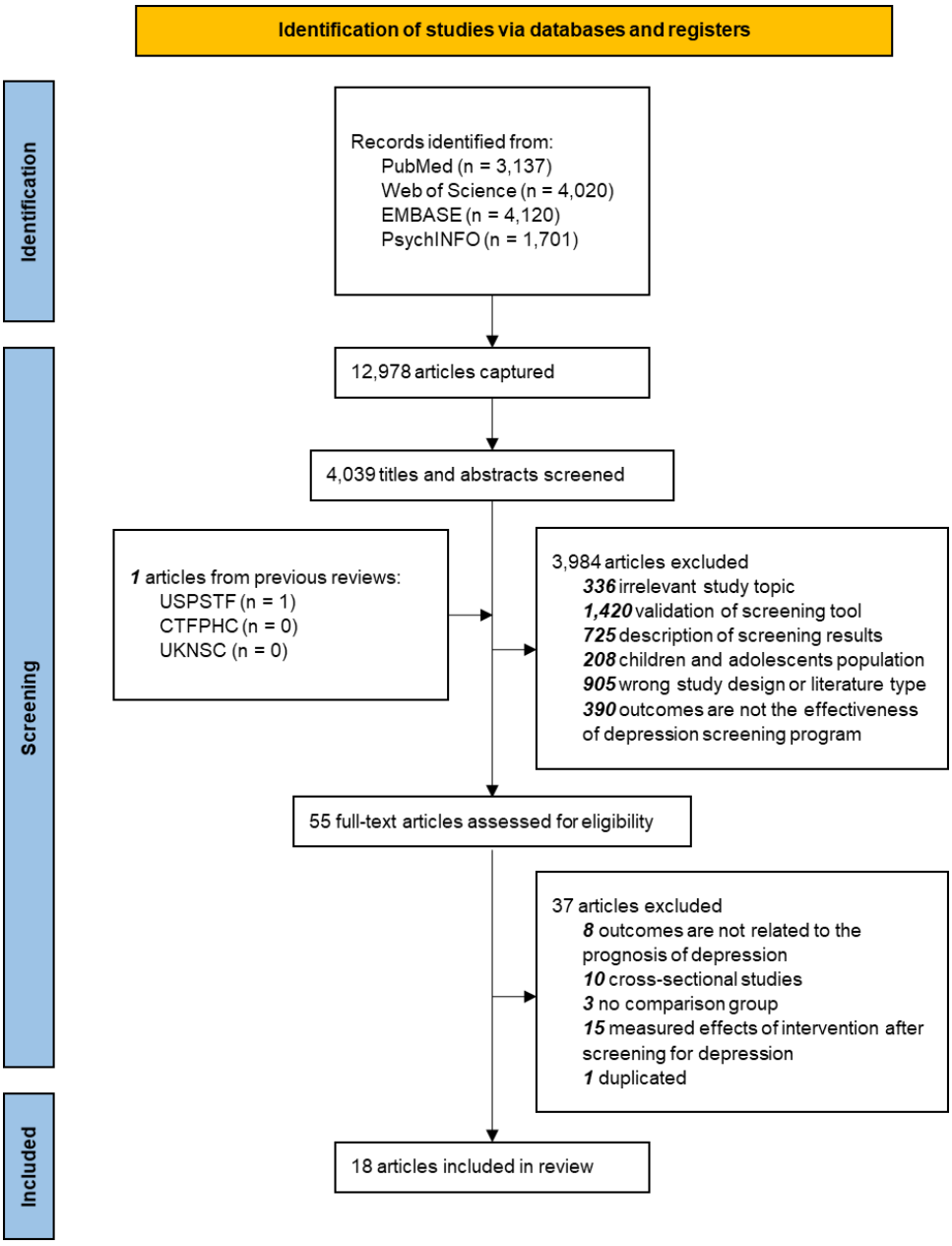


Figure 6. PRISMA flow diagram

3.1.2. Study characteristics and risk of bias assessments

The characteristics of included studies are presented in Table 3. Eight studies were led by the same author group in Aomori Prefecture, Japan. [86-93] Three studies were conducted in the USA[94-96], three in the UK[97-99], and others in Canada, Spain, China, and Taiwan. [100-103] Of the RCT studies, three focused on the general population in primary care or community settings[90,100,101], two examined patients with specific illnesses such as acute coronary syndromes and osteoarthritis[94,98], one on military veterans[97], and two on the postpartum mothers. [99,102] All NRSs were focused on the general population aged at least 40 years or older. [86-93,96,103] Total number of participants was 10,098,035 (RCT *n* = 10,084,165, NRS *n* = 13,870), with women comprising an average of 59.5%. The follow-up period varied between 2 and 120 months. Regarding outcomes of depression screening, seven studies assessed depressive symptoms using self-reported tools[90,94,97-100,102], five focused on psychiatric diagnoses or hospitalizations[95-97,101,103], and eight examined suicide. [86-89,91-93,103] Detailed summaries of the included studies are presented in Appendix 4.

In the evaluation of risk of bias using RoB 2, four studies were classified as having a low risk of bias[94,97,98,102], two studies were noted to have some concerns regarding bias[95,99], and the remaining three studies were classified as having a high risk of bias[90,100,101] (Table 4). In the evaluation of risk of bias using ROBINS-I, all studies, except for one study conducted by Chen[103], were identified as having a serious bias in their studies (Table 5). [86-89,91-93,96]

Table 3. Characteristics of studies included in the review

Author, Year	Country	Population	Sample size	Women, %	FUP, months	Outcomes
RCT						
Kronish, 2020[94]	USA	Patients with specific illness (acute coronary syndromes)	*N* = 999 Intervention *n* = 499 Control *n* = 500	28.3	18	Depressive symptoms (PHQ-8 score)
Silverstone, 2017[100]	Canada	General population in primary care	*N* = 530 Intervention *n* = 255 Control *n* = 275	N/I	3	Depressive symptoms (PHQ-9 score)
Rona, 2017[97]	UK	Military veterans	*N* = 6,365 Intervention *n* = 3,996 Control *n* = 2,369	2.8	17	Depressive and anxiety symptoms (PHQ-9 and GAD-7 score) and any mental disorder diagnosis
Mallen, 2017[98]	UK	Patients with specific illness (osteoarthritis)	*N* = 1,412 Intervention *n* = 501 Control *n* = 911	56.7	12	Depressive symptoms (PHQ-8 score)
Oyama, 2014[90]	Japan	Aged 40 to 64 years (Shichinohe Township)	*N* = 1,856 Intervention *n* = 690 Control *n* = 1,166	51.5	60	Moderate-to-severe depressive symptoms (CES-D≥24)
Romera, 2013[101]	Spain	General population in primary care	*N* = 525 Intervention *n* = 257 Control *n* = 268	73.9	2	Recognition and treatment of depression (randomly selected records)
Leung, 2011[102]	China	Postpartum mothers	*N* = 462 Intervention *n* = 231 Control *n* = 231	100	6	Depressive symptoms (EPDS≥10)
MacArthur, 2002[99]	UK	Postpartum mothers	*N* = 1,503 Intervention *n* = 801 Control *n* = 702	100	4	Depressive symptoms (EPDS≥13)
Williams, 1999[95]	USA	General population in primary care	*N* = 218 Intervention *n* = 153 Control *n* = 65	71	3	Diagnosis of depression

Author, Year	Country	Population	Sample size	Women, %	FUP, months	Outcomes
NRS						
Chen, 2024[103]	Taiwan	Aged 40 and over	*N* = 9,944,456 Intervention *n* = 4,972,228 Control *n* = 4,972,228	55	84	Psychiatric hospitalization, treatment of depression, and suicide
Rhee, 2018[96]	USA	Adults aged 65 or older	*N* = 15,596 Intervention *n* = 323 Control *n* = 15,273	58.3	48	Mental disorder diagnosis (macro-level)
Oyama, 2017[88]	Japan	Aged 36-64 in Aomori Prefecture	*N* = 16,097 Intervention *n* = 8,010 Control *n* = 8,087	N/I	48	Suicide (macro-level)
Oyama, 2016[89]	Japan	Aged 60 and over (Hachinohe)	*N* = 24312 Intervention *n* = 11,710 Control *n* =12,602	57.5	72	Suicide (macro-level)
Oyama, 2010[87]	Japan	Aged 60 and over (Minami)	*N* = 28,435 Intervention *n* = 14,504 Control *n* = 13,931	59	24	Suicide (macro-level)
Oyama, 2006 (1)[91]	Japan	Aged 65 and over (Matsudai)	*N* = 15,260 Intervention *n* = 6,015 Control *n* = 9,245	N/I	120	Suicide (macro-level)
Oyama, 2006 (2)[86]	Japan	Aged 65 and over (Yasuzuka)	*N* = 14,031 Intervention *n* = 4,940 Control *n* = 9,091	N/I	120	Suicide (macro-level)
Oyama, 2006 (3)[92]	Japan	Aged 65 and over (Minami)	*N* = 4,985 Intervention *n* = 1,685 Control *n* = 3,300	N/I	72	Suicide (macro-level)
Oyama, 2004[93]	Japan	Aged 65 and over (Joboji)	*N* = 20,993 Intervention *n* = 9,721 Control *n* = 11,272	N/I	120	Suicide (macro-level)

Abbreviation: RCT, randomized controlled trial; NRS, non-randomized observational studies; FUP, follow-up; OR, odds ratio; CI, confidence interval; N/I, no information; PHQ, Patients Health Questionnaire; GAD-7, Generalized Anxiety Disorder-7; CES-D, Center for Epidemiological Studies-Depression Scale; EPDS, Edinburgh Postnatal Depression Scale.

Table 4. Risk of bias assessments for RCT studies using RoB 2

Reference	Randomization process	Deviations from intended interventions	Missing outcome data	Outcome measurement	Selection of the reported results	Overall bias
Kronish I.M., 2020[94]	Low	Low	Low	Low	Low	Low
Silverstone., 2017[100]	High	High	Some concerns	Low	Low	High
Rona., 2017[97]	Low	Low	Low	Low	Low	Low
Mallen C.D., 2017[98]	Low	Low	Low	Low	Low	Low
Oyama H., 2014[90]	Some concerns	Some concerns	High	Some concerns	Some concerns	High
Romera I., 2013[101]	Low	Some concerns	High	High	Low	High
Leung S.S.L., 2011[102]	Low	Low	Low	Low	Low	Low
MacArthur., 2002[99]	Low	Some concerns	Some concerns	Low	Some concerns	Some concerns
Williams Jr. J.W. 1999[95]	Low	Some concerns	Some concerns	Low	Low	Some concerns

Notes. Dark gray represents a high risk of bias in each study domain, light gray indicates some concerns, and white reflects a low risk of bias, as assessed using the RoB 2 guideline. Abbreviation: RCT, randomized controlled trial; RoB 2, Revised Cochrane Risk-of-Bias Tool for Randomized Trials.

Table 5. Risk of bias assessments for observational studies using ROBINS-I

Reference	Confounding	Selection of participants	Classification of intervention	Deviations from intended interventions	Missing data	Outcome measurement	Selection of the reported results	Overall bias
Chen Y.-L., 2024[103]	Moderate	Low	Low	Moderate	Low	Low	Low	Moderate
Rhee T.G., 2018[96]	Serious	Moderate	Moderate	Moderate	Serious	Serious	Moderate	Serious
Oyama H., 2017[88]	Serious	Serious	Low	Serious	Low	Low	Low	Serious
Oyama H., 2016[89]	Serious	Serious	Low	Serious	Low	Low	Low	Serious
Oyama H., 2010[87]	Serious	Serious	Low	Serious	Low	Low	Low	Serious
Oyama H., 2006 (1) [91]	Serious	Serious	Low	Serious	Low	Low	Low	Serious
Oyama H., 2006 (2) [86]	Serious	Serious	Low	Serious	Low	Low	Low	Serious
Oyama H., 2006 (3) [92]	Serious	Serious	Low	Serious	Low	Low	Low	Serious
Oyama H., 2004[93]	Serious	Serious	Low	Serious	Low	Low	Low	Serious

Notes. Dark gray represents a serious risk of bias in each study domain, light gray indicates moderate risk of bias, and white reflects a low risk of bias, as assessed using the ROBINS-I guideline. Abbreviation: NRSs, non-randomized observational studies; ROBINS-I, Risk of Bias in Non-Randomized Studies - of Interventions.

3.1.3. Findings on the effectiveness of depression screening in previous studies

Combined findings from nine RCTs and nine NRSs on depression screening suggest that evidence for its effectiveness in the general population is inconsistent, with low-quality evidence overall. None of the studies achieved all six criteria for evaluating the level of evidence (Table 6). While three studies did not impose eligibility restrictions on the depression screening, they lacked detailed descriptions of participant characteristics and demographics. [95,100,101] A study from Taiwan evaluated the effectiveness of nationwide depression screening but excluded individuals in their 20s and 30s. [103] Only three RCTs and one NRS excluded individuals who had been treated or were currently undergoing treatment for depression and psychosis before screening. [94,98,102,103] Seven RCTs and one NRS minimized post-randomization exclusions by defining study eligibility prior to follow-up or screening. [94-100,102,103] Three studies included interventions other than depression screening, categorization of high-risk group, and associated medication and non-medication treatments. The UK study of veterans included screening and treatment for PTSD, alcohol use, and anxiety[97]; the UK study of osteoarthritis patients included screening and treatment for anxiety and pain[98]; and the Taiwanese study of the general population included health check-up factors other than depression screening. [103] With one exception, the studies implemented validated screening instruments for depression. The U.S. study, which used a difference-in-differences design to evaluate changes in prevalence of mental disorder diagnosis post-national depression screening guidelines, did not clarify the screening tools participants utilized. [96] Eight studies followed participants over time to identify differences in outcomes between screened and non-screened groups[94-100,102,103], while the remaining studies relied on suicide registries or random sampling from the settings participants were associated with. It should be noted that Oyama's studies repeatedly evaluated the outcome of nearly identical depression screening programs across different regions, resulting in shared similarities in study design, target population, and program content. [86-93]

Table 6. Evaluation of the level of evidence for the effectiveness of depression screening using six predetermined criteria

Reference	1st criteria	2nd criteria	3rd criteria	4th criteria	5th criteria	6th criteria
RCT						
Kronish I.M.., 2020[94]	Yes	Yes	Yes	No	Yes	Yes
Silverstone., 2017[100]	No	No	Yes	No	Yes	Yes
Rona., 2017[97]	Yes	No	Yes	Yes[1]	Yes	Yes
Mallen C.D., 2017[98]	Yes	Yes	Yes	Yes[2]	Yes	Yes
Oyama H., 2014[90]	Yes	No	No	No	Yes	Partially
Romera I., 2013[101]	No	No	No	No	Yes	Partially
Leung S.S.L., 2011[102]	Yes	Yes	Yes	No	Yes	Yes
MacArthur., 2002[99]	Yes	No	Yes	No	Yes	Yes
Williams Jr. J.W., 1999[95]	No	No	Yes	No	Yes	Yes
NRS						
Chen Y.-L., 2024[103]	Partially	Yes	Yes	Yes[3]	Yes	Yes
Rhee T.G., 2018[96]	Yes	No	No	No	No	No
Oyama H., 2017[88]	Yes	No	No	No	Yes	No
Oyama H., 2016[89]	Yes	No	No	No	Yes	No
Oyama H., 2010[87]	Yes	No	No	No	Yes	No
Oyama H., 2006 (1) [91]	Yes	No	No	No	Yes	No
Oyama H., 2006 (2) [86]	Yes	No	No	No	Yes	No
Oyama H., 2006 (3) [92]	Yes	No	No	No	Yes	No
Oyama H., 2004[93]	Yes	No	No	No	Yes	No

Notes. Criteria for assessing each study include: (1) targeting specific population, (2) exclusion of prevalent cases of depression, (3) eligibility determined before follow-up, (4) inclusion of additional components besides screening, (5) use of a validated screening tool, and (6) analysis of outcomes using individual-level data. White cells correspond to the ideal criteria components, light gray marks designs that are partially ideal, and dark gray signifies non-ideal design components.

[1] Screening for PTSD, alcohol use abuse, and anxiety. [2] Screening and treatment for anxiety and pain. [3] Other health check-up. Abbreviation: RCT, randomized controlled trial; NRS, non-randomized observational study; PHQ-2, Patient Health Questionnaire-2; GAD-2, Generalized Anxiety Disorder-2; N/I, no information; N/S, not significant; FUP, follow-up.

These variations in study design, population, and the quality of prior research on depression screening have been transitioned into the results of meta-analyses. Figure 7 depicts the results of meta-analysis stratified by the study design. The meta-analysis revealed that depression screening reduced the odds of depression-related outcomes by 26% (OR 0.74, 95% CI 0.62–0.87). The I^2 statistic was 72.4% ($p < 0.001$), and the 95% PI ranged from 0.40 to 1.34, indicating considerable heterogeneity within studies. The pooled OR in RCTs was 0.84 (95% CI 0.70–1.00; I^2 73.4%, $p < 0.001$; 95% PI 0.48–1.47), and in

NRSs, it was 0.58 (95% CI 0.43–0.78; I^2 73.0%, $p < 0.001$; 95% PI 0.25–1.37). The funnel plot showed an asymmetrical figure, indicating an overestimation of the intervention effect due to publication bias (Figure 8). The Egger's test showed that the small-study effect was statistically significant (bias coefficient -1.27, $p = 0.008$).

