Exploring the impact of socially assistive robots on health and wellbeing across the lifespan: An umbrella review and Meta-analysis

Bethany NICHOL, Jemma McCREADY, Goran ERFANI, Dania COMPARCINI, Valentina SIMONETTI, Giancarlo CICOLINI, Kristina MIKKONEN, Miyae YAMAKAWA, Marco TOMIETTO

PII: S0020-7489(24)00042-7
DOI: https://doi.org/10.1016/j.ijnurstu.2024.104730
Reference: NS 104730

To appear in:

Received date: 30 October 2023
Revised date: 14 February 2024
Accepted date: 15 February 2024

Please cite this article as: B. NICHOL, J. McCREADY, G. ERFANI, et al., Exploring the impact of socially assistive robots on health and wellbeing across the lifespan: An umbrella review and Meta-analysis, (2024), https://doi.org/10.1016/j.ijnurstu.2024.104730

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 The Author(s). Published by Elsevier Ltd.
Title: Exploring the impact of Socially Assistive Robots on health and wellbeing across the lifespan: an Umbrella review and Meta-Analysis

Short running title: Socially Assistive Robots in improving health and wellbeing: umbrella review and meta-analysis.

List of authors: Bethany NICHOL*, Jemma McCREADY*, Goran ERFANI, Dania COMPARCINI, Valentina SIMONETTI, Giancarlo CICOLINI, Kristina MIKKONEN, Miyae YAMAKAWA, Marco TOMIETTO**.
* equally contributed
** corresponding author

Authors:

1. Bethany NICHOL, MSc, Research Assistant, Department of Social Work, Education and Community Wellbeing, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom; bethany.nichol@northumbria.ac.uk; ORCID: https://orcid.org/0000-0002-7642-1448; Twitter: beth_research
2. Jemma McCREADY, MSc, BSc (Hons), Research Assistant, Department of Social Work, Education and Community Wellbeing, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom. j.mccready@northumbria.ac.uk; ORCID: https://orcid.org/0000-0002-9997-6956; Twitter: @JemmaMcC90
3. Goran ERFANI, PhD, Research Fellow, Department of Nursing, Midwifery and Health, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom; email: goran.erfani@northumbria.ac.uk; ORCID: https://orcid.org/0000-0002-8135-2846; Twitter: @GoranErfani
4. Dania COMPARCINI, PhD, MSc, RN, Research fellow, Section of Nursing, Department of Precision and Regenerative Medicine and Ionian Area, University of Bari "Aldo Moro", Bari, Italy; dania.comparcini@uniba.it; ORCID: https://orcid.org/0000-0003-3622-6370; Twitter: @d_comparcini.
5. Valentina SIMONETTI, PhD, MSc, RN, Associate Professor, University “LUM” Giuseppe Degennaro, Casamassima, Bari, Italy; simonetti@lum.it. ORCID: https://orcid.org/0000-0002-7185-4850; Twitter: @v_simonetti77.

6. Giancarlo CICOLINI, PhD, RN, Associate Professor, Section of Nursing, Department of Precision and Regenerative Medicine and Ionian Area, University of Bari "Aldo Moro", Bari, Italy; giancarlo.cicolini@uniba.it; ORCID: https://orcid.org/0000-0002-2736-1792; Twitter: @g_cicolini

7. Kristina MIKKONEN, Research Unit of Health Science and Technology, University of Oulu, Oulu, Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland; kristina.mikkonen@oulu.fi; ORCID: https://orcid.org/0000-0002-4355-3428; Twitter: @Kristinamikkon;

8. Miyae YAMAKAWA, PhD, RN, Associate Professor, Division of Health Sciences, Graduate School of Medicine, Osaka University, Suita City, Japan; miyatabu@sahs.med.osaka-u.ac.jp; ORCID: https://orcid.org/0000-0002-2399-8366;

9. Marco TOMIETTO*, PhD, MSc, RN, Professor of Nursing, Department of Nursing, Midwifery and Health, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom; Adjunct Professor, Research Unit of Health Science and Technology, University of Oulu, Oulu (Finland); marco.tomietto@northumbria.ac.uk; ORCID: https://orcid.org/0000-0002-3813-1490; Twitter: @marco_tomietto.

**Corresponding Author:** Marco Tomietto, Department Nursing, Midwifery and Health, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom; e-mail: marco.tomietto@northumbria.ac.uk; telephone number: +44 (0191) 215 6060
Abstract

**Background:** Socially assistive robots offer an alternate source of connection for interventions within health and social care amidst a landscape of technological advancement and reduced staff capacity. There is a need to summarise the available systematic reviews on the health and wellbeing impacts to evaluate effectiveness, explore potential moderators and mediators, and identify recommendations for future research and practice.

**Objective:** To explore the effect of socially assistive robots within health and social care on psychosocial, behavioural, and physiological health and wellbeing outcomes across the lifespan (International Prospective Register of Systematic Reviews (PROSPERO) registration: (blinded for peer review)).

**Design:** An umbrella review utilising meta-analysis, narrative synthesis, and vote counting by direction of effect.

**Methods:** 14 databases were searched (ProQuest Health Research Premium collection, Scopus, PubMed, Web of Science, ASM Digital Library, IEEE Xplore, Cochrane Reviews, and EPISTEMONIKOS) from 2005 to May 4, 2023. Systematic reviews including the effects of socially assistive robots on health outcomes were included and a pooled meta-analysis, vote counting by direction of effect, and narrative synthesis were applied. The second version of A MeaSurement Tool to Assess systematic Reviews (AMSTAR-2) was applied to assess quality of included reviews.

**Results:** 35 reviews were identified, most focusing on older adults with or without dementia (n = 24). Pooled meta-analysis indicated no effect of socially assistive robots on quality of life (standard mean difference (SMD) = 0.43), anxiety (SMD = -0.02), or depression (SMD = 0.21), although vote counting identified significant improvements in social interaction, mood, positive affect, loneliness, stress, and pain across the lifespan, and narrative synthesis
identified an improvement in anxiety in children. However, some reviews reported no significant difference between the effects of socially assistive robots and a plush toy, and there was no effect of socially assistive robots on psychiatric outcomes including agitation, neuropsychiatric symptoms, and medication use.

**Discussion:** Socially assistive robots show promise for improving non-psychiatric outcomes such as loneliness, positive affect, stress, and pain, but exert no effect on psychiatric outcomes such as depression and agitation. The main mechanism of effect within group settings appeared to be the stimulation of social interaction with other humans. Limitations include the low quality and high amount of overlap between included reviews.

**Conclusion:** Socially assistive robots may help to improve loneliness, social interaction, and positive affect in older adults, decrease anxiety and distress in children, and improve mood, stress, and reduce pain across the lifespan. However, before recommendations for socially assistive robots can be made, a cost-effectiveness analysis of socially assistive robots to improve mood across the lifespan, and a quantitative analysis of the effects on pain, anxiety, and distress in children are required.

**Keywords:** socially assistive robots, health outcomes, wellbeing, children, older adults, nursing.
What is already known

- Social interaction is a key determinant of health and wellbeing across the lifespan and socially assistive robots thus offer a potentially cost-effective source of social interaction.
- Many existing systematic reviews have evaluated the impact of socially assistive robots and report positive health and wellbeing impacts including reduced depression and social isolation in adults, and reduced distress and pain in children.
- The sparse available literature that has explored the mechanism/s of effect indicates that socially assistive robots exert positive health and wellbeing outcomes through an empathic relationship and emotional bond, much like with an animal or human.

