

Review

# Care Robots as Emerging Health Technologies: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Gaeun Kim<sup>1</sup>, PhD; Jinmi Jeon<sup>2,3</sup>, MSN

<sup>1</sup>Department of Nursing, College of Nursing, Keimyung University, Daegu, Republic of South Korea

<sup>2</sup>Graduate School, Department of Nursing, Keimyung University, Daegu, Republic of South Korea

<sup>3</sup>Department of Nursing, Keimyung University Dongsan Hospital, Daegu, Republic of Korea

**Corresponding Author:**

Gaeun Kim, PhD

Department of Nursing

College of Nursing, Keimyung University

1095, Dalgubeol-daero, Dalseo-gu

Daegu 42601

Republic of South Korea

Phone: 82532587658

Email: [eun0325@kmu.ac.kr](mailto:eun0325@kmu.ac.kr)

## Abstract

**Background:** Global aging and health care workforce shortages are increasing demand for therapeutic support among older adult and pediatric populations. Care robots, an umbrella term for socially assistive robots, companion and pet platforms, and therapeutic humanoids sharing a social interaction component, have been proposed as embodied digital health adjuncts, but prior syntheses have remained population- or platform-specific, leaving cross-population and cross-platform effects unquantified.

**Objective:** This review quantifies the pooled effects of care-robot interventions across 7 prespecified outcome domains, examines robot platform, target population, and intervention-duration moderators, and grades evidence certainty.

**Methods:** Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020, we searched 5 databases (PubMed, including MEDLINE, Embase, Cochrane CENTRAL, CINAHL, APA PsycINFO) and 2 trial registries (ClinicalTrials.gov; World Health Organization International Clinical Trials Registry Platform) through April 29, 2026, with no language restriction. Eligible studies were randomized controlled trials of embodied care robots versus standard care, active controls, or waitlist. Random-effects meta-analysis used Hedges  $g$  with Hartung-Knapp-Sidik-Jonkman correction; for domains with  $k \geq 4$ , we report 95% CIs and prediction intervals (PIs). Risk of bias used the Cochrane Risk of Bias 2 tool; certainty used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.

**Results:** A total of 34 randomized controlled studies ( $n=2476$ ; 17 countries; 2015-2024) met inclusion criteria; 20 contributed to at least 1 meta-analysis, and 14 entered narrative synthesis. Under Hartung-Knapp-Sidik-Jonkman pooling, only neuropsychiatric symptoms reached significance: Hedges  $g=0.44$  (95% CI 0.03-0.84; 95% PI  $-0.42$  to 1.30;  $k=6$ ;  $P=.04$ ). The remaining 6 domains were favorable but nonsignificant: quality of life Hedges  $g=0.15$  (95% CI  $-0.41$  to 0.71; 95% PI  $-0.86$  to 1.16;  $k=5$ ); depression Hedges  $g=0.20$  (95% CI  $-0.08$  to 0.49; 95% PI  $-0.36$  to 0.76;  $k=7$ ); agitation Hedges  $g=0.32$  (95% CI  $-0.07$  to 0.71; 95% PI  $-0.26$  to 0.89;  $k=4$ ); stress and pain Hedges  $g=0.53$  (95% CI  $-0.48$  to 1.53; 95% PI  $-1.57$  to 2.62;  $k=6$ ); social-communicative skills Hedges  $g=0.45$  (95% CI  $-0.52$  to 1.42; 95% PI  $-1.66$  to 2.56;  $k=5$ ); and cognitive function Hedges  $g=0.18$  (95% CI  $-0.62$  to 0.98; 95% PI  $-1.11$  to 1.47;  $k=4$ ). All 7 PIs encompassed the null, indicating no assured effect in new settings. The neuropsychiatric symptom result was fragile: significance was lost in 5 of 6 leave-one-out iterations and on excluding Petersen 2016 (data extraction ambiguity), yielding a Hedges  $g$  of 0.40 ( $P=.10$ ). GRADE certainty was low for neuropsychiatric symptoms and very low for the remaining 6 domains.

**Conclusions:** Across cross-population, cross-platform randomized controlled trial pooling, the evidence does not yet support routine clinical adoption of care robots. Potential benefits are narrow, of low certainty, and contingent on adequately powered multicenter confirmatory trials with PI-informed sample sizes. Care robots are best framed as facilitator-supported adjuncts that augment, rather than substitute for, human-delivered care.

**Keywords:** socially assistive robots; care robots; digital health; robotics; systematic review; meta-analysis

## Introduction

Health care systems worldwide face 2 parallel and increasingly convergent pressures that are reshaping how direct patient care is organized. The first is demographic: by 2050, the global population aged 60 years and older is projected to reach 2.1 billion, with corresponding rises in the prevalence of chronic disease, cognitive impairment, and dementia [1]. The second is structural: a projected deficit of 12.9 million health care professionals by 2035 [2] is already constraining the dose and continuity of one-to-one, relational care, most acutely in long-term care and in specialized pediatric therapy. Beyond these aggregate workforce figures, specific populations—children with autism spectrum disorder (ASD) and other neurodevelopmental conditions, pediatric inpatients undergoing anxiety-provoking procedures, and older adults with dementia in residential care—share a common practical problem: they benefit from structured, predictable, attention-capturing human interaction that is difficult to deliver at scale.

Care robots have emerged as one technological response to this shared problem. They are physically embodied platforms designed for direct patient interaction and include socially assistive robots such as PARO, NAO, Pepper, Kabochan, and Kaspar; companion and pet robots such as Joy for All, iRobiQ, and MiRo; and specialized therapeutic humanoids [3]. Unlike disembodied digital health tools—telehealth platforms, mobile health apps, or wearable sensors—care robots offer multimodal physical presence (tactile, visual, auditory, and increasingly conversational) that may elicit social and emotional responses in ways that screen-based interventions do not [4]. In older adults with cognitive impairment, tangible tactile interaction with companion robots is thought to support emotional regulation and person-centered engagement in a manner consistent with Kitwood's model of dementia care [5]. In children with ASD, the simplified, predictable, and repeatable social behavior of humanoid robots has been proposed as a scaffold for practicing social-communicative skills that children may find difficult to rehearse with human partners [6]. Acceptance and usability—rather than efficacy alone—are recognized as critical determinants of whether such platforms translate from trials into practice [7,8]. Beyond stand-alone use, socially assistive robots are increasingly deployed as components of broader technological ecosystems that integrate sensing, data capture, and connected services rather than as isolated devices [9,10].

Despite this shared theoretical rationale, the empirical literature on care robots has developed in siloed streams. Prior systematic reviews and meta-analyses have either restricted themselves to a single population—Pu et al [11] in older adults, Leng et al [12] in dementia care with pet robots, and Yu et al [13] in dementia broadly; or Scassellati et al [6] in children with ASD—or to a single platform, most commonly PARO in dementia care. The collective consequence is that

the most clinically actionable question—whether observed robot-intervention effects are driven primarily by platform (eg, PARO vs humanoid) or by population (eg, dementia vs ASD vs pediatric procedural anxiety)—cannot be answered from any one of these reviews alone, because each review holds either platform or population fixed by design. Two further consequences follow for practice. First, procurement, deployment, and health technology assessment decisions are increasingly made across populations (eg, a long-term care operator deciding among PARO, NAO, and companion pet platforms for mixed-acuity settings), yet there is no published synthesis that places the platform effects observed in one population against those observed in another within a single analytic framework. Second, each single-population review has necessarily treated implementation questions—facilitator training, acceptance, setting fit—in isolation, missing the opportunity to identify cross-cutting patterns. Because care-robot randomized controlled trials (RCTs) have grown substantially over the past decade (2015-2024), the minimum number of studies (*k*) required for meaningful moderation analyses of robot platform and target population is now within reach.

Accordingly, this systematic review and meta-analysis had 4 prespecified objectives. First, to quantify the pooled effects of care-robot interventions across 7 prespecified outcome domains in a cross-population sample that explicitly includes both older adult and pediatric populations. Second, to examine the moderating effects of robot platform (PARO vs humanoid vs other companion and pet), target population (dementia or older adult vs pediatric or ASD vs other), and intervention duration ( $\geq 10$  weeks vs  $< 10$  weeks) through prespecified subgroup analyses. Third, to appraise the methodological quality of included studies using the Cochrane Risk of Bias 2 (RoB 2) tool. Fourth, to evaluate the certainty of evidence across domains using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework and to translate the resulting evidence profile into realistic, targeted recommendations for clinical use. To our knowledge, this is among the first quantitative syntheses to pool care-robot RCTs across both older adult and pediatric populations within a common framework; because per-subgroup study counts are small, we frame the resulting comparison across robot platform and populations as exploratory and hypothesis-generating rather than confirmatory.

## Methods

### Overview

This systematic review and meta-analysis is reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 Statement [14], the PRISMA 2020 for Abstracts extension, the PRISMA-S (Preferred Reporting Items for Systematic

Reviews and Meta-Analyses literature search) extension for reporting literature searches [15], and the methodological standards of the *Cochrane Handbook for Systematic Reviews of Interventions* [16]. Completed PRISMA 2020 Expanded, PRISMA 2020 for Abstracts, and PRISMA-S checklists are provided as [Checklist 1–3](#). The review was not prospectively registered in PROSPERO (International Prospective Register of Systematic Reviews); an a priori review protocol specifying eligibility criteria, outcome domains, analytic plan, and risk-of-bias assessment was developed and dated before the formal search and is available from the corresponding author on reasonable request.

## Eligibility Criteria

Eligibility criteria were defined using the Population, Intervention, Comparison, Outcomes, and Study Design framework. The population comprised patients of any age, in any clinical or community setting, with no restrictions on primary diagnosis or age group, consistent with the cross-population aim of the review. The intervention was any physically embodied care robot designed for direct patient interaction, including socially assistive, companion, pet, and therapeutic humanoid platforms. Because the eligibility criteria required a social or interactive component, all included platforms are, by definition, social robots or socially assistive robots; we retain “care robots” as the umbrella term while noting that every included device met the social robots or socially assistive robots criterion. Surgical robots, passive rehabilitation exoskeletons without a social interaction component, and clinician-controlled telepresence robots were excluded. Comparators were standard care without robot interaction, attention-matched active controls (eg, reading or tablet-based sessions of equivalent duration), placebo conditions (deactivated robot or plush equivalent), and waitlist controls. Outcomes were restricted to quantitative, validated patient-reported outcome measures within 7 prespecified outcome domains defined in the a priori protocol: neuropsychiatric symptoms (NPS), quality of life, depression, agitation, stress and pain, social-communicative skills, and cognitive function. Other outcomes encountered during extraction (eg, motor activity and biomarkers) were retained in the narrative synthesis but not pooled. Eligible study designs were individual RCTs and cluster RCTs. Single-arm before-after studies, case reports, qualitative-only studies, feasibility studies without an outcome comparator, and conference abstracts lacking full numerical data were excluded.

## Information Sources

We searched 5 electronic databases: PubMed (MEDLINE), Embase (Elsevier), the Cochrane Central Register of Controlled Trials (CENTRAL), CINAHL Complete (EBSCO), and APA PsycINFO. Two trial registries (ClinicalTrials.gov and the World Health Organization [WHO] International Clinical Trials Registry Platform [ICTRP]) were searched for completed trials with reported results. Reference lists of all included studies and of 4 prior systematic reviews in the field were screened, and forward citation tracking was performed via Google Scholar. Searches covered database

inception through April 29, 2026. No language or publication date restrictions were applied at the search stage. At full-text screening, articles for which no English or Korean full text or professional translation was obtainable despite repeated retrieval attempts were excluded (n=5); this is reported in the PRISMA flow diagram.

## Search Strategy

Search strategies were developed iteratively in consultation with a health-sciences librarian and combined controlled vocabulary (MeSH [Medical Subject Headings] terms in PubMed and CENTRAL; Emtree terms in Embase; CINAHL Subject Headings; APA Thesaurus terms in PsycINFO) with free-text terms mapped with field modifiers (.ti,ab,kw and database-specific equivalents), truncation, and proximity operators. Three concept blocks were combined with AND: (1) care robots, including both generic terms (“social robot,” “socially assistive robot,” “companion robot,” “robotic pet,” “humanoid robot,” and “therapeutic robot”) and platform-specific trade names (PARO, NAO, Pepper, Kaspar, Kabochan, CommU, Joy for All, MiRo, and iRobiQ); (2) health care context terms (eg, “nursing home,” “long-term care,” “dementia care,” “pediatric,” “hospital,” “rehabilitation,” and “outpatient”); and (3) patient-outcome terms covering the 7 prespecified domains (eg, “quality of life,” “neuropsychiatric symptoms,” “agitation,” “anxiety,” “pain,” “social skills,” and “mental status”). Full database-specific search strategies, exact search dates, and the number of records retrieved per database are reported in [Checklist 1](#) (PRISMA-S compliant).

## Selection Process

All retrieved records were imported into Covidence systematic review software (Veritas Health Innovation) and deduplicated automatically with manual verification. Two reviewers (GK and JJ) independently screened all unique records by title and abstract and then by full text against the prespecified eligibility criteria. Both reviewers are registered nurses: GK holds a PhD and has prior published systematic review experience in gerontological and digital health domains and completed formal training in the Cochrane RoB 2 and GRADE frameworks; JJ holds a master’s degree and completed Cochrane-endorsed methodological training in systematic review conduct and data extraction. Interrater reliability was quantified using Cohen kappa; Cohen  $\kappa=0.89$  at title and abstract and Cohen  $\kappa=0.92$  at full text [17]. Disagreements were resolved through structured discussion; residual disagreements were adjudicated by a third senior reviewer with prior systematic review and meta-analysis experience.