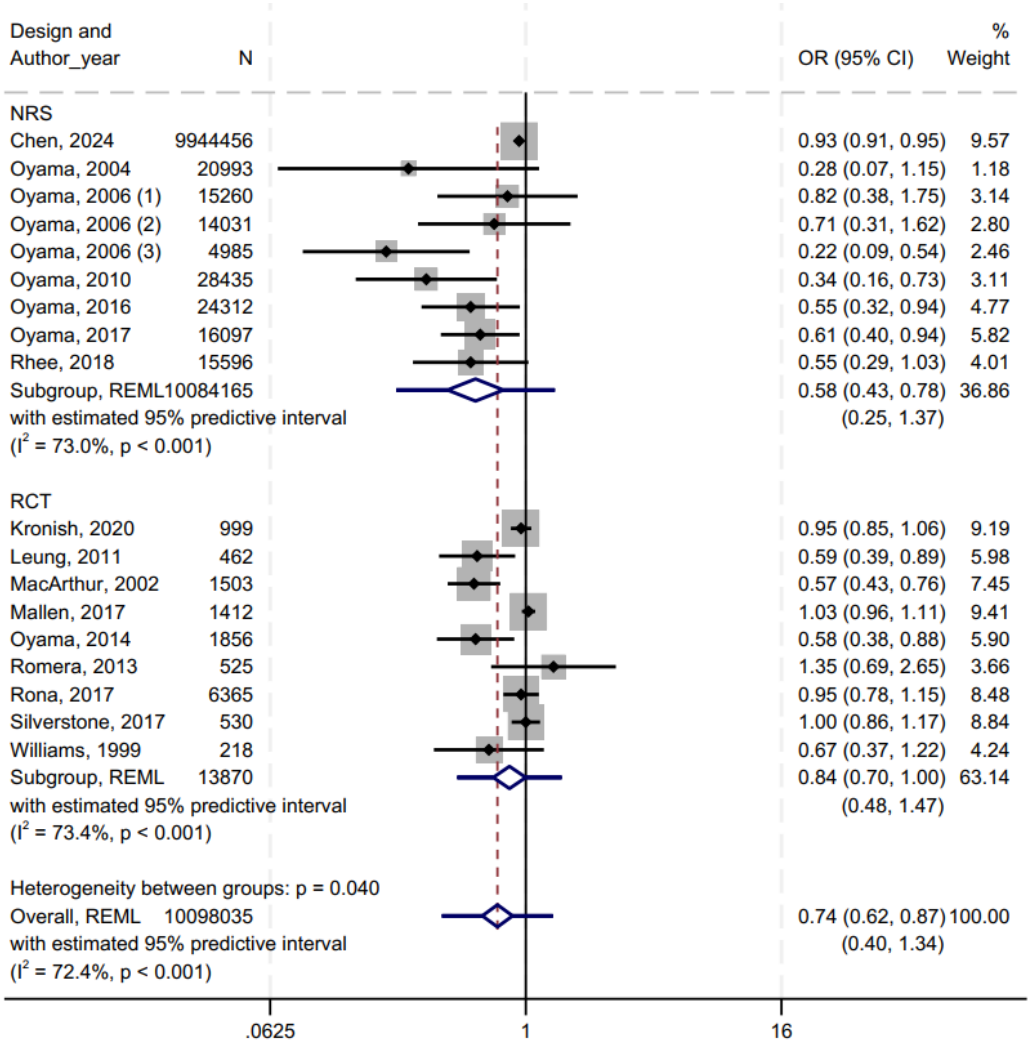


Figure 7. Forest plot of depression screening outcomes

Abbreviation: NRS, non-randomized observational studies; RCT, randomized controlled trial; REML, restricted maximum likelihood; OR, odds ratio; CI, confidence intervals

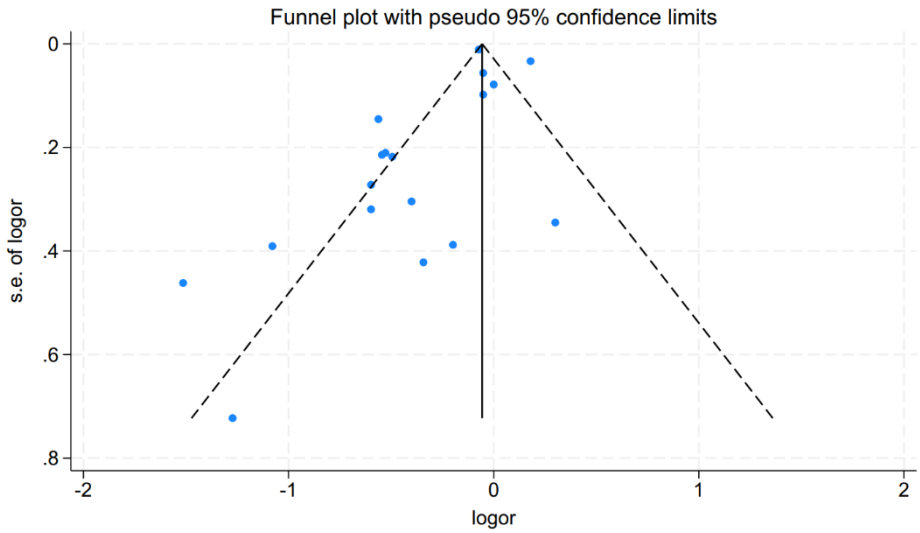


Figure 8. Funnel plot and publication bias assessment

When the outcomes were categorized into depressive symptoms, psychiatric diagnosis and treatment, and suicide, all three pooled ORs showed statistical significance (OR for depressive symptoms: 0.88, 95% CI: 0.67–0.99; OR for psychiatric diagnosis and treatment: 0.93, 95% CI: 0.91–0.96; OR for suicide: 0.58, 95% CI: 0.40–0.83). The pooled OR was not statistically significant for men (OR 1.00; 95% CI 0.98–1.03), but it was significant for women (OR 0.57; 95% CI 0.43–0.75). When stratified by the geographic regions, the effect of the depression screening was significant in Asia (OR 0.60; 95% CI 0.46–0.77). Findings from both America and Europe did not show statistical significance. The pooled estimate became marginally significant (OR 0.85, 95% CI: 0.70–0.99) when studies identified as having a high risk of bias were excluded (Table 7).

Table 7. Subgroup meta-analysis on outcomes of depression screening

Stratification group		Overall			RCT			NRS		
		N	OR (95% CI) (95% PI)	I^2	*N*	OR (95% CI) (95% PI)	I^2	*N*	OR (95% CI) (95% PI)	I^2
Outcome definition	**Depressive symptoms**	7	0.88 (0.67, 0.99) (0.43, 1.56)	78.4	7	0.88 (0.67, 0.99) (0.43, 1.56)	78.4	-	-	-
	Psychiatric diagnosis / hospitalization	5	0.93 (0.91, 0.96) (0.90, 0.96)	21.2	3	0.94 (0.79, 1.13) (0.30, 2.96)	14.8	2	0.79 (0.49, 1.27)	62.9
	Suicide	8	0.58 (0.40, 0.83) (0.20, 1.69)	78.3	-	-	-	8	0.58 (0.40, 0.83) (0.20, 1.69)	78.3
Sex	**Men**	10	1.00 (0.98, 1.03) (0.97, 1.04)	9.6	3	0.96 (0.85, 1.07) (0.46, 2.00)	35.1	7	0.92 (0.85, 1.07) (0.57, 1.49)	1.3
	Women	11	0.57 (0.43, 0.75) (0.25, 1.28)	70.9	4	0.67 (0.52, 0.85) (0.27, 1.67)	52.9	7	0.41 (0.23, 0.73) (0.08, 2.05)	71.6
Region	**Asia**	10	0.60 (0.46, 0.77) (0.29, 1.22)	75.2	2	0.59 (0.44, 0.79)	0	8	0.57 (0.41, 0.81) (0.21, 1.54)	74.1
	America	4	0.95 (0.87, 1.04) (0.78. 1.15)	35.7	3	0.96 (0.88, 1.05) (0.54, 1.70)	0	1	0.55 (0.29, 1.03)	0
	Europe	4	0.89 (0.65, 1.22) (0.22, 3.59)	81.9	4	0.89 (0.65, 1.22) (0.22, 3.59)	81.9	-	-	-
Risk of Bias	**Excluded high-risk studies**	7	0.85 (0.70, 1.00) (0.49, 1.46)	75.8	6	0.81 (0.66, 1.01) (0.40, 1.66)	78.5	1	0.93 (0.91, 0.95)	0

Abbreviation: RCT, randomized controlled trial; NRS, non-randomized observational study; CI, confidence interval; PI, predictive interval.

3.1.4. Summary

Nine RCTs and nine non-randomized observational studies were identified to examine the effectiveness of depression screening in the general adult population within primary care, non-mental health practices, or community settings. There is some evidence suggesting that the depression screening program might help prevent worse prognoses in depression-related outcomes; however, this should be interpreted cautiously. Depression screening yielded a pooled OR of 0.74 (95% CI 0.62–0.87) for depression-related outcomes such as depressive symptoms, psychiatric diagnoses, and suicide, though significant heterogeneity and publication bias undermine the accuracy of this estimate. The pooled estimate reached marginal significance (OR 0.85, 95% CI: 0.70–0.99) after studies with a high risk of bias were removed from the analysis. A pronounced effect of depression screening was observed in women (OR 0.57, 95% CI 0.43–0.75) and in Asian populations (OR 0.60, 95% CI 0.46–0.77). However, most of these studies showed a high risk of bias and considerable heterogeneity, highlighting the persistent lack of strong evidence supporting depression screening. Of the 18 studies, only seven studies were categorized as having low, moderate, or some concerns regarding bias. None of the studies met all six criteria for assessing the evidence level.

3.2. Target Trial Emulation using NHIS Database

3.2.1. Response rate of depression screening among eligible group

After randomly selecting 2,123,299 individuals from all citizens alive in 2019, 457,119 with preexisting psychiatric disorders and 50,834 previously screened for depression were excluded. Of the remaining individuals, 138,107 aged 20, 30, 40, 50, 60, or 70 were eligible for depression screening in 2019, and 26,844 were actually screened (overall response rate in 2019: 19.4%). Screened and non-screened individuals were matched by age, sex, and subscription type, with age intervals set two years apart at the upper and lower bounds, creating a study population aged 18 to 72 years. Figure 9 shows the distribution of individuals eligible for depression screening, separated into participants and non-participants across subgroups. Women had a response rate of 40.1%, which was higher than men's 34.5%. Younger adults aged 20–30 had lower response rates than older adults, with rates of 12.3% for individuals in their 20s and 36.7% for those in their 30s. Individuals with employee-based insurance had the highest response rate (41.3%), while those with self-employed insurance and medical aid showed lower rates. Urbanicity of residence showed minimal differences in response rates. Higher participation in prior health check-ups was associated with higher response rates in the 2019 depression screening. Additionally, individuals with two or more comorbidities had greater response rates compared to those with one or no comorbidities.

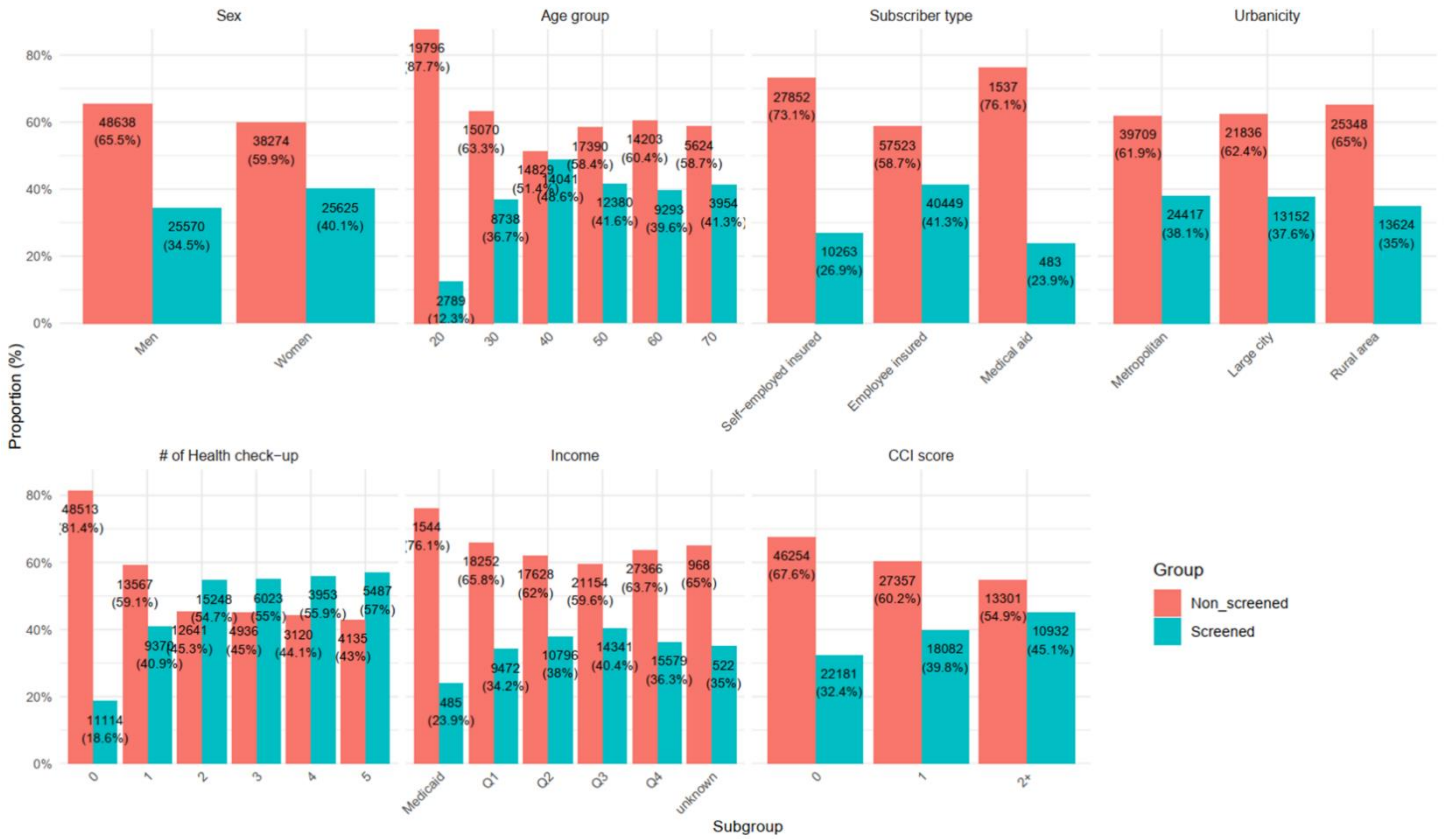


Figure 9. Patterns of depression screening response rates among subgroups eligible in 2019 (*n* = 138,107)

3.2.2. Diagnostic performance of PHQ-9 for diagnosis of mood disorders

The nationwide depression screening program conducted in 2019 used the PHQ-9 tool with a cut-off score of 10 to identify individuals with clinically significant depressive symptoms. Out of 255,153 participants without preexisting psychiatric diagnoses, 15,173 (5.94%) were classified as positive based on PHQ-9 results. Of 678 patients diagnosed with mood disorders within 3 months, 129 had positive PHQ-9 results, reflecting a sensitivity of 19%. Among 254,475 individuals without mood disorders, 239,131 tested negative, corresponding to a specificity of 94.1%. The screening demonstrated a positive predictive value of 0.85% and a negative predictive value of 99.8%. The positive likelihood ratio was 3.22, meaning that a positive test result was 3.2 times more likely to occur in individuals with mood disorders compared to those without mood disorders (Table 8).

Table 8. Depression screening results in 2019 (PHQ-9) for diagnosis of mood disorders within 3-months

PHQ-9 results	Mood disorder Dx	No mood disorder Dx	*Total*
Positive	129	15,044	15,173
Negative	549	239,431	239,980
Total	678	254,475	255,153

Abbreviation: PHQ-9, Patients Health Questionnaire-9; Dx, diagnosis.

3.2.3. Depression-related outcomes among participants of screening

Over the course of 4 years, 14,019 out of a total of 255,153 participants in depression screening in 2019 (5.5%) were diagnosed with mood disorders. Mood disorder incidence varied by PHQ-9 score category, with proportion of 22.3% for those reporting severe depressive symptoms, 12.2% for moderate symptoms, 7% for mild symptoms, and 4.5% for no symptoms (Figure 10).

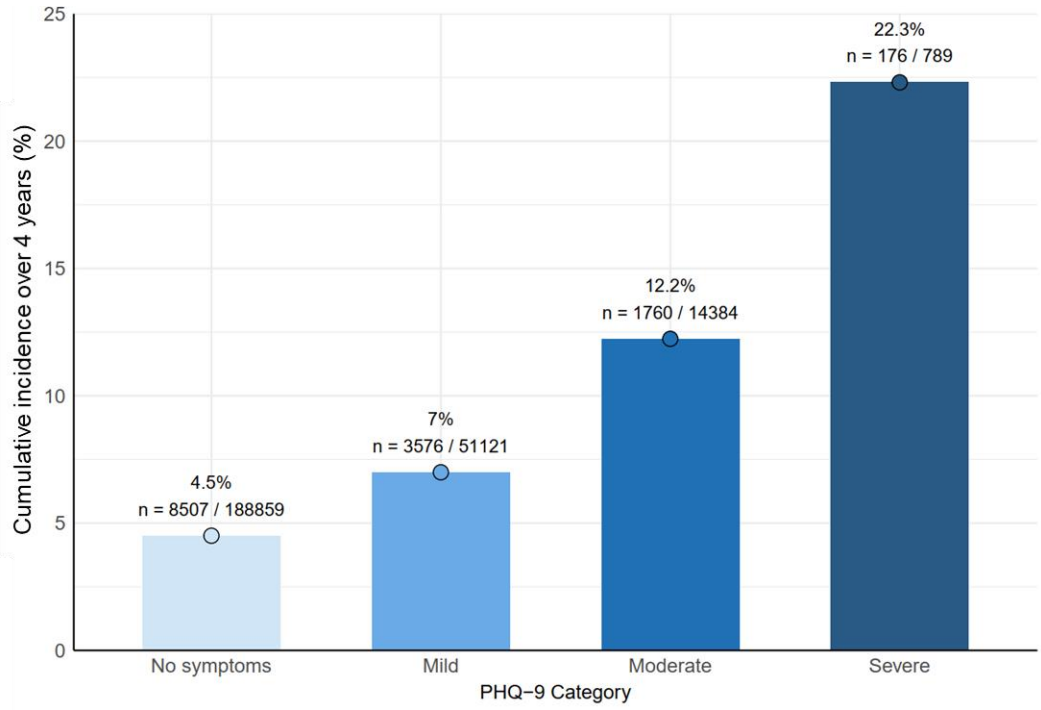


Figure 10. Incidence of mood disorders by PHQ-9 category among participants of depression screening in 2019 (*n* = 255,153)

The cumulative incidence of hospitalization for mood disorders over 4 years increased from 0.19% among individuals without symptoms to 1.27% among those with severe depressive symptoms. Similarly, the incidence of emergency department visits for mood disorders rose from 0.34% in individuals without symptoms to 1.52% in those with severe depressive symptoms. Additionally, the incidence of suicide and suicidal behaviors increased from 0.07% for individuals without symptoms to 0.38% for those with severe depressive symptoms (Figure 11).