What this paper adds

- Socially assistive robots may exert some positive impact on non-psychiatric outcomes including social interaction, positive affect, loneliness, stress, and pain, but have no impact on psychiatric outcomes including agitation, depression, neuropsychiatric symptoms, and medication use.
- The mechanism of effect within group settings appears not to be through an emotional bond with the socially assistive robot, but through encouragement of social interaction amongst other humans.
- This research identifies gaps in the literature including a need for quantitative analysis on the cost-effectiveness of socially assistive robots, their impact on pain, anxiety, and distress in children, and the mediating effect of social interaction.
Introduction

Digital solutions are reshaping human interactions; both in terms of human-human and human-machine interaction. This new landscape also affects the concepts of social connection and belonging, as these human needs are now located in a virtual and technological space which do not necessarily require physical interaction. It has been argued that humans evolved with a tendency towards social connection, facilitated by genetic, neural, and hormonal changes that promote bonding and companionship (1). According to Maslow’s theory, achieving higher levels of self-actualisation in life means building a strong foundation of physical and psychological safety, followed by social belonging (2). If a higher need is compromised, the previous ones are also at risk. For example, if social connection or belonging are compromised, physical and psychological safety and basic needs are in danger too. Empirical evidence demonstrates the consistency of these theoretical assumptions: lack of social connection is associated with negative health outcomes across the lifespan including depression and anxiety in older adults (3), dementia risk (4), and physiological changes including disturbed sleep, increased body weight, and reduced immune functioning (5). Thus, non-pharmacological interventions often utilise social connection such as befriending (6), although this requires resources to coordinate and face-to-face interaction is not always feasible (7). The argument for interventions to increase social connection is aligned to the recent global drives towards person-centred care which encourages a holistic approach to health and wellbeing (8) and provides a theoretical and empirical foundation to promote health and wellbeing from a wider perspective rather than through pharmacological intervention only. Within health and social care, the nursing profession in particular has long reflected such a holistic approach to the health and wellbeing of patients driven by the provision of care and by creating caring relationship between the nurse and the patient (9).
Technological advancement allows for alternative sources of social connection (10). Socially assistive robots (SARs), often resembling a human or pet, work with the help of AI-supported software (or a similar formulation) and are able to interact with humans and form social relationships through a variety of sensors, without the need to touch (11). SARs may provide a potentially cost-effective intervention to promote wellbeing within health and social care settings which acknowledges already limited staff capacity. Whilst SARs were initially perceived as the antithesis of the quality of care provided by nurses (12), perceptions are slowly moving towards SARs as a potential tool in nursing care to provide companionship (13). SARs have been widely applied across the lifespan, although existing research mainly clusters around children and older adults (14). Namely, SARs have been reported to improve independent living, depression and social isolation in older adults (14, 15), and demonstrate effectiveness in managing distress and pain management in children with medical conditions such as cancer (16, 17) or during medical procedures such as venipuncture (18) or intravenous insertion (19). Although intuitively the mechanism of effect of SARs during medical procedures may be distraction, a pilot study of intravenous insertion in paediatric patients found SARs to be effective in reducing pain only when programmed to display empathy in comparison to distraction behaviour (19). Indeed, within other health and social care contexts across the lifespan, observations suggest that SARs exert positive health and wellbeing outcomes through an empathic relationship and emotional bond (10, 14, 20, 21), much like with a human or animal.

To date, there is no comparison of the effectiveness of SARs across the lifespan, as primary studies and existing reviews have largely been conducted in silos according to children and adolescents or older adults. Thus, it is uncertain whether the aforementioned clusters of findings related to effectiveness represent distinct mechanisms of and resulting outcomes according to age group, or rather a dichotomy of research areas that are different across age
groups. For example, it is uncertain whether SARs only reduce anxiety in children, or whether studies that measure the effectiveness of SARs on anxiety levels focus on children only. Given the vast amount of existing literature on SARs within a range of settings and populations, there is a need to synthesise the literature to identify when SARs are effective for which individuals and for which outcomes, through comparison of outcomes, sub-populations, and types of socially assistive robot, of which an umbrella review methodology allows (22). An umbrella review also takes into account methodological quality of included primary studies and reviews, allowing findings to be weighted accordingly (22). Consequently, umbrella reviews have been previously applied to amass evidence on effectiveness of technological interventions on health outcomes (23-25), although this methodology has not yet been applied to assess SARs.

**Methods**

*Research objectives*

Based on the existing literature, the aims of the current umbrella review were to investigate the effectiveness of SARs in improving the health and wellbeing of service users within health and social care contexts, compare their effectiveness with standard care and other available therapies on health and wellbeing outcomes within health and social care contexts, and investigate the interactions between outcomes and other factors as moderators or mediators of any identified effects, including sub-population. Thus, the primary outcomes were the psychosocial, behavioural, and physiological effects on health and wellbeing, and the secondary outcomes were any moderators or mediators of these effects or their interaction with each other. Reporting of subsequent methods adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (see Supplementary Material 1).
Protocol and registration

The protocol was pre-registered prior to the search via the International Prospective Register of Systematic Reviews (PROSPERO, registration number: *(blinded for peer review)*). Deviations from this protocol are outlined within the section ‘Deviations from the registered protocol’ alongside justifications. Searches of the International Prospective Register of Systematic Reviews (PROSPERO), Joanna Briggs Institute Registries, and Open Science Framework confirmed that there were no existing protocols for an umbrella review with a similar scope.

Eligibility criteria

Inclusion criteria was formed around the PICOS (population, intervention, comparator, outcome, study type/setting) criteria (26). Any recipients of the intervention including all ages of the population were included, whilst reviews that reported effects on those not directly targeted by the intervention such as staff, parents, and guardians were excluded. The primary included outcomes were psychosocial (e.g. isolation, depression), behavioural (e.g. sleep, medication adherence), and physiological (e.g. pain, stress) effects on health and wellbeing, the latter included as an objective indicator of psychosocial and behavioural effects. Thus, reviews that only included non-health related psychological outcomes (including memory, cognitive functioning and learning), health related knowledge (e.g. Type 1 diabetes knowledge), emotions that do not directly relate to health (e.g. anger), and satisfaction with the socially assistive robot as outcomes were excluded. The secondary outcomes, namely moderators or mediators of the effect of SARs on health and wellbeing outcomes, or of interactions between outcomes, did not inform the inclusion criteria. Studies including all types of comparator groups were included; there were no exclusion criteria based on comparator. Included interventions were SARs with or without pharmacological
support, defined as artificial intelligence in an embodied and physical form that is able to receive information from humans and respond accordingly, thus excluding virtual reality and voice activated speakers. Inclusion criteria for study type and setting were systematic reviews with or without meta-analysis across any country, of studies within health and social care settings. Criteria for systematic reviews were fulfilment at least two of the following; 1) pre-specified eligibility criteria (e.g registration on the International Prospective Register of Systematic Reviews (PROSPERO) or protocol), 2) systematic methods (i.e PRISMA) that are explicit and reproducible, and 3) risk of bias or quality assessment. Reviews including qualitative studies were excluded unless the results of quantitative studies were presented separately. Where reviews include studies from outside of health and social care contexts, only the findings from within health and social care contexts were included. Conference abstracts were excluded as insufficient reporting would not allow for full data extraction relating to study characteristics, findings, quality, and overlap (27).