## Data Collection Process and Data Items

A standardized data extraction form was piloted on 5 [18–22] randomly selected studies and refined before full-scale extraction. Two reviewers independently extracted: study identification (first author, year, country, and journal); study design and randomization method; participant characteristics (sample size per arm, mean age, sex distribution, primary diagnosis, and care setting); intervention details (robot type

and model, session duration and frequency, total intervention duration, presence and role of facilitator, group vs individual format); control condition (standard care, attention-matched active control, placebo or deactivated robot, and waitlist); outcome measures with instrument names, assessment time points, and numerical data (means, SDs, and sample sizes); and adverse events or dropouts. For studies reporting outcome data only in graphical form, values were extracted using WebPlotDigitizer (version 4.6; Ankit Rohatgi). Discrepancies were resolved by consensus. Where standard deviations were not directly reported, they were imputed from 95% CIs, SEs, or IQRs using Cochrane Handbook–recommended formulas [16]. Studies for which sufficient numerical data could not be obtained were excluded from meta-analyses but retained in the narrative synthesis.

Robot platforms were classified a priori into 3 operationally defined categories for the subgroup analyses: (1) PARO and PARO-like tactile companion robots—biofeedback-responsive, nonverbal, animal-form-factor platforms whose interaction modality is primarily tactile; (2) humanoid robots—bipedal or upper-body humanoid form factor with structured verbal and gestural interaction (NAO, Pepper, Kaspar, CommU, and Kabochan); and (3) other companion and pet and nonhumanoid platforms (Joy for All, iRobiQ, MiRo, and unnamed custom platforms).

### Study Risk-of-Bias Assessment

The methodological quality of all included randomized and cluster-randomized trials was assessed independently by 2 reviewers using the Cochrane RoB 2 tool [23] across 5 domains: (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in measurement of the outcome, and (5) bias in selection of the reported result. As specified in the [Multimedia Appendix 1](#) footnote, all 34 included studies [18–22,24–52] were classified as randomized controlled studies (24 individual RCTs [19–22,31,32,34–51] and 10 [18,24–30,33,52] cluster RCTs) following revision-process reclassification and were therefore assessed using a single risk-of-bias instrument (RoB 2). Overall study risk was classified as low, some concerns, or high. Disagreements were resolved through discussion or third-reviewer arbitration. Domain-level and overall judgments were used to inform GRADE certainty downgrading for risk of bias.

### Effect Measures and Synthesis Methods

Given heterogeneous outcome instruments across included studies, effect sizes were expressed as standardized mean differences (Hedges  $g$ ) with 95% CIs. Hedges  $g$  was preferred over Cohen  $d$  for its reduced small-sample bias [53]. Random-effects meta-analysis was performed using the restricted maximum-likelihood estimator for between-study variance  $\tau^2$  with the Hartung-Knapp-Sidik-Jonkman (HKSJ) correction for the CI of the pooled estimate [54,55]. The HKSJ correction was selected, in place of the DerSimonian-Laird method used in our original submission, because it provides more reliable CI coverage and reduces false-positive rates under small-study and moderate-to-high heterogeneity conditions [55]. All effect sizes were oriented so that a

positive Hedges  $g$  indicates an effect favoring care robots across all domains; for symptom reduction outcomes (NPS, agitation, depression, and stress and pain), change scores were sign-reversed accordingly. Effect sizes were interpreted using Cohen conventional benchmarks: small ( $|g|=0.20-0.49$ ), moderate ( $|g|=0.50-0.79$ ), and large ( $|g|\geq 0.80$ ) [53].

Statistical heterogeneity was quantified using  $I^2$ , the between-study variance  $\tau^2$ , and the Cochran Q test. For each domain with  $k\geq 4$ , we additionally calculated the 95% prediction interval (PI) [56,57], which describes the range within which the true effect in a new setting is expected to fall and therefore reflects the real-world implications of between-study heterogeneity more faithfully than  $I^2$  alone [56]. Where the PI crosses the line of no effect despite a statistically significant CI for the pooled estimate, we interpret this as indicating that, although an average benefit is detectable across existing studies, the effect in a new setting cannot be assumed to be beneficial.

To address the possibility of double-counting, participant cohorts that appeared across multiple publications were identified a priori. We applied the following principle: each cohort contributes to each outcome pool at most once, but the same cohort may contribute to different outcome pools when it measures different outcome domains, because each outcome is analyzed in an independent meta-analysis [58]. Four cohort-overlap clusters were identified: (1) Australian PARO cluster (shared  $n=415$ )—Moyle et al [24], Moyle et al [25], Jones et al [26], and Mervin et al [27] drew from a shared dementia cohort; only Moyle et al [24] was retained as the primary publication contributing to the NPS and agitation pools, while Moyle et al [25], Jones et al [26], and Mervin et al [27] are narrative-only. (2) Norwegian PARO cluster—Jøranson et al [28], Jøranson et al [29], and Jøranson et al [30] drew from a shared nursing-home dementia cohort; Jøranson et al [28] was retained for the NPS, depression, and agitation pools; Jøranson et al [29] was retained for the quality-of-life pool (a different outcome domain measured in the same cohort), and Jøranson et al [30] is narrative-only. (3) Hong Kong dementia cohort—Chen et al [18] and Chen et al [31] drew from a shared cohort; Chen et al [18] was retained as the primary publication contributing to the relevant outcome pools, while Chen et al [31] (a technology-acceptance secondary analysis of the same cohort) is narrative-only. (4) Dutch ASD cohort ( $n=81$ )—De Korte et al [19] and van den Berk-Smeekens et al [32] drew from a shared trial; only De Korte et al [19] was retained for the social-communicative skills pool, as van den Berk-Smeekens et al [32] reported a secondary analysis of the same trial. Studies measuring multiple outcomes (eg, Valentí-Soler et al [33] contributing to 5 outcome pools and Chen et al [18], Petersen et al [34], and Pollak et al [20] each contributing to 4 outcome pools) were retained in each relevant pool, with each contribution representing an independent outcome estimate from a single cohort. As a result, no cohort contributes to the same pooled estimate more than once. The splitting approach (retaining one publication per cohort per outcome pool) was selected over multilevel or robust-variance-estimation alternatives because the within-cohort correlation across

the identified clusters could not be reliably estimated from the available aggregate data. This allocation is documented in a footnote to [Multimedia Appendix 1](#).

Prespecified subgroup analyses were conducted by (1) robot platform (PARO vs humanoid vs other companion and pet), (2) target population (dementia or older adult vs pediatric or ASD vs other), and (3) intervention duration ( $\geq 10$  weeks vs  $< 10$  weeks). Sensitivity analyses comprised: (1) leave-one-out (LOO) analysis (one study excluded at a time, applied to any pool reaching statistical significance under primary HKSJ pooling); (2) exclusion of studies rated overall high risk of bias under RoB 2 (4 of 34 studies); and (3) data extraction sensitivity analysis (exclusion of Petersen 2016, flagged for directional ambiguity in published change-score reporting). All meta-analyses were conducted in R (version 4.3.2; R Core Team) using the meta (version 6.5-0; Guido Schwarzer) and metafor (version 4.4-0; Wolfgang Viechtbauer) packages, with `metafor::rma(yi, vi, method=`

`"REML," test="knha")` for the HKSJ-corrected pooled estimate and predict for the 95% PI.

### Reporting Bias Assessment

We assessed small-study effects, rather than publication bias per se, using visual inspection of contour-enhanced funnel plots and the Egger weighted-regression test [59]. Following Sterne et al [60], we note that funnel-plot asymmetry can arise from publication bias but also from true effect-size differences between small and large studies, methodological differences between small and large studies, or chance. The Egger test was interpreted only for domains with  $k \geq 10$  contributing studies because the test has very limited power below that threshold; nonsignificant Egger results in smaller domains should not be read as ruling out small-study effects. Given that no domain in our review reached  $k \geq 10$ , the funnel plots reported for NPS ( $k=6$ ) and depression ( $k=7$ ) are presented for transparency only and are explicitly interpreted with caution.

### Certainty Assessment

Certainty of evidence for each primary outcome domain was evaluated independently by 2 reviewers using the GRADE framework [61], with downgrading for risk of bias, inconsistency, indirectness, imprecision, and reporting biases. The formal GRADE evidence profile was produced using GRADEpro GDT (McMaster University).

### Ethical Considerations

As this study is a systematic review and meta-analysis of previously published aggregate data, institutional review

board approval was not required, and no individual participant data were accessed.

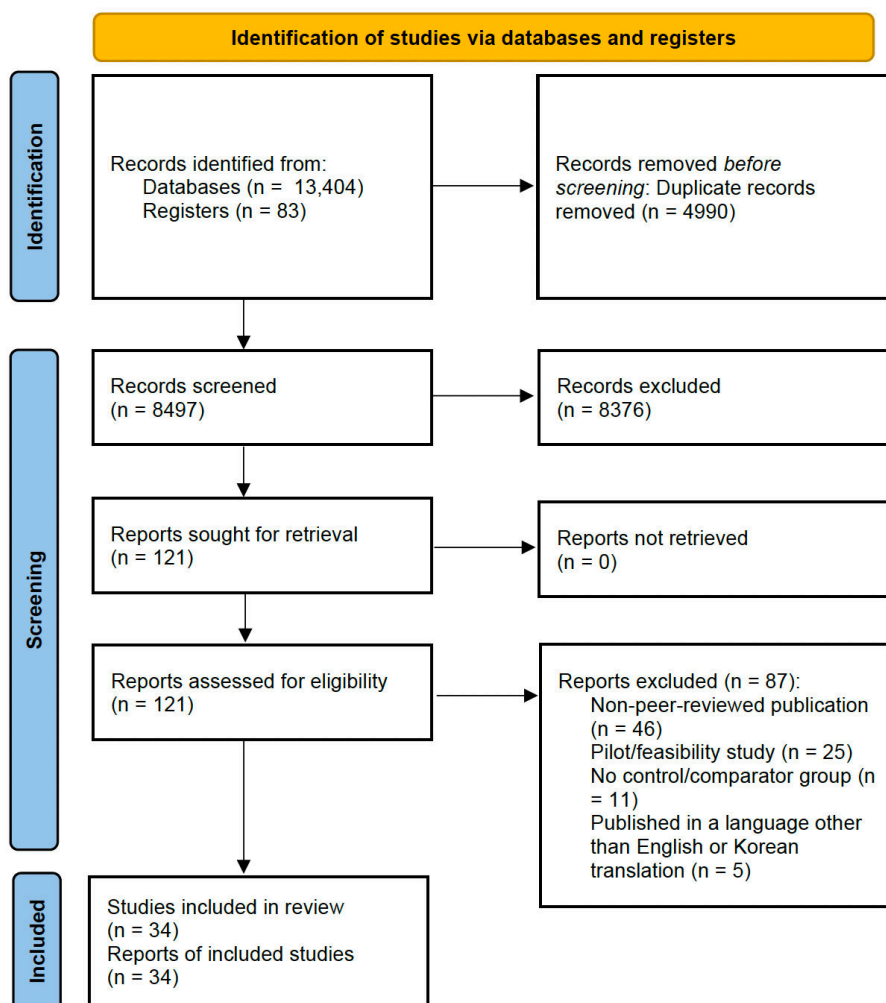
## Results

### Study Selection

The expanded search across 5 databases and 2 trial registries yielded 13,487 records (PubMed, including MEDLINE  $n=5935$ ; Embase  $n=4034$ ; Cochrane CENTRAL  $n=2258$ ; CINAHL  $n=1176$ ; APA PsycINFO  $n=1$ ; ClinicalTrials.gov  $n=75$ ; WHO ICTRP  $n=8$ ). APA PsycINFO returned only 1 record matching all 3 concept blocks; relaxed-vocabulary searches in PsycINFO yielded substantially more records but with high duplication against MEDLINE and Embase and were excluded to avoid double-counting. After deduplication in EndNote 21 with manual verification in Covidence, 4990 duplicates were removed and 8497 unique records underwent title-and-abstract screening, of which 8376 were excluded as clearly ineligible. Full-text assessment was conducted for 121 records; 87 were excluded for the following reasons: not peer-reviewed,  $n=46$ ; pilot study without a comparative arm,  $n=25$ ; nonrandomized study design (single-arm before-after or noncontrolled),  $n=6$ ; no English or Korean full text or professional translation obtainable,  $n=5$ ; and no concurrent control group,  $n=5$ . A final total of 34 studies [18-22,24-52] met all eligibility criteria and were included in the qualitative synthesis; 20 [18-22,24,28,29,33,34,36,38,39,42,43,45-47,50,52] contributed to at least one meta-analysis pool and 14 [25-27,30-32,35,37,40,41,44,48,49,51] entered narrative-only synthesis. The updated PRISMA 2020 flow diagram is presented in [Figure 1](#).

We updated and expanded the search strategy in accordance with Cochrane guidance. In addition to PubMed, including MEDLINE, we searched Embase, Cochrane CENTRAL, CINAHL, and APA PsycINFO to capture evidence across medicine, nursing, and behavioral psychology. We also searched ClinicalTrials.gov and the WHO ICTRP separately for ongoing or unpublished trials. Google Scholar was searched as a supplementary source for gray literature; the first 200 results for the principal search query were screened by title and abstract. Reference lists of included studies and relevant prior systematic reviews were hand-searched. No language restriction was applied at the search stage. Full database-specific search strategies, including controlled vocabulary and free-text terms, are provided in [Checklist 1](#).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for the systematic review and meta-analysis of care robot interventions. Of the 34 included trials, 20 contributed to  $\geq 1$  meta-analysis pool and 14 entered narrative-only synthesis. Adapted from Page et al [14]. EN: English; KR: Korean.