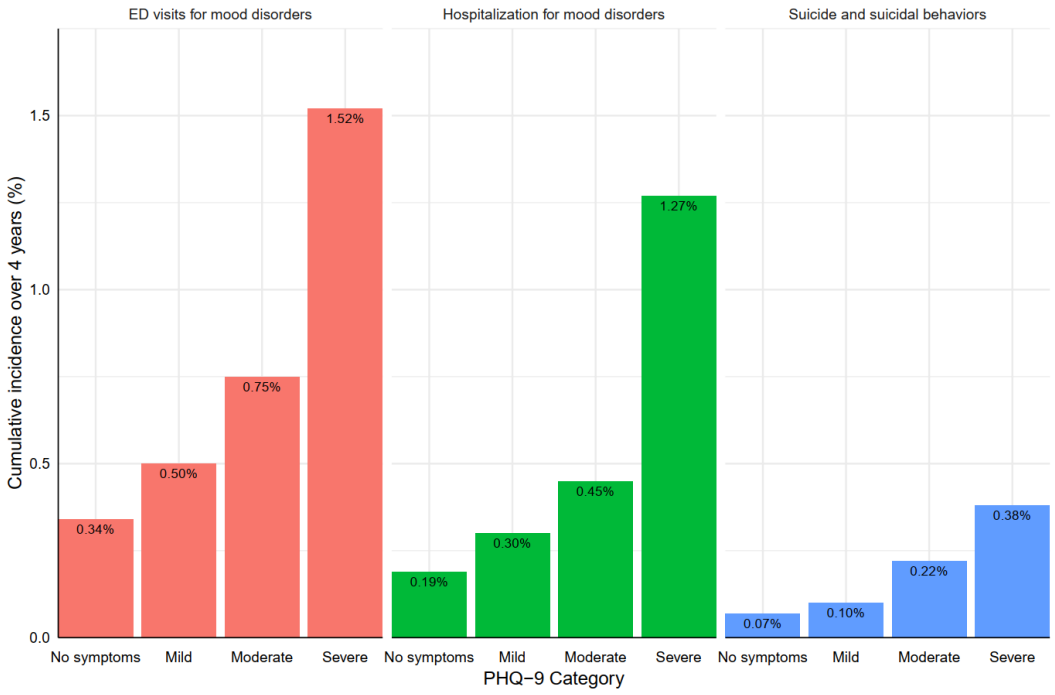


Figure 11. Incidence of emergency department visits, hospitalizations for mood disorders, and suicide or suicidal behaviors by PHQ-9 category among participants of depression screening in 2019 (*n* = 255,153)

Among the 255,153 participants of depression screening in 2019, 12,802 participants (5%) initiated antidepressant therapy during four years of follow-up. Based on PHQ-9 score categories, first-time antidepressant prescriptions were given to 12.9% of individuals with severe depressive symptoms, 8.8% with moderate symptoms, 5.7% with mild symptoms, and 4.5% with no depressive symptoms. When examining the data by type of antidepressant—SSRIs, SNRIs, and TCAs—comparable linear trends were evident. SSRIs were most commonly initiated, followed by TCAs and SNRIs, respectively (Figure 12).

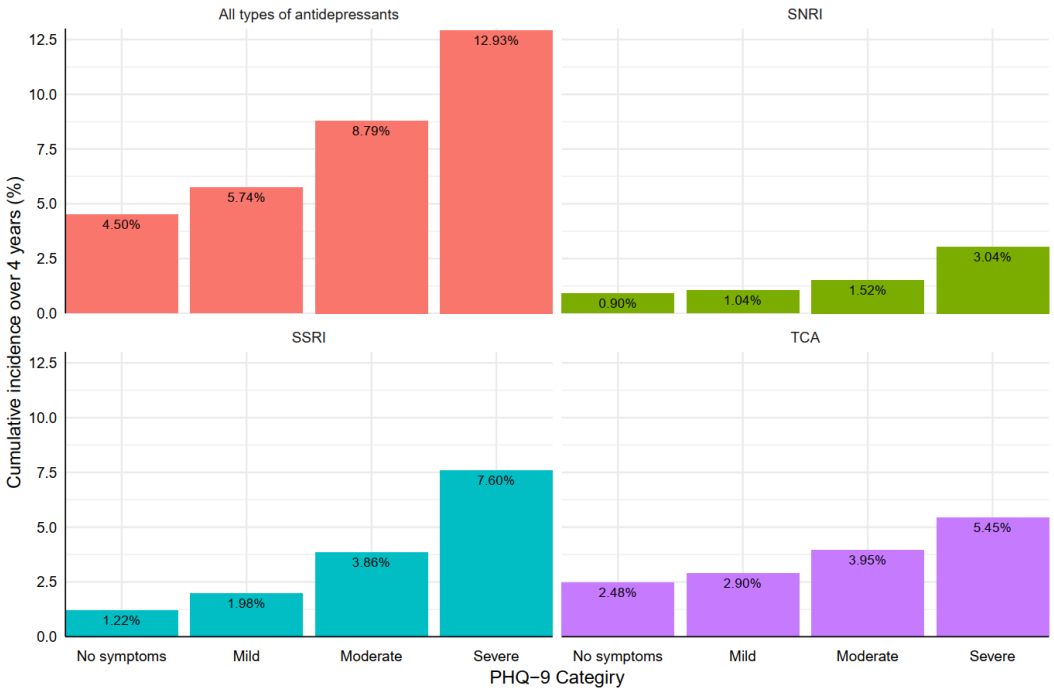


Figure 12. Incidence of first use of antidepressant by PHQ-9 category among participants of depression screening in 2019 (*n* = 255,153)

Abbreviation: SSRI, selective serotonin reuptake inhibitors; SNRI, serotonin-norepinephrine reuptake inhibitors; TCA, tricyclic antidepressants.

The UpSet plot examined antidepressant usage patterns in a cohort of 15,173 individuals who screened positive on the PHQ-9 scale (Figure 13). A significant proportion received SSRIs alone (4.1%), with TCA monotherapy close behind (4%). Combination therapies were relatively uncommon in PHQ-9 positive individuals without a prior history of psychiatric conditions.

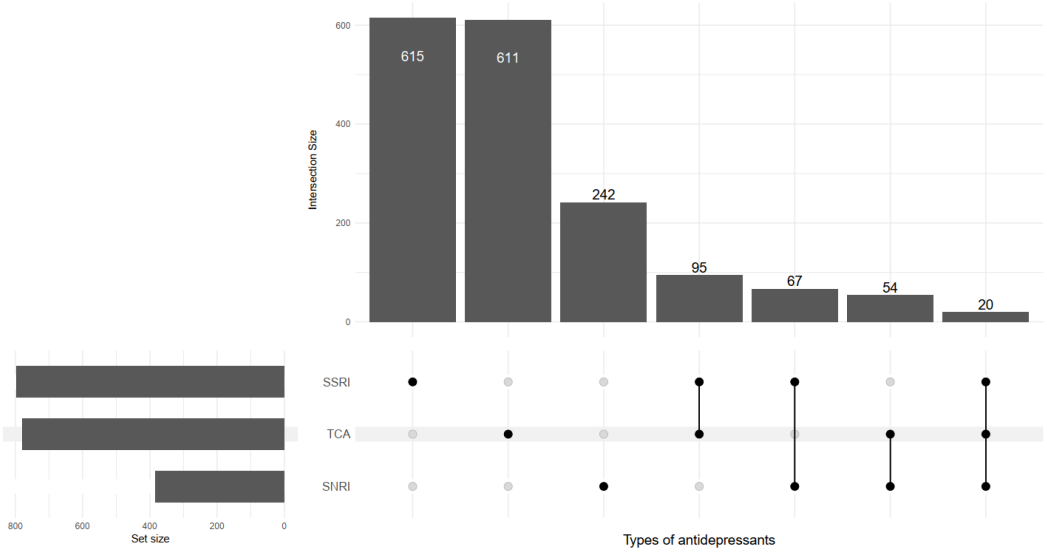


Figure 13. Distribution of antidepressant usage patterns among those who tested positive on the PHQ-9 (*n* = 15,173)

Notes. This UpSet plot illustrates the frequencies and overlap of antidepressant usage patterns among individuals who tested positive on the PHQ-9 scale. Connected dots below each bar indicate the specific antidepressant(s) included in each group. Abbreviation: SSRI, selective serotonin reuptake inhibitors; SNRI, serotonin-norepinephrine reuptake inhibitors; TCA, tricyclic antidepressants.

3.2.4. Baseline characteristics of study population

In the traditional ITT analysis, individuals eligible for depression screening in 2019 were those aged 20, 30, 40, 50, 60, or 70. Individuals in non-screening group were matched with individuals in screening group by age (±2 years), sex, and subscriber type. Screened individuals shared similar characteristics with non-screened individuals, except for differences in the number of health check-ups, income, and medical comorbidities (Table 9). In the modified ITT analysis, which excluded major protocol violations, screened individuals were slightly more likely to reside in large cities compared to metropolitan or rural areas. They also had a higher number of health check-ups in the preceding 5 years, slightly higher income levels, and a marginally greater burden of diagnosed medical comorbidities, as measured by the CCI. Finally, non-screened individuals experienced higher mortality than those who were screened (Table 10).

The Appendix 6 outlines the baseline characteristics of 568,180 participants in the as-treated approach. The data showed a slight gender imbalance, with men constituting 54% of the sample. A high proportion of participants were in their 40s and 50s, and many had employee insurance coverage. Those screened tended to live in large cities more often than in metropolitan or rural areas. Additionally, the screened group reported more health check-ups in the past 5 years and slightly higher income levels than the non-screened group. The CCI suggested a slightly higher burden of medical comorbidities among the screened individuals.

In Table 11, the characteristics of PHQ-9 positive and negative individuals from the 2019 depression screening are described. The positive group was defined as participants with PHQ-9 scores of 10 or more and was limited to a 3-point range (scores of 10, 11, or 12) to maintain comparability between groups. The negative group consisted of those with PHQ-9 scores of 7, 8, or 9. No significant differences were found between the groups in terms of death, sex, age, urbanicity, income, or medical comorbidities. Minor differences emerged in the number of previous health check-ups, and the positive group included a higher proportion of self-employed insured individuals.

Table 9. Characteristics of screened and non-screened groups in the traditional ITT analysis (*n* = 276,214, follow-up: 2019/1/1 – 2022/12/31).

		Screened		Non-screened		
		n = 138,107		*n* = 138,107		
Variables		***n***	**(%)**	***n***	**(%)**	***p*-value**
Death		1158	(0.84)	1219	(0.88)	0.21
Sex	Men	74208	(53.7)	74208	(53.7)	–
	Women	63899	(46.3)	63899	(46.3)	
Age group	20s	22585	(16.4)	22585	(16.4)	–
	30s	23808	(17.2)	23808	(17.2)	
	40s	28870	(20.9)	28870	(20.9)	
	50s	29770	(21.6)	29770	(21.6)	
	60s	23496	(17.0)	23496	(17.0)	
	70s	9578	(6.9)	9578	(6.9)	
Subscriber types	Self-employed insured	38115	(27.6)	38115	(27.6)	–
	Employee insured	97972	(70.9)	97972	(70.9)	
	Medical aid	2020	(1.5)	2020	(1.5)	
Urbanicity	Metropolitan	64126	(46.4)	64340	(46.6)	0.83
	Large city	34986	(25.3)	34952	(25.3)	
	Rural area	38972	(28.2)	38790	(28.1)	
	Unknown	23	(0.0)	25	(0.0)	
# of health check-up (5 yr)	0	59627	(43.2)	59476	(43.1)	<.0001
	1	22937	(16.6)	23253	(16.8)	
	2	27889	(20.2)	24276	(17.6)	
	3	10959	(7.9)	14819	(10.7)	
	4	7073	(5.1)	7125	(5.2)	
	5	9622	(7.0)	9158	(6.6)	
Income	Medicaid	2029	(1.5)	2029	(1.5)	0.04
	Q1	27724	(20.1)	27971	(20.3)	
	Q2	28424	(20.6)	28762	(20.8)	
	Q3	35495	(25.7)	35549	(25.7)	
	Q4	42945	(31.1)	42220	(30.6)	
	Unknown	1490	(1.1)	1576	(1.1)	
CCI score	0	68435	(49.6)	69507	(50.3)	0.0002
	1	45439	(32.9)	44657	(32.3)	
	2+	24233	(17.5)	23943	(17.3)	

Notes. Coarsened exact matching was used to match the screened and non-screened groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Abbreviation: ITT, intent-to-treat; CCI, Charlson comorbidities index.

Table 10. Characteristics of screened and non-screened groups in the modified ITT analysis (n = 53,688, follow-up: 2019/1/1 – 2022/12/31).

		Screened *n* = 26,844		Non-screened *n* = 26,844		
Variables		***n***	**(%)**	***n***	**(%)**	***p*-value**
Death		148	(0.55)	438	(1.63)	<.0001
Sex	Men	13,490	(50.3)	13,490	(50.3)	−
	Women	13,354	(49.7)	13,354	(49.7)	
Age group	20s	2,064	(7.7)	2,064	(7.7)	−
	30s	4,462	(16.6)	4,462	(16.6)	
	40s	6,977	(26.0)	6,977	(26.0)	
	50s	6,208	(23.1)	6,208	(23.1)	
	60s	4,757	(17.7)	4,757	(17.7)	
	70s	2,376	(8.9)	2,376	(8.9)	
Subscriber types	Self-employed insured	6,637	(24.7)	6,637	(24.7)	−
	Employee insured	19,847	(73.9)	19,847	(73.9)	
	Medical aid	360	(1.3)	360	(1.3)	
Urbanicity	Metropolitan	12,728	(47.4)	12,853	(47.9)	0.001
	Large city	6,955	(25.9)	6,647	(24.8)	
	Rural area	7,160	(26.7)	7,335	(27.3)	
	Unknown	1	(0.0)	9	(0.0)	
# of health check-up (5 yr)	0	6,708	(25.0)	11,064	(41.2)	<.0001
	1	5,038	(18.8)	5,435	(20.2)	
	2	8,126	(30.3)	5,046	(18.8)	
	3	2,830	(10.5)	2,312	(8.6)	
	4	1,785	(6.6)	1,335	(5.0)	
	5	2,357	(8.8)	1,652	(6.2)	
Income	Medicaid	362	(1.3)	362	(1.3)	<.0001
	Q1	4,974	(18.5)	5,960	(22.2)	
	Q2	5,513	(20.5)	5,807	(21.6)	
	Q3	7,514	(28.0)	6,686	(24.9)	
	Q4	8,227	(30.6)	7,727	(28.8)	
	Unknown	254	(0.9)	302	(1.1)	
CCI score	0	11,622	(43.3)	13,498	(50.3)	<.0001
	1	9,438	(35.2)	8,315	(31.0)	
	2+	5,784	(21.5)	5,031	(18.7)	

Notes. Coarsened exact matching was used to match the screened and non-screened groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Abbreviation: modified ITT, modified intent-to-treat; CCI, Charlson comorbidities index.

Table 11. Characteristics of PHQ-9 positive and negative individuals who participated in the depression screening in 2019 (*n* = 17,202, follow-up: 2019/1/1 – 2022/12/31)

		Positive[a] *n* = 8,601		Negative[b] *n* = 8,601		
Variables		***n***	**(%)**	***n***	**(%)**	***p*-value**
Death		45	(0.52)	39	(0.45)	0.51
Sex	Men	3,644	(42.4)	3,644	(42.4)	−
	Women	4,957	(57.6)	4,957	(57.6)	
Age group	20s	686	(8.0)	686	(8.0)	−
	30s	2,152	(25.0)	2,152	(25.0)	
	40s	2,560	(29.8)	2,560	(29.8)	
	50s	1,782	(20.7)	1,782	(20.7)	
	60s	1,077	(12.5)	1,077	(12.5)	
	70s	344	(4.0)	344	(4.0)	
Subscriber type	Self-employed insured	1,935	(22.5)	1,715	(19.9)	0.0003
	Employee insured	6,545	(76.1)	6,765	(78.7)	
	Medical aid	121	(1.4)	121	(1.4)	
Urbanicity	Metropolitan	4,433	(51.5)	4,430	(51.5)	0.78
	Large city	2,160	(25.1)	2,148	(25.0)	
	Rural area	2,008	(23.3)	2,022	(23.5)	
	Unknown	0	(0.0)	1	(0.0)	
# of health check-up (5 yr)	0	2,429	(28.2)	2,265	(26.3)	0.03
	1	1,761	(20.5)	1,746	(20.3)	
	2	2,239	(26.0)	2,244	(26.1)	
	3	898	(10.4)	965	(11.2)	
	4	555	(6.5)	589	(6.8)	
	5	719	(8.4)	792	(9.2)	
Income	Medicaid	121	(1.4)	121	(1.4)	0.08
	Q1	2,401	(27.9)	2,286	(26.6)	
	Q2	1,849	(21.5)	1,802	(21.0)	
	Q3	2,477	(28.8)	2,524	(29.3)	
	Q4	1,700	(19.8)	1,827	(21.2)	
	Unknown	53	(0.6)	41	(0.5)	
CCI score	0	5,569	(64.7)	5,613	(65.3)	0.42
	1	2,291	(26.6)	2,294	(26.7)	
	2+	741	(8.6)	694	(8.1)	

***Notes*.** Coarsened exact matching was used to match the positive and negative groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. [a] The positive group was defined as those scoring 10 or higher on the PHQ-9. To ensure comparability between groups, the dataset was restricted to a 3-point interval, so the positive group in this analysis consisted of participants with scores of 10, 11, or 12. [b] The negative group was defined as those with PHQ-9 socres of 7, 8, or 9. Abbreviation: PHQ-9, Patient Health Questionnaire-9; CCI, Charlson comorbidities index.

3.2.5. Effects of depression screening on the primary and secondary outcomes

The effects of depression screening on primary, secondary, and negative control outcomes were evaluated using traditional ITT and modified ITT approaches, as illustrated in Figure 14. The traditional ITT analysis showed no statistically significant results. In contrast, the modified ITT analysis indicated that hospitalization rates for mood disorders decreased from 0.84 per 1,000 person-years in the non-screened group to 0.52 in the screened group. The unadjusted cause-specific HR for mood disorder hospitalizations was 0.61 (95% CI: 0.44–0.86), and after adjusting for confounders, the HR was 0.56 (95% CI: 0.41–0.76; E-value: 2.97, 95% CI: 1.96). Screened participants had a higher rate of antidepressant initiation compared to non-screened participants (adjusted HR: 1.21, 95% CI: 1.10–1.33; E-value: 1.71, 95% CI: 1.43). Additionally, the hazards of emergency department visits were lower in the screened group than the non-screened group (adjusted HR: 0.71, 95% CI: 0.51–0.98; E-value: 2.17, 95% CI: 1.16). There was no significant association between depression screening and suicide or suicidal behaviors (adjusted HR: 0.62, 95% CI: 0.32–1.21). The negative control outcome, hospitalizations for cancers, showed no statistically significant associations.

According to the as-treated approach, the 2019 depression screening was linked to an increased rate of first-time antidepressant use and a reduced rate of suicide and suicidal behaviors (adjusted HR for antidepressant use: 1.22, 95% CI: 1.17–1.27; adjusted HR for suicide and suicidal behaviors: 0.65, 95% CI: 0.47–0.89) (Appendix 7).