Search strategy
Scoping searches were completed prior to the review search to develop the preliminary search strategy (Table 1), which was peer-reviewed by a librarian at Northumbria University using the Peer Review of Electronic Search Strategies checklist (PRESS) (28). A search based on title and abstract was employed on May 4, 2023 (search alerts were created thereafter) on the following databases; the Health Research Premium collection via ProQuest (Consumer Health Database, Health & Medical Collection, Healthcare Administration Database, MEDLINE®, Nursing & Allied Health Database, Psychology Database, and Public Health Database), Scopus, PubMed, Web of Science, ASM Digital Library, IEEE Xplore, Cochrane Reviews within the Cochrane library, and EPISTEMONIKOS. Given that no primary literature on SARs has been identified before 2005, the search as applied across all databases
was limited to 2005 onwards. Only reviews published in English were included, given that the primary reviewers are English language speakers with no translation services available. The overall search strategy was modified accordingly to fit each database’s requirements (see Supplementary Material 2 for the exact search applied for each database). Additionally, forward and backward citation searching of included reviews, and colleagues and other academics provided further sources of reviews. As SARs are an increasingly emergent topic, multiple sources were employed to identify grey literature, including hand searching, OpenGrey, Google Scholar, and preprint servers such as MedRxiv.

Selection of reviews

Search results were first exported via a RIS file into Endnote, where duplicates were recorded and removed. Remaining articles were uploaded onto Covidence (a systematic review management tool) (29) where they were screened against the pre-defined inclusion table by two independent reviewers (blinded for peer review), first for title and abstract, then for full text. For both stages, inter-rater reliability was calculated using Cohen’s kappa statistic (30) and assessed using the conservative parameters from Altman (31). Any discrepancies were resolved through discussion and consultation with a third reviewer if necessary. Screening and resolution of discrepancies were conducted and logged via Covidence.

Data extraction and assessment of methodological quality

The data extraction form (Supplementary Material 3) was built using Cochrane guidance (32) and tailored in accordance with the research question and methodology. Extracted data
included information about the review, characteristics and findings of included primary studies, reports of any meta-analyses or reasoning for not conducting one, indications of included primary study and review quality, and record of any correspondence required and received from review authors. Again, data extraction was performed in duplicate by two independent reviewers (*blinded for peer review*), with any disagreements resolved through discussion. Only data relevant to the aims of the current umbrella review were extracted from included reviews.

Review quality was assessed using the AMSTAR (A MeaSurement Tool to Assess systematic Reviews) 2 (33) and performed in duplicate by two independent reviewers (*blinded for peer review*). Again, any discrepancies were resolved through discussion or consultation with a third reviewer (*blinded for peer review*), if required. The AMSTAR 2 checklist assesses whether the review set a clear research question with a pre-registered protocol and clearly defined inclusion criteria, whether a comprehensive search strategy was employed with independent study selection and extraction by a second reviewer, key information was clearly described, and an appropriate method was selected and applied to synthesise evidence which assessed and accounted for risk of bias. Overall score was calculated using the following criteria; no = -2, yes = 2, and partial yes = 1. Results of AMSTAR 2 assessments were used to inform confidence in included reviews through application of scoring criteria previously applied within a similar umbrella review (23). Namely, AMSTAR 2 items are divided into six critical weaknesses (protocol registration, adequacy of search strategy, risk-of-bias assessment, appropriateness of meta-analytical methods, use of risk of bias during interpretation, and assessment of publication bias) and 10 non-critical weaknesses (the remaining items). Reviews were categorised into high (<3 non-critical weaknesses, 0 critical weaknesses) moderate (<3 non-critical weaknesses 1 critical weakness), low (<3 non-critical weaknesses, >1 critical weakness), or critically low (≥3 non-critical weaknesses, >1 critical weakness).
weakness) confidence. Although quality of included reviews was taken into account when interpreting the results, no reviews were excluded based on quality, retaining a comprehensive overview.

Deviations from the registered protocol

There were no deviations from the registered protocol.

Strategy for data synthesis

Overlap was calculated using the formula for corrected covered area (34), and categorised into slight (0–5%), moderate (6–10%), high (11–15%), or very high (>15%) overlap. Only relevant primary studies to the research question were entered into the overlap table, and one overall overlap statistic was calculated given that included primary studies ranged across the lifespan.

Given the low number of included reviews that applied a meta-analysis, narrative synthesis was applied to synthesise findings of included reviews where a meta-analysis was not conducted. This was assisted by vote counting by direction of effect (35), tested for significance using a two-tailed binomial test with the null assumption of positive effects at a 50% proportion (35). It is noteworthy that vote counting by direction of effect rather than significance was dependent on the reporting of reviews, as the primary study could not be included if reviews reported no significant differences without providing direction of effect. Secondary outcomes of moderators and mediators of the effect of SARs on health and wellbeing outcomes, or of interactions between outcomes, were also reported.

A pooled meta-analysis was applied, including the results of primary studies within the included meta-analyses. A meta-meta-analysis was judged to be inappropriate for two
reasons. Firstly, there was a high proportion of overlap amongst the included primary studies within the included meta-analyses, such that a meta-meta-analysis would inflate the effects of the duplicated primary studies (36). Secondly, during data extraction it was apparent there were major inconsistencies in the reporting of values for primary studies. Namely, the results for the control and intervention groups were often switched, and some values were reported as negative values when the result reported by the primary study was positive, inconsistencies that became apparent due to the variability in reporting of values for the same primary study (even when accounting for multiple timepoints of primary studies). Subsequently, data was extracted from the primary studies to ensure accurate extraction. Nonetheless, a sensitivity analysis was performed using the meta-meta-analysis to demonstrate how the outcome may be affected by the overlap and misreporting. A random effects model was applied for all pooled meta-analyses.

The total effect and heterogeneity indexes were estimated using SPSS v28 (37). Random-effect estimates were based on the effect size calculation from each study and the total estimate of the effect size. In detail, the Profile Likelihood (PL) random-effects model was adopted to estimate the overall effect and to include an estimation of heterogeneity in the weighting (38).

Heterogeneity was determined using the Q-statistic in the \( \chi^2 \) distribution and the p-value (39): a significant p-value indicated that heterogeneity could affect the results. Heterogeneity was further assessed by calculating the \( I^2 \) statistic (40). According to the Cochrane standards, heterogeneity is not important if \( I^2 \) ranges from 0% to 40%, moderate from 30% to 60%, substantial from 50% to 90% and considerable from 75% to 100% (41). The \( \tau^2 \) statistic was also determined, as an estimate of the amount of variation between the included studies.

A funnel plot was included to visually represent the estimation of the treatment’s effect in the studies included in a meta-analysis and to represent a possible publication bias: if a
publication bias exists, the funnel plot is affected by an asymmetrical appearance and the meta-analysis could overestimate the treatment’s effect (42).

Findings of the quantitative analyses and vote counting and narrative synthesis are presented in terms of behavioural, psychosocial, or physiological outcomes. Finally, a figure was created to summarise the strength of the available evidence to support the effect of SARs on each outcome, adhering to good practice for umbrella reviews (43). Findings were triangulated from pooled meta-analysis, meta-analysis of included reviews, vote counting by direction of effect, and narrative synthesis where required. In the absence of standardised guidance (43), each outcome was rated according to amount (well researched (≥10 studies), moderately researched (7-9 studies), or insufficiently researched (<7 studies)), direction (positive, negative, or no overall effect), and consistency (consistent corroborated by quantitative analysis (meta-analysis or pooled meta-analysis), consistent across individual studies (vote counting only), mostly consistent, or inconsistent) of evidence to give an overall assessment of ‘strong positive’ (moderately or well researched, positive, and consistent corroborated by quantitative analysis), ‘weak positive’ (moderately or well researched, positive, and consistent across individual studies or mostly consistent), ‘suggestive positive’ (insufficiently researched, positive, and consistent across individual studies or mostly consistent), or ‘non-significant’ (no overall effect, inconsistent).