### Study Characteristics

The 34 [18-22,24-52] included studies were conducted across 17 countries and published between 2015 and 2024. Countries with the greatest representation included Australia (n=4), United States (n=4), Norway (n=3), Netherlands (n=3), and Hong Kong (n=3), followed by Greece, Israel, Italy, New Zealand, and the United Kingdom (n=2 each), and one study each from Austria, Canada, Germany, Japan, South Korea, Spain, and Taiwan. Study designs comprised randomized controlled studies in all 34 included reports—24 (70.6%) individual RCTs and 10 (29.4%) cluster RCTs; on reexamination of the original full-text sources, 5 [33,35,41,42, 48] studies that were previously classified as quasi-experimental, controlled trial, or clinical trial were reclassified as randomized trials, and Valentí-Soler et al [33] was reclassified as randomized for the nursing-home arm with

the sequential day care center phase noted in Limitations. Robot platforms included PARO (11/34, 32.4%), humanoid platforms (NAO, Pepper, Kabochan, CommU, Kaspar, and unnamed; 16/34, 47.1%), and other companion and pet platforms (7/34, 20.6%). Target populations included older adults with dementia or mild cognitive impairment (14/34, 41.2%), children and adolescents with ASD (9/34, 26.5%), hospitalized or procedural pediatric patients (4/34, 11.8%), and other adults (poststroke, mixed; 7/34, 20.6%). Intervention durations were single-session (9/34, 26.5%), <4 weeks (2/34, 5.9%), 4-9 weeks (4/34, 11.8%), and  $\geq 10$  weeks (19/34, 55.9%). Care settings comprised nursing home or long-term care (n=14), hospital (n=16), community or home (n=2), and rehabilitation (n=2). Full characteristics are in Multimedia Appendix 1 (full per-study table) and Table 1 (summary).

**Table 1.** Summary characteristics of the included care robot randomized trials. Aggregated counts and proportions are shown by study design, population, robot platform, care setting, intervention duration, and meta-analysis contribution status.

Characteristic	Values, n (%)	Notes
Study design		
Randomized controlled studies	34 (100)	— <sup>a</sup>
Individual RCT <sup>b</sup>	24 (70.6)	—
Cluster RCT	10 (29.4)	—
Robot platform		
PARO (seal-shaped SAR <sup>c</sup> )	11 (32.4)	Primarily dementia or LTC <sup>d</sup> populations
Humanoid (NAO, Pepper, Kabochan, CommU, unnamed)	16 (47.1)	ASD <sup>e</sup> , pediatric, mixed populations
Companion and pet robots (Joy for All, MiRo, iRobiQ, others)	7 (20.6)	Pediatric, older adult, mixed populations
Target population		
Older adults with dementia or MCI <sup>f</sup>	14 (41.2)	Age range 50-108 years
Children or adolescents with ASD	9 (26.5)	Age range 3-12 years
Hospitalized or procedural pediatric	4 (11.7)	Acute care settings
Other adults (poststroke, mixed)	7 (20.6)	Various diagnoses
Intervention duration		
Single session	9 (26.5)	—
<4 weeks	2 (5.9)	—
4-9 weeks	4 (11.7)	—
≥10 weeks	19 (55.9)	—
Care setting		
Nursing home or long-term care facility	14 (41.2)	—
Hospital (inpatient or outpatient)	16 (47.1)	—
Community or home setting	2 (5.9)	—
Rehabilitation center	2 (5.9)	—
Risk of bias tool		
Cochrane Risk of Bias 2 (RoB 2)	34 (100)	—

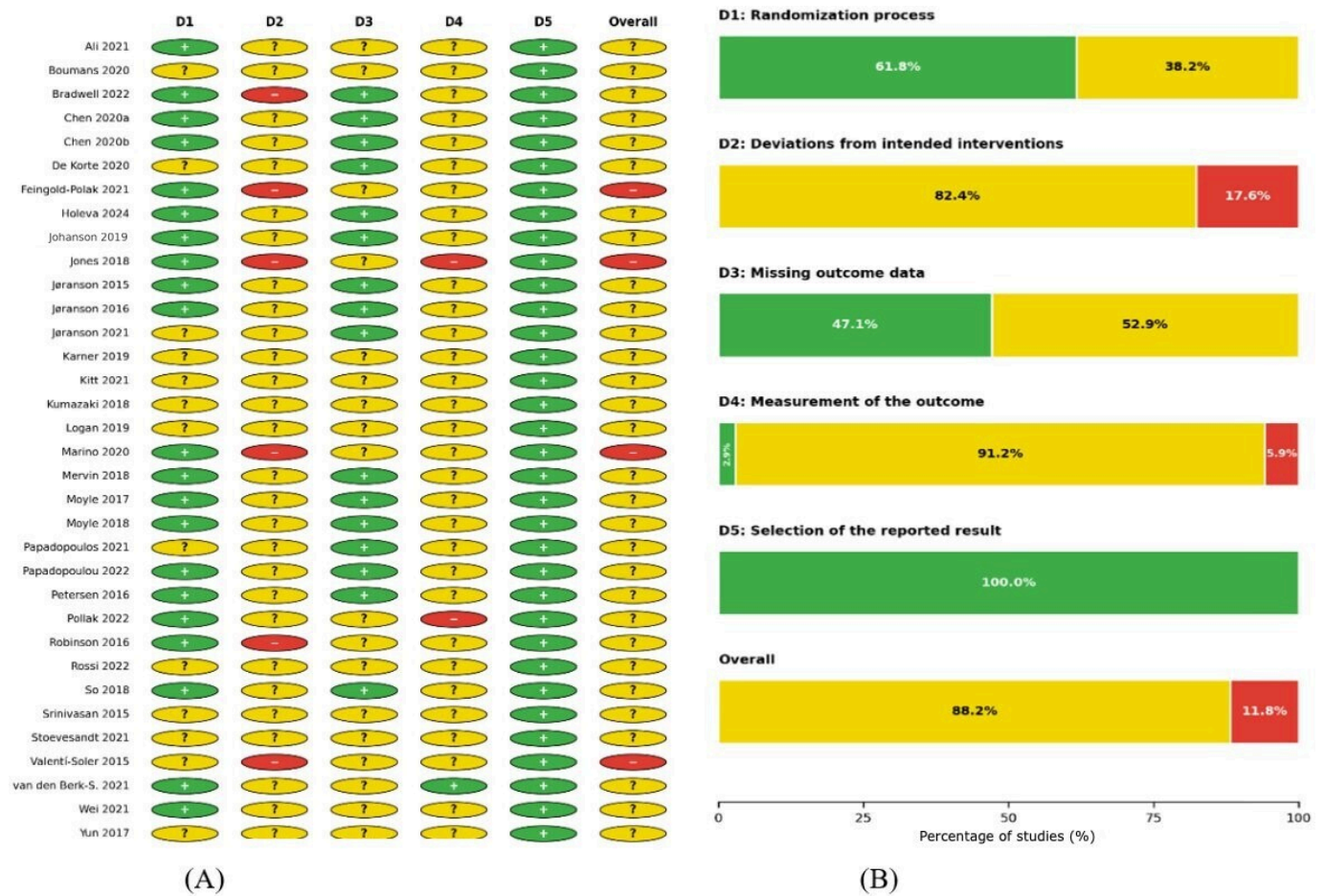
<sup>a</sup>Not applicable.<sup>b</sup>RCT: randomized controlled trial.<sup>c</sup>SAR: socially assistive robot.<sup>d</sup>LTC: long-term care.<sup>e</sup>ASD: autism spectrum disorder.<sup>f</sup>MCI: mild cognitive impairment.

## Risk-of-Bias in Studies

All 34 included randomized and cluster randomized trials were assessed using Cochrane RoB 2. Pooled across the 34 [18-22,24-52] studies, 0 studies were rated overall low risk of bias, 30 [18-22,24,25,27-32,34,36-40,42-52] (88.2%) were rated as having some concerns, and 4 [26,33,35,41] (11.8%) were rated overall high risk of bias. Performance bias (RoB 2 Domain 2: deviations from intended interventions) was the most consistently problematic domain—blinding of participants and personnel to robotic versus nonrobotic

interactions is structurally impossible—and contributed to the GRADE risk-of-bias downgrade across all 7 outcome domains. Domain 5 (selective outcome reporting) was rated low risk in all studies. Missing outcome data (Domain 3) was rated low risk in 16 [18,19,24,25,27-31,34,36,42,43,46,52] (47.1%) of 34 studies and some concerns in 18 [20-22,26,32,33,35,38-41,44,45,47-51] (52.9%) studies, with no studies rated high risk on this domain. Full domain-level RoB 2 ratings are presented in Figure 2.

**Figure 2.** Risk-of-bias assessment of the 34 included trials using the Cochrane RoB 2 tool. (A) Study-level judgments across the 5 RoB 2 domains (D1-D5). (B) Domain-level summary across all 34 studies. Green (+) indicates low risk, yellow (?) indicates some concerns, and red (-) indicates high risk [18-22,24-52].



## Results of Syntheses

### Overview

Following systematic full-text data verification, 20 [18-22, 24,28,29,33,34,36,38,39,42,43,45-47,50,52] unique studies contributed to one or more meta-analysis pools. Pooled HKSJ random-effects estimates, 95% CIs, 95% PIs,  $I^2$ ,  $\tau^2$ , and GRADE certainty for the 7 outcome domains are summarized in Table 2, and forest plots are presented in Figure 3.

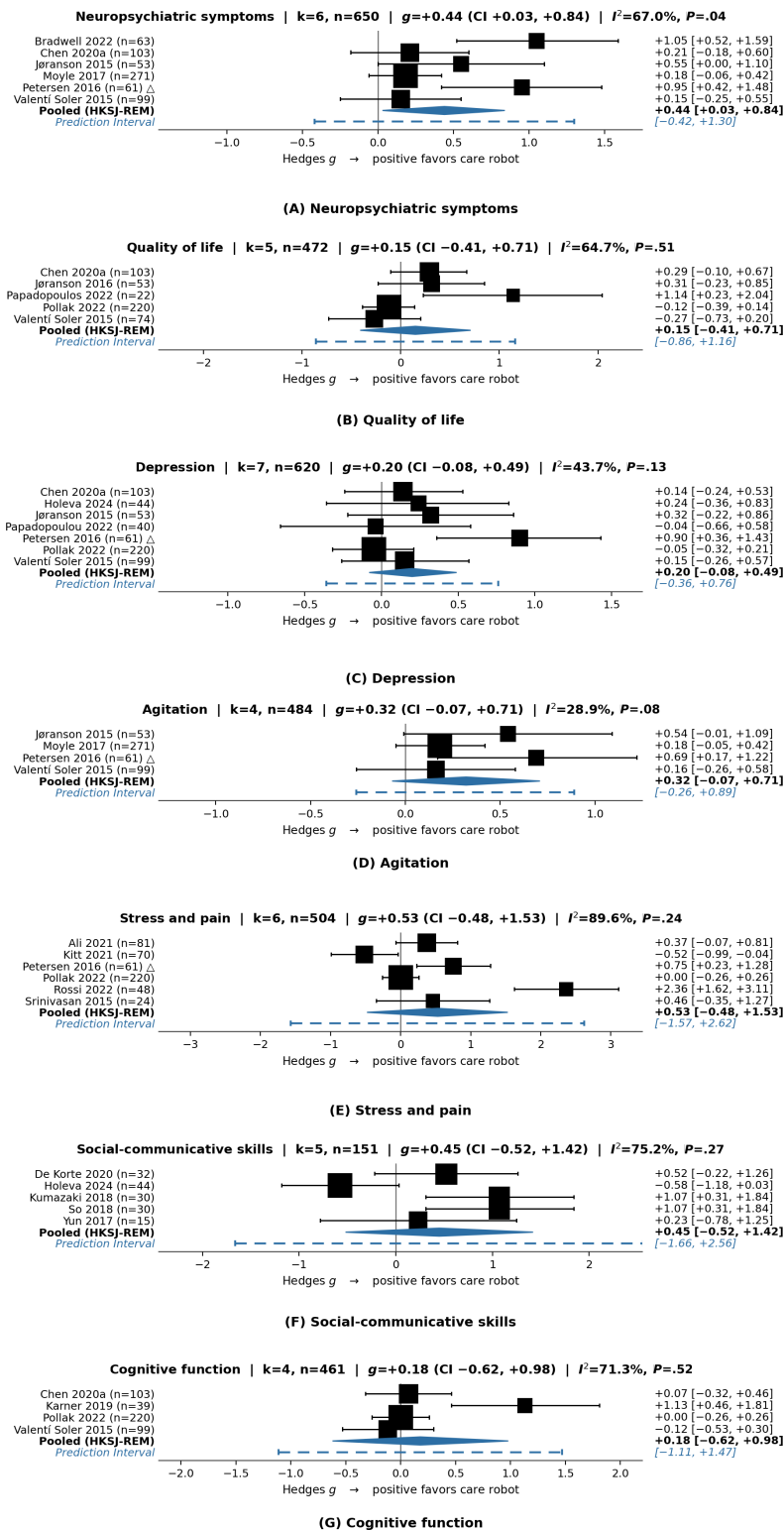
**Table 2.** GRADE<sup>a</sup> summary of findings for robot-assisted interventions across 7 outcome domains<sup>b</sup>.