Estimands	Non−screened			Screened			HR (95% CI)	aHR (95% CI)
	Case N	Person−year	Incidence rate (per 1,000 py)	Case N	Person−year	Incidence rate (per 1,000 py)		
Traditional ITT (n = 276,214)								
Hospitalization for mood disorders	399	549186.15	0.73	360	549320.75	0.66	0.90 (0.78, 1.04)	0.89 (0.77, 1.04)
Use of antidepressants	5680	539209.67	10.53	5781	538953.33	10.73	1.01 (0.98, 1.05)	1.01 (0.97, 1.05)
ED visit for mood disorders	544	548904.46	0.99	539	548980.3	0.98	0.99 (0.88, 1.12)	0.97 (0.85, 1.09)
Suicide and suicidal behaviors	145	549817.07	0.26	133	549901.38	0.24	0.91 (0.72, 1.15)	0.99 (0.77, 1.27)
Negative control: Hospitalization for cancer	2582	545380.15	4.73	2571	545403.5	4.71	0.99 (0.94, 1.05)	0.99 (0.94, 1.05)
Modified ITT (n = 53,688)								
Hospitalization for mood disorders	89	106366.08	0.84	56	106984.51	0.52	0.61 (0.44, 0.86)	0.56 (0.41, 0.76)
Use of antidepressants	1111	104411.72	10.64	1365	104500.29	13.06	1.22 (1.12, 1.32)	1.21 (1.10, 1.33)
ED visit for mood disorders	122	106298.51	1.15	95	106917.34	0.89	0.76 (0.62, 0.92)	0.71 (0.51, 0.98)
Suicide and suicidal behaviors	42	106493.3	0.42	22	107078.96	0.21	0.50 (0.30, 0.83)	0.62 (0.32, 1.21)
Negative control: Hospitalization for cancer	676	105344.12	6.42	575	105980.79	5.43	0.84 (0.75, 0.94)	0.88 (0.76, 1.01)

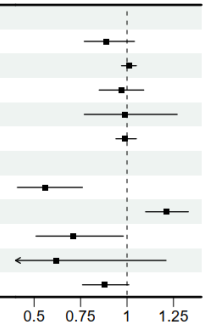


Figure 14. Effects of the 2019 depression screening on the primary, secondary, and negative control outcomes based on the traditional and modified ITT approaches

Notes. Coarsened exact matching was used to match the screened and non-screened groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Participant's urbanicity of residence, the number of previous health check-ups, income, and CCI were adjusted in the final models. Cancer hospitalizations were excluded in the CCI calculation when the diagnosis of cancer was used as a negative control outcome. Abbreviation: modified ITT, modified intent-to-treat; ED, emergency department.

The sensitivity analyses produced results consistent with the main models based on the modified ITT approach (Table 12). Similar findings were obtained using subdistributional hazard models, which aligned closely with those from the cause-specific hazard models. HRs for mood disorder hospitalizations became stronger when the lag time was extended to 180 days or 1 year. Similarly, more pronounced HRs were observed when continuous antidepressant use was defined as lasting 30 or 90 days. The effects of the 2019 depression screening remained unchanged regardless of the lag time for antidepressant use.

Among antidepressant classes, TCAs showed the strongest association, followed by SSRIs, while SNRIs were not statistically significant (Table 13). Age-stratified analysis revealed a significantly greater increase in the first-time use of SSRI among younger adults who participated in depression screening compared to the non-screened younger adults. Conversely, first-time TCA use increased significantly in both younger and older adults, with a stronger association observed in older adults. SNRI use did not differ by depression screening in either age group.

Table 12. Sensitivity analysis of the modified ITT approach based on the regression model and the definition of the outcomes (*n* = 53,688)

	Non-screened			Screened			
	Case *n*	Person-year	Incidence rate*	Case *n*	Person-year	Incidence rate*	aHR (95% CI)
Use of subdistributional hazard models							
Hospitalization for mood disorders	89	106366.08	0.84	56	106984.51	0.52	0.56 (0.41, 0.76)
Use of antidepressants	1111	104411.72	10.64	1365	104500.29	13.06	1.19 (1.10, 1.29)
Differential lag times of the first hospitalization for mood disorders							
Lag time: 180-day	88	106371.15	0.83	53	106995.87	0.50	0.53 (0.34, 0.82)
Lag time: 365-day	78	106399.13	0.73	43	107026.84	0.40	0.42 (0.25, 0.69)
Differential durations of the first use of antidepressants							
Duration: 30-day	492	105610.44	4.66	605	105992.22	5.71	1.24 (1.08, 1.44)
Duration: 90-day	265	106014.24	2.50	340	106445.67	3.19	1.29 (1.07, 1.57)
Differential lag times of the first use of antidepressants							
Lag time: 30-day	1103	104455.85	10.56	1350	104563.06	12.91	1.19 (1.09, 1.29)
Lag time: 90-day	1081	104550.94	10.34	1323	104682.84	12.64	1.19 (1.09, 1.29)

Notes. Coarsened exact matching was used to match the screening and non-screening groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Participant's urbanicity of residence, the number of previous health check-ups, income, and CCI were adjusted in the final models. Hospitalizations for mood disorders were examined using lag times of 180 days and 1 year, while antidepressant use was analyzed with varying durations and lag times. Incidence rate were calculated per 1,000 person-years. Abbreviation: ITT, intent-to-treat; Py, person-year; aHR, adjusted hazards ratio.

Table 13. Effects of the 2019 depression screening on the first use of antidepressants by their types and the age groups based on the modified ITT approach (*n* = 53,688)

		Non-screened			Screened			
Age groups	**Types**	**Case *n***	**Person-year**	**Incidence rate**	**Case *n***	**Person-year**	**Incidence rate**	**aHR (95% CI)**
All adults (*n* = 53,688)	SSRI	376	105872.86	3.55	431	106381.39	4.05	1.20 (1.02, 1.42)
	SNRI	226	106062.54	2.13	266	106598.25	2.5	1.16 (0.93, 1.45)
	TCA	539	105472.54	5.11	715	105689.48	6.77	1.32 (1.15, 1.51)
Younger adults (*n* = 39,422)	SSRI	251	78128.11	3.21	309	78223.74	3.95	1.24 (1.04, 1.48)
	SNRI	117	78333.02	1.49	129	78511.29	1.64	1.13 (0.85, 1.48)
	TCA	330	77935.74	4.23	419	77947.9	5.38	1.24 (1.06, 1.45)
Older adults (*n* = 14,266)	SSRI	125	27744.75	4.51	122	28157.65	4.33	1.05 (0.76, 1.45)
	SNRI	109	27729.52	3.93	137	28086.97	4.88	1.06 (0.77, 1.44)
	TCA	209	27536.8	7.59	296	27741.58	10.67	1.30 (1.06, 1.60)

Notes. Younger adults refer to adults aged 20 to 50, while older adults refer to seniors aged 60 to 70. Coarsened exact matching was used to match the screening and non-screening groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Participant's urbanicity of residence, the number of previous health check-ups, income, and CCI were adjusted in the final models. Antidepressant classifications analyzed included SSRIs, SNRIs, and TCAs as key categories. Incidence rate were calculated per 1,000 person-years. Abbreviation: ITT, intent-to-treat; Py, person-year; aHR, adjusted hazards ratio; SSRI, selective serotonin reuptake inhibitor; SNRI, serotonin-norepinephrine reuptake inhibitor; TCA, tricyclic antidepressant.

3.2.6. Subgroup analysis of hospitalization for mood disorders and antidepressant use

Subgroup analyses were conducted based on sex, age group (over 60 years vs. younger), and income level (Q3 and Q4 vs. Medicaid aid, Q1, and Q2). The results of subgroup analysis of modified ITT approach were presented in Figure 15. In the modified ITT analysis, sex-specific analysis showed a reduction in the hazards of hospitalization for mood disorders only in women (adjusted HR 0.47, 95% CI: 0.23–0.96; E-value: 3.68, 95% CI: 1.25). Age stratification indicated that the decreased hazards of hospitalization was confined to older adults (adjusted HR 0.43, 95% CI: 0.22–0.86; E-value: 4.08, 95% CI: 1.6). However, no significant associations were found between income levels and the incidence of hospitalization for mood disorders. For antidepressant initiation, depression screening had similar effects across subgroups, except by income level. Increased antidepressant initiation was observed in both men and women, as well as in both younger and older adults. When stratified by income, higher antidepressant initiation was only observed in screened individuals with higher income compared to their non-screened counterparts (adjusted HR 1.35, 95% CI: 1.20–1.52; E-value: 2.04, 95% CI: 1.69).

Subgroup analysis using the as-treated approach yielded results that were consistent with those of the modified ITT approach. However, unlike the modified ITT analysis, the as-treated approach revealed a statistically significant association between depression screening and antidepressant use among individuals with higher income (adjusted HR 1.30, 95% CI: 1.18–1.42; E-value: 1.92, 95% CI: 1.64) (Appendix 8).

Subgroups	Non-screened			Screened			aHR (95% CI)
	Case N	Person-year	Incidence rate (per 1,000 py)	Case N	Person-year	Incidence rate (per 1,000 py)	
Hospitalization for mood disorders							
Sex							
Men (n = 26,980)	45	53349.43	0.84	33	53707.43	0.61	0.57 (0.31, 1.06)
Women (n =26,708)	44	53016.65	0.83	23	53277.08	0.43	0.47 (0.23, 0.96)
Age							
Younger adults (n = 39,422)	40	78491.47	0.51	23	78698.78	0.29	0.53 (0.26, 1.1)
Older adults (n = 14,266)	49	27874.61	1.76	33	28285.73	1.17	0.43 (0.22, 0.86)
Income							
Lower income (n = 22,978)	49	47975.36	1.02	23	43197.78	0.53	0.71 (0.21, 2.37)
Higher income (n = 30,154)	39	57195.94	0.68	31	62777.49	0.49	1.86 (0.66, 5.28)
Use of antidepressants							
Sex							
Men (n = 26,980)	438	52592.44	8.33	561	52705.39	10.64	1.24 (1.11, 1.39)
Women (n =26,708)	673	51819.28	12.99	804	51794.9	15.52	1.19 (1.09, 1.3)
Age							
Younger adults (n = 39,422)	674	77290.4	8.72	814	77214.85	10.54	1.2 (1.09, 1.31)
Older adults (n = 14,266)	437	27121.32	16.11	551	27285.45	20.19	1.22 (1.09, 1.39)
Income							
Lower income (n = 22,978)	521	47061.69	11.07	536	42251.98	12.69	0.92 (0.8, 1.06)
Higher income (n = 30,154)	573	56176.33	10.2	814	61264.54	13.29	1.35 (1.2, 1.52)

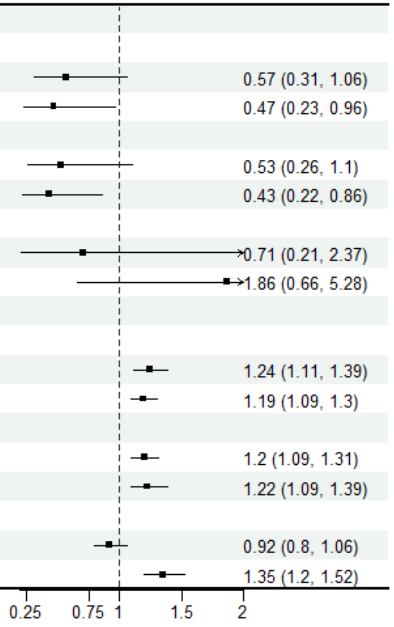


Figure 15. Subgroup analysis for hospitalization for mood disorders and use of antidepressants in the modified ITT analysis

Notes. Younger adults refer to adults aged 20 to 50, while older adults refer to seniors aged 60 to 70. Coarsened exact matching was used to match the screening and non-screening groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Participant's urbanicity of residence, the number of previous health check-ups, income, and CCI were adjusted in the final models. Abbreviation: aHR, adjusted hazards ratio; ITT; intent-to-treat; py, person-year; CI, confidence intervals

3.2.7. Exploratory analysis of the effect of post-screening intervention

To explore the effectiveness of a post-screening intervention for depression, the dataset was limited to the screened population, and depression-related outcomes were compared between the positive and negative groups. A 3-point bandwidth was applied to account for differences in depressive symptom severity. Consequently, 8,601 positive individuals (PHQ-9 scores of 10–12) and 8,601 negative individuals (PHQ-9 scores of 7–9) were included in the analysis.

No statistically significant differences were observed in the outcomes between the PHQ-9 positive and negative groups (Table 14). The hazards of hospitalization for mood disorders was 47% lower in the positive group, while antidepressant use was 1.12 times higher; however, neither finding reached statistical significance. ED visits for mood disorders were 19% higher in the positive group. Suicide-related outcomes could not be evaluated due to model convergence issues. Similarly, the composite outcomes, which included hospitalizations, ED visits, and suicide or suicidal behaviors, did not differ significantly based on PHQ-9 results. Cancer hospitalizations, used as a negative control, showed no significant association with the PHQ-9 results.

Table 14. Effects of the post-screening intervention on the depression-related outcomes between positive and negative screening groups (1:1 matched, *n* = 17,202)

	N	Person-year	Case *n*	Incidence rate (per 1,000 py)	Unadjusted HR (95% CI)	Adjusted HR (95% CI)
Hospitalization for mood disorders						
Negative	8,601	34279.67	29	0.85		
Positive	8,601	34286.34	26	0.76	0.90 (0.53, 1.52)	0.53 (0.21, 1.37)
Use of antidepressants						
Negative	8,601	33165.40	570	17.19		
Positive	8,601	33029.31	651	19.71	1.15 (1.03, 1.29)	1.12 (1.00, 1.26)
ED visit for mood disorders						
Negative	8,601	34247.53	43	1.26		
Positive	8,601	34231.31	57	1.67	1.33 (0.89, 1.97)	1.19 (0.70, 2.01)
Suicide and suicidal behaviors						
Negative	8,601	34315.16	13	0.38		
Positive	8,601	34311.36	11	0.32	0.85 (0.38, 1.90)	– [a]
Composite outcomes[b]						
Negative	8,601	34220.02	63	1.84		
Positive	8,601	34216.98	72	2.10	1.14 (0.81, 1.60)	1.04 (0.68, 1.59)
Negative control outcome: Hospitalization for cancer						
Negative	8,601	34019.08	165	4.85		
Positive	8,601	34001.47	180	5.29	1.11 (0.90, 1.38)	1.10 (0.88, 1.38)

Notes. The positive group was defined by PHQ-9 scores of 10–12, while scores of 7–9 defined the negative group. Coarsened exact matching was used to match the screening and non-screening groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Participant's urbanicity of residence, the number of previous health check-ups, income, and CCI were adjusted in the final models. Cancer hospitalizations were excluded in the CCI calculation when the diagnosis of cancer was used as a negative control outcome. [a] The results are not reported as the model failed to converge. [b] The composite outcomes included the hospitalization for mood disorders, ED visit for mood disorders, and suicide and suicidal behaviors. Abbreviation: PHQ-9, Patients Health Questionnaire-9; ED, emergency department; HR, hazards ratio; py, person-year; CI, confidence interval.

4. DISCUSSION

4.1. Summary of Study Findings

This dissertation sought to assess the effectiveness of depression screening in the general adult population through two methodological approaches: a literature review using systematic review and meta-analysis to evaluate current evidence from RCTs and observational studies, and a target trial emulation study using a population-based retrospective cohort from the NHIS database. This systematic review and meta-analysis of nine RCTs and nine observational studies found that depression screening yielded a pooled OR of 0.74 (95% CI 0.62–0.87) for the depresion-related outcomes such as depressive symptoms, psychiatric diagnoses, and suicide. Stronger effects were noted in women (OR 0.57, 95% CI 0.43–0.75) and Asian populations (OR 0.60, 95% CI 0.46–0.77). Despite these findings, most studies had a high risk of bias and considerable heterogeneity, limiting evidence for robust conclusions about the effectiveness of depression screening. Only seven of the 18 included studies demonstrated low, moderate, or some concerns regarding bias. None of the included studies fulfilled all six pre-specified criteria, meaning they did not target an unrestricted general population, exclude cases of existing depression before follow-up, determine eligibility prior to follow-up, account for factors beyond depression screening, utilize validated screening tools, or assess outcomes at the individual level.

The target trial emulation study using the NHIS database evaluated the impact of the nationwide depression screening program in South Korea. In 2019, the screening response rate among eligible individuals aged 20, 30, 40, 50, 60, or 70 years was 19.4%. Women, those aged 40–50, individuals with the employed subscriber type, prior health check-up participants, and those with more medical comorbidities exhibited higher participation rates in the depression screening. With a PHQ-9 score of 10 as the cut-off for positive depressive symptoms, approximately 5.9% of the screened participants were classified as positive. For mood disorder diagnoses within three months, the sensitivity was 19%, specificity was 94.1%, the positive predictive value was 0.85%, and the positive likelihood ratio was 3.22. The cumulative 4-year incidence of mood disorders was 4.5% for individuals with no symptoms and rose to 22.3% for those experiencing severe depressive symptoms in 2019.

Participation of the depression screening in 2019 was associated with the lower rates of the first hospitalization for mood disorders (modified ITT HR 0.56, 95% CI: 0.44–0.86). This reduction in the hospitalization for mood disorders was observed specifically in

women and in older adults aged 60 and 70. Furthermore, the screened group showed higher rates of initiating antidepressant use (modified ITT HR 1.21, 95% CI: 1.10–1.33) and fewer emergency department visits for mood disorders (modified ITT HR 0.71, 95% CI: 0.51–0.98) compared to the non-screened group. However, the depression screening had no significant effect on suicide or suicidal behaviors (modified ITT HR 0.62, 95% CI: 0.32–1.21), potentially due to insufficient statistical power (64 cases at total in the modified ITT approach). Robustness of the findings was confirmed using subdistributional hazards models in place of cause-specific hazards models, testing different lag periods for hospitalization and antidepressant initiation, and considering various lengths of sustained antidepressant use. Among the antidepressant types, TCAs had the strongest association with the participation of the depression screening, followed by SSRIs, whereas SNRIs showed no statistical significance. The exploratory analysis comparing PHQ-9 positive (scores 10–12) and negative groups (scores 7–9) did not yield statistically significant results in hospitalization and emergency department visit for mood disorders, and suicide or suicidal behaviors, with the exception of a marginally higher rate of antidepressant initiation in the positive group.

4.2. Discussion of Study Findings

Recognizing the potential harms and benefits based on scientific evidence is essential when it comes to decide implementing organized mass screening program at population level. [36] Without robust evidence on the effectiveness of screening for depression in healthy population, there could be potential harms that the target population being exposed to unnecessary treatment and additional testing. Here, 18 studies investigating the impact of depression screening were identified. Evidence supported the effectiveness of screening (pooled OR: 0.74, 95% CI 0.62–0.87); however, most studies had low quality and exhibited high heterogeneity. Thus, additional studies are required to evaluate whether depression screening effectively alters the trajectory of depression, alongside basic research on the natural history of the condition.