**Results**

*Selection of reviews*
The selection process is displayed within the PRISMA (44) diagram (Figure 1). The search identified a total number of 9936 articles, of which 35 reviews were included for analysis in the current umbrella review (see Supplementary Material 4 for references and reason for exclusion for articles excluded based on full text), with ten eligible for quantitative analysis. The main characteristics of included reviews are shown in Supplementary Material 5. The scope of most included reviews were older adults (n = 24), with most of those further specifying people with dementia (n = 13). Only five focused on children (45-49), and six either focused on adults or did not specify inclusion criteria based on population. Seven reviews focused on animal-like SARs (50-56) (two specifically on PARO (55, 56)) and three on humanoid SARs (46, 57, 58) (one of each specifically on Telenoid (57) and NAO (58)). Of those reviews that did not exclude studies based on type of socially assistive robot, PARO was the most common socially assistive robot within the literature, followed by NAO (see Table 2 for a description of the most commonly utilised SARs by primary studies). SARs acted as a companion, distraction, educator, or motivator. Most commonly, SARs adopted a companion role to older adults or hospitalised children during their stay. However, SARs were also utilised as a distraction for children during medical procedures such as IV insertion (48, 59), as educators of disease management such as type 1 diabetes to children (46, 58), and motivators to encourage exercise (60, 61), healthy habits (58), or rehabilitation (58). Frequency of SARs interventions ranged from a single session (45, 46, 49-51, 62-64) to five times weekly (59, 65-67) and over a 15-week period (68), duration from 3 minutes (45, 67) to being there constantly (63, 69), both group and individual formats. Most timepoints of measurement were immediately after exposure, although some follow-ups occurred as long as 8 months (50) after the start of the intervention. SARs interventions that included a control group were most commonly compared to usual care, although other control groups included usual activities such as reading, music, and art therapy, a live dog, and plush toys or SARs.
with their batteries removed. Psychosocial outcomes were the most common outcomes discussed by reviews, specifically depression, anxiety, agitation, and quality of life. Reviews focusing on children were more likely to report on pain and distress, and reviews focused on older adults more likely to discuss agitation, loneliness, apathy, and neuropsychiatric symptoms. 11 reviews only included RCTs (52, 53, 55, 56, 59, 65, 70-74). Searches ran from inception to August 2022 (70, 72), with sample sizes of included primary studies ranging from 2 to 455. Primary studies were published between 2001 and 2021 and conducted across a range of high-income countries including the US, Australia, Japan, Italy, Netherlands, New Zealand, and Denmark.

*Quality of included evidence*

Interrater reliability of quality appraisal was moderate (46.22%). After a consensus was reached through discussion, quality appraisal scores ranged from -26 (the minimum for a systematic review with no meta-analysis) to 13 (see Table 3 for scoring of individual items), with only ten reviews achieving more AMSTAR-2 checklist items than not (a positive overall score). Most reviews conducted independent screening (85%), although independent extraction was less frequent (47%). No reviews reported on the funding source of included primary studies, and only one provided a list of excluded studies based on full text screening or explained their inclusion criteria for study design. Very few reviews used a comprehensive literature search (15%), with many reviews scored negatively for not explaining language restrictions. Also, few reviews considered the influence of risk of bias when analysing (29%) or discussing (26%) results, with the few reviews that did achieve this item excluding studies based on risk of bias. Subsequently, confidence in all included reviews was rated as critically low, aside from four which achieved low confidence (49, 62, 67, 75), and one which achieved moderate confidence (69).
All but six reviews (46, 48, 57, 58, 60, 63) evidenced assessment for quality or risk of bias of included primary studies using established tools. Most commonly, risk of bias was assessed using Cochrane Risk of Bias tools (47, 51-53, 56, 59, 66, 69-74, 76). The most commonly applied quality appraisal tools included the Critical Appraisal Skills Programme (CASP) checklist (50), Downs and Black checklist (49, 68, 77), Joanna Briggs Institute tools (55, 75, 78), Mixed Methods Appraisal Tool (67, 79), and Physiotherapy Evidence Database (PEDro) scale (54, 65). A key confounding factor to accurate quality and risk of bias was underreporting and missing information within many of the included primary studies (51, 52, 55, 71, 75, 78), meaning many items were scored as ‘unclear’. Only 10 reviews reported mostly moderate to high quality primary studies relevant to the current review (45, 49, 50, 53-55, 61, 65, 68, 79), three of which excluded based on quality (50, 54, 65). Quality of primary studies was most commonly mixed, ranging from mostly high quality (61) to mostly high risk of bias (66).

Study overlap
Corrected covered area showed only slight overlap across primary studies (2.9%). However, overlap was particularly high within included meta-analyses, likely as meta-analyses were only performed within the sub population of older adults. For example, 10 meta-analyses for depression had been performed utilising a total of 10 primary studies.

Results of quantitative analysis
Psychosocial outcomes
All pooled meta-analyses included older adults living with or without dementia and related to psychosocial outcomes. Data on 469 individuals pooled from ten primary studies (Figure 2)
showed no significant impact of socially assistive robot intervention on depression (standard mean difference (SMD) = 0.21, 95% CI [-0.26, 0.68], Z = 0.86, p = 0.39), with considerable heterogeneity ($I^2 = 83\%$) (references to the primary studies included in the pooled meta-analyses are included in Supplementary Material 6). Results indicate one study (Peterson et al., 2017, in Supplementary Material 6) reporting a statistically significant difference to be a questionable study. For the anxiety outcome, data on 725 individuals were pooled from seven studies (Figure 3). The results showed no significant effect of socially assistive robot intervention (SMD = -0.02, 95% CI [-1.32, 1.27], Z = -0.04, p = 0.97), with extreme heterogeneity ($I^2 = 98\%$). Data on 447 individuals pooled from seven studies (Figure 4) showed no significant effect of socially assistive robot intervention on quality of life (SMD = 0.43, 95% CI [-0.12, 0.98], Z = 1.52, p = 0.13), with considerable heterogeneity ($I^2 = 87\%$). Again, meta-analysis results indicate one primary study (Moyle et al., 2013, in Supplementary Material 6) to be questionable. Asymmetrical funnel plots for depression, anxiety, and quality of life indicate potential publication bias (See Figure 5).

For outcomes not suitable for pooled meta-analyses, results of each meta-analysis are provided in Supplementary Material 7. Meta-analyses consistently reported small but significant effects of SARs on social interaction and loneliness, and most reviews reported a small significant effect on positive affect. Meta-analyses did not support an effect on apathy or neuropsychiatric symptoms, and there were mixed results regarding a significant effect on agitation.

Results of vote counting and narrative synthesis

Psychosocial outcomes
The findings from vote counting by direction of effect are shown in Table 4. Loneliness, social interaction, mood, and positive affect all showed strong evidence to support an effect, regardless of effect size. Furthermore, the effect of SARs on anxiety was significant, although a minority of primary studies reported an increase in anxiety after exposure to SARs. Meanwhile, the evidence for depression, agitation, neuropsychiatric symptoms, apathy, and quality of life was inconsistent, indicating a weak level of evidence to support the effect of SARs on these outcomes. Results for motivation, and knowledge of one’s disease were promising but insufficient to reach statistical significance. Assessment of non-significant findings reported without detail of the direction of effect support vote counting results, namely that there are numerous non-significant findings of primary studies of effects on quality of life, neuropsychiatric symptoms, and agitation.