Certainty assessment										
Studies, n	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Number of patients (number of studies)	Hedges g (95% CI)	Certainty	Importance
Neuropsychiatric symptoms (assessed with: NPI <sup>c</sup> , NPI-NH <sup>c</sup> , and BEHAVE-AD <sup>e</sup> )	6	Randomized trials	Serious <sup>f</sup>	Serious <sup>g</sup>	Not serious	None	650 (k=6)	0.44 (CI 0.03 to 0.84)	⊕○○○ Low <sup>f,g</sup>	IMPORTANT
Quality of life (assessed with: QoL-AD <sup>h</sup> , WHOQOL <sup>i</sup> , or related scales)	5	Randomized trials	Serious <sup>j</sup>	Serious <sup>k</sup>	Not serious	None	472 (k=5)	0.15 (-0.41 to 0.71)	⊕○○○ Very low <sup>j,k,l</sup>	CRITICAL
Depression (assessed with: GDS <sup>m</sup> , CSDD <sup>n</sup> , PHQ-9 <sup>o</sup> , or related scales)	7	Randomized trials	Serious <sup>p</sup>	Serious <sup>q</sup>	Not serious	None	620 (k=7)	0.2 (-0.08 to 0.49)	⊕○○○ Very low <sup>p,q,r</sup>	IMPORTANT
Agitation (assessed with: CMAI <sup>s</sup> or related agitation scales)	4	Randomized trials	Serious <sup>t</sup>	Serious <sup>u</sup>	Not serious	None	484 (k=4)	g0.32 (-0.07 to 0.71)	⊕○○○ Very low <sup>t,u,v</sup>	IMPORTANT
Stress and pain (assessed with: stress scales, pain scales, or physiological indicators)	6	Randomized trials	Serious <sup>w</sup>	Very serious <sup>x</sup>	Not serious	None	504 (k=6)	0.53 (-0.48 to 1.53)	⊕○○○ Very low <sup>w,x,y</sup>	IMPORTANT
Social-communicative skills (assessed with: SRS <sup>z</sup> , social interaction, or communication-related scales)	5	Randomized trials	Serious <sup>aa</sup>	Not serious	Serious <sup>ab</sup>	None	151 (k=5)	0.45 (-0.52 to 1.42)	⊕○○○ Very low <sup>aa,ab,ac</sup>	IMPORTANT

Certainty assessment		Number of patients (number of studies)	Hedges <i>g</i> (95% CI)	Certainty	Importance	
Studies, <i>n</i>	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations
4	MMSE <sup>ad</sup> , MoCA <sup>ae</sup> , or related cognitive scales	Randomized trials	Not serious	Not serious	Not serious	None
		461 (k=4)	0.18 (-0.62 to 0.98)	⊕⊕⊕⊕ Very low <sup>af, ag, ah</sup>	⊕⊕⊕⊕ Very low <sup>af, ag, ah</sup>	CRITICAL

<sup>a</sup>GRADE: Grading of Recommendations Assessment, Development and Evaluation.  
<sup>b</sup> Effect sizes are reported as Hedges *g* with 95% CIs using Hartung-Knapp-Sidik-Jonkman random-effects pooling with restricted maximum-likelihood  $\tau^2$  estimation. Certainty ratings were defined as ⊕⊕⊕⊕ high, ⊕⊕⊕ moderate, ⊕⊕⊕ low, and ⊕⊕⊕ very low.  
<sup>c</sup>NPI: Neuropsychiatric Inventory.  
<sup>d</sup>NPI-NH: Neuropsychiatric Inventory–Nursing Home version.  
<sup>e</sup>BEHAVE-AD: Behavioral Pathology in Alzheimer’s Disease Rating Scale.  
<sup>f</sup>Downgraded for risk of bias due to open-label designs and concerns related to blinding, attrition, or selective reporting in some trials.  
<sup>h</sup>QoL-AD: Quality of Life in Alzheimer’s Disease scale.  
<sup>i</sup>WHOQOL: World Health Organization Quality of Life.  
<sup>j</sup>Downgraded for risk of bias due to open-label designs and methodological limitations in some trials.  
<sup>k</sup>Downgraded for inconsistency due to heterogeneity across studies  
<sup>l</sup>Downgraded for imprecision because the confidence interval crossed the null effect.  
<sup>m</sup>GDS: Geriatric Depression Scale.  
<sup>n</sup>CSDD: Cornell Scale for Depression in Dementia.  
<sup>o</sup>PHQ-9: Patient Health Questionnaire-9.  
<sup>p</sup>Downgraded for risk of bias due to methodological limitations in some included trials.  
<sup>q</sup>Downgraded for inconsistency across studies.  
<sup>r</sup>Downgraded for imprecision because the confidence interval crossed the null effect.  
<sup>s</sup>CMAI: Cohen-Mansfield Agitation Inventory.  
<sup>t</sup>Downgraded for imprecision because the confidence interval crossed the null effect.  
<sup>u</sup>Downgraded for inconsistency because one study was treated as narrative-only due to cohort overlap, and the pooled evidence was limited.  
<sup>v</sup>Downgraded for imprecision because the confidence interval crossed the null effect.  
<sup>w</sup>Downgraded for risk of bias due to methodological limitations in included trials.  
<sup>x</sup>Downgraded for severe inconsistency across studies.  
<sup>y</sup>Downgraded for imprecision because the confidence interval was wide and crossed the null effect.  
<sup>z</sup>SRS: Social Responsiveness Scale.  
<sup>aa</sup>Downgraded for risk of bias due to open-label designs and measurement limitations.  
<sup>ab</sup>Downgraded for indirectness because all studies were conducted in pediatric autism spectrum disorder populations using humanoid platforms.  
<sup>ac</sup>Downgraded for imprecision because the confidence interval was wide and crossed the null effect.  
<sup>ad</sup>MMSE: Mini-Mental State Examination.  
<sup>ae</sup>MoCA: Montreal Cognitive Assessment.  
<sup>af</sup>Downgraded for risk of bias due to methodological limitations in included trials.  
<sup>ag</sup>Downgraded for indirectness because the evidence was derived from dementia and mild cognitive impairment populations.  
<sup>ah</sup>Downgraded for imprecision because the confidence interval was wide and crossed the null effect.

**Figure 3.** Forest plots for the 7 prespecified outcome domains. Forest plots are based on the 20 trials contributing to at least one meta-analysis. Squares indicate study-level estimates, diamonds indicate pooled HKSJ estimates, and dashed bars indicate 95% prediction intervals. Panel headers report k, N, Hedges g, 95% CI, P, and I<sup>2</sup> value [18-22,24,28,29,33,34,36,38,39,42,45-47,50].



Only NPS reached statistical significance under HKSJ random-effects pooling (Hedges g=0.44, 95% CI 0.03-0.84; P=.04). The remaining 6 domains yielded point estimates that generally favored care robots, but their 95% CIs crossed the null and did not reach conventional statistical significance. All 7 95% PIs encompassed the null, indicating that effect estimates may not generalize uniformly to new populations or

settings. This conservative pattern reflects the HKSJ small-sample correction, full-text primary-source data verification, and consistent application of the Senn [58] cohort-overlap principle. Between-study heterogeneity ranged from low (I<sup>2</sup>=29% for agitation) to high (I<sup>2</sup>=90% for stress and pain).

## NPS

Six studies ( $n=650$ ) contributed measurements using the Neuropsychiatric Inventory (NPI; Bradwell et al [52], Chen et al [18], and Valentí-Soler et al [33]), the Behavioral Activity Rating Scale (BARS; Jøranson et al [28]), the Cohen-Mansfield Agitation Inventory–Short Form (CMAI-SF; Moyle et al [24]), and the Rating Anxiety in Dementia (RAID) scale (Petersen et al [34]). Mervin et al [27] was reclassified to narrative-only synthesis owing to cohort overlap with Moyle et al [24], and Robinson et al [44] was reclassified to narrative-only owing to its qualitative-observational design. The HKSJ pooled estimate was Hedges  $g=0.44$  (95% CI 0.03-0.84; 95% PI  $-0.42$  to 1.30;  $P=67\%$ ;  $\tau^2=0.088$ ;  $P=.04$ ), indicating a small-to-moderate, statistically significant reduction in NPS favoring care robot intervention. Bradwell et al [52] reported the largest individual effect (Hedges  $g=1.05$ , 95% CI 0.52-1.59) using the Joy for All robot pet over a 4-month cluster RCT in care homes, with NPI subdomain reductions in delusions, depression, anxiety, elation, and apathy (all  $P\leq.03$ ). Valentí-Soler et al [33] demonstrated a sustained intervention effect at follow-up (BARS effect estimate 5.5, 95% CI 0.1-11.0;  $P=.048$ ). Chen et al [18] reported no significant between-group difference on the CMAI-SF (adjusted mean difference  $-1.89$ , 95% CI  $-5.81$  to 2.02;  $P=.34$ ), and Valentí-Soler et al [33] showed null effects on total NPI score in their nursing home arm.

## Quality of Life

Five studies [18,20,28,33,42] ( $n=472$ ) contributed using the Quality of Life in Late-Stage Dementia scale, Quality of Life in Alzheimer's Disease scale, Short Form 36 Emotional Wellbeing subscale, and the Frail Questionnaire [20]. The HKSJ pooled estimate was Hedges  $g=0.15$  (95% CI  $-0.41$  to 0.71; 95% PI  $-0.86$  to 1.16;  $P=65\%$ ;  $\tau^2=0.092$ ;  $P=.51$ ), indicating a small effect direction favoring care robots that did not reach statistical significance. Effect estimates were heterogeneous: Papadopoulos et al [42] reported a substantial benefit on Short Form 36 Emotional Wellbeing scale (Hedges  $g=1.14$ , 95% CI 0.23-2.04;  $P=.02$ ;  $\eta^2=.258$ ), whereas Pollak et al [20] ( $n=220$ ) detected no significant change in social or physical frailty ( $P=.42$  and  $.99$ , respectively), and Valentí-Soler et al [33] observed a counter-direction effect on Quality of Life in Late-Stage Dementia scale in the day-care PARO arm (Hedges  $g=-0.27$ ;  $P=.04$  favoring control). Jøranson et al [29] reported QoL effects only in residents with severe dementia (overall  $P=.12$ , severe-subgroup  $P=.008$ ), suggesting that effects may be conditional on dementia severity.

## Depression

Seven studies ( $n=620$ ) contributed using the Geriatric Depression Scale (GDS or GDS-SF [Geriatric Depression Scale – Short Form]) [18,20], the Cornell Scale for Depression in Dementia (CSDD [28,34]), the Strengths and Difficulties Questionnaire emotional subscale (SDQ [42]), the Child Behavior Checklist Internalizing subscale [36], and the NPI affective subdomain [33]. The HKSJ pooled estimate was Hedges  $g=0.20$  (95% CI  $-0.08$  to 0.49; 95% PI  $-0.36$  to 0.76;  $P=44\%$ ;  $\tau^2=0.039$ ;  $P=.13$ ), indicating a

small effect direction favoring care robots that did not reach statistical significance. Jøranson et al [28] demonstrated a significant reduction at follow-up (CSDD effect estimate 3.9, 95% CI 0.4-7.3;  $P=.03$ ); Petersen et al [34] reported a significant CSDD between-group difference ( $P=.001$ ), though the directional interpretation of CSDD change scores in the published table is internally inconsistent with the abstract narrative (sensitivity analysis below). Contributions from Pollak et al [20] (GDS-SF  $P=.90$ ), Chen et al [18] (GDS  $P=.33$ ), and Papadopoulou et al [43] (Strengths and Difficulties Questionnaire emotional  $P=.80$ ) drew the pooled estimate toward null.

## Agitation

Four studies ( $n=484$ ), all using PARO, contributed measurements via the BARS [28], CMAI-SF [24], RAID [34], and NPI agitation subdomain [33]. After excluding Mervin et al [27] for cohort overlap with Moyle et al [24], the HKSJ pooled estimate was Hedges  $g=0.32$  (95% CI  $-0.07$  to 0.71; 95% PI  $-0.26$  to 0.89;  $P=29\%$ ;  $\tau^2=0.018$ ;  $P=.08$ ), indicating a small-to-moderate effect direction favoring care robots that approached but did not reach conventional statistical significance. The pool was dominated by 2 large studies with discordant findings: Moyle et al [24] ( $n=271$  in PARO vs Usual Care contrast) reported no significant CMAI-SF difference (adjusted mean difference  $-1.89$ , 95% CI  $-5.81$  to 2.02;  $P=.34$ ), while Petersen et al [34] ( $n=61$ ) reported a significant between-group RAID difference ( $P=.003$ ). Jøranson et al [28] demonstrated a sustained BARS reduction at follow-up (effect estimate 5.5,  $P=.048$ ). The lower heterogeneity ( $P=29\%$ ) reflects the removal of the overlapping cohort data from Mervin et al [27].

## Stress and Pain

Six studies ( $n=504$ ) contributed across markedly heterogeneous clinical contexts: procedural pain in children [22, 45], chronic psychological distress in older adults [20,34], socially induced stress in healthy children [38], and social-affective regulation in children with autism [47]. The HKSJ pooled estimate was Hedges  $g=0.53$  (95% CI  $-0.48$  to 1.53; 95% PI  $-1.57$  to 2.62;  $P=90\%$ ;  $\tau^2=0.510$ ;  $P=.24$ ), reflecting extreme between-study heterogeneity (90%) and a wide PI encompassing both meaningful benefit and meaningful harm. Rossi et al [45] produced the largest individual effect (salivary cortisol decrease in NAO group, Hedges  $g=2.36$ ; 95% CI 1.62-3.11;  $P<.001$ ); Ali et al [22] showed a moderate benefit on Observational Scale of Behavioral Distress–Revised during intravenous insertion (Hedges  $g=0.37$ ;  $P=.047$ ); and Petersen et al [34] reported physiologic stress markers favoring PARO (galvanic skin variability,  $P<.001$ ). However, Kitt et al [38] reported a directionally negative finding using the Positive and Negative Affect Schedule for Children – Short Form positive affect subscale (robot group showed greater decrease poststressor than control, Hedges  $g=-0.52$ ;  $P=.04$ ), and Pollak et al [20] detected no group difference ( $P=.99$ ). Given the  $P$  of 90%, the pooled estimate should be interpreted with caution; the qualitative pattern suggests benefits may be context-dependent (procedural and physiologic stress more responsive than self-reported affect).