The inconsistency in meta-analysis findings on depression screening has been a point of criticism and has been associated with conflicting recommendations on primary care depression screening from author groups in the USA, UK, and Canada. [41] The USPSTF's systematic review in 2023 focused on RCTs that investigated the benefits of screening programs for depression or suicide risk. [38] Unlike this review, USPSTF’s review considered studies that did not have non-screened control groups or those where the control group

underwent depression screening, but the results were not communicated to the primary care physicians. As a result of meta-analysis, they included seventeen trials and found that the screening interventions were associated with a lower prevalence of depression or depressive symptoms at 6 months post-baseline (synthesized OR: 0.60 [95% CI: 0.50-0.73]). [38] In a systematic review conducted by CTFPHC in 2022[60], they included three RCTs that were also included in the current review. [94,98,102] Similarly, they identified that only one RCT with moderate certainty found no difference between the screened and non-screened groups[94], concluding that the evidence is insufficient to determine the effectiveness of depression screening. The UKNSC conducted a systematic review in 2020 to update their recommendations. [40] This review included two studies published between 2014 and 2020, which were also incorporated into the current review. [100,104] UKNSC also found a lack of evidence to make a conclusion about the effectiveness of screening for depression in primary care.

Among the criteria used in this systematic review, the criteria related to excluding pre-existing depression, determining eligibility before follow-up, and ensuring equal treatment access were proposed by Thombs and Ziegelstein (2014) as essential for evaluating the effectiveness of depression screening. [46] These criteria were also used by systematic reviews from the UKNSC and the CTFPHC. As a result, the conclusions of this review are consistent with those of the USPSTF review, which relied on a limited number of methodologically robust studies from diverse populations. However, a cautious interpretation is warranted, as drawing definitive conclusions remains challenging. It is important to emphasize that this review aimed to provide a broad overview of the related studies. Therefore, the point estimates derived from the meta-analysis should not be considered inherently valid but must be interpreted in light of the significant heterogeneity among the included studies. This discrepancy arises possibly because 76% of the studies included in the USPSTF review (n = 13) compared outcomes using screened controls rather than unscreened controls. [38] We excluded these studies from our analysis, as their results are more likely attributable to post-screening interventions rather than the effects of depression screening itself.

Goodyear-Smith and colleagues explored the ongoing conflicting results over the past 20 years among reviews that address the same research questions concerning the effectiveness of depression screening. [49] They analyzed two meta-analyses that reached contradictory conclusions despite having many overlapping studies. [105,106] They found that the selection of studies could be affected by confirmation bias, where authors may seek or interpret evidence in a way that aligns with their existing beliefs, expectations, or hypotheses. [49,107] All types of screening, not just for depression, involve various

stakeholders such as insurance companies, employers, and healthcare professionals, and some screenings are often recommended without sufficient evidence. [36,108] Given that systematic reviews and meta-analyses require many subjective decisions, future literature reviews should involve a diverse team of authors, including a third independent group from various health systems, to evaluate the evidence on the effectiveness of depression screening.

In the second study, the effectiveness of the nationwide depression screening implemented in 2019 for all adults aged 20 to 70 in South Korea was evaluated using the NHIS database. Individuals who undertook the screening for depression in 2019 had 11–44% lower incidence of hospitalization due to mood disorders (range of HRs by estimands: 0.56–0.89). The findings are consistent with a prior study conducted in Taiwan that used a target trial emulation framework with electronic health records, though the effect size reported in that study was relatively smaller (HR for psychiatric hospitalization: 0.93 [95% CI: 0.91, 0.95]). [103] This discrepancy may be explained by the fact that Chen and colleagues' study (2024) used hospitalization for any psychiatric disorder as the outcome measure, which is a valid approach due to the frequent comorbidity among psychiatric disorders. However, since psychiatric disorders includes a wide range of cognitive, behavioral, and emotional conditions, such as dementia and schizophrenia, screening for depression may have had a more significant preventive effect specifically on severe episodes that require hospitalization for mood disorders, including depression. The results of this study are also consistent with a randomized controlled trial conducted by Leung and colleagues, which demonstrated the preventive effect of a screening program for postpartum depressive symptoms in Hong Kong, China. [102] They found that the relative risk of being positive for depression at a 6-month follow-up, as measured by the EPDS assessment tool, was lower among participants who received depression screening followed by a post-screening intervention compared to those who did not (risk ratio 0.59, 95% CI: 0.39–0.89). The estimate is very similar with the HR observed in the modified ITT analysis of this study (adjusted HR 0.56, 95% CI: 0.41–0.76). However, it is also important to note that three other RCTs, which targeted patients with acute coronary syndromes, patients with osteoarthritis, and military platoons recently returned from deployment in Afghanistan, identified as having a low risk of bias, reported no efficacy of depression screening. [94,97,98]

The causal pathway from screening to a prevention of severe episodes of mood disorders is not clear. DSM-5 and ICD-11 define psychiatric disorders based on commonly co-occurring symptoms and not on etiology, psychobiology or prognosis. [35] Therefore, since defining the sojourn time of depression is challenging, it is difficult to determine the precise stage in the natural history of depression where the impact of screening would occur.

So far, the proposed mechanism for depression screening is that detecting previously undiagnosed depression may reduce the duration of untreated depression. [38] Previous studies have found that a longer duration of untreated depression is associated with increased depression-related disability and a worse prognosis, particularly during the first episode of depression. [22,23,109,110] However, although this study found improved depression-related outcomes in the screened individuals, supplementary analyses indicate that these improvements were likely due to non-specific effects of the depression screening rather than the intended mechanism. Notably, the 2019 depression screening in South Korea reported a significantly lower positive predictive value for identifying new mood disorders within three months than reported in psychometric studies of the PHQ-9 test itself[79], which raises serious concerns about a high rate of false negatives. Furthermore, the exploratory analysis showed no associations between PHQ-9 outcomes (positives versus negatives) and a reduction in depression-related outcomes, except for a marginally higher rate of antidepressant initiation in the positive group. Although this analysis is limited by a small sample size, it suggests that other mechanisms, not addressed in the introduction of this dissertation, may have contributed to the improvement in depression-related outcomes.

One possibility is that it is possible that risk group classification and early intervention, the central components of screening, might not be functioning as intended, and the benefit may instead stem from the act of providing depression screening itself. Non-specific effects, such as empathy and attentiveness from physicians, have been identified as significant factors influencing the effectiveness of antidepressants and psychotherapies. [111,112] Since depression and mental health are not yet broadly accepted as medical conditions in South Korea, the depression screening offered alongside psychoeducation within the healthcare system could have fostered future help-seeking behaviors. However, understanding the mechanisms underlying the effectiveness of depression screening requires fundamental research into the various phenotypes and the natural history of depression. Therefore, the conclusion regarding the effectiveness of depression screening should be approached with caution, as it requires further in-depth investigation from a long-term perspective.

This study also found that the effects of depression screening were only significant in women and older adults. These observed effects in women and older adults align with previous studies that were also deemed to have an acceptable level of risk of bias in the systematic review and meta-analysis conducted in the first part of this dissertation. [98,102,103] Depression screening is likely to be effective in identifying and addressing depressive episodes that might otherwise go unnoticed until patients actively seek treatment. Therefore, this effectiveness may be particularly pronounced in older populations, where untreated depression is more prevalent. [28,113] Additionally, the guideline for this screening program

recommends referral to psychiatric clinic if moderate or severe depression is detected. However, if these referrals do not lead to appropriate diagnosis and treatment, the screening's effectiveness in preventing severe depressive episodes and reducing psychiatric hospitalizations may be limited. Additionally, studies have shown that women tend to have higher adherence to antidepressant treatment compared to men. [114] Therefore, the difference in screening effectiveness between genders in this study may be partially explained by the difference in adherence to antidepressants.

This study found no evidence of depression screening effectiveness in younger adults aged 20–50, despite young adulthood being a critical period for the onset and burden of depression, making it a key target for prevention programs. [10,14,15] Figure 9 illustrates participation rates in screening across subgroups, revealing particularly low rates among individuals in their 20s (response rate: 12.3%). Similar trends of low participation of preventive health check-up among young adults were observed internationally, including in Germany and Austria. [115,116] In South Korea, the Occupational Safety and Health Act mandates biennial general health checkups for employees insured individuals, but response rates among young adults may be low due to the high proportion of unemployed and the relative novelty of these checkups for this age group. [117] To enhance the effectiveness of depression screening in young adults, efforts should focus on improving participation in general health checkups that include depression screening.

4.3. Limitations and Strengths

The systematic review and meta-analysis highlighted substantial heterogeneity and publication bias, as the research aimed to comprehensively review prior studies on depression screening. Although the pooled OR was estimated at 0.74 (95% CI: 0.62–0.87), the high I^2 value of 72.4% ($p < 0.001$) and a wide 95% PI (0.40–1.34) indicated significant variability across studies. Notably, the study populations differed greatly, and none of the included studies evaluated the effectiveness of depression screening in the general adult population, particularly individuals aged 20–30. Moreover, the definition of a screening program varied widely among included studies. Some studies included only the administration of self-report assessments and referrals[94,98], while others incorporated extensive post-screening interventions. [100,102,104] As a result, the variability in the content of screening programs, target populations, and study quality complicates the interpretation and applicability of the pooled effect size calculated in this meta-analysis. [118] Also, studies not available in the English and Korean language, abstracts and poster presentations were

not included. Despite these limitations, the current review had several strengths. The discrepant findings between existing reviews and meta-analysis studies was addressed by re-evaluating the existing evidence as an independent group of author. Additionally, the current review included both RCTs and observational studies. Observational studies, in contrast to RCTs, provide greater insight into the practical effectiveness of interventions within real-world public health settings. [71] By using information from observational studies, this review revealed that observational studies reported a more pronounced preventive effect of depression screening compared to RCTs, thereby validating the interpretation of results of subsequent observational study using NHIS database.

Several limitations of the target trial emulation study should be considered. First, despite efforts to minimize biases using the target trial emulation framework, this study is inherently non-randomized observational. Unmeasured confounders, such as education level, marital status, and health behaviors, may have influenced the results. While adjusting for confounders based on a pre-specified directed acyclic graph and using a negative control outcome and E-values helped demonstrate the robustness of the findings, the potential impact of confounding cannot be entirely excluded, and caution is advised when interpreting the study results. Second, the potential harms associated with the nationwide depression screening were not addressed in this study. These harms could stem from false-positive results leading to unnecessary referrals, labeling, and stigma or unnecessary diagnostic investigation. [76] Although the direct cost of the depression screening is relatively low (approximately $3 USD), the opportunity costs due to overdiagnosis and excessive medical expenses could become significant as the number of false positives accumulates over time at the population level. Further studies should assess the balance between the potential benefits and harms associated with screening programs. As a result, this study cannot definitively recommend for or against nationwide depression screening for adults in the general population. Third, there is no direct measure of whether the physicians conducting the depression screening adhered to the guidelines. Simply providing feedback on test results has shown suboptimal outcomes compared to programs that include effective follow-up and treatment in line with the published guidelines for depression screening. [119] Consequently, it is unclear whether the observed effectiveness of the depression screening can be attributed precisely to the screening program as implemented according to the guidelines, or if other unmeasured factors in the NHIS database influenced the results. Furthermore, the preventive effect of depression screening may differ depending on the physicians who conducted the health check-up with depression screening and the psychiatrists involved when a referral to a psychiatric clinic occurred, potentially violating the consistency assumption of causal inference. Fourth, while both pharmacological and

psychological interventions for mood disorders have proven effective, this study only captured the pharmacological treatments reimbursed by NHIS. Therefore, non-pharmacological treatment including psychotherapy and counseling could not be assessed in the NHIS database.

However, the use of the target trial emulation framework and a causal diagram helped address threats to the internal validity of observational studies, such as immortal time bias, selection bias, and confounding factors, thereby strengthening the causal inference of the study results. [72,120] Specifically, aligning the time zero of follow-up with the specification of eligibility criteria and treatment assignment in the modified ITT analyses prevented immortal time bias. [121,122] By employing a pre-specified directed acyclic graph, it was able to explicitly account for unmeasured confounding effects when interpreting the results. Additionally, the directed acyclic graph clarified the presence of an exogenous variable used for intervention assignment, namely the specific age in 2019. Secondly, the study population included young adults in their 20s and 30s. Assessing the effectiveness of depression screening in this age group is crucial, as the initial onset and burden of depression are highest during this period. While the effectiveness of depression screening in individuals under 60 was found to be negligible, it is important to investigate the reasons for this low effectiveness in this group in the further studies. This exploration is particularly crucial, as earlier onset of major depression is associated with poorer prognoses, including lower quality of life, higher rates of comorbidity (both medical and psychiatric), a greater number of depressive episodes, increased symptom severity, and elevated suicide-related behaviors. [15]

5. CONCLUSION

The systematic review and meta-analysis found significant but weak evidence that depression screening for the general adult population in primary care or at the national level is effective (pooled OR 0.74, 95% CI 0.62–0.87). However, the included studies were mostly of poor methodological quality, characterized by significant bias and considerable heterogeneity across studies. Findings from the target trial emulation study based on NHIS data revealed that the 2019 nationwide depression screening in South Korea was associated with a lower rate of hospitalizations for mood disorders (aHR for modified ITT 0.56, 95% CI 0.41–0.76), notably among women and older adults. It was also linked to increased first-time antidepressant use and decreased emergency department visits for mood disorders, with no statistically significant association with suicide or suicidal behaviors. Although South Korea conducts nationwide depression screening annually, it is still unclear whether this approach over the long term will result in unnecessary medical costs or effectively prevent severe depression, thereby enhancing quality of life and reducing suicide rates. Furthermore, differences observed between screened and non-screened individuals in this study may be attributed to non-specific effects of the screening program, underscoring the need for continuous evaluation to inform evidence-based decisions. Additionally, there is a need to enhance young adults' participation in health check-ups and to address the psychological barriers that discourage them from seeking psychiatric and psychological help.

APPENDICES

Appendix 1. Items of Patient Health Questionnaire-9

Over the last 2 weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things				
2. Feeling down, depressed or hopeless				
3. Trouble falling asleep, staying asleep, or sleeping too much				
4. Feeling tired or having little energy				
5. Poor appetite or overeating				
6. Feeling bad about yourself - or that you're a failure or have let yourself or your family down				
7. Trouble concentrating on things, such as reading the newspaper or watching television				
8. Moving or speaking so slowly that other people could have noticed. Or, the opposite - being so fidgety or restless that you have been moving around a lot more than usual				
9. Thoughts that you would be better off dead or of hurting yourself in some way				

Note. Cited from Kroenke et al. [77]

Appendix 2. List of antidepressants used in this study

Classification	Antidepressant	ATC code
TCA	Imipramine	1737
TCA	Amitriptyline	1075
TCA	Amoxapine	1080
TCA	Clomipramine	1363
TCA	Doxepin	1492, 4519
SSRI	Citalopram	4283
SSRI	Escitalopram	4748, 5211
SSRI	Fluoxetine	1615
SSRI	Fluvoxamine	1625
SSRI	Paroxetine	2093
SSRI	Sertraline	2270
SNRI	Desvenlafaxine	6264, 6876, 6877
SNRI	Duloxetine	4955
SNRI	Milnacipran	3558
SNRI	Venlafaxine	2475
Others	Bupropion	4281
Others	Mirtazapine	1962
Others	Trazodone	2429

Abbreviation: ATC, anatomical therapeutic chemical classification system; TCA, tricyclic antidepressants; SNRI, serotonin–norepinephrine reuptake inhibitors; SSRI, selective serotonin reuptake inhibitors.

Appendix 3. Diagnostic categories and ICD-10 codes for calculating the Charlson Comorbidity Index

Condition	Weights	ICD-10 codes
Acute myocardial infarction	1	I21, I22, I252
Congestive heart failure	1	I50
Peripheral vascular disease	1	I71, I790, I739, R02, Z958, Z959
Cerebral vascular accident	1	I60, I61, I62, I63, I65, I66,G450, G451, G452, G458, G459, G46, I64, G454, I670, I671, I672, I674, I675, I676, I677 I678, I679, I681, I682, I688, I69
Dementia	1	F00, F01, F02, F051
Pulmonary disease	1	J40, J41, J42, J44, J43, J45, J46, J47, J67, J44, J60, J61, J62, J63, J66, J64, J65
Connective tissue disorder	1	M32, M34, M332, M053, M058, M059, M060, M063, M069, M050, M052, M051, M353
Peptic ulcer	1	K25, K26, K27, K28
Liver disease	1	K702, K703, K73, K717, K740, K742, K746, K743, K744, K745
Diabetes	1	E109, E119, E139, E149, E101, E111, E131, E141, E105, E115, E135, E145
Diabetes complications	2	E102, E112, E132, E142 E103, E113, E133, E143 E104, E114, E134, E144
Paraplegia	2	G81 G041, G820, G821, G822
Renal disease	2	N03, N052, N053, N054, N055, N056, N072, N073, N074, N01, N18, N19, N25
Cancer	2	C0, C1, C2, C3, C40, C41, C43, C45, C46, C47, C48, C49, C5, C6, C70, C71, C72, C73, C74, C75, C76, C80, C81, C82, C83, C84, C85, C883, C887, C889, C900, C901, C91, C92, C93, C940, C941, C942, C943, C9451, C947, C95, C96
Metastatic cancer	3	C77, C78, C79, C80
Severe liver disease	3	K729, K766, K767, K721
HIV	6	B20, B21, B22, B23, B24

Note. Cited from Sundararajan et al. (2004) [83]

Appendix 4. Summary of included studies in the literature review

References (most recent)	Description
Chen Y.-L., 2024	**Title:** Effectiveness of Health Checkup with Depression Screening on Depression Treatment and Outcomes in Middle-Aged and Older Adults: A Target Trial Emulation Study **Country:** Taiwan **Study design:** Non-randomized **Population:** Adults aged 40 years and above **Data source:** Taiwan's National Health Insurance claims database **Eligibility:** ·Not yet received health checkups with depression screening ·Enrolled in the national health insurance program between 2013 and 2019. ·Excluded if they have not data on birth year, sex, monthly income, or residential area. ·Excluded if they were diagnosed with depressive disorders or bipolar disorder before assignment **Analyzed sample size:** ·Intervention: 4,792,228 ·PS matched control: 4,792,228 **Intervention:** ·Intervention: nationwide health check-up with depression screening ·Control: TAU **Follow-up / Study period:** 100 months (Aug 2011 to Dec 2019) **Outcomes and main findings (95% CI):** Incidence of depression HR 1.63 (1.62, 1.64), psychiatric hospitalization HR 0.93 (0.91, 0.95) **Adjusted covariates:** Residential area, monthly income, Charlson comorbidity index, # of outpatient visits, and common medical and psychiatric comorbidities.