**Physiological outcomes**

Whilst pain showed strong evidence to support an effect of SARs, the little available evidence indicated an effect on distress although was insufficient to reach statistical significance. Although the findings from vote counting indicated strong evidence for SARs in decreasing self-reported stress, there was insufficient evidence to support the efficacy of SARs in reducing biochemical indicators of stress. There was too little evidence to apply vote counting to Body Mass Index although the available evidence indicated no significant effect (62, 74).

**Behavioural outcomes**

Sleep, medication use, and activity all showed inconsistent evidence and thus indicated weak support for the effect of SARs on behavioural outcomes. Adherence to medication possessed too little evidence to apply vote counting although indicated no significant effect (74), and one primary study reported an increase in activities of daily living (75). Although the direction of effect was not provided, numerous non-significant effects on number of
medications were reported. As a secondary outcome, one review reported on four studies that identified an improvement in disease management (46). Triangulation from vote counting and quantitative analysis findings are displayed in Table 4.

Results according to quality of included reviews and population subgroups

There was a trend within one review that lower quality studies were more likely to report an effect regarding agitation, anxiety, depression, and quality of life (67). No other reviews assessed for an effect of quality of primary studies. Although some effects including reduced pain were retained across different measurement tools (59), one review reported a significant effect on agitation only through video observation and not using the Cohen-Mansfield Agitation Inventory Short Form (53). Similarly, larger effects on depression were observed when the Cornell Scale for Depression in Dementia was applied within adults with dementia, (71) and when the Geriatric Depression Scale was applied within adults without dementia, (69), although both reviews reported no effect at follow-up (69, 71).

Whilst SARs reduced the need for pain and psychotropic medication (51, 56, 68, 74, 75), no effect was found for medication for sleep (51, 56, 74), depression (51, 56, 74), or dementia (75). Similarly, whilst there was evidence that SARs reduced the neuropsychiatric symptoms of delusions (50, 58) and irrational beliefs, SARs also reportedly increased hallucinations (64, 68). Reviews reported on different aspects relating to sleep suggested a significant decrease in hours of daytime sleep (56, 65) and nighttime activity (53, 56, 70, 73, 75), although a decrease in daytime activity too (53), perhaps explaining mixed results identified within vote counting. Regarding biochemical indicators of stress, one primary study reporting a decrease in pulse oximetry and pulse rate was identified as questionable within quantitative analysis of independent outcomes (see ‘results of quantitative analysis).
There was some evidence of a greater reduction in depression for adults without dementia compared to those with dementia from the review with the lowest risk of bias (69), possibly linked to the findings that dementia (58) and behaviour or mood problems (62) negatively influenced socially assistive robot interactions. However, dementia measurements were either self-reported or rated by intervention staff, which could sometimes conflict (75). There were some instances of SARs as most effective for those with the most severe symptoms, including for loneliness (60) and psychotropic medication use for those with severe rather than mild or moderate dementia (56). Whilst there was no effect of SARs and anxiety overall, all reviews within children specifically reported a reduction in anxiety (47-49), which were mostly significant. Contrastingly, the reduction in pain was identified both within samples of children (48, 49, 59, 62) and as reducing neuropathic pain in adults (59). Regarding cultural differences as a moderator of engagement with SARs, Italian (46) and Japanese (62) children were more engaged with SARs than Dutch (46) and Serbian (62) children respectively, and engagement was amplified for American older adults (60).

SARs encouraged social interaction with the socially assistive robot itself (50), but also other participants (50, 60, 62-64), setting staff (48, 63, 64, 69, 73), family and friends (48, 60), and the deliverer of the intervention (48, 50, 60). Whilst some reviews reported no effect of SARs interventions delivered as a group compared to individually (53, 75) including on the effects of SARs on depression (76), group interventions were more effective for reducing anxiety in people with dementia (76), the only study in one review that found direction of an effect on depression sampled women living alone (78), and the presence of a parent reduced pain (48, 49) and anxiety (48) compared to the child interacting with the socially assistive robot alone.

One confounder to this potential moderator is that adherence to the socially assistive intervention was lower when using PARO at home (67). Few reviews reported effect of type of socially assistive robot, although one reported no effect on depression and anxiety
outcomes (76). Whilst length of each session and total weekly exposure time to the socially assistive robot positively predicted a reduction in depression, frequency of sessions and overall duration of the intervention negatively impacted depression scores (73), indicating longer and fewer sessions as optimal. In contrast, one review suggested two to three 15-20 minute sessions weekly to avoid participants losing focus and as similar levels of interaction were reported across differing numbers of sessions (57). Duration may also be a confounding factor, given that one review reported that sessions for children tended to be longer than sessions for older adults with dementia (48).

There was generally inconsistency in what was described as usual care within primary studies (76), making comparison difficult. When compared to a live dog, SARs generally showed no difference in their effectiveness on agitation, depression, quality of life (54), or anxiety (49). When compared to virtual reality, SARs were reported as more effective for reducing anxiety but not depression, of which there was no significant difference (77). Perhaps explicably, the relationship built with a socially assistive robot was stronger than with a virtual avatar, which in turn predicted increased use (46, 48). For reviews that compared the effectiveness of SARs separately for comparison groups, reviews reported improvements in health outcomes compared to usual care (51, 68, 79) but not a plush toy (51, 68, 79) or live dog (50). Perhaps relatedly, literature was mixed as to whether SARs elicited increased engagement compared to a plush toy (48, 67, 79) or live dog (51, 62, 73) or no difference compared to a plush toy (51, 60, 62) or live dog (50, 67), with a minority of evidence to suggest increased engagement compared to usual care (51).

SARs generally demonstrated a good acceptability across the lifespan (46, 49, 58, 62, 64) including being engaging and enjoyable, and participants felt they were non-judgemental (58, 63). Furthermore, some reviews reported a preference of SARs or increased engagement to humans (61, 62, 67), particularly within those with Autism Spectrum Disorder (58, 62).
Technical issues were a barrier to SARs interventions (57, 67, 69), including issues with movement (57), SARs perceived as childish or making strange noises (55, 67), and internet connection and sufficient training to operate socially assistive robot as necessary (57). Furthermore, some reviews reported issues with accessibility (57, 58, 69), including SARs as less appropriate for those with intellectual disability and hearing or vision impairment (58).

**Discussion**

A summary of strength of evidence for each outcome is displayed in Table 5. There is mixed evidence surrounding psychosocial outcomes; whilst there is promising evidence to suggest an effect of SARs on loneliness, social interaction, mood, and positive affect, more research is needed to validate effects on motivation, and knowledge of one’s disease. The available evidence indicates no effect on depression, quality of life, agitation, apathy, or neuropsychiatric symptoms. Whilst quantitative analysis suggests no effect on anxiety, comparison across sample groups indicate an effect of SARs on anxiety in children but not older adults. The evidence for physiological outcomes is promising specifically for pain and stress, although more research is required to validate effects particularly for distress and biochemical indicators of stress such as pulse oximetry and cortisol. There is limited evidence to support an effect of SARs on behavioural outcomes, indicating no effect on sleep, medication use, and activity.