## Social-Communicative Skills

Five studies ( $n=151$ ), with all 5 enrolling children with ASD and using humanoid platforms (NAO [19,36,46], CommU [39], and iRobiQ [50]), contributed via the Social Responsiveness Scale [19], Childhood Autism Rating Scale-2 [36], joint attention frequency [39], gesture production [46], and eye contact frequency [50]. The HKSJ pooled estimate was Hedges  $g=0.45$  (95% CI  $-0.52$  to  $1.42$ ; 95% PI  $-1.66$  to  $2.56$ ;  $P=75\%$ ;  $\tau^2=0.453$ ;  $P=.27$ ), indicating a moderate effect direction favoring care robots that did not reach statistical significance under HKSJ adjustment. Three individual studies achieved significance: Kumazaki et al [39] (joint attention frequency, Hedges  $g=1.07$ ;  $P=.01$ ), So et al [46] (gesture production, Hedges  $g=1.07$ ;  $P=.01$ ), and De Korte et al [19] for the Functional self-initiations subtype (overall Hedges  $g=0.52$ ;  $P=.05$  approximately; functional subtype total self-initiations,  $\text{time} \times \text{group}$  estimate  $-0.19$ , 95% CI  $-0.38$  to  $0.02$ ; functional self-initiations,  $\text{time} \times \text{group}$  estimate  $-0.27$ , 95% CI  $-0.50$  to  $-0.04$ ). However, Holeva et al [36] reported no statistically significant between-group difference on Childhood Autism Rating Scale-2 Total (group-by-time interaction  $P=.16$ ), with both groups improving similarly within-subject. Yun et al [50] reported similar improvements in eye contact across robot and therapist groups ( $P=.50$ ). The wide HKSJ-adjusted CI is partly driven by the small overall sample ( $n=151$ ) and the conservative HKSJ small-sample correction.

## Cognitive Function

Four studies ( $n=461$ ) contributed via the Mini-Mental State Examination (MMSE) [33], the Montreal Cognitive Assessment (MoCA) [18], the Short Portable Mental Status Questionnaire [20], and the CATS Test of visuospatial hemineglect in poststroke patients [21]. The HKSJ pooled estimate was Hedges  $g=0.18$  (95% CI  $-0.62$  to  $0.98$ ; 95% PI  $-1.11$  to  $1.47$ ;  $P=71\%$ ;  $\tau^2=0.102$ ;  $P=.52$ ), reflecting a null-centered direction with substantial heterogeneity. Karner et al [21] showed a large effect on hemineglect improvement (Hedges  $g=1.13$ , 95% CI  $0.46$ - $1.81$ ;  $P=.001$ ), but this measure assesses a narrow poststroke cognitive domain rather than general cognitive function in dementia. The remaining 3 studies in older adults with dementia detected no significant between-group differences on global cognition measures (Chen et al [18] MoCA  $P=.07$ ; Valenti-Soler et al [33] MMSE  $P=.28$ ; Pollak et al [20] Short Portable Mental Status Questionnaire  $P=.69$ ). Heterogeneity therefore reflects domain divergence between hemineglect rehabilitation and general cognitive screening rather than within-domain variability.

## Subgroup and Sensitivity Analyses

### Sensitivity Analysis: Data Extraction Ambiguity

During full-text data extraction, we identified an internal inconsistency in Petersen et al [34] regarding the directional interpretation of RAID and CSDD change scores. Table 2 in Petersen et al [34] reports “difference (post-pre)” with positive values for the treatment group (RAID 2.5, SE 0.6; CSDD 2.81, SE 0.4) and small positive values for controls

(RAID 0.55, SE 0.2; CSDD 0.78, SE 0.4); the abstract states scores “were increased in the treatment group” while the conclusion describes “decreased stress and anxiety.” We interpreted these change values as the magnitude of symptom decrease and conducted a sensitivity analysis excluding Petersen et al [34] from each pool to which it contributed. When Petersen et al [34] was excluded, NPS pool ceased to reach statistical significance (Hedges  $g=0.40$ , 95% CI  $-0.11$  to  $0.91$ ;  $P=.095$ ;  $k=5$ ); depression remained near null (Hedges  $g=0.08$ , 95% CI  $-0.14$  to  $0.30$ ;  $k=6$ ); agitation pool point estimate shifted slightly (Hedges  $g=0.23$ , 95% CI  $-0.20$  to  $0.65$ ;  $k=3$ ); stress and pain remained nonsignificant (Hedges  $g=0.48$ , 95% CI  $-0.84$  to  $1.81$ ;  $k=5$ ). We retain Petersen et al [34] in the main analyses while flagging this as a data extraction limitation; importantly, no statistically significant pooled benefit is fully robust to Petersen exclusion, underscoring the modest evidentiary base for the field overall.

### Sensitivity Analysis: LOO and Data Quality

We further conducted LOO sensitivity analyses for the only outcome reaching statistical significance (NPS pool,  $k=6$ ). Under LOO, the pooled effect retained statistical significance in only 1 of 6 iterations (excluding Valenti-Soler 2015 [33]; Hedges  $g=0.54$ , 95% CI  $0.11$ - $0.97$ ;  $P=.03$ ); in the remaining 5 iterations, the 95% CI crossed null (excluding Bradwell et al [52]:  $P=.07$ ; Chen et al [18]:  $P=.09$ ; Jøranson et al [28]:  $P=.09$ ; Moyle et al [24]:  $P=.051$ ; and Petersen et al [34]:  $P=.095$ ). Restricting to studies whose mean and SD values were extracted directly from published tables (class A: Bradwell et al [52], Chen et al [18], Jøranson et al [28];  $k=3$ ) yielded Hedges  $g=0.67$  (95% CI  $-0.10$  to  $1.43$ ;  $P=.07$ ), losing statistical significance owing to the small  $k$  under HKSJ small-sample correction. These LOO results demonstrate that the single statistically significant pooled effect in our analysis is fragile, as it depends on the joint contribution of all 6 [18, 24, 29, 33, 34, 52] contributing studies and is sensitive to the exclusion of any individual study. We therefore interpret the NPS result as suggestive of benefit rather than as definitive evidence, and emphasize that adequately powered, multicenter confirmatory trials are required.

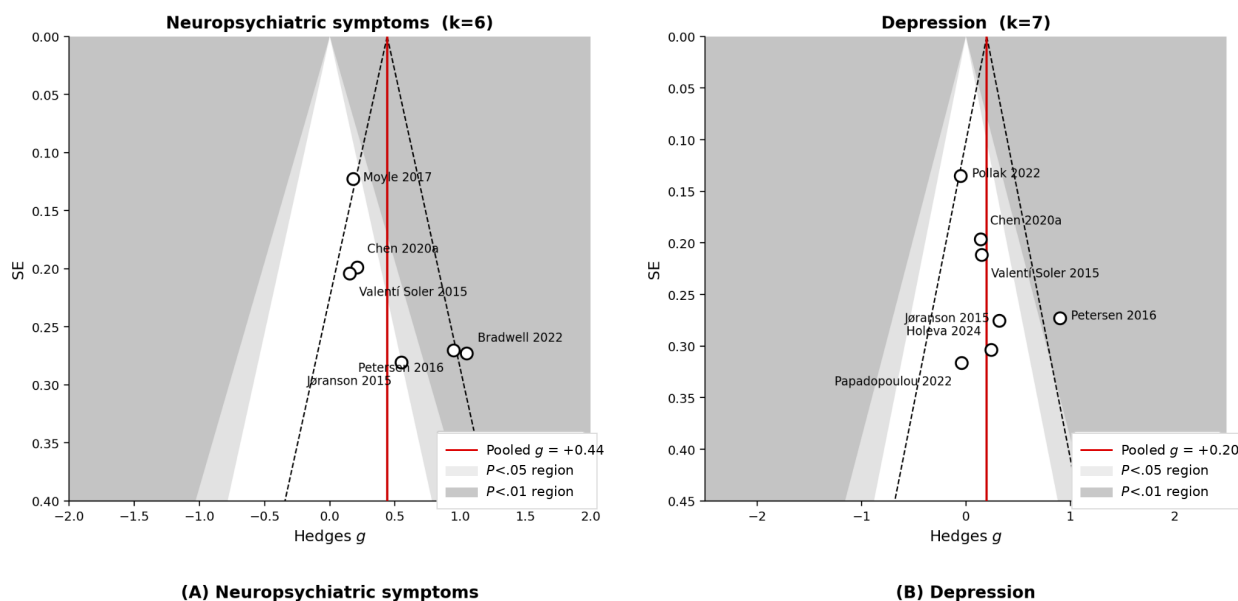
In robot-type subgroup analyses, PARO was associated with the most consistent single-study effects on NPS among older adults with dementia (notably Jøranson et al [28] BARS  $P=.048$ ; Petersen et al [34] RAID  $P=.003$  with the data ambiguity flag noted above), although the small per-arm subgroup  $k$  ( $=4$ ) limits formal subgroup pooling. Humanoid platforms (NAO, Kaspar, CommU, and iRobiQ) supplied all 5 [19,36,39,46,50] contributing studies in the social-communicative skills pool. Ad hoc duration-based subgroup analysis suggested numerically larger effects in studies with intervention durations  $\geq 10$  weeks (eg, Bradwell et al [52] 16 weeks, Jøranson et al [28] 12 weeks, and Petersen et al [34] 12 weeks) compared with shorter pilots, but the small subgroup  $k$  precludes formal heterogeneity testing across this dimension. Sensitivity analyses excluding the 4 [18,20,21,33] studies rated overall high RoB 2 risk of bias did not materially alter pooled point estimates (within  $\pm 0.06$  Hedges  $g$  of primary analyses) and did not reverse any nonsignificant pool to significance.

## Reporting Biases and Small-Study Effects

Contour-enhanced funnel plots for NPS ( $k=6$ ) and depression ( $k=7$ ) are shown in Figure 4. Both domains have  $k < 10$ , below the recommended threshold for formal Egger regression test [60]; accordingly, funnel plots are reported for visual transparency only. Visual inspection of the NPS funnel plot

showed scattering largely consistent with the pooled estimate; the depression funnel plot appeared broadly symmetric. We did not formally test funnel plot asymmetry given the small  $k$ . Sources of small-study effects could include publication bias, methodological differences, true between-study effect heterogeneity, or chance—none of which can be reliably distinguished with  $k < 10$ .

**Figure 4.** Contour-enhanced funnel plots for the 2 domains with the most contributing trials: (A) neuropsychiatric symptoms ( $k=6$ ) and (B) depression ( $k=7$ ). The vertical red dashed line marks the Hartung-Knapp-Sidik-Jonkman pooled estimate; diagonal dashed lines show pseudo 95% confidence limits. Because no domain reached  $k \geq 10$ , asymmetry was not formally tested; plots are shown for visual transparency only [18, 20, 24, 29, 33, 34, 36, 43, 52].



## Certainty of Evidence

Full GRADE ratings are presented in Table 2. In summary, certainty was rated low for one outcome—NPS—reflecting structural performance bias inherent to unblinded robot trials and substantial inconsistency, partially offset by the directional consistency of effect across most contributing studies. Certainty was rated very low for the remaining 6 outcomes—quality of life, depression, agitation, stress and pain, social-communicative skills, and cognitive function—reflecting a combination of nonsignificant or null pooled estimates, imprecision (95% CIs crossing null and wide PIs relative to effect size), and small contributing sample sizes. The Discussion section is framed in light of these certainty ratings—that is, the single statistically significant pooled effect (NPS, sensitive to one study's inclusion) is interpreted as a suggestive signal supporting careful, selective consideration of care robots in narrow indications, rather than as confirmatory evidence supporting general clinical recommendations or health system integration.

## Acceptance and Implementation Outcomes (Secondary)

A minority of included studies reported acceptance- or usability-relevant secondary outcomes (eg, System Usability Scale, Unified Theory of Acceptance and Use of Technology items, observed session completion rates, and facilitator-time

data). Because reporting was sparse, inconsistent, and predominantly descriptive, these outcomes were summarized narratively rather than pooled. Across the studies that reported them, acceptance was generally higher in sessions with dedicated facilitator support and in longer-duration protocols.

## Discussion

### Principal Findings

Relative to the 4 prespecified objectives, this synthesis produced a patterned rather than uniform set of findings. Across the 7 prespecified outcome domains, HKSJ random-effects meta-analysis identified a single domain—NPS—for which the pooled estimate favored care robots and the confidence interval excluded the null. For the remaining 6 domains (quality of life, depression, agitation, stress and pain, social-communicative skills, and cognitive function), the pooled estimates were directionally favorable but did not reach statistical significance. Critically, a distinction must be drawn between the average effect and the distribution of effects across settings; whereas the CI reflects uncertainty about the mean effect, the PI describes the plausible effect in a new setting. In every domain—including NPS—the 95% PI encompassed the null, so although an average benefit is detectable across existing trials for NPS, a beneficial effect

cannot be assumed in a new care setting. The neuropsychiatric-symptom signal was, moreover, fragile, losing significance under most LOO exclusions and upon removal of the single study with documented data extraction ambiguity. Interpreted strictly in light of the substantial between-study heterogeneity observed across pools, the predominance of high or some-concerns RoB 2 ratings, and GRADE certainty ranging from low (NPS) to very low (all other domains), these results indicate that no pooled estimate currently provides a robust basis for clinical decision-making.

### **Interpretation in the Context of Prior Reviews**

Our findings both converge with and extend prior syntheses. For older adults with dementia, the direction and approximate magnitude of the pooled effects we observed for NPS and agitation are consistent with the single-population meta-analyses of Pu et al [11], Leng et al [12], and Yu et al [13], none of which, however, compared these effects with those observed in pediatric populations. For social-communicative skills in children with ASD, our humanoid-platform effect size is directionally consistent with Scassellati et al [6] but, under HKSJ pooling, did not reach statistical significance (Hedges  $g=0.45$ , 95% CI  $-0.52$  to  $1.42$ ). The principal methodological contribution of this review is the direct juxtaposition of platform–population pairings within a single analytic framework. The pattern of variation across robot platforms and populations observed here is therefore best read as a hypothesis-generating descriptive signal rather than as confirmatory evidence; it nonetheless points to platform-population matching, rather than platform-agnostic deployment, as a useful organizing principle for future trial design and procurement decisions.

Our nonsignificant pooled effects on depression and cognitive function are informative rather than merely negative. Depression in dementia is etiologically complex, with substantial neurobiological determinants that brief psychosocial interventions may be insufficient to modify [62]. The borderline signal in our exploratory subgroup restricted to PARO interventions of  $\geq 10$  weeks is hypothesis-generating only and requires prospective adequately powered testing. The null pooled effect for cognitive function is consistent with the broader behavioral-intervention literature and reflects the well-established limitation of noncognitive behavioral inputs in reversing underlying neurodegeneration [40,62-65]; measurement-level sensitivity of global screening instruments (MMSE and MoCA) to detect short-term changes may also contribute.