References (most recent)	Description
Kronish I.M., 2020	**Title:** Effect of Depression Screening After Acute Coronary Syndromes on Quality of Life: The CODIACS-QoL Randomized Clinical Trial **Country:** USA **Study design:** Randomized **Population:** Patients with acute coronary syndromes without a history of depression **Data source:** Recruited from Minnesota, North Carolina, Oregon, and New York. **Eligibility:** ·Declined participation ·Ineligible for medical reason ·Ineligible for psychiatric reason ·Ineligible per provider review ·Non-English or Spanish speaking **Analyzed sample size:** ·Intervention 1: 498 ·Intervention 2: 501 ·Control: 498 **Intervention:** ·Intervention 1: Screening and notification ·Intervention 2: Screening, notification, and treatment ·Control: TAU **Follow-up / Study period:** 18 months **Outcomes and main findings (*p*-value):** Mean difference of depression-free days (p =.63) ·Intervention 1: 341.1 days ·Intervention 2: 351.3 days ·Control: 339.0 days **Adjusted covariates:** N/A

References (most recent)	Description
Rhee T.G., 2018	**Title:** Effects of the 2009 USPSTF Depression Recommendation on Diagnosing and Treating Mental Health Conditions in Older Adults: A Difference-in-Differences Analysis **Country:** USA **Study design:** Non-randomized **Population:** Adults aged 65 or older who visited office-based outpatient primary care. **Data source:** National Ambulatory Medical Care Survey (NAMCS) 2006-2012. **Eligibility:** ·Age 65 or older ·Complete data for all covariates **Analyzed sample size:** ·Pre-recommendation: 6,283 ·Post-recommendation: 9,313 **Intervention:** ·Intervention: Answers "yes" to the question, "Was the depression screening exam ordered or provided at the visit?" ·Control: Answers "no" to the question, "Was the depression screening exam ordered or provided at the visit?" **Follow-up / Study period:** ·Pre-recommendation: 2006-2009 ·Post-recommendation: 2010-2012 **Outcomes and main findings (95% CI):** Prevalence difference of any mental health diagnosis -14.4% (-28.2, -0.6) **Adjusted covariates:** Age, gender, race/ethnicity, region, primary source of payment, reason for visit, and repeat of visits within the past 12 months.
Silverstone., 2017	**Title:** Depression Outcomes in Adults Attending Family Practice Were Not Improved by Screening, Stepped-Care, or Online CBT during a 12-Week Study when Compared to Controls in a Randomized Trial **Country:** Canada **Study design:** Randomized **Population:** Patients in primary care. **Data source:** Recruited from the two clinics containing 18 primary care physicians from Nov 2013 to Dec 2014. **Eligibility:** N/A **Analyzed sample size:** ·Intervention 1: 286 ·Intervention 2: 255 ·Intervention 3: 73 ·Control: 275 **Intervention:** ·Intervention 1: Screening with notification of results to their physicians ·Intervention 2: Screening with result notification and CBT-based self-help program ·Intervention 3: Screening with stepped-care pathway ·Control: Screening without disclosing the results (unless with indication of suicide risk) **Follow-up / Study period:** 12 weeks **Outcomes and main findings (95% CI):** Positive for PHQ-9 RR 0.96 (0.67, 1.34) * **Adjusted covariates:** N/A

References (most recent)	Description
Rona., 2017	**Title:** Post-Deployment Screening for Mental Disorders and Tailored Advice about Help-Seeking in the UK Military: A Cluster Randomized Controlled Trial **Country:** UK **Study design:** Randomized (unit: individual) **Population:** Platoon members who recently returned from deployment in Afghanistan. **Data source:** Recruited from Royal Marines and Army personnel in the UK military between Oct 24, 2011 and Feb 15, 2013. **Eligibility:** ·Excluded if they did not deploy ·Excluded if they deployed but moved to another location before randomization ·Excluded if they were reserved personnel ·Excluded if they were formed specifically for deployment and had dispersed upon return home **Analyzed sample size:** ·Intervention: 3,996 ·Control:2,369 **Intervention:** ·Intervention: Screening + tailored help-seeking advice ·Control: General mental health advice **Follow-up / Study period:** 10-24 months **Outcomes and main findings (95% CI):** Depression and anxiety OR 0.91 (0.91, 1.16) **Adjusted covariates:** Rank, age, and date of deployment, which were associated with probability of missingness or nonresponse

References (most recent)	Description
Mallen C.D., 2017	**Title:** The Effects of Implementing a Point-of-Care Electronic Template to Prompt Routine Anxiety and Depression Screening in Patients Consulting for Osteoarthritis (the Primary Care Osteoarthritis Trial): A Cluster Randomised Trial in Primary Care **Country:** UK **Study design:** Randomized (unit: practice) **Population:** Older patients (>= 34y) consulting for osteoarthritis. **Data source:** Recruited from 45 English general practices. **Eligibility:** ·Excluded if they don't have osteoarthritis ·Excluded if they have current mental health problem ·Excluded if they are vulnerable patient (dementia or terminal illness) ·Excluded if they are resident of nursing home ·Excluded if they have medical comorbidities **Analyzed sample size:** ·Intervention: 501 ·Control: 911 **Intervention:** ·Intervention: Point-of-care by GP prompted by the electronic template containing information on screening for depression/anxiety/pain and treatment ·Control: Point-of-care by GP prompted by the electronic template containing information on screening for pain intensity **Follow-up / Study period:** 12 months **Outcomes and main findings (95% CI):** Mean difference of PHQ-8 score at 12 months -0.36 (-0.41, 1.14) **Adjusted covariates:** Age, sex, and time between consultation date and mailing response date.

References (most recent)	Description
Oyama H., 2017	**Title:** Community-Based Screening Intervention for Depression Affects Suicide Rates among Middle-Aged Japanese Adults **Country:** Japan **Study design:** Non-randomized **Population:** All middle-aged adults aged 36-64 in 2009-2012 in districts (Aomori Prefecture) with a history of high suicide rates. **Data source:** National registry data. **Eligibility (district level):** ·Population large enough to deliver mental health services and interventions. ·Had a suicide rate higher than the prefectural average. ·Excluded when it had recently depression-related intervention ·Assignment was determined based on preference of each district. ·Criteria for control districts: Comparable in terms of average population size, average annual gross income, immigration rate, emigration rate, unemployment rate, and suicide rates. **Analyzed sample size:** ·Intervention: 5 districts (average *n* of individuals: 8,010) ·Control: 6 districts (average *n* of individuals: 8,087) **Intervention:** ·Intervention: Two-stage screening and referral to psychiatrist ·Control: TAU **Follow-up / Study period:** ·Pre-implementation: 2005-2008 ·Implementation: 2009-2012 **Outcomes and main findings (95% CI):** Suicide IRR 0.57 (0.41, 0.78) **Adjusted covariates:** Age and sex

References (most recent)	Description
Oyama H., 2016	**Title:** Long-Term Effects of a Screening Intervention for Depression on Suicide Rates among Japanese Community-Dwelling Older Adults **Country:** Japan **Study design:** Non-randomized **Population:** All adults over 60 years olds in districts (Aomori Prefecture) with a history of high suicide rates. **Data source:** National registry data. **Eligibility (district level):** ·Criteria for control district: Comparable in terms of population size, unemployment rate, average annual gross income, immigration rates, emigration rates, accessibility to health services, socioeconomic status, and suicide rates. **Analyzed sample size:** ·Intervention: 3 districts (*n* of individuals: 11,710) ·Comparison: 3 districts (*n* of individuals: 12,602) **Intervention:** ·Intervention: Screening and health education ·Control: TAU **Follow-up / Study period:** Pre-implementation (1999 – 2004), post-implementation (2005 – 2010, implementation: 2005 – 2006) **Outcomes and main findings (95% CI):** Suicide IRR: 0.55 (0.32, 0.93) **Adjusted covariates:** Age and sex
Oyama H., 2014	**Title:** Effects of Universal Screening for Depression Among Middle-Aged Adults in a Community With a High Suicide Rate **Country:** Japan **Study design:** Randomized (unit: district) **Population:** All residents aged 42 to 65 years who were provided with the opportunity to participate in the intervention program during 2005 to 2008 in districts (Aomori Prefecture) with a history of high suicide rates. **Data source:** Repeated cross-sectional surveys of residents in the study districts. **Eligibility:** ·Excluded if they had missing data on marital status, recent contact with a general practitioner, and employment status. **Analyzed sample size:** ·Intervention: (1st survey) 543, (2nd survey) 586 in 4 districts ·Control: (1st survey) 973, (2nd survey) 1010 in 6 districts **Intervention:** ·Intervention: Screening and education about depression ·Control: TAU **Follow-up / Study period:** Baseline (2004) and follow-up (2009) **Outcomes and main findings (95% CI):** Prevalence ratio of moderate-to-severe depressive symptoms (CES-D>=24) 0.58 (0.38, 0.88) **Adjusted covariates:** Age and sex

References (most recent)	Description
Romera I., 2013	**Title:** Systematic Depression Screening in High-Risk Patients Attending Primary Care: A Pragmatic Cluster-Randomized Trial **Country:** Spain **Study design:** Randomized (unit: practice) **Population:** Primary care physicians were recruited from public healthcare system. Patients who visit the primary care practice who participated in this study. **Data source:** Patient's medical records reviewed by the participating physicians. **Eligibility:** ·Excluded if they are already following the recommendations or planning to do so ·Excluded if they were absent from their practice for long time during the study period ·Excluded if they could not ensure effective management of DEP **Analyzed sample size:** ·Intervention: 257 patients of 30 physicians ·Control: 268 patients of 32 physicians **Intervention:** ·Intervention: Physicians received training in screening and diagnostic interview. ·Control: TAU **Follow-up / Study period:** ·Randomization: Sep 2009 ·Intervention: Oct 2009 - Mar 2010 ·Outcome assessment: Apr - Jul 2010 **Outcomes and main findings (95% CI):** Underrecognition of depression OR 1.40 (0.73, 2.68) **Adjusted covariates:** Number of patients attending the practice daily, physician’s shift, gender, age, education, work status, medical comorbidities, non-psychiatric treatment, and time since previous visit.

References (most recent)	Description
Leung S.S.L., 2011[102]	**Title:** Outcome of Postnatal Depression Screening Programme Using the Edinburgh Postnatal Depression Scale: A Randomized Controlled Trial **Country:** China **Study design:** Randomized (unit: individual) **Population:** Mothers of 2-month-old babies who visited the centre for routine child health services. **Data source:** Recruited from Maternal and Child Health Centre of four selected communities in Hong Kong (2005-2006). **Eligibility:** ·Excluded non-local residents ·Excluded if they don't speak Chinese ·Excluded if they participated in other DS program ·Excluded if they received psychiatric treatment. **Analyzed sample size:** ·Intervention: 231 ·Control: 231 **Intervention:** ·Intervention: Screening and counseling/management/referral by nurse ·Control: TAU **Follow-up / Study period:** 6 months **Outcomes and main findings (95% CI):** Positive EPDS RR 0.59 (0.39, 0.89) **Adjusted covariates:** Marital relationship at 2 months, history of psychiatric illness, depression during pregnancy and relationship with mother-in-law.
Oyama H., 2010	**Title:** A Community-Based Survey and Screening for Depression in the Elderly: The Short-Term Effect on Suicide Risk in Japan **Country:** Japan **Study design:** Non-randomized **Population:** Adults over 60 years olds in Sannohe, Takko, and Nanbu in Aomori Prefecture. **Data source:** National registry data. **Eligibility:** N/A **Analyzed sample size:** ·Intervention: 14,504 in 3 districts ·Control: 13,931 in 3 districts **Intervention:** ·Intervention: Two-stage screening and education about depression ·Control: TAU **Follow-up / Study period:** Baseline (2003-2004) and intervention period (2005-2006) **Outcomes and main findings (95% CI):** Suicide IRR in men 0.39 (0.18, 0.87), suicide IRR in women: 0.49 (0.19, 1.22) **Adjusted covariates:** Age

References (most recent)	Description
Oyama H.., 2006 (1)	**Title:** Preventing Elderly Suicide Through Primary Care by Community-Based Screening for Depression in Rural Japan **Country:** Japan **Study design:** Non-randomized **Population:** Residents aged 65 and older in Matsudai **Data source:** Registry data from public health centers. **Eligibility for control district (district level):** ·Population 4000 - 10000 ·Average percentage of elderly aged 65 and over is greater than 15% ·Elderly suicide rate over 150 in both women and men **Analyzed sample size:** ·Intervention: 1 district ·Control: 1 district **Intervention:** ·Intervention: Public health education/two step screening with referral by nurse ·Control: TAU **Follow-up / Study period:** Baseline (1978-1988) and intervention period (1988-1998) **Outcomes and main findings (95% CI):** Age-adjusted IRR for suicide Men ·Intervention: 1.02 (0.49, 2.13) ·Control: 0.69 (0.32, 1.50), p-value for Wald test 0.47* Women ·Intervention: 0.30 (0.14, 0.67) ·Control: 0.68 (0.32, 1.45), p-value for Wald test 0.07* **Adjusted covariates:** Age

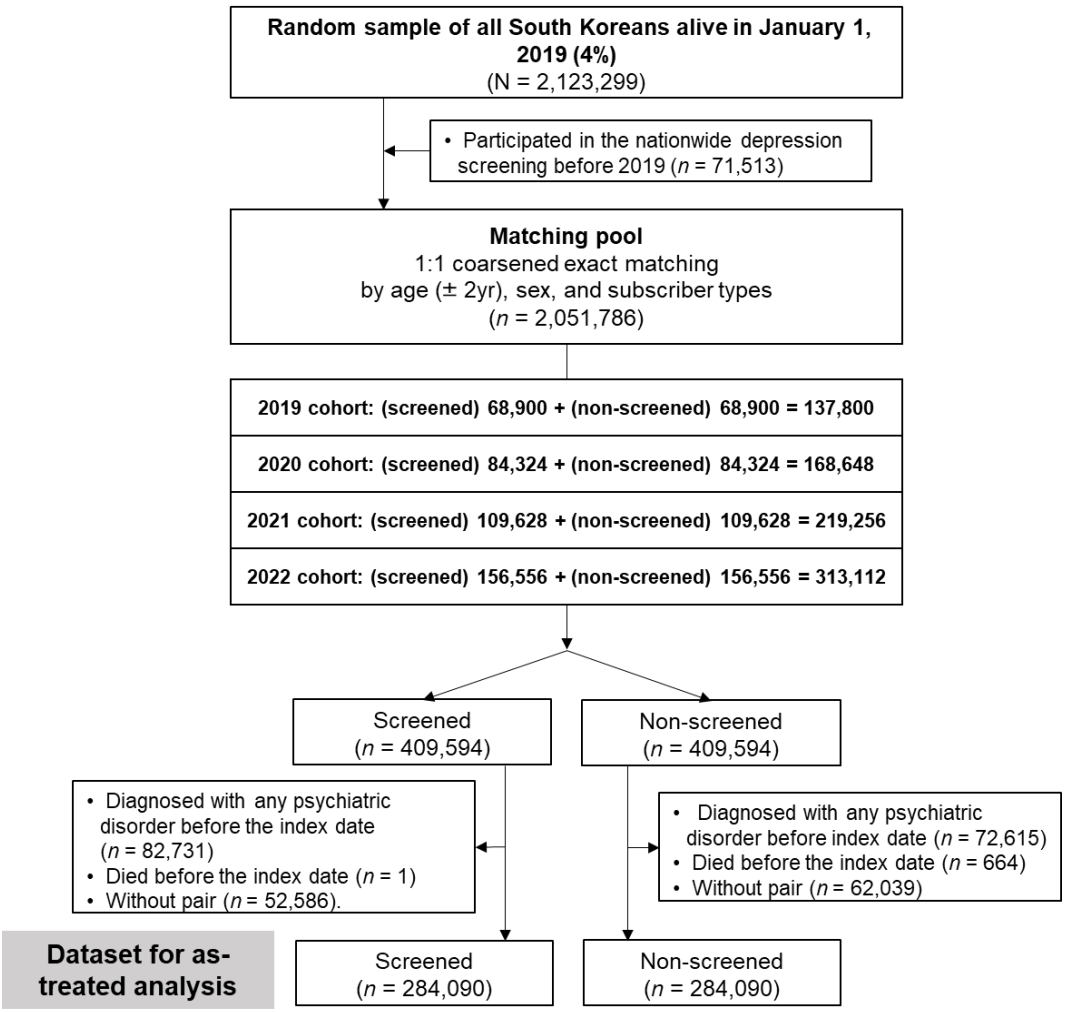
References (most recent)	Description
Oyama H.., 2006 (2)	**Title:** Outcomes of Community-Based Screening for Depression and Suicide Prevention Among Japanese Elders **Country:** Japan **Study design:** Non-randomized **Population:** Residents aged 65 and older in Yasuzuka **Data source:** Registry data from public health centers **Eligibility for control districts (district level):** ·Population 4000 - 10000 ·Average percentage of elderly aged 65 and over is greater than 15% ·Elderly suicide rate over 150 in both women and men **Analyzed sample size:** ·Intervention: 1 district ·Control: 1 district **Intervention:** ·Intervention: Public health education/two step screening with referral by nurse ·Control: TAU **Follow-up / Study period:** Baseline (1981-1990) and intervention period (1991-2000) **Outcomes and main findings (95% CI):** Age-adjusted IRR for suicide Men ·Intervention: 0.51 (0.22, 1.19) ·Control: 0.62 (0.29, 1.30), p-value for Wald test 0.73* Women ·Intervention: 0.36 (0.14, 0.93) ·Control: 0.60 (0.28, 1.30), p-value for Wald test 0.73* **Adjusted covariates:** Age

References (most recent)	Description
Oyama H.., 2006 (3)	**Title:** Local Community Intervention Through Depression Screening and Group Activity for Elderly Suicide Prevention **Country:** Japan **Study design:** Non-randomized **Population:** All residents aged 65 and older in Minami. **Data source:** Registry data from public health centers **Eligibility for control district (district level):** ·Population <=10000 ·Average percentage of elderly aged 65 and over >= 20% ·Elderly suicide rate over 150 in both women and men **Analyzed sample size:** ·Intervention: 1 district ·Control: 2 districts **Intervention:** ·Intervention: Public health education/two step screening with referral by nurse ·Control: TAU **Follow-up / Study period:** Baseline (1993-1998) and follow-up (1999-2004) **Outcomes and main findings (95% CI):** Age-adjusted IRR for suicide Men ·Intervention: 0.48 (0.10, 2.31) ·Control1: 0.61 (0.24, 1.59), p-value for Wald test 0.78* ·Control2: 0.75 (0.27, 2.10), p-value for Wald test 0.76* Women ·Intervention: 0.26 (0.07, 0.98) ·Control1: 1.86 (0.75, 4.48), p-value for Wald test 0.01* ·Control2: 0.91 (0.28, 2.96), p-value for Wald test 0.15* **Adjusted covariates:** Age and sex