Overall, the effects of SARs observed relate to non-psychiatric outcomes including positive affect, loneliness, and mood, and an inability of SARs to treat psychiatric symptoms. This pattern is generally reflected in the available literature on animal-assisted therapy (80), with those finding an improvement in psychiatric symptoms may be attributed to a holistic effect of animal-assisted therapy within nature (81). Although the effect of animal-assisted therapy on loneliness is more mixed than for SARs (80, 82), the positive impact of both SARs and
animal-assisted therapy on loneliness is not mediated by attachment (82). As SARs and animal-assisted therapies exert similar effects through similar mechanisms, the decision of intervention should be dependent on their relevant costs. It is also noteworthy that primary studies to support both interventions are inherently biased by virtue that only participants who like, accept, or are not fearful of the socially assistive robot (50, 55) or animal (81) are included. This is a particular issue for SARs given that some reviews identified a lack of acceptability (63), including that some residents showed fear or disgust that did not dissipate with repeated interaction (57). Furthermore, given that similar effects were observed between SARs and plush toys, it is questionable whether an interaction with participants is required at all to deliver the same effects. Thus, a cost-benefit analysis is needed to explore whether the significant but small in magnitude effects of SARs justify their price.

Findings also suggest an influence of the presence of dementia and other psychological disorders, possibly mediating or moderating the effect of stive robots on health and wellbeing. This is particularly noteworthy given that most of the available literature within adults includes people with dementia. It is conceivable that cognitive impairment may predict different interactions with SARs, and a distortion of reality observed in people with dementia may affect the ‘uncanny valley’ - by which uncomfortable feelings arise when interacting with something resembling but not quite reaching a human form. Whilst there is no available literature on ‘uncanny valley’ within cognitive impairment, people with Autism Spectrum Disorder have been found to show a higher threshold to its effect (83). Thus, further exploration is needed on the effect of SARs within adults without cognitive impairment.

Although a mechanism of SARs on mood related outcomes was not formally explored by included reviews, findings suggest that SARs stimulate social interaction, either with staff, other participants, or parents of children. The established positive effects of social interaction on wellbeing (84) supports social interaction as a feasible mediator. However, SARs have
also been reported as effective for increasing wellbeing in individual settings, where the socially assistive robot may act as the companion, rather than stimulating companionship with others within group sessions. This would align with literature that finds attachment to SARs does not account for the improvement in loneliness in nursing homes (82). Subsequently, a mediation analysis would be useful to confirm this suggested mechanism in group settings, and further explore the mechanism within individual interventions.

**Strengths and limitations**

This umbrella review possesses several strengths and limitations. No additional studies were added after extensive searching of other sources outside of the search, indicating a comprehensive search strategy. An umbrella review methodology allowed us to assess the current literature in this area from a high-level perspective, identifying the copious amounts of reviews within older adults which have overstated the low number of primary studies, and errors in reporting these primary studies that could significantly mislead their data analyses and subsequent interpretation. The errors identified highlight an urgent need for improved accuracy of reporting of primary studies, raising an important issue of building in error within reviews and umbrella reviews that indicate a need for vigilance. Limitations that affect the generalisability of findings include the English language limitation of the search and that all primary evidence is from high income countries. The findings of the current review are also limited by the considerable risk of bias of included reviews and the overlap of primary studies across meta-analyses, as reviews particularly within older adults do not add anything to existing knowledge. Additionally, quantitative analysis identified one questionable primary study that should be taken with caution when interpreting findings.
Conclusion

SARs are effective on mood related outcomes and encourage social interaction, although the magnitude of these effects is small and may be comparable to considerably more cost-effective plush toys. Within group settings, positive effects appear to be exerted through an increase in social interaction with staff, friends and family, and other participants, although a formal mediation analysis is required to confirm this. Thus, the main mechanism of SARs appears not to replace social connection, but to encourage it amongst other humans. Future research would first benefit from agreeing a consensus definition of SARs to ensure consistent comparison. More research is required to explore the effect of SARs on physiological outcomes, and the evidence of SARs on behavioural outcomes is generally weak. Future research should therefore quantitatively assess the overall effect of SARs on physiological outcomes through meta-analysis, as the current magnitude of the identified positive effects aggregated across primary studies is uncertain. More broadly, the current umbrella review identified that although the literature on pain, anxiety, and distress in children is promising, a meta-analysis is needed to explore the magnitude of effects.

Funding

This study is funded by Northumbria University (Project ID: 502620).

Conflict of interest

The authors declare no conflict of interest.
References

72. Hsu TC, Hsieh CJ. Socially Assistive Robots Assisting Older Adults in an Internet and Smart Healthcare Era: A Literature Review. 2022:591-4.
**CRediT statement:** Bethany Nichol: Writing-Original draft preparation, Investigation, Formal Analysis; Jemma McCready: Writing- Review and Editing, Investigation, Formal Analysis; Goran Erfani: Investigation, Formal Analysis; Dania Compacini: Conceptualization, Writing- Review and Editing; Valentina Simonetti: Conceptualization, Writing- Review and Editing; Giancarlo Cicolini: Supervision, Conceptualization, Writing- Review and Editing; Kristina Mikkonen: Writing- Review and Editing, Software, Methodology; Miyae Yamakawa: Writing- Review and Editing, Validation; Marco Tomietto: Conceptualization, Methodology, Writing- Review and Editing, Resources, Project Administration, Funding acquisition.

**Authors’ contributions:** The authors have checked to make sure that our submission conforms as applicable to the Journal’s statistical guidelines. The authors affirm that the methods used in the data analyses are suitably applied to their data within their study design and context, and the statistical findings have been implemented and interpreted correctly. The authors agree to take responsibility for ensuring that the choice of statistical approach is appropriate and is conducted and interpreted correctly as a condition to submit to the Journal. A statistician is part of the research team.

**Conflict of Interest Statement:** The authors have no conflicts of interest to declare.
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
Figure 1: PRISMA diagram delineating the article selection process for the current systematic review.

Figure 2: Forest plot for pooled meta-analysis for the outcome depression.
Figure 3: Forest plot for pooled meta-analysis for the outcome Anxiety. SeR = self-reported anxiety, Str = staff reported anxiety, PT = when compared to a plush toy, UC = when compared to usual care.

Figure 4: Forest plots for pooled meta-analysis for the outcome Quality of life. SeR = self-reported anxiety, Str = staff reported anxiety. C = when compared to the control group, D = when compared to a live dog.
Figure 5: Funnel plots for pooled meta-analysis for the outcomes depression, anxiety, and quality of life.
Table 1: Search strategy

<table>
<thead>
<tr>
<th>SARs</th>
<th>Outcome</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>“social robot***” OR “socially assistive robot***” OR “socially assistive humanoid robot***” OR humanoid* OR “artificial intelligence robot***” OR “humanoid assistive robot***” OR robopet* OR “robotic pet***” OR “robotic animal***” OR “pet robot***” OR aibo OR paro OR pepper OR hyodol OR silbot OR pleo OR huggable OR iCat</td>
<td>“quality of life” OR QoL OR happiness OR social* OR mortality OR morbidity OR survival OR aging OR function* OR physical OR satisfaction OR self-esteem OR purpos* OR depression OR isolation OR loneliness OR anxiety OR pain OR distress* OR fear* OR worry* OR psychosocial OR psychological OR physiological OR health* OR wellbeing OR well-being OR “negative emotion***” OR mental* OR rehabilitation OR recover* OR dementia OR alzheimer* OR symptoms OR stress OR disease* OR cancer OR diabetes OR ASD OR “autism spectrum disorder” OR sleep OR “medica* adherence” OR “medication compliance” OR “activities of daily living” OR ADL OR ADLs OR “end of life care” OR “palliative care”</td>
<td>systematic* OR review OR meta-analysis* OR metaanalysis* OR “meta analysis***” OR meta-synthesis OR metasynthesis OR Synthesis OR Literature OR meta-review OR meta-analytic OR meta-narrative OR meta-study</td>
</tr>
</tbody>
</table>

Table 2: Description of the most commonly included socially-assistive robots.