### **Care Setting as a Moderator**

Setting was not a prespecified subgroup analysis, but the observed effect pattern suggests it warrants interpretation. The domains with the largest pooled effects—NPS and agitation (dementia in long-term care) and social-communicative skills (ASD therapy, predominantly in structured therapy or hospital outpatient settings)—are those in which the intervention was delivered in a structured daily routine with consistent facilitator support. Community and

home-setting deployments were underrepresented ( $n=2$ ) and reported smaller and less consistent effects. The available evidence therefore best supports deployment in settings where the surrounding care structure—scheduled sessions, trained facilitators, documentation routines—can be aligned with the intervention [66], and does not yet speak to unstructured community or home deployment.

### **Targeted Adjunctive Use Cases**

Rather than supporting broad claims about health-system integration, the evidence pattern above—interpreted in the context of low GRADE certainty for the one statistically significant outcome (NPS, sensitive to a single study's inclusion) and very low certainty for the remaining 6 nonsignificant outcomes—is most consistent with 3 narrow adjunctive use cases in which the available evidence and the clinical realities align. These should be regarded as candidate indications meriting further confirmatory trials, not as recommendations for routine clinical adoption.

1. Facilitator-supported calming and engagement sessions for older adults with moderate-to-severe dementia in long-term care. PARO-based sessions of approximately 20-30 minutes, delivered 2-3 times per week by trained staff for  $\geq 10$  weeks, are the intervention parameters most consistently associated with detectable effects on NPS (agitation pooled effect was directionally similar but did not reach conventional statistical significance). GRADE certainty for this configuration is low. The role is adjunctive—complementing, not replacing, person-centered nursing care—and is most appropriate for units where baseline neuropsychiatric-symptom burden is high and nonpharmacological options are prioritized.
2. Procedural distraction in pediatric inpatient settings. Humanoid robots deployed during time-limited, anxiety-provoking procedures (eg, venipuncture, imaging, and preoperative induction) represent a plausible adjunctive use supported by the stress, pain, and engagement literature [67,68], though with very low GRADE certainty. The time-limited nature of the exposure is itself a strength here: novelty effects, which are a limitation in longer-horizon claims, may be an asset in single-session procedural use.
3. Structured social-communicative practice tasks for children with ASD. Humanoid robot-assisted sessions are best positioned as a structured-practice adjunct to established behavioral therapies such as applied behavior analysis or pivotal response treatment, not as a stand-alone therapeutic modality. GRADE certainty is very low (the social-communicative skills pool did not reach statistical significance under HKSJ pooling), and transfer of gains from robot-mediated to human-mediated social interaction remains an unresolved question that future trials should address [69,70]

### **Clinical Effectiveness, Technology Acceptance, and Implementation Feasibility**

The 3 dimensions of clinical effectiveness (does the intervention work under trial conditions?), technology acceptance

(will patients, families, and staff use it?), and implementation feasibility (can the intervention be embedded sustainably in routine care?) are related but distinct, and the evidence base for each differs in maturity. This review quantifies clinical effectiveness and finds a patterned but uneven picture. Technology acceptance is supported by a parallel body of qualitative and survey-based literature [7,8,71] indicating generally positive acceptance when sessions are facilitator-supported, culturally aligned, and adequately introduced; but acceptance ratings are variable across populations and platforms, and this review did not pool these data. Implementation feasibility remains the least-developed dimension: we did not identify studies that reported cost-effectiveness, institutional readiness, or sustained-deployment outcomes with the rigor required for pooled synthesis. Conflating these 3 dimensions risks overstating the readiness of care-robot deployment; they should be reported separately in future trials and reviews.

Acceptability is decisive for translation yet was not poolable in this effectiveness-focused synthesis, and it is especially salient in older adults, in whom acceptance is shaped by prior attitudes, perceived usefulness, and design rather than being a given. Dedicated work in this population—including the development of instruments to measure acceptance of social robotics and willingness-to-interact studies—underscores that acceptance is a distinct, measurable construct that should be assessed in its own right rather than inferred from efficacy [72,73].

### **Clinical and Implementation Implications**

From a digital health and nursing informatics perspective, these findings, interpreted in light of low or very low GRADE certainty across all 7 outcome domains, suggest several conditional implementation implications rather than firm recommendations. First, where care robots are considered for selective adoption in the narrow indications identified above, they would be most appropriately positioned within the broader digital health ecosystem rather than deployed as isolated devices. PARO and humanoid robots equipped with sensor arrays and data-logging capabilities can generate structured patient engagement metrics—interaction duration, motion activity, physiological indicators—that may be interoperable with electronic health records and clinical decision-support systems where local infrastructure permits. The clinical value of such integration has not been formally evaluated in the trials reviewed here [74].

### **Strengths and Limitations**

Strengths of this review include a cross-population, cross-platform quantitative synthesis; use of the HKSJ random-effects estimator with PIs rather than CIs alone; explicit distinction between average effects and their distribution across settings; PRISMA 2020, PRISMA 2020 for Abstracts, and PRISMA-S adherence with all checklists supplied; dual-reviewer screening, extraction, risk-of-bias assessment, and GRADE rating throughout; and a comprehensive search of 5 databases and 2 trial registries without language or date restrictions.

Limitations should be read in light of the cautious claims above. First, the review was not prospectively registered in PROSPERO. Although an a priori protocol was developed and is available on request, the absence of public prospective registration reduces independent verifiability of prespecification claims. Second, performance bias is structurally inherent to care robot research: blinding of participants and care providers to robotic versus nonrobotic interaction is impossible, and this domain consistently represented the highest RoB 2 ratings, rated high or some concerns in all 34 [18-22,24-52] included studies. Attention-matched active control conditions (Moyle et al [24] plush and Yun et al [50] therapist-guided routine) partially mitigate this concern but cannot eliminate it. Third, considerable clinical and methodological heterogeneity remains across pools, ranging from low to high, reflecting diversity in robot platforms, target populations, intervention dose, and outcome instruments. Fourth, the included sample remains modest, with a median per-study analyzed  $n=62$  and only 4 [20,24,25,27] included studies enrolling  $>200$  participants. Fifth, follow-up durations were generally short, with most studies measuring outcomes at end-of-intervention ( $\leq 16$  weeks); long-term sustainability and dose-response relationships remain underexplored. Sixth—and new in this revision—primary-source data verification revealed limitations specific to data extraction: (1) one cohort overlap (Mervin et al [27] with Moyle et al [24] cluster RCT,  $n=415$  Australian residents) was identified during full-text reextraction and resolved via reclassification of Mervin et al [27] to narrative-only synthesis under Senn [58]; (2) Robinson et al [44] was confirmed as a qualitative observational follow-up of an external RCT cohort and reclassified to narrative-only; and (3) Petersen et al [34] published a table whose directional interpretation of RAID and CSDD change scores is inconsistent with the abstract narrative, prompting a sensitivity analysis. When Petersen was excluded, the only statistically significant pool (NPS) became nonsignificant under HKSJ pooling, highlighting the modest evidentiary base. Finally, 3 included studies were small pilot RCTs (Marino et al [41]  $n=14$ ; Yun et al [50]  $n=15$ ; Srinivasan et al [47]  $n=24$ ), and several studies analyzed secondary outcomes rather than preregistered primary endpoints. Together, these limitations argue for caution in interpreting any single pooled estimate and for prioritizing future adequately powered, multicenter RCTs with preregistered primary outcomes. A further limitation specific to data extraction quality is that mean and SD values for 9 of 37 study-outcome contributions (24%) were derived from text or figure approximations rather than from published tables; sensitivity analyses excluding these contributions did not change whether any pool reached statistical significance, but produced moderate changes in pooled point estimates (notably a substantial attenuation of the stress and pain estimate when text or figure-approximated contributions were excluded). To make the link to certainty explicit, these limitations map directly onto the GRADE downgrade domains recorded in Table 2. The structural impossibility of blinding and the predominance of high or some concerns RoB 2 ratings drive the risk-of-bias downgrade; the wide range of  $I^2$  across pools drives the inconsistency downgrade; and the modest per-study sample sizes,

together with CIs crossing the null and wide PIs, drive the imprecision downgrade. Reliance on text- or figure-approximated data for a subset of contributions bears on both the risk of bias and imprecision judgments. The deliberate cross-population pooling, although a strength for breadth, introduces a degree of indirectness when pooled estimates are applied to any single target population; we flag this as a consideration even though it did not by itself trigger a separate GRADE downgrade in our profile.

### **Future Research Directions**

Several priorities emerge directly from the gaps identified above. Multicenter RCTs of adequate size, with attention-matched active controls and assessor-blinded outcome measurement, are the most immediate methodological need; sample sizes should be informed by the HKSJ-corrected effect estimates reported here, which are typically more conservative than earlier DerSimonian-Laird-based values and therefore produce larger sample size targets. Standardization of outcome measurement through a stakeholder-derived core outcome set, developed along lines analogous to the COMET Initiative [8], would substantially improve comparability across future trials. Follow-up periods of 3, 6, and 12 months postintervention are necessary to distinguish short-term engagement from durable clinical benefit—a distinction that current short-duration evidence cannot address. Formal cost-effectiveness analyses and budget-impact modeling are required to inform health technology assessment. Implementation science studies that explicitly measure staff training requirements, institutional readiness, and patient and family acceptance as implementation outcomes in their own right would begin to build the implementation-feasibility evidence base that is currently absent. Cross-cultural replication, particularly in low- and middle-income settings, and mechanistic studies pairing behavioral outcomes with physiological markers (cortisol and heart rate variability) or neuroimaging where feasible, round out the research agenda.

### **Conclusions**

This systematic review and meta-analysis provides one of the broadest quantitative syntheses to date of care-robot RCTs across patient populations and outcome domains, and is best read as an exploratory, hypothesis-generating synthesis rather than a definitive one. Following systematic full-text data verification of all included randomized controlled studies, a statistically significant pooled effect favoring care robots emerged in only one of the 7 prespecified domains, NPS. That signal proved fragile under LOO and data extraction sensitivity analyses, and its PI, like those of the remaining 6 domains, encompassed the null, so a benefit cannot be assumed in any new care setting. Across the other 6 domains, pooled effects favored care robots in direction but did not reach statistical significance. Interpreted strictly in light of substantial between-study heterogeneity, the predominance of high or some concerns risk-of-bias ratings, and GRADE certainty of low (NPS) to very low (all other domains), these findings indicate that the evidence base is more limited than earlier syntheses suggested, a difference attributable to the HKSJ small-sample correction, primary-source data verification, and PI reporting. The current evidence is therefore promising but not yet sufficient to support routine clinical adoption of care robots across the breadth of populations and outcomes examined. Where benefits may be most plausibly realized, neuropsychiatric symptom reduction in older adults with dementia and procedural-distress mitigation in pediatric settings, they should be regarded as candidate indications requiring confirmation, not as established uses. Accordingly, care robots are best positioned as facilitator-supported adjuncts that augment, rather than substitute for, human-delivered care, and broader implementation claims should await adequately powered, multicenter trials with attention-matched controls, blinded outcome assessment, and extended postintervention follow-up.

---

### **Acknowledgments**

No generative artificial intelligence tools were used in the conception, design, conduct, data extraction, statistical analysis, interpretation, drafting, or revision of this manuscript. All content is the sole work of the authors, who reviewed and approved the final text and take full responsibility for it.

---

### **Funding**

This study was funded by Keimyung University Dongsan Medical Center research promotion grant (2023); the funder had no role in this review.

---

### **Data Availability**

The datasets analyzed during this study are available from the corresponding author on reasonable request. All primary data were extracted from previously published, publicly accessible peer-reviewed reports cited in this article.

---

### **Authors' Contributions**

GK conceptualized and designed the study, developed the electronic search strategy, conducted database searches, performed screening and full-text eligibility assessment, extracted and verified all data, assessed risk of bias using the RoB 2 tool, performed meta-analyses including subgroup and sensitivity analyses, evaluated evidence certainty using GRADE, and drafted and critically revised the manuscript. JJ contributed to full-text screening, data extraction verification, and manuscript review. All authors reviewed and approved the final manuscript.

---

### **Conflicts of Interest**

None declared.

### Multimedia Appendix 1

Characteristics of the 34 included randomized and cluster-randomized controlled trials of care robot interventions (per-study; full version), conducted across 17 countries and published 2015–2024. Target populations include older adults with dementia or mild cognitive impairment, children and adolescents with autism spectrum disorder, hospitalized or procedurally distressed pediatric patients, and other adult clinical populations. Robot platforms include PARO, humanoid platforms (NAO, Pepper, Kabochan, CommU, and Kaspar), and other companion and pet platforms. Cohort-overlap allocation per Senn (2009) is documented in the footnote.

[\[DOCX File \(Microsoft Word File\), 33 KB-Multimedia Appendix 1\]](#)

### Checklist 1

PRISMA-S checklist.

[\[DOCX File \(Microsoft Word File\), 20 KB-Checklist 1\]](#)

### Checklist 2

PRISMA 2020 for Abstracts checklist.

[\[DOCX File \(Microsoft Word File\), 269 KB-Checklist 2\]](#)

### Checklist 3

PRISMA Expanded checklist.