References (most recent)	Description
Oyama H.., 2004	**Title:** Community-Based Prevention for Suicide in Elderly by Depression Screening and Follow-up **Country:** Japan **Study design:** Non-randomized **Population:** Elderly aged 65 and over in the Japanese rural town of Joboji ·With two neighboring control areas (Kunohe and Yamagata) **Data source:** Registry data from public health centers **Eligibility for control districts:** ·Average population 4000 - 10000 ·Average percentage of elderly aged 65 and over is greater than 10% ·Average rate of suicidal mortality among residents aged 65 and over **Analyzed sample size:** ·Intervention: 1 district ·Control: 2 districts **Intervention:** ·Intervention: Screening, follow-up care and education about depression ·Control: TAU **Follow-up / Study period:** Preparation stage (1985-1989), intervention stage (1990-1994), and maintenance stage (1995-1999) **Outcomes and main findings (95% CI):** Age-adjusted OR for cumulative number of suicide Men ·Intervention: 0.27 (0.08, 0.94) ·Control1: 0.92 (0.14, 5.99), p-value for Wald test 0.28* ·Control2: 0.91 (0.25, 3.33), p-value for Wald test 0.18* Women ·Intervention: 0.24 (0.10, 0.59) ·Control1: 0.89 (0.17, 4.60), p-value for Wald test 0.17* ·Control2: 0.51 (0.14, 1.92), p-value for Wald test 0.35* **Adjusted covariates:** Age

References (most recent)	Description
MacArthur., 2002	**Title:** Effects of Redesigned Community Postnatal Care on Women's' Health 4 Month after Birth: A Cluster Randomised Controlled Trial **Country:** USA **Study design:** Randomized (unit: practice) **Population:** Women who were in 34 weeks' gestation and having postnatal care **Data source:** From 36 randomly selected practice in the West Midlands health region of the UK between Oct 1997 and Apr 1999 **Eligibility:** ·Excluded if the practice is managed by midwife ·Excluded if they expected to move out of the general practice **Analyzed sample size:** ·Intervention: 1087 ·Control: 977 **Intervention:** ·Intervention: Midwife-led home visit including screening and referral ·Control: TAU **Follow-up / Study period:** 4 months **Outcomes and main findings (p-value):** Positive for EPDS OR 0.57 (0.43, 0.76) **Adjusted covariates:** (practice level) GP partners, # of midwives, Townsend score, midwife qualification score, cluster size; (individual level) parity, maternal age, mode of delivery, perineal trauma, other adults in house, education, social support, home ownership, Townsend quartiles.
Williams Jr. J.W., 1999	**Title:** Case-Finding for Depression in Primary Care: A Randomized Trial **Country:** USA **Study design:** Randomized (unit: individual) **Population:** Adults who were attending community clinics **Data source:** Recruited from a community-based family medicine clinic, a Veterans-Affairs general internal medicine clinic, and university-affiliated general internal medicine clinics. **Eligibility:** ·Excluded if they had no telephone or stable address. **Analyzed sample size:** ·Intervention: 153 ·Control: 65 **Intervention:** ·Intervention: Case-finding +notification to their physician ·Control: Usual care **Follow-up / Study period:** 3 months **Outcomes and main findings (p-value):** Prevalence of depression RR 0.79 (0.57, 1.11) **Adjusted covariates:** N/A

Appendix 5. Flow chart of the study population (as-treated approach)



Note. The flow chart illustrates the selection of the study population for the as-treated approach, where participants were categorized as the screened group if they completed depression screening at any point during the study. Abbreviation: DS, depression screening; yr, year.

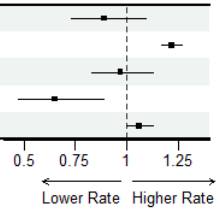
Appendix 6. Characteristics of screened and non-screened groups in the as-treated analysis (n = 568,180, follow-up: date of screening – 2022/12/31).

		Screened *n* = 284,090		Non-screened *n* = 284,090		
Variables		***n***	**(%)**	***n***	**(%)**	***p*-value**
Death		484	(0.17)	1,342	(0.47)	<.0001
Sex	Men	152,806	(53.8)	152,806	(53.8)	−
	Women	131,284	(46.2)	131,284	(46.2)	
Age group	20	18,608	(6.6)	18,608	(6.6)	−
	30	57,469	(20.2)	57,469	(20.2)	
	40	66,081	(23.3)	66,081	(23.3)	
	50	67,043	(23.6)	67,043	(23.6)	
	60	52,605	(18.5)	52,605	(18.5)	
	70	22,284	(7.8)	22,284	(7.8)	
Subscriber types	Self-employed insured	64,832	(22.8)	64,832	(22.8)	−
	Employee insured	217,550	(76.6)	217,550	(76.6)	
	Medical aid	1,704	(0.6)	1,704	(0.6)	
Urbanicity	Metropolitan	132,493	(46.6)	136,249	(48.0)	<.0001
	Large city	73,027	(25.7)	68,672	(24.2)	
	Rural area	78,560	(27.7)	79,149	(27.9)	
	Unknown	10	(0.0)	20	(0.0)	
# of health check-up (5 yr)	0	63,214	(22.3)	116,237	(40.9)	<.0001
	1	54,412	(19.2)	56,914	(20.0)	
	2	86,191	(30.3)	49,622	(17.5)	
	3	29,145	(10.3)	29,264	(10.3)	
	4	20,878	(7.3)	14,044	(4.9)	
	5	30,250	(10.6)	18,009	(6.3)	
Income	Medicaid	1,704	(0.6)	1,704	(0.6)	<.0001
	Q1	49,517	(17.4)	64,093	(22.6)	
	Q2	60,922	(21.4)	64,469	(22.7)	
	Q3	78,195	(27.5)	69,374	(24.4)	
	Q4	87,557	(30.8)	78,510	(27.6)	
	unknown	6,195	(2.2)	5,940	(2.1)	
CCI score	0	194,470	(68.5)	202,573	(71.3)	<.0001
	1	66,676	(23.5)	60,705	(21.4)	
	2+	22,944	(8.1)	20,812	(7.3)	

***Notes*.** Coarsened exact matching was used to match the screened and non-screened groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Abbreviation: CCI, Charlson comorbidities index.

Appendix 7. Results of primary and secondary outcomes in depression screening analyzed using the as-treated approaches

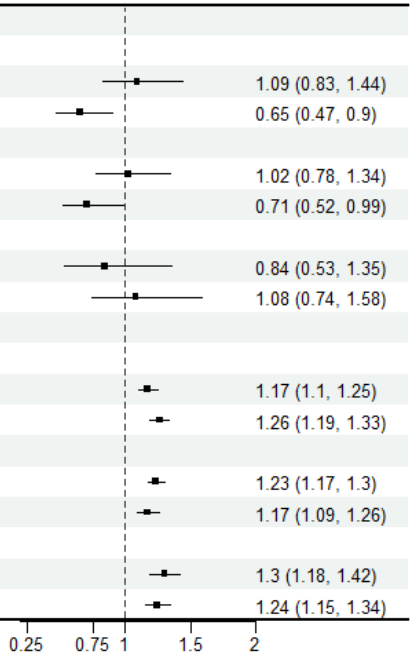
	Non-screened			Screened				
Outcomes	Case N	Person-year	Incidence rate (per 1,000 py)	Case N	Person-year	Incidence rate (per 1,000 py)	HR (95% CI)	aHR (95% CI)
Hospitalization for mood disorders	283	408544.95	0.69	238	409604.54	0.58	0.84 (0.70, 0.99)	0.89 (0.73, 1.09)
Antidepressant use	4776	403766.44	11.83	5875	403452.53	14.56	1.23 (1.18, 1.28)	1.22 (1.17, 1.27)
ED visit for mood disorders	427	408414.11	1.05	407	409388.25	1	0.95 (0.83, 1.09)	0.97 (0.83, 1.13)
Suicide and suicidal behaviors	158	408794.62	0.39	91	409818.7	0.22	0.58 (0.45, 0.75)	0.65 (0.47, 0.89)
Hospitalization for cancer	2714	406121.92	6.68	2672	406739.17	6.57	0.99 (0.94, 1.04)	1.06 (1.00, 1.13)



Notes. Coarsened exact matching was used to match the screened and non-screened groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Participant's urbanicity of residence, the number of previous health check-ups, income, and CCI were adjusted in the final models. The hospitalization for cancer was used as a negative control outcome. Cancer diagnosis was excluded in the CCI calculation when the diagnosis of cancer was used as a negative control outcome. Abbreviation: ED, emergency department. HR, hazard ratio; aHR, adjusted hazard ratio.

Appendix 8. Subgroup analysis for hospitalization for mood disorders and use of antidepressants in the as-treated analysis

Subgroups	Non-screened			Screened			aHR (95% CI)
	Case N	Person-year	Incidence rate (per 1,000 py)	Case N	Person-year	Incidence rate (per 1,000 py)	
Hospitalization for mood disorders							
Sex							
Men (n = 305,612)	151	218732.87	0.69	142	219516.42	0.64	1.09 (0.83, 1.44)
Women (n =262,568)	132	189812.08	0.7	96	190088.13	0.51	0.65 (0.47, 0.9)
Age							
Younger adults (n = 209,201)	146	305041.73	0.48	131	305373.61	0.43	1.02 (0.78, 1.34)
Older adults (n = 74,889)	137	103503.22	1.32	107	104230.93	1.03	0.71 (0.52, 0.99)
Income							
Lower income (n = 242,409)	136	186799.23	0.69	106	160927.19	0.57	0.84 (0.53, 1.35)
Higher income (n = 313,636)	139	213117.14	0.58	127	240510.3	0.42	1.08 (0.74, 1.58)
Use of antidepressants							
Sex							
Men (n = 305,612)	2082	216696.14	9.61	2523	216924.64	11.63	1.17 (1.1, 1.25)
Women (n =262,568)	2694	187070.31	14.4	3352	186527.9	17.97	1.26 (1.19, 1.33)
Age							
Younger adults (n = 209,201)	3079	301932.47	10.2	3721	301508.49	12.34	1.23 (1.17, 1.3)
Older adults (n = 74,889)	1697	101833.98	10.2	2154	101944.04	21.13	1.17 (1.09, 1.26)
Income							
Lower income (n = 242,409)	2202	184565.27	11.93	2470	158370.32	15.6	1.3 (1.18, 1.42)
Higher income (n = 313,636)	2482	210662.8	11.78	3281	237009.86	13.84	1.24 (1.15, 1.34)



Notes. Younger adults refer to adults aged 20 to 50, while older adults refer to seniors aged 60 to 70. Coarsened exact matching was used to match the screening and non-screening groups in a 1:1 ratio based on age (± 2-year intervals), sex, and subscriber types. Participant's urbanicity of residence, the number of previous health check-ups, income, and CCI were adjusted in the final models. Abbreviation: aHR, adjusted hazards ratio; py, person-year; CI, confidence interval

REFERENCES

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5: American psychiatric association Washington, DC; 2013.
2. 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.
3. Patten SB, Williams JV, Lavorato DH, Wang JL, Bulloch AG, Sajobi T. The association between major depression prevalence and sex becomes weaker with age. Social psychiatry and psychiatric epidemiology 2016;51:203-10.
4. Luppa M, Sikorski C, Luck T, Ehreke L, Konnopka A, Wiese B, et al. Age-and gender-specific prevalence of depression in latest-life–systematic review and meta-analysis. Journal of affective disorders 2012;136:212-21.
5. Angst J, Gamma A, Gastpar M, Lépine J-P, Mendlewicz J, Tylee A. Gender differences in depression: Epidemiological findings from the European DEPRES I and II studies. European archives of psychiatry and clinical neuroscience 2002;252:201-9.
6. Moreno-Agostino D, Wu Y-T, Daskalopoulou C, Hasan MT, Huisman M, Prina M. Global trends in the prevalence and incidence of depression: a systematic review and meta-analysis. Journal of affective disorders 2021;281:235-43.
7. Freeman A, Tyrovolas S, Koyanagi A, Chatterji S, Leonardi M, Ayuso-Mateos JL, et al. The role of socio-economic status in depression: results from the COURAGE (aging survey in Europe). BMC public health 2016;16:1-8.
8. Miech RA, Shanahan MJ. Socioeconomic status and depression over the life course. Journal of health and social behavior 2000:162-76.
9. Lorant V, Deliège D, Eaton W, Robert A, Philippot P, Ansseau M. Socioeconomic inequalities in depression: a meta-analysis. American journal of epidemiology 2003;157:98-112.
10. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. The lancet 2013;382:1575-86.

11. Turecki G, Brent DA. Suicide and suicidal behaviour. The Lancet 2016;387:1227-39.
12. Li Z, Page A, Martin G, Taylor R. Attributable risk of psychiatric and socio-economic factors for suicide from individual-level, population-based studies: a systematic review. Social science & medicine 2011;72:608-16.
13. Kessler RC. The costs of depression. Psychiatric Clinics 2012;35:1-14.
14. Grayson L, Thomas A. A systematic review comparing clinical features in early age at onset and late age at onset late-life depression. Journal of affective disorders 2013;150:161-70.
15. Zisook S, Lesser I, Stewart JW, Wisniewski SR, Balasubramani G, Fava M, et al. Effect of age at onset on the course of major depressive disorder. American Journal of Psychiatry 2007;164:1539-46.
16. Sadock BJ. Kaplan & Sadock's synopsis of psychiatry: behavioral sciences/clinical psychiatry: Wolters Kluwer Philadelphia, PA; 2015.
17. Keller MB, Lavori PW, Mueller TI, Endicott J, Coryell W, Hirschfeld RM, et al. Time to recovery, chronicity, and levels of psychopathology in major depression: a 5-year prospective follow-up of 431 subjects. Archives of general psychiatry 1992;49:809-16.
18. Mueller TI, Keller MB, Leon AC, Solomon DA, Shea MT, Coryell W, et al. Recovery after 5 years of unremitting major depressive disorder. Archives of general psychiatry 1996;53:794-9.
19. Monroe SM, Harkness KL. Major depression and its recurrences: life course matters. Annual review of clinical psychology 2022;18:329-57.
20. Monroe SM, Harkness KL. Life stress, the" kindling" hypothesis, and the recurrence of depression: considerations from a life stress perspective. Psychological review 2005;112:417.
21. Kraus C, Kadriu B, Lanzenberger R, Zarate Jr CA, Kasper S. Prognosis and improved outcomes in major depression: a review. Translational psychiatry 2019;9:127.
22. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. Journal of affective disorders 2014;152:45-51.
23. Hung C-I, Liu C-Y, Yang C-H. Untreated duration predicted the severity of depression at the two-year follow-up point. PloS one 2017;12:e0185119.
24. Bukh JD, Bock C, Vinberg M, Kessing LV. The effect of prolonged duration of untreated depression on antidepressant treatment outcome. Journal of affective disorders 2013;145:42-8.

25. Kautzky A, Dold M, Bartova L, Spies M, Kranz GS, Souery D, et al. Clinical factors predicting treatment resistant depression: affirmative results from the European multicenter study. Acta Psychiatrica Scandinavica 2019;139:78-88.
26. Lecrubier Y. Widespread underrecognition and undertreatment of anxiety and mood disorders: results from 3 European studies. Journal of Clinical Psychiatry 2007;68:36-41s.
27. Pelletier L, O'Donnell S, Dykxhoorn J, McRae L, Patten S. Under-diagnosis of mood disorders in Canada. Epidemiology and psychiatric sciences 2017;26:414-23.
28. Faisal-Cury A, Ziebold C, de Oliveira Rodrigues DM, Matijasevich A. Depression underdiagnosis: prevalence and associated factors. A population-based study. Journal of psychiatric research 2022;151:157-65.
29. Wang PS, Angermeyer M, Borges G, Bruffaerts R, Chiu WT, De Girolamo G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World psychiatry 2007;6:177.
30. Martín-Moreno BS, Guija J, Blanco M, Porras-Segovia A, Pereira-Sánchez V, Baca-García E, et al. Underdiagnosis, Undertreatment, and Noncompliance With Treatment in People Who Died by Suicide. The Journal of Clinical Psychiatry 2024;85:55925.
31. O'connor DW, Rosewarne R, Bruce A. Depression in primary care 1: Elderly patients' disclosure of depressive symptoms to their doctors. International Psychogeriatrics 2001;13:359-65.
32. Kessler D, Heath I, Lloyd K, Lewis G, Gray DP. General PracticeCross sectional study of symptom attribution and recognition of depression and anxiety in primary careCommentary: There must be limits to the medicalisation of human distress. Bmj 1999;318:436-40.
33. Schnyder N, Panczak R, Groth N, Schultze-Lutter F. Association between mental health-related stigma and active help-seeking: systematic review and meta-analysis. The British Journal of Psychiatry 2017;210:261-8.
34. O'Connor EA, Perdue LA, Coppola EL, Henninger ML, Thomas RG, Gaynes BN. Depression and suicide risk screening: updated evidence report and systematic review for the US preventive services task force. JAMA 2023;329:2068-85.
35. Fish FJ, Casey P, Kelly B. Fish's clinical psychopathology: signs and symptoms in psychiatry: Cambridge University Press; 2024.
36. Raffle AE, Mackie A, Gray JM. Screening: evidence and practice: Oxford University Press, USA; 2019.

37. Otto MW, Birk JL, Fitzgerald HE, Chauvin GV, Gold AK, Carl JR. Stage models for major depression: Cognitive behavior therapy, mechanistic treatment targets, and the prevention of stage transition. Clinical psychology review 2022;95:102172.
38. Barry MJ, Nicholson WK, Silverstein M, Chelmow D, Coker TR, Davidson KW, et al. Screening for depression and suicide risk in adults: US Preventive Services Task Force recommendation statement. Jama 2023;329:2057-67.
39. Lang E, Colquhoun H, LeBlanc JC, Riva JJ, Moore A, Traversy G, et al. Recommendation on instrument-based screening for depression during pregnancy and the postpartum period. CMAJ 2022;194:E981-E9.
40. UK National Screening Committee. Screening for depression in adults: External review against programme appraisal criteria for the UK National Screening Committee. UK National Screening Committee; 2020.
41. Thombs BD, Saadat N, Riehm KE, Karter JM, Vaswani A, Andrews BK, et al. Consistency and sources of divergence in recommendations on screening with questionnaires for presently experienced health problems or symptoms: a comparison of recommendations from the Canadian Task Force on Preventive Health Care, UK National Screening Committee, and US Preventive Services Task Force. BMC medicine 2017;15:1-17.
42. Royal Australian College of General Practitioners. Preventive activities in general practice: Mental health - Depression. Available at: https://www.racgp.org.au/clinical-resources/clinical-guidelines/key-racgp-guidelines/view-all-racgp-guidelines/preventive-activities-in-general-practice/mental-health/depression#ref-num-4 [Accessed December 7 2024]
43. Park SM, Yun J-M, Son JS, Jeong S-M, Lee G, Kim K, et al. Evaluation of the Validity of Adopting the National Health Screening Program for Koreans in their Twenties and Thirties. Ministry of Health & Welfare; 2017.
44. Japanese Ministry of Health Labour and Welfare. Stress Check System and Occupational Safety and Health. Available at: https://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei12/ [Accessed December 7 2024]
45. Health Promotion Administration; Taiwanese Ministry of Health and Welfare. Mental health promotion strategies and resources. Available at: https://www.hpa.gov.tw/EngPages/Detail.aspx?nodeid=1076&pid=18516 [Accessed December 7 2024]
46. Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? Bmj 2014;348.