<table>
<thead>
<tr>
<th>Socially assistive robot</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humanoid</td>
<td></td>
</tr>
<tr>
<td>NAO</td>
<td>Boxy bipedal robot with shiny accents and rounded features resembling an action figure. Able to walk and talk. Height of around 57.3 cm.</td>
</tr>
<tr>
<td>Telenoid</td>
<td>Human-like torso and face with partial limbs. Able to talk. Height ranges from 50-80 cm across models.</td>
</tr>
<tr>
<td>Animal</td>
<td></td>
</tr>
<tr>
<td>PARO</td>
<td>Robotic seal with soft fur resembling a plush toy. Able to move and blink.</td>
</tr>
<tr>
<td>PLEO</td>
<td>Robotic colourful dinasoar resembling a toy. Able to move and walk.</td>
</tr>
<tr>
<td>AiBO</td>
<td>A series of robot dogs with no fur. Able to move and walk.</td>
</tr>
<tr>
<td>NeCoRo</td>
<td>Robotic tabby cat with fur and collar resembling a plush toy. Able to move.</td>
</tr>
<tr>
<td>Huggable</td>
<td>Robotic soft and colourful bear resembling a plush toy. Able to move and talk.</td>
</tr>
</tbody>
</table>

Table 3: AMSTAR quality ratings and an overall quality judgment for each review.

| Study (citation number) | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | Q12 | Q13 | Q14 | Q15 | Q16 | Overall Score |
|-------------------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|----------------|
| Aarskog 2019            | Y  | N  | N  | N  | N  | Y  | Y  | N  | Y  | P   | Y   | N   | N/A | N/A | Y   | Y   | Y/N/A          | 1              |
| Abbott 2019             | N  | N  | N  | Y  | Y  | Y  | N  | Y  | Y  | N   | Y   | N   | N   | N   | N   | N/Y/A          | 0              |
| Anjum 2021              | N  | N  | N  | Y  | N  | N  | P  | Y  | N  | N   | N/A | N/A | N   | N/A | N   | Y   | N/A            | -11            |
| Chan 2021               | Y  | N  | N  | N  | Y  | N  | Y  | N  | Y  | N   | N/A | N/A | N   | N   | Y   | N   | N/              | -6             |
| Name            | Y | N | N | N | Y | Y | N | N | Y | N | N | N | N | Y | Y | Y | N | N | N | A | A | A | A | A | N | N | N | Y | -2 |
| Cheng 2021      | Y | N | N | N | Y | Y | N | Y | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y | -3 |
| Chen 2018       | N | Y | N | N | Y | Y | N | Y | P | Y | Y | N | N | N | N | N | Y | Y | N | N | N | A | A | A | A | A | N | N | N | A | N | -11|
| Demiralay 2020  | N | N | N | N | Y | N | N | P | Y | Y | N | N | N | N | N | N | Y | N | N | Y | N | N | N | N | A | A | A | A | A | N | A | -1|
| He 2023         | Y | Y | N | Y | Y | N | P | Y | Y | N | N | N | N | N | Y | N | Y | N | N | N | Y | N | A | A | A | A | A | N | N | N | N | -12 |
| Hirt 2021       | Y | Y | N | N | N | Y | N | Y | N | Y | N | N | N | N | Y | Y | N | N | Y | N | N | N | A | A | A | A | A | Y | Y | N | Y | -13 |
| Hsueh 2023      | Y | P | Y | N | N | Y | N | N | N | Y | N | Y | N | N | N | N | N | N | Y | Y | N | Y | N | Y | A | A | A | A | A | N | N | -14 |
| Hsu 2022        | N | N | N | N | N | N | N | N | Y | N | N | N | N | N | Y | Y | Y | Y | N | N | Y | N | N | N | A | A | A | A | A | N | N | -15 |
| Jones 2020      | Y | Y | N | N | Y | N | P | P | P | Y | N | N | N | N | Y | N | N | N | N | N | N | Y | N | N | A | A | A | A | A | N | N | -16 |
| Kachouie 2014   | N | N | Y | Y | Y | N | P | P | P | Y | N | N | N | N | N | N | Y | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -17 |
| Kang 2020       | N | N | N | Y | N | Y | N | P | P | P | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -18 |
| Koroupa 2022    | Y | Y | N | N | Y | N | N | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -19 |
| Lau 2020        | Y | N | N | N | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -20|
| Leng 2019       | Y | Y | N | N | Y | Y | N | P | Y | Y | N | Y | N | Y | N | N | N | N | Y | N | Y | N | N | N | A | A | A | A | A | N | N | -21|
| Littler 2021    | Y | N | N | Y | Y | Y | N | P | Y | P | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -22|
| Loveys 2022     | Y | Y | N | P | Y | Y | Y | N | Y | Y | N | N | N | N | N | N | N | N | N | Y | Y | N | N | N | A | A | A | A | A | N | N | -23|
| Lu 2021         | Y | N | N | N | Y | Y | N | Y | Y | Y | N | Y | Y | Y | Y | Y | Y | Y | Y | N | N | N | N | N | A | A | A | A | A | N | N | -24|
| Moerman 2019    | N | N | N | N | Y | N | N | P | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -25|
| Moyle 2022      | N | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -26|
| Ong 2021        | Y | N | N | N | Y | N | N | N | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -27|
| Park 2020       | N | N | N | P | Y | Y | N | N | N | Y | Y | N | Y | N | Y | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -28|
| Pu 2019         | Y | Y | N | N | Y | Y | N | Y | Y | Y | N | N | N | N | Y | Y | Y | Y | N | Y | Y | N | N | N | A | A | A | A | A | N | N | -29|
| Riches 2021     | N | Y | N | P | Y | N | N | N | P | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -30|
| Robaczewski 2020| N | N | N | N | N | N | N | N | P | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -31|
| Saragih 2021    | Y | Y | N | Y | N | Y | Y | N | Y | Y | N | N | N | Y | Y | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -32|
| Scoglio 2019    | N | N | N | N | Y | N | N | P | Y | N | N | N | N | N | N | N | Y | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -33|
| Store 2022      | Y | P | Y | N | N | Y | N | N | Y | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -34|
| Trost 2019      | N | Y | N | Y | N | Y | Y | Y | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -35|
| Vandemeulbroeke 2021 | N | N | Y | Y | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -36|
| Vogan 2020      | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | A | A | A | A | A | N | N | -37|
| Year  | Y | N | N | N | N | N | Y | Y | N | N | N | N | Y | -17 |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|
| Wang 2021 | | | | | | | | | | | | | | |     |
| Yu 2022 | N | Y | N | Y | Y | Y | N | N | Y | Y | Y | Y | N | Y | 8   |

Q1: Did the research questions and inclusion criteria for the review include the components of PICO?
Q2: Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol review?
Q4: Did the review authors use a comprehensive literature search strategy?
Q5: Did the review authors perform study selection in duplicate?
Q6: Did the review authors perform data extraction in duplicate?
Q7: Did the review authors provide a list of excluded studies and justify the exclusions?
Q8: Did the review authors describe the included studies in adequate detail?
Q9: Did the review authors use a satisfactory technique for assessing the risk of bias in individual studies that were included in the review?
Q10: Did the review authors report on the sources of funding for the studies included in the review?
Q11: If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?
Q12: If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?
Q13: Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?
Q14: Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?
Q15: If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?
Q16: Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?
Table 4: Vote counting by direction of effect rather than statistical significance, where reported within included reviews. Code for describing ‘strength of evidence’: all (or above 90%) in favour and significant; Strong, most in favour and significant; Moderate, all in favour but non-significant; insufficient evidence, highly mixed; Inconsistent findings.