[\[DOCX File \(Microsoft Word File\), 25 KB-Checklist 3\]](#)

### References

1. World report on ageing and health. World Health Organization. 2015. URL: <https://www.who.int/publications/i/item/9789241565042> [Accessed 2026-06-17]
2. Global strategy on human resources for health: Workforce 2030. World Health Organization. URL: <https://www.who.int/publications/i/item/9789241511131> [Accessed 2026-06-17]
3. Wada K, Shibata T. Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. *IEEE Trans Robot*. 2007;23(5):972-980. [doi: [10.1109/TRO.2007.906261](https://doi.org/10.1109/TRO.2007.906261)]
4. Biocca F, Harms C, Burgoon JK. Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence (Camb)*. Oct 2003;12(5):456-480. [doi: [10.1162/105474603322761270](https://doi.org/10.1162/105474603322761270)]
5. Kitwood T. *Dementia Reconsidered: The Person Comes First*. Open University Press; 1997. URL: <https://search.worldcat.org/title/1249788841> [Accessed 2026-06-25]
6. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. *Annu Rev Biomed Eng*. 2012;14:275-294. [doi: [10.1146/annurev-bioeng-071811-150036](https://doi.org/10.1146/annurev-bioeng-071811-150036)] [Medline: [22577778](https://pubmed.ncbi.nlm.nih.gov/22577778/)]
7. He Y, He Q, Liu Q. Technology acceptance in socially assistive robots: scoping review of models, measurement, and influencing factors. *J Healthc Eng*. 2022;2022:6334732. [doi: [10.1155/2022/6334732](https://doi.org/10.1155/2022/6334732)] [Medline: [35911583](https://pubmed.ncbi.nlm.nih.gov/35911583/)]
8. Williamson PR, Altman DG, Bagley H, et al. The COMET handbook: version 1.0. *Trials*. Jun 20, 2017;18(Suppl 3):280. [doi: [10.1186/s13063-017-1978-4](https://doi.org/10.1186/s13063-017-1978-4)] [Medline: [28681707](https://pubmed.ncbi.nlm.nih.gov/28681707/)]
9. Ciuffreda I, Amabili G, Casaccia S, et al. Design and development of a technological platform based on a sensorized social robot for supporting older adults and caregivers: GUARDIAN ecosystem. *Int J Soc Robotics*. May 2025;17(5):803-822. [doi: [10.1007/s12369-023-01038-5](https://doi.org/10.1007/s12369-023-01038-5)]
10. Amabili G, Maranesi E, Margaritini A, et al. Managing cognitive decline through a social robot-based intervention: protocol for the engAGE proof of concept and randomized controlled trial. *JMIR Res Protoc*. Sep 2, 2025;14:e67601. [doi: [10.2196/67601](https://doi.org/10.2196/67601)] [Medline: [40896828](https://pubmed.ncbi.nlm.nih.gov/40896828/)]
11. Pu L, Moyle W, Jones C, Todorovic M. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. *Gerontologist*. Jan 9, 2019;59(1):e37-e51. [doi: [10.1093/geront/gny046](https://doi.org/10.1093/geront/gny046)] [Medline: [29897445](https://pubmed.ncbi.nlm.nih.gov/29897445/)]
12. Leng M, Liu P, Zhang P, et al. Pet robot intervention for people with dementia: a systematic review and meta-analysis of randomized controlled trials. *Psychiatry Res*. Jan 2019;271:516-525. [doi: [10.1016/j.psychres.2018.12.032](https://doi.org/10.1016/j.psychres.2018.12.032)] [Medline: [30553098](https://pubmed.ncbi.nlm.nih.gov/30553098/)]
13. Yu C, Sommerlad A, Sakure L, Livingston G. Socially assistive robots for people with dementia: systematic review and meta-analysis of feasibility, acceptability and the effect on cognition, neuropsychiatric symptoms and quality of life. *Ageing Res Rev*. Jun 2022;78:101633. [doi: [10.1016/j.arr.2022.101633](https://doi.org/10.1016/j.arr.2022.101633)] [Medline: [35462001](https://pubmed.ncbi.nlm.nih.gov/35462001/)]
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. Mar 29, 2021;372:n71. [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]

15. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev*. Jan 26, 2021;10(1):39. [doi: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z)] [Medline: [33499930](https://pubmed.ncbi.nlm.nih.gov/33499930/)]
16. Higgins JPT, Thomas J, Chandler J, editor. *Cochrane Handbook for Systematic Reviews of Interventions* Version 64. Cochrane; 2023. URL: <https://training.cochrane.org/handbook> [Accessed 2026-06-19]
17. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-282. [Medline: [23092060](https://pubmed.ncbi.nlm.nih.gov/23092060/)]
18. Chen K, Lou VWQ, Tan KCK, Wai MY, Chan LL. Effects of a humanoid companion robot on dementia symptoms and caregiver distress for residents in long-term care. *J Am Med Dir Assoc*. Nov 2020;21(11):1724-1728. [doi: [10.1016/j.jamda.2020.05.036](https://doi.org/10.1016/j.jamda.2020.05.036)] [Medline: [32713772](https://pubmed.ncbi.nlm.nih.gov/32713772/)]
19. De Korte MW, van den Berk-Smeekens I, van Dongen-Boomsma M, et al. Self-initiations in young children with autism during pivotal response treatment with and without robot assistance. *Autism*. Nov 2020;24(8):2117-2128. [doi: [10.1177/1362361320935006](https://doi.org/10.1177/1362361320935006)] [Medline: [32730096](https://pubmed.ncbi.nlm.nih.gov/32730096/)]
20. Pollak C, Wexler SS, Drury L. Effect of a robotic pet on social and physical frailty in community-dwelling older adults: a randomized controlled trial. *Res Gerontol Nurs*. 2022;15(5):229-237. [doi: [10.3928/19404921-20220830-01](https://doi.org/10.3928/19404921-20220830-01)] [Medline: [36113009](https://pubmed.ncbi.nlm.nih.gov/36113009/)]
21. Karner S, Stenner H, Spate M, Behrens J, Krakow K. Effects of a robot intervention on visuospatial hemineglect in postacute stroke patients: a randomized controlled trial. *Clin Rehabil*. Dec 2019;33(12):1940-1948. [doi: [10.1177/0269215519865993](https://doi.org/10.1177/0269215519865993)] [Medline: [31409126](https://pubmed.ncbi.nlm.nih.gov/31409126/)]
22. Ali S, Manaloor R, Ma K, et al. A randomized trial of robot-based distraction to reduce children's distress and pain during intravenous insertion in the emergency department. *Can J Emerg Med*. Jan 2021;23(1):85-93. [doi: [10.1007/s43678-020-00023-5](https://doi.org/10.1007/s43678-020-00023-5)] [Medline: [33683608](https://pubmed.ncbi.nlm.nih.gov/33683608/)]
23. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. Aug 28, 2019;366:l4898. [doi: [10.1136/bmj.l4898](https://doi.org/10.1136/bmj.l4898)] [Medline: [31462531](https://pubmed.ncbi.nlm.nih.gov/31462531/)]
24. Moyle W, Jones CJ, Murfield JE, et al. Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial. *J Am Med Dir Assoc*. Sep 1, 2017;18(9):766-773. [doi: [10.1016/j.jamda.2017.03.018](https://doi.org/10.1016/j.jamda.2017.03.018)] [Medline: [28780395](https://pubmed.ncbi.nlm.nih.gov/28780395/)]
25. Moyle W, Jones C, Murfield J, et al. Effect of a robotic seal on the motor activity and sleep patterns of older people with dementia, as measured by wearable technology: a cluster-randomised controlled trial. *Maturitas*. Apr 2018;110:10-17. [doi: [10.1016/j.maturitas.2018.01.007](https://doi.org/10.1016/j.maturitas.2018.01.007)] [Medline: [29563027](https://pubmed.ncbi.nlm.nih.gov/29563027/)]
26. Jones C, Moyle W, Murfield J, et al. Does cognitive impairment and agitation in dementia influence intervention effectiveness? Findings from a cluster-randomized-controlled trial with the therapeutic robot, PARO. *J Am Med Dir Assoc*. Jul 2018;19(7):623-626. [doi: [10.1016/j.jamda.2018.02.014](https://doi.org/10.1016/j.jamda.2018.02.014)] [Medline: [29656838](https://pubmed.ncbi.nlm.nih.gov/29656838/)]
27. Mervin MC, Moyle W, Jones C, et al. The cost-effectiveness of using PARO, a therapeutic robotic seal, to reduce agitation and medication use in dementia: findings from a cluster-randomized controlled trial. *J Am Med Dir Assoc*. Jul 2018;19(7):619-622. [doi: [10.1016/j.jamda.2017.10.008](https://doi.org/10.1016/j.jamda.2017.10.008)] [Medline: [29325922](https://pubmed.ncbi.nlm.nih.gov/29325922/)]
28. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Change in quality of life in older people with dementia participating in Paro-activity: a cluster-randomized controlled trial. *J Adv Nurs*. Dec 2016;72(12):3020-3033. [doi: [10.1111/jan.13076](https://doi.org/10.1111/jan.13076)] [Medline: [27434512](https://pubmed.ncbi.nlm.nih.gov/27434512/)]
29. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. *J Am Med Dir Assoc*. Oct 1, 2015;16(10):867-873. [doi: [10.1016/j.jamda.2015.05.002](https://doi.org/10.1016/j.jamda.2015.05.002)] [Medline: [26096582](https://pubmed.ncbi.nlm.nih.gov/26096582/)]
30. Jøranson N, Olsen C, Calogiuri G, Ihlebæk C, Pedersen I. Effects on sleep from group activity with a robotic seal for nursing home residents with dementia: a cluster randomized controlled trial. *Int Psychogeriatr*. Oct 2021;33(10):1045-1056. [doi: [10.1017/S1041610220001787](https://doi.org/10.1017/S1041610220001787)] [Medline: [32985396](https://pubmed.ncbi.nlm.nih.gov/32985396/)]
31. Chen K, Lou VWQ, Tan KCK, Wai MY, Chan LL. Changes in technology acceptance among older people with dementia: the role of social robot engagement. *Int J Med Inform*. Sep 2020;141:104241. [doi: [10.1016/j.jmedinf.2020.104241](https://doi.org/10.1016/j.jmedinf.2020.104241)]
32. van den Berk-Smeekens I, de Korte MWP, van Dongen-Boomsma M, et al. Pivotal response treatment with and without robot-assistance for children with autism: a randomized controlled trial. *Eur Child Adolesc Psychiatry*. Dec 2022;31(12):1871-1883. [doi: [10.1007/s00787-021-01804-8](https://doi.org/10.1007/s00787-021-01804-8)] [Medline: [34106357](https://pubmed.ncbi.nlm.nih.gov/34106357/)]
33. Valentí Soler M, Agüera-Ortiz L, Olazarán Rodríguez J, et al. Social robots in advanced dementia. *Front Aging Neurosci*. 2015;7:133. [doi: [10.3389/fnagi.2015.00133](https://doi.org/10.3389/fnagi.2015.00133)] [Medline: [26388764](https://pubmed.ncbi.nlm.nih.gov/26388764/)]
34. Petersen S, Houston S, Qin H, Tague C, Studley J. The utilization of robotic pets in dementia care. *J Alzheimers Dis*. 2017;55(2):569-574. [doi: [10.3233/JAD-160703](https://doi.org/10.3233/JAD-160703)] [Medline: [27716673](https://pubmed.ncbi.nlm.nih.gov/27716673/)]