47. García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. European journal of epidemiology 2017;32:495-500.
48. Nallamothu BK, Hayward RA, Bates ER. Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. Circulation 2008;118:1294-303.
49. Goodyear-Smith FA, van Driel ML, Arroll B, Del Mar C. Analysis of decisions made in meta-analyses of depression screening and the risk of confirmation bias: a case study. BMC Med Res Methodol 2012;12:76.
50. Kim GE, Jo M-W, Shin Y-W. Increased prevalence of depression in South Korea from 2002 to 2013. Scientific reports 2020;10:16979.
51. OECD. Suicide rates. 2020.
52. Chang SM, Hong J-P, Cho MJ. Economic burden of depression in South Korea. Social psychiatry and psychiatric epidemiology 2012;47:683-9.
53. Park S-Y, Bernstein KS. Depression and Korean American immigrants. Archives of psychiatric nursing 2008;22:12-9.
54. An S, Lee H, Lee J, Kang S. Social stigma of suicide in South Korea: A cultural perspective. Death studies 2023;47:259-67.
55. Lee J, Choi K-S, Yun J-A. The effects of sociodemographic factors on help-seeking for depression: Based on the 2017–2020 Korean Community Health Survey. PLoS One 2023;18:e0280642.
56. Shin DW, Cho J, Park JH, Cho B. National General Health Screening Program in Korea: history, current status, and future direction. Precision and Future Medicine 2022;6:9-31.
57. Chun CB, Kim SY, Lee JY, Lee SY, Organization WH. Republic of Korea: health system review. 2009.
58. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj 2021;372.
59. Joffres M, Jaramillo A, Dickinson J, Lewin G, Pottie K, Shaw E, et al. Recommendations on screening for depression in adults. Cmaj 2013;185:775-82.
60. Beck A, Hamel C, Thuku M, Esmaeilisaraji L, Bennett A, Shaver N, et al. Screening for depression among the general adult population and in women during pregnancy or the first-year postpartum: two systematic reviews to inform a guideline of the Canadian Task Force on Preventive Health Care. Systematic Reviews 2022;11:176.

61. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Statistics in medicine 2000;19:3127-31.
62. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. bmj 2019;366.
63. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. bmj 2016;355.
64. Tompsett D, Zylbersztejn A, Hardelid P, De Stavola B. Target trial emulation and bias through missing eligibility data: an application to a study of palivizumab for the prevention of hospitalization due to infant respiratory illness. American Journal of Epidemiology 2023;192:600-11.
65. Fu EL. Target trial emulation to improve causal inference from observational data: what, why, and how? Journal of the American Society of Nephrology 2023;34:1305-14.
66. Dobrow MJ, Hagens V, Chafe R, Sullivan T, Rabeneck L. Consolidated principles for screening based on a systematic review and consensus process. Cmaj 2018;190:E422-E9.
67. Sutton AJ, Kendrick D, Coupland CA. Meta-analysis of individual-and aggregate-level data. Statistics in medicine 2008;27:651-69.
68. Cooper H, Patall EA. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. Psychological methods 2009;14:165.
69. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to meta-analysis: John Wiley & Sons; 2021.
70. Harris RJ, Deeks JJ, Altman DG, Bradburn MJ, Harbord RM, Sterne JA. Metan: fixed-and random-effects meta-analysis. The Stata Journal 2008;8:3-28.
71. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. American journal of epidemiology 2016;183:758-64.
72. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. Journal of clinical epidemiology 2016;79:70-5.
73. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. Jama 2022;328:2446-7.
74. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. Clinical trials 2012;9:48-55.

75. Dossing A, Tarp S, Furst DE, Gluud C, Wells GA, Beyene J, et al. Modified intention-to-treat analysis did not bias trial results. Journal of Clinical Epidemiology 2016;72:66-74.
76. Thombs BD, Markham S, Rice DB, Ziegelstein RC. Does depression screening in primary care improve mental health outcomes? bmj 2021;374.
77. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. Journal of general internal medicine 2001;16:606-13.
78. Choi HS, Choi JH, Park KH, Joo KJ, Ga H, Ko HJ, et al. Standardization of the Korean Version of Patient Health Questionnaire-9 as a Screening Instrument for Major Depressive Disorder. Journal of the Korean Academy of Family Medicine 2007;28:114-9.
79. Levis B, Benedetti A, Thombs BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. bmj 2019;365.
80. Imbens GW, Lemieux T. Regression discontinuity designs: A guide to practice. Journal of econometrics 2008;142:615-35.
81. Stahl SM. Stahl's essential psychopharmacology: neuroscientific basis and practical applications: Cambridge university press; 2021.
82. Wang HR, Bahk W-M, Park Y-M, Lee HB, Song HR, Jeong J-H, et al. Korean medication algorithm for depressive disorder: comparisons with other treatment guidelines. Psychiatry Investigation 2014;11:1.
83. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. Journal of clinical epidemiology 2004;57:1288-94.
84. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med 2017;167:268-74.
85. Team RC. R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria 2013.
86. Oyama H, Fujita M, Goto M, Shibuya H, Sakashita T. Outcomes of community-based screening for depression and suicide prevention among Japanese elders. The Gerontologist 2006;46:821-6.
87. Oyama H, Sakashita T, Hojo K, Ono Y, Watanabe N, Takizawa T, et al. A community-based survey and screening for depression in the elderly. Crisis 2010.
88. Oyama H, Sakashita T. Community-based screening intervention for depression affects suicide rates among middle-aged Japanese adults. Psychological medicine 2017;47:1500-9.

89. Oyama H, Sakashita T. Long-term effects of a screening intervention for depression on suicide rates among Japanese community-dwelling older adults. The American Journal of Geriatric Psychiatry 2016;24:287-96.
90. Oyama H, Sakashita T. Effects of universal screening for depression among middle-aged adults in a community with a high suicide rate. The Journal of Nervous and Mental Disease 2014;202:280-6.
91. Oyama H, Goto M, Fujita M, Shibuya H, Sakashita T. Preventing elderly suicide through primary care by community-based screening for depression in rural Japan. Crisis 2006;27:58-65.
92. Oyama H, Ono Y, Watanabe N, Tanaka E, Kudoh S, Sakashita T, et al. Local community intervention through depression screening and group activity for elderly suicide prevention. Psychiatry and clinical neurosciences 2006;60:110-4.
93. Oyama H, Koida J, Sakashita T, Kudo K. Community-based prevention for suicide in elderly by depression screening and follow-up. Community mental health journal 2004;40:249-63.
94. Kronish IM, Moise N, Cheung YK, Clarke GN, Dolor RJ, Duer-Hefele J, et al. Effect of depression screening after acute coronary syndromes on quality of life: the CODIACS-QoL randomized clinical trial. JAMA internal medicine 2020;180:45-53.
95. Williams Jr JW, Mulrow CD, Kroenke K, Dhanda R, Badgett RG, Omori D, et al. Case-finding for depression in primary care: a randomized trial∗. The American journal of medicine 1999;106:36-43.
96. Rhee TG, Capistrant BD, Schommer JC, Hadsall RS, Uden DL. Effects of the 2009 USPSTF depression screening recommendation on diagnosing and treating mental health conditions in older adults: a difference-in-differences analysis. Journal of managed care & specialty pharmacy 2018;24:769-76.
97. Rona RJ, Burdett H, Khondoker M, Chesnokov M, Green K, Pernet D, et al. Post-deployment screening for mental disorders and tailored advice about help-seeking in the UK military: a cluster randomised controlled trial. The Lancet 2017;389:1410-23.
98. Mallen CD, Nicholl BI, Lewis M, Bartlam B, Green D, Jowett S, et al. The effects of implementing a point-of-care electronic template to prompt routine anxiety and depression screening in patients consulting for osteoarthritis (the Primary Care Osteoarthritis Trial): a cluster randomised trial in primary care. PLoS medicine 2017;14:e1002273.

99. MacArthur C, Winter H, Bick D, Knowles H, Lilford R, Henderson C, et al. Effects of redesigned community postnatal care on womens' health 4 months after birth: a cluster randomised controlled trial. The lancet 2002;359:378-85.
100. Silverstone PH, Rittenbach K, Suen VY, Moretzsohn A, Cribben I, Bercov M, et al. Depression outcomes in adults attending family practice were not improved by screening, stepped-care, or online CBT during a 12-week study when compared to controls in a randomized trial. Frontiers in psychiatry 2017;8:32.
101. Romera I, Montejo Á L, Aragonés E, Arbesú JÁ , Iglesias-Garcí a C, López S, et al. Systematic depression screening in high-risk patients attending primary care: a pragmatic cluster-randomized trial. Bmc Psychiatry 2013;13:1-9.
102. Leung SS, Leung C, Lam T, Hung S, Chan R, Yeung T, et al. Outcome of a postnatal depression screening programme using the Edinburgh Postnatal Depression Scale: a randomized controlled trial. Journal of Public Health 2011;33:292-301.
103. Chen Y-L, Wu M-S, Wang S-H, Lien Y-J, Liao S-C, Chang C-M, et al. Effectiveness of health checkup with depression screening on depression treatment and outcomes in middle-aged and older adults: a target trial emulation study. The Lancet Regional Health–Western Pacific 2024;43.
104. Oyama H, Sakashita T. Differences in specific depressive symptoms among community-dwelling middle-aged Japanese adults before and after a universal screening intervention. Social psychiatry and psychiatric epidemiology 2014;49:251-8.
105. Gilbody S, House A, Sheldon T. Screening and case finding instruments for depression. Cochrane Database of Systematic Reviews 2005.
106. US Preventive Services Task Force. Screening for depression: recommendations and rationale. Annals of internal medicine 2002;136:760-4.
107. Nickerson RS. Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology 1998;2:175-220.
108. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer “epidemic”—screening and overdiagnosis. New England Journal of Medicine 2014;371:1765-7.
109. Ghio L, Gotelli S, Cervetti A, Respino M, Natta W, Marcenaro M, et al. Duration of untreated depression influences clinical outcomes and disability. Journal of affective disorders 2015;175:224-8.
110. de Diego-Adeliño J, Portella MJ, Puigdemont D, Pérez-Egea R, Á lvarez E, Pérez V. A short duration of untreated illness (DUI) improves response outcomes in first-depressive episodes. Journal of affective disorders 2010;120:221-5.

111. Naudet F, Maria AS, Falissard B. Antidepressant response in major depressive disorder: a meta-regression comparison of randomized controlled trials and observational studies. PLoS One 2011;6:e20811.
112. Palpacuer C, Gallet L, Drapier D, Reymann J-M, Falissard B, Naudet F. Specific and non-specific effects of psychotherapeutic interventions for depression: Results from a meta-analysis of 84 studies. Journal of Psychiatric Research 2017;87:95-104.
113. Briggs R, Tobin K, Kenny RA, Kennelly SP. What is the prevalence of untreated depression and death ideation in older people? Data from the Irish Longitudinal Study on Aging. International psychogeriatrics 2018;30:1393-401.
114. Krivoy A, Balicer RD, Feldman B, Hoshen M, Zalsman G, Weizman A, et al. The impact of age and gender on adherence to antidepressants: a 4-year population-based cohort study. Psychopharmacology 2015;232:3385-90.
115. Schülein S, Taylor KJ, Schriefer D, Blettner M, Klug SJ. Participation in preventive health check-ups among 19,351 women in Germany. Preventive medicine reports 2017;6:23-6.
116. Brunner-Ziegler S, Rieder A, Stein KV, Koppensteiner R, Hoffmann K, Dorner TE. Predictors of participation in preventive health examinations in Austria. BMC Public Health 2013;13:1-9.
117. Shin H-Y, Kang H-T, Lee JW, Lim H-J. The association between socioeconomic status and adherence to health check-up in Korean adults, based on the 2010–2012 Korean National Health and Nutrition Examination Survey. Korean Journal of Family Medicine 2018;39:114.
118. Kriston L. Dealing with clinical heterogeneity in meta-analysis. Assumptions, methods, interpretation. International journal of methods in psychiatric research 2013;22:1-15.
119. U.S. Preventive Services Task Force. Screening for depression: recommendations and rationale. Ann Intern Med 2002;136:760-4.
120. De Stavola BL, Gomes M, Katsoulis M. Transparency and Rigor: Target Trial Emulation Aims to Achieve Both. Epidemiology 2023;34:624-6.
121. Kuehne F, Arvandi M, Hess LM, Faries DE, Gothe RM, Gothe H, et al. Causal analyses with target trial emulation for real-world evidence removed large self-inflicted biases: systematic bias assessment of ovarian cancer treatment effectiveness. Journal of Clinical Epidemiology 2022;152:269-80.
122. Farjat AE, Virdone S, Thomas LE, Kakkar AK, Pieper KS, Piccini JP. The importance of the design of observational studies in comparative effectiveness

research: lessons from the GARFIELD-AF and ORBIT-AF registries. American Heart Journal 2022;243:110-21.

ABSTRACT IN KOREAN

전국민 대상 우울증 선별검사의 효과성 평가: 체계적 문헌 고찰 및 국민건강보험공단 데이터베이스를 이용한 표적 시험 에뮬레이션 연구

[연구배경] 지난 20 년간 1 차 진료나 국가 차원에서 조직적으로 시행된 우울증 선별검사의 효과에 대한 이전 연구들은 서로 다른 결론을 내었다. 따라서 본 연구에서는 체계적 문헌 고찰 및 메타분석과 국민건강보험공단 데이터베이스를 이용한 표적 시험 에뮬레이션 연구를 수행하여, 전국민 대상 우울증 선별검사의 효과성을 평가하였다.

[방법] 체계적 문헌 검토 및 메타분석은 일반 성인 인구를 대상으로 시행된 우울증 선별검사의 효과성을 평가하는 무작위 임상 시험(RCT)과 관찰 연구들의 결과들을 종합하였다. PubMed, EMBASE, PsychINFO 및 Web of Science 와 같은 데이터베이스를 검색하여 2024 년 8 월까지 발표된 관련 논문들을 추출했다. 두 명의 독립적인 연구자가 연구 검색, 적합성 검토, 데이터 추출, 편향 위험 평가 및 증거 수준의 질적 평가를 수행했다. 메타 분석은 제한된 최대 가능도 추정법을 적용해 랜덤 효과 모델을 사용하여 수행했다. 연구 간 이질성은 I^2 통계와 예측 구간을, 출판 비뚤림은 깔때기 도표와 Egger's test 를 이용해 평가했다. 국내에 도입된 2019 년 우울증 선별검사의 효과성을 평가하기 위해 국민건강보험공단 데이터베이스를 사용하여 표적 시험 에뮬레이션 연구를 수행했다. 추적 관찰 기간은 2019 년 1 월 1 일에 시작하여 2022 년 12 월 31 일에 종료되었다. 주요 결과는 기분 장애로 인한 입원 발생이고, 부차적 결과는 (1) 첫 항우울제 사용, (2) 기분 장애로 인한 응급실 방문의 발생, (3) 자살 또는 자살 관련 행동의 발생이다. 2019 년에 생존한 18 세 이상 성인 중 4%를 랜덤 추출한 후, 이전에 정신과 진단을 받거나

우울증 선별검사에 참여했던 사람들은 제외했고, 수검자격인 특정 연령을 기준으로 유자격자와 무자격자를 분류했다. 그 다음, 연령(±2 세), 성별, 보험 자격에 따라 1:1 비율로 유자격자와 무자격자를 대략화된 일치 매칭(coarsened exact matching)하였다. 그 다음 중대한 프로토콜 위반 사례를 제외한 수검자와 비수검자의 데이터를 이용해 수정된 치료의향분석(intent-to-treat, ITT) 효과를 추정하였다(n=53,688). 선별검사 효과는 매칭된 쌍에 따라 층화된 다변량 원인-특정 Cox 회귀 모델로 추정했다. 보정된 모델에서는 거주지의 도시화 수준, 과거 건강 검진 횟수, 소득 수준, Charlson 동반 질환 지수와 같은 교란 변수를 보정했다. 하위 그룹 분석은 성별, 연령군, 소득 수준으로 층화해 분석했다. 음성 대조군 분석과 E-value 로 측정되지 않은 교란 효과를 평가했다.

[결과] 체계적 검토 및 메타분석에서는 일반 성인 인구를 대상으로 한 우울증 선별검사가 우울 관련 결과를 예방할 수 있다는 증거를 발견했으나(통합된 odds ratio 0.74, 95% CI: 0.62–0.87), 포함된 연구 대부분이 낮은 품질을 보였으며 연구 간 이질성이 컸다. 수정된 ITT 접근법을 기반으로 한 표적 시험 에뮬레이션 연구 결과, 2019 년 우울증 선별검사 수검자는 비수검자에 비해 기분 장애로 인한 입원 위험이 44% 낮았고(보정된 위험 비[adjusted hazard ratio, aHR]: 0.56, 95% CI: 0.41–0.76), 이는 여성과 고령층에서만 나타났다. 또한 수검자는 비수검자에 비해 항우울제 치료 시작이 많고(aHR: 1.21, 95% CI: 1.10–1.33), 응급실 방문 위험은 낮았으나(aHR: 0.71, 95% CI: 0.51–0.98), 자살 및 자살 행동에서는 유의한 차이가 없었다(aHR: 0.62, 95% CI: 0.32–1.21).

[결론] 국내에 도입된 전국민 대상 우울증 선별검사는 기분장애로 인한 입원과 응급실 방문의 감소 및 항우울제 치료의 활성화와 연관이 있다는 증거를 발견하였다. 그러나 이러한 수검자와 비수검자의 차이는 선별검사의 비특이적 효과로 인한 것으로 보이며, 향후 지속적인 효과성 평가를 통해 불필요한 비용을 감소할 필요가 있다.

핵심되는 말: 전국적 우울증 선별검사, 이차 예방, 체계적 문헌 검토, 메타 분석, 표적 시험 에뮬레이션, 국민건강보험공단 데이터베이스