<table>
<thead>
<tr>
<th>Coding of outcome</th>
<th>Outcome</th>
<th>Review findings</th>
<th>Vote counting evidence to support benefit</th>
<th>Strength of meta-analysis evidence to support benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychosocial</td>
<td>Depression</td>
<td>23 reviews with 26 unique studies were identified. Findings were inconsistent (18/26; 69%, p = .076)</td>
<td>Inconsistent</td>
<td>Pooled meta-analysis indicated no significant effect of SARs</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>22 reviews with 21 unique studies were identified. A statistically significant majority of studies supported a positive effect (16/21; 76%, p = .027)</td>
<td>Moderate</td>
<td>Pooled meta-analysis indicated no significant effect of SARs</td>
</tr>
<tr>
<td></td>
<td>Agitation</td>
<td>18 reviews with 12 unique studies were identified. Findings were inconsistent (8/12; 67%, p = .388)</td>
<td>Inconsistent</td>
<td>10 meta-analyses, 40% significantly favoured SAR intervention</td>
</tr>
<tr>
<td></td>
<td>Quality of life</td>
<td>17 reviews with 20 unique studies were identified. Findings were inconsistent (14/20; 70%, p = .115)</td>
<td>Inconsistent</td>
<td>Pooled meta-analysis indicated no significant effect of SARs</td>
</tr>
<tr>
<td></td>
<td>Apathy</td>
<td>14 reviews with 7 unique studies were identified. Findings were inconsistent (4/7; 57%, p = 1.000)</td>
<td>Inconsistent</td>
<td>2 meta-analyses, 0% significantly favoured SAR intervention</td>
</tr>
<tr>
<td></td>
<td>Neuropsychiatric symptoms</td>
<td>9 reviews with 7 unique studies were identified. Findings were inconsistent (4/7; 57%, p = 1.000)</td>
<td>Inconsistent</td>
<td>6 meta-analyses, 0% significantly favoured SAR intervention</td>
</tr>
<tr>
<td></td>
<td>Loneliness</td>
<td>6 reviews with 7 unique studies were identified. All studies supported a positive effect, which was statistically significant (7/7; 100%, p = .016)</td>
<td>Strong</td>
<td>1 meta-analyses, 100% non-significant difference favouring the intervention</td>
</tr>
<tr>
<td></td>
<td>Social interaction</td>
<td>10 reviews with 13 unique studies were identified. All studies supported a positive effect (10/13; 77%, p = .023)</td>
<td>Strong</td>
<td>3 meta-analyses, 100% significantly favoured SAR</td>
</tr>
<tr>
<td>Category</td>
<td>Supported Effect</td>
<td>Evidence Level</td>
<td>Additional Notes</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------------</td>
<td>----------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>Mood</td>
<td>supported a positive effect, which was statistically significant (13/13; 100%, p &lt; .001)</td>
<td>Strong</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Positive affect</td>
<td>11 reviews with 10 unique studies were identified. All studies supported a positive effect, which was statistically significant (10/10; 100%, p = .002)</td>
<td>Strong</td>
<td>3 meta-analyses, 67% significantly favoured SAR intervention</td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>2 reviews with 2 unique studies were identified. All studies supported a positive effect, although non-significant (2/2; 100%, p = .500)</td>
<td>Insufficient evidence</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Disease knowledge</td>
<td>2 reviews with 2 unique studies were identified. All studies supported a positive effect, although non-significant (2/2; 100%, p = .500)</td>
<td>Insufficient evidence</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Behavioural Sleep</td>
<td>3 reviews with 5 unique studies were identified. Findings were inconsistent (3/5; 60%, p = 1.00)</td>
<td>Inconsistent</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Medication use</td>
<td>9 reviews with 5 unique studies were identified. Findings were inconsistent (2/5; 40%, p = 1.000)</td>
<td>Inconsistent</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>4 reviews with 6 unique studies were identified. Findings were inconsistent (4/6; 67%, p = .688)</td>
<td>Inconsistent</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Physiological Pain</td>
<td>4 reviews with 7 unique studies were identified. All studies supported a positive effect, which was statistically significant (7/7; 100%, p</td>
<td>Strong</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Number of reviews/unique studies</td>
<td>Summary of findings</td>
<td>Evidence strength</td>
<td>Notes</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Distress</td>
<td>2/3</td>
<td>All studies supported a positive effect, although non-significant (3/3; 100%, p = .250)</td>
<td>Insufficient evidence</td>
<td>N/A</td>
</tr>
<tr>
<td>Biochemical indicators of stress</td>
<td>8/5</td>
<td>All studies supported a positive effect, although non-significant (5/5; 100%, p = .063)</td>
<td>Insufficient evidence</td>
<td>N/A</td>
</tr>
<tr>
<td>Stress</td>
<td>5/7</td>
<td>All studies supported a positive effect, which was statistically significant (7/7; 100%, p = .016)</td>
<td>Strong</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 5: A summary of the available evidence for each outcome. Ticks represent the sample groups of which there is data to support an effect. Anxiety is the only outcome explored across the lifespan but only effective within children.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Outcome</th>
<th>Explanation</th>
<th>Older adults (55+)</th>
<th>Children (&lt;18)</th>
<th>Lifespan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychosocial</td>
<td>Depression</td>
<td>Well researched, inconsistent data suggesting no overall effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>Well researched, inconsistent data suggesting only an effect within children</td>
<td></td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agitation</td>
<td>Well researched, inconsistent data suggesting no overall effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quality of life</td>
<td>Well researched, inconsistent data suggesting no overall effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Apathy</td>
<td>Moderately researched, inconsistent data suggesting no overall effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neuropsychiatric symptoms</td>
<td>Moderately researched, inconsistent data suggesting no overall effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loneliness</td>
<td>Moderately researched, consistent evidence to suggest a positive effect</td>
<td></td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social interaction</td>
<td>Well researched, consistent evidence to suggest a positive effect</td>
<td></td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mood</td>
<td>Well researched, consistent evidence from individual studies to suggest a positive effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive affect</td>
<td>Well researched, most evidence suggests a positive effect</td>
<td></td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Motivation</td>
<td>Insufficiently researched, available individual studies suggest a positive effect</td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>Disease knowledge</td>
<td>Insufficiently researched, available individual studies suggest a positive effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioural</td>
<td>Sleep</td>
<td>Insufficiently researched, inconsistent data suggesting no overall effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medication use</td>
<td>Insufficiently researched, inconsistent data suggesting no overall effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Activity</td>
<td>Insufficiently researched,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Effect Description</td>
<td>Significance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physiological</td>
<td>available individual studies suggest no overall effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td>Moderately researched, consistent evidence from individual studies to suggest a positive effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>Moderately researched, consistent evidence from individual studies to suggest a positive effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distress</td>
<td>Insufficiently researched, available individual studies suggest a positive effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biochemical indicators of stress</td>
<td>Insufficiently researched, available individual studies suggest a positive effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Strong positive**
- **Weak positive**
- **Suggestive positive**
- **Non-significant**