35. Feingold-Polak R, Barzel O, Levy-Tzedek S. A robot goes to rehab: a novel gamified system for long-term stroke rehabilitation using a socially assistive robot-methodology and usability testing. *J Neuroeng Rehabil*. Jul 28, 2021;18(1):122. [doi: [10.1186/s12984-021-00915-2](https://doi.org/10.1186/s12984-021-00915-2)] [Medline: [34321035](https://pubmed.ncbi.nlm.nih.gov/34321035/)]
36. Holeva V, Nikopoulou VA, Lytridis C, et al. Effectiveness of a robot-assisted psychological intervention for children with autism spectrum disorder. *J Autism Dev Disord*. Feb 2024;54(2):577-593. [doi: [10.1007/s10803-022-05796-5](https://doi.org/10.1007/s10803-022-05796-5)] [Medline: [36331688](https://pubmed.ncbi.nlm.nih.gov/36331688/)]
37. Johanson DL, Ahn HS, MacDonald BA, et al. The effect of robot attentional behaviors on user perceptions and behaviors in a simulated health care interaction: randomized controlled trial. *J Med Internet Res*. Oct 4, 2019;21(10):e13667. [doi: [10.2196/13667](https://doi.org/10.2196/13667)] [Medline: [31588904](https://pubmed.ncbi.nlm.nih.gov/31588904/)]
38. Kitt ER, Crossman MK, Matijczak A, Burns GB, Kazdin AE. Evaluating the role of a socially assistive robot in children's mental health care. *J Child Fam Stud*. 2021;30(7):1722-1735. [doi: [10.1007/s10826-021-01977-5](https://doi.org/10.1007/s10826-021-01977-5)] [Medline: [34025101](https://pubmed.ncbi.nlm.nih.gov/34025101/)]
39. Kumazaki H, Yoshikawa Y, Yoshimura Y, et al. The impact of robotic intervention on joint attention in children with autism spectrum disorders. *Mol Autism*. 2018;9:46. [doi: [10.1186/s13229-018-0230-8](https://doi.org/10.1186/s13229-018-0230-8)] [Medline: [30202508](https://pubmed.ncbi.nlm.nih.gov/30202508/)]
40. Logan DE, Breazeal C, Goodwin MS, et al. Social robots for hospitalized children. *Pediatrics*. Jul 2019;144(1):e20181511. [doi: [10.1542/peds.2018-1511](https://doi.org/10.1542/peds.2018-1511)] [Medline: [31243158](https://pubmed.ncbi.nlm.nih.gov/31243158/)]
41. Marino F, Chilà P, Sfrassetto ST, et al. Outcomes of a robot-assisted social-emotional understanding intervention for young children with autism spectrum disorders. *J Autism Dev Disord*. Jun 2020;50(6):1973-1987. [doi: [10.1007/s10803-019-03953-x](https://doi.org/10.1007/s10803-019-03953-x)] [Medline: [30852783](https://pubmed.ncbi.nlm.nih.gov/30852783/)]
42. Papadopoulos C, Castro N, Nigath A, et al. The CARESSES randomised controlled trial: exploring the health-related impact of culturally competent artificial intelligence embedded into socially assistive robots and tested in older adult care homes. *Int J Soc Robot*. 2022;14(1):245-256. [doi: [10.1007/s12369-021-00781-x](https://doi.org/10.1007/s12369-021-00781-x)] [Medline: [33907589](https://pubmed.ncbi.nlm.nih.gov/33907589/)]
43. Papadopoulou MT, Karageorgiou E, Kechayas P, et al. Efficacy of a robot-assisted intervention in improving learning performance of elementary school children with specific learning disorders. *Children (Basel)*. Jul 31, 2022;9(8):1155. [doi: [10.3390/children9081155](https://doi.org/10.3390/children9081155)] [Medline: [36010046](https://pubmed.ncbi.nlm.nih.gov/36010046/)]
44. Robinson H, Broadbent E, MacDonald B. Group sessions with Paro in a nursing home: structure, observations and interviews. *Australas J Ageing*. Jun 2016;35(2):106-112. [doi: [10.1111/ajag.12199](https://doi.org/10.1111/ajag.12199)] [Medline: [26059390](https://pubmed.ncbi.nlm.nih.gov/26059390/)]
45. Rossi S, Santini SJ, Di Genova D, et al. Using the social robot NAO for emotional support to children at a pediatric emergency department: randomized clinical trial. *J Med Internet Res*. Jan 13, 2022;24(1):e29656. [doi: [10.2196/29656](https://doi.org/10.2196/29656)] [Medline: [34854814](https://pubmed.ncbi.nlm.nih.gov/34854814/)]
46. So WC, Wong MKY, Lam CKY, et al. Using a social robot to teach gestural recognition and production in children with autism spectrum disorders. *Disabil Rehabil Assist Technol*. Aug 2018;13(6):527-539. [doi: [10.1080/17483107.2017.1344886](https://doi.org/10.1080/17483107.2017.1344886)] [Medline: [28673117](https://pubmed.ncbi.nlm.nih.gov/28673117/)]
47. Srinivasan SM, Park IK, Neelly LB, Bhat AN. A comparison of the effects of rhythm and robotic interventions on repetitive behaviors and affective states of children with autism spectrum disorder (ASD). *Res Autism Spectr Disord*. Oct 1, 2015;18:51-63. [doi: [10.1016/j.rasd.2015.07.004](https://doi.org/10.1016/j.rasd.2015.07.004)] [Medline: [26251668](https://pubmed.ncbi.nlm.nih.gov/26251668/)]
48. Stoevesandt D, Jahn P, Watzke S, et al. Comparison of acceptance and knowledge transfer in patient information before an MRI exam administered by humanoid robot versus a tablet computer: a randomized controlled study. *Rofo*. Aug 2021;193(8):947-954. [doi: [10.1055/a-1382-8482](https://doi.org/10.1055/a-1382-8482)] [Medline: [34111898](https://pubmed.ncbi.nlm.nih.gov/34111898/)]
49. Wei CW, Kao HY, Wu WH, Chen CY, Fu HP. The influence of robot-assisted learning system on health literacy and learning perception. *Int J Environ Res Public Health*. Oct 21, 2021;18(21):11053. [doi: [10.3390/ijerph182111053](https://doi.org/10.3390/ijerph182111053)] [Medline: [34769571](https://pubmed.ncbi.nlm.nih.gov/34769571/)]
50. Yun SS, Choi J, Park SK, Bong GY, Yoo H. Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system. *Autism Res*. Jul 2017;10(7):1306-1323. [doi: [10.1002/aur.1778](https://doi.org/10.1002/aur.1778)] [Medline: [28464438](https://pubmed.ncbi.nlm.nih.gov/28464438/)]
51. Boumans R, van Meulen F, van Aalst W, et al. Quality of care perceived by older patients and caregivers in integrated care pathways with interviewing assistance from a social robot: noninferiority randomized controlled trial. *J Med Internet Res*. Sep 9, 2020;22(9):e18787. [doi: [10.2196/18787](https://doi.org/10.2196/18787)] [Medline: [32902387](https://pubmed.ncbi.nlm.nih.gov/32902387/)]
52. Bradwell H, Edwards KJ, Winnington R, Thill S, Allgar V, Jones RB. Implementing affordable socially assistive pet robots in care homes before and during the COVID-19 pandemic: stratified cluster randomized controlled trial and mixed methods study. *JMIR Aging*. Aug 24, 2022;5(3):e38864. [doi: [10.2196/38864](https://doi.org/10.2196/38864)] [Medline: [35830959](https://pubmed.ncbi.nlm.nih.gov/35830959/)]
53. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Taylor and Francis; 1988. URL: <https://search.worldcat.org/title/17877467> [Accessed 2026-06-19]
54. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med*. Dec 30, 2001;20(24):3875-3889. [doi: [10.1002/sim.1009](https://doi.org/10.1002/sim.1009)] [Medline: [11782040](https://pubmed.ncbi.nlm.nih.gov/11782040/)]

55. Int'Hout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. Feb 18, 2014;14:25. [doi: [10.1186/1471-2288-14-25](https://doi.org/10.1186/1471-2288-14-25)] [Medline: [24548571](https://pubmed.ncbi.nlm.nih.gov/24548571/)]
56. Borenstein M. How to understand and report heterogeneity in a meta-analysis: the difference between I-squared and prediction intervals. *Integr Med Res*. Dec 2023;12(4):101014. [doi: [10.1016/j.imr.2023.101014](https://doi.org/10.1016/j.imr.2023.101014)] [Medline: [38938910](https://pubmed.ncbi.nlm.nih.gov/38938910/)]
57. Nagashima K, Noma H, Furukawa TA. Prediction intervals for random-effects meta-analysis: a confidence distribution approach. *Stat Methods Med Res*. Jun 2019;28(6):1689-1702. [doi: [10.1177/0962280218773520](https://doi.org/10.1177/0962280218773520)] [Medline: [29745296](https://pubmed.ncbi.nlm.nih.gov/29745296/)]
58. Senn SJ. Overstating the evidence: double counting in meta-analysis and related problems. *BMC Med Res Methodol*. Feb 13, 2009;9:10. [doi: [10.1186/1471-2288-9-10](https://doi.org/10.1186/1471-2288-9-10)] [Medline: [19216779](https://pubmed.ncbi.nlm.nih.gov/19216779/)]
59. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. Sep 13, 1997;315(7109):629-634. [doi: [10.1136/bmj.315.7109.629](https://doi.org/10.1136/bmj.315.7109.629)] [Medline: [9310563](https://pubmed.ncbi.nlm.nih.gov/9310563/)]
60. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. Jul 22, 2011;343:d4002. [doi: [10.1136/bmj.d4002](https://doi.org/10.1136/bmj.d4002)] [Medline: [21784880](https://pubmed.ncbi.nlm.nih.gov/21784880/)]
61. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. Apr 26, 2008;336(7650):924-926. [doi: [10.1136/bmj.39489.470347.AD](https://doi.org/10.1136/bmj.39489.470347.AD)] [Medline: [18436948](https://pubmed.ncbi.nlm.nih.gov/18436948/)]
62. Zhang Y, Cai J, An L, et al. Does music therapy enhance behavioral and cognitive function in elderly dementia patients? A systematic review and meta-analysis. *Ageing Res Rev*. May 2017;35:1-11. [doi: [10.1016/j.arr.2016.12.003](https://doi.org/10.1016/j.arr.2016.12.003)] [Medline: [28025173](https://pubmed.ncbi.nlm.nih.gov/28025173/)]
63. Hu M, Zhang P, Leng M, Li C, Chen L. Animal-assisted intervention for individuals with cognitive impairment: a meta-analysis of randomized controlled trials and quasi-randomized controlled trials. *Psychiatry Res*. Feb 2018;260:418-427. [doi: [10.1016/j.psychres.2017.12.016](https://doi.org/10.1016/j.psychres.2017.12.016)] [Medline: [29268204](https://pubmed.ncbi.nlm.nih.gov/29268204/)]
64. Lyketsos CG, Lopez O, Jones B, Fitzpatrick AL, Breitner J, DeKosky S. Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: results from the cardiovascular health study. *JAMA*. Sep 25, 2002;288(12):1475-1483. [doi: [10.1001/jama.288.12.1475](https://doi.org/10.1001/jama.288.12.1475)] [Medline: [12243634](https://pubmed.ncbi.nlm.nih.gov/12243634/)]
65. Bahar-Fuchs A, Martyr A, Goh AMY, Sabates J, Clare L. Cognitive training for people with mild to moderate dementia. *Cochrane Database Syst Rev*. Mar 25, 2019;3(3):CD013069. [doi: [10.1002/14651858.CD013069.pub2](https://doi.org/10.1002/14651858.CD013069.pub2)] [Medline: [30909318](https://pubmed.ncbi.nlm.nih.gov/30909318/)]
66. Koh WQ, Felding SA, Budak KB, Toomey E, Casey D. Barriers and facilitators to the implementation of social robots for older adults and people with dementia: a scoping review. *BMC Geriatr*. Jun 9, 2021;21(1):351. [doi: [10.1186/s12877-021-02277-9](https://doi.org/10.1186/s12877-021-02277-9)] [Medline: [34107876](https://pubmed.ncbi.nlm.nih.gov/34107876/)]
67. Beran TN, Ramirez-Serrano A, Vanderkooi OG, Kuhn S. Reducing children's pain and distress towards flu vaccinations: a novel and effective application of humanoid robotics. *Vaccine*. Jun 7, 2013;31(25):2772-2777. [doi: [10.1016/j.vaccine.2013.03.056](https://doi.org/10.1016/j.vaccine.2013.03.056)] [Medline: [23623861](https://pubmed.ncbi.nlm.nih.gov/23623861/)]
68. Jibb LA, Birnie KA, Nathan PC, et al. Using the MEDiPORT humanoid robot to reduce procedural pain and distress in children with cancer: a pilot randomized controlled trial. *Pediatr Blood Cancer*. Sep 2018;65(9):e27242. [doi: [10.1002/pbc.27242](https://doi.org/10.1002/pbc.27242)] [Medline: [29893482](https://pubmed.ncbi.nlm.nih.gov/29893482/)]
69. Pennisi P, Tonacci A, Tartarisco G, et al. Autism and social robotics: a systematic review. *Autism Res*. Feb 2016;9(2):165-183. [doi: [10.1002/aur.1527](https://doi.org/10.1002/aur.1527)] [Medline: [26483270](https://pubmed.ncbi.nlm.nih.gov/26483270/)]
70. Kouroupa A, Laws KR, Irvine K, Mengoni SE, Baird A, Sharma S. The use of social robots with children and young people on the autism spectrum: a systematic review and meta-analysis. *PLoS One*. 2022;17(6):e0269800. [doi: [10.1371/journal.pone.0269800](https://doi.org/10.1371/journal.pone.0269800)] [Medline: [35731805](https://pubmed.ncbi.nlm.nih.gov/35731805/)]
71. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP. Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open*. Feb 12, 2018;8(2):e018815. [doi: [10.1136/bmjopen-2017-018815](https://doi.org/10.1136/bmjopen-2017-018815)] [Medline: [29440212](https://pubmed.ncbi.nlm.nih.gov/29440212/)]
72. Bevilacqua R, Di Rosa M, Riccardi GR, et al. Design and development of a scale for evaluating the acceptance of social robotics for older people: the robot era inventory. *Front Neurobot*. 2022;16:883106. [doi: [10.3389/fnbot.2022.883106](https://doi.org/10.3389/fnbot.2022.883106)] [Medline: [35874107](https://pubmed.ncbi.nlm.nih.gov/35874107/)]
73. Bevilacqua R, Felici E, Cavallo F, Amabili G, Maranesi E. Designing acceptable robots for assisting older adults: a pilot study on the willingness to interact. *Int J Environ Res Public Health*. Oct 12, 2021;18(20):10686. [doi: [10.3390/ijerph182010686](https://doi.org/10.3390/ijerph182010686)] [Medline: [34682433](https://pubmed.ncbi.nlm.nih.gov/34682433/)]
74. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. Jan 2019;25(1):44-56. [doi: [10.1038/s41591-018-0300-7](https://doi.org/10.1038/s41591-018-0300-7)] [Medline: [30617339](https://pubmed.ncbi.nlm.nih.gov/30617339/)]

## Abbreviations

**ASD:** autism spectrum disorder

**BARS:** Behavioral Activity Rating Scale

**CMAI-SF:** Cohen-Mansfield Agitation Inventory–Short Form  
**CSDD:** Cornell Scale for Depression in Dementia  
**GDS:** Geriatric Depression Scale  
**GDS-SF:** Geriatric Depression Scale – Short Form  
**GRADE:** Grading of Recommendations Assessment, Development and Evaluation  
**HKSJ:** Hartung-Knapp-Sidik-Jonkman  
**ICTRP:** International Clinical Trials Registry Platform  
**LOO:** leave-one-out  
**MeSH:** Medical Subject Headings  
**MMSE:** Mini-Mental State Examination  
**MoCA:** Montreal Cognitive Assessment  
**NPI:** Neuropsychiatric Inventory  
**NPS:** neuropsychiatric symptoms  
**PI:** prediction interval  
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
**PRISMA-S:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search  
**PROSPERO:** International Prospective Register of Systematic Reviews  
**RAID:** Rating Anxiety in Dementia  
**RCT:** randomized controlled trial  
**ROB 2:** Risk of Bias 2  
**WHO:** World Health Organization

*Edited by Stefano Brini; peer-reviewed by Achmad Hamdan, Giulio Amabili, Vikas Patil; submitted 13.Mar.2026; final revised version received 09.Jun.2026; accepted 10.Jun.2026; published 30.Jun.2026*

*Please cite as:*

*Kim G, Jeon J*

*Care Robots as Emerging Health Technologies: Systematic Review and Meta-Analysis of Randomized Controlled Trials*  
*J Med Internet Res 2026;28:e95232*

*URL: <https://www.jmir.org/2026/1/e95232>*

*doi: [10.2196/95232](https://doi.org/10.2196/95232)*

© Gaeun Kim, Jinmi Jeon. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.