Supplementary Material A

Table A.1 Risk of bias assessment using the RoB 2 tool for RCTs

Study	D1		D2		D3		D4		D5		Overall	
	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
Oh et al. (2025)	!	+	!	+	+	+	-	-	!	-	-	-
Collombon et al. (2024)	+	+	+	+	-	+	+	+	!	!	-	!
Wlasak et al. (2023)	!	!	+	+	+	!	+	!	!	!	!	-
Hajna et al. (2021)	!	+	!	!	-	+	+	!	-	!	-	!
Bickmore et al. (2013)	!	+	+	!	+	+	-	!	!	+	-	!

Table A.1 (caption): Risk of bias (RoB) judgments for each domain (D1–D5) and overall assessment for included RCTs, evaluated using the revised tool to assess risk of bias in randomized trials (RoB 2). Ratings were independently provided by two reviewers (R1 and R2). Domains: D1: Bias arising from the randomization process, D2: Bias due to deviations from intended interventions, D3: Bias due to missing outcome data, D4: Bias in measurement of the outcome, D5: Bias in selection of the reported result. Risk of bias judgments are reported as Low risk (+), Some concerns (!), or High risk (-) for each study and domain.

Table A.2 RoB 2 signalling questions and responses for included RCTs.

Domain		Question	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
			Oh et al. (2025)		Collombon et al. (2024)		Wlasak et al. (2023)		Hajna et al. (2021)		Bickmore et al. (2013)	
D1	1.1	Was the allocation sequence random?	PY	Y	Y	Y	NI	Y	PY	Y	Y	Y
	1.2	Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	NI	PY	Y	PY	NI	NI	NI	PY	NI	PY
	1.3	Did baseline differences between	N	NI	N	N	NI	NI	NI	NI	NI	NI

		intervention groups suggest a problem with the randomization process?										
D2	2.1	Were participants aware of their assigned intervention during the trial?	Y	PY	Y	PY	PY	NI	Y	NI	Y	PN
	2.2	Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	Y	PY	PY	PY	PY	PY	PY	PY	PY	PY
	2.3	If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the experimental context?	NI	PN	N	PN	PN	PN	PY	NI	PN	NI
	2.4	If Y/PY to 2.3: Were these deviations likely to have affected the outcome?	NA	NA	NA	NA	NA	NA	NI	NA	NA	NA
	2.5	If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?	NA	NA	NA	NA	NA	NA	PY	NA	NA	NA
	2.6	Was an appropriate analysis used to estimate the effect of assignment to intervention?	PY	PY	Y	PY	PY	PY	Y	Y	Y	PY

	2.7	If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
D3	3.1	Were data for this outcome available for all, or nearly all, participants randomized?	Y	PN	N	PN	N	PN	PN	PN	NI	PY
	3.2	If N/PN/NI to 3.1: Is there evidence that result was not biased by missing outcome data?	NA	PN	PN	PN	PY	N	PN	PN	PY	NA
	3.3	If N/PN to 3.2: Could missingness in the outcome depend on its true value?	NA	PN	Y	PN	NA	NI	PY	PN	NA	NA
	3.4	If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	NA	NA	Y	NA	NA	PN	PY	NA	NA	NA
D4	4.1	Was the method of measuring the outcome inappropriate?	PN	PN	N	PN	N	PN	PN	N	PN	N
	4.2	Could measurement or ascertainment of the outcome have differed between intervention groups?	PN	PY	N	PN	N	NI	PN	NI	N	NI

	4.3	Were outcome assessors aware of the intervention received by study participants?	Y	NA	PY	PY	N	PY	N	NI	Y	PY
	4.4	If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	PY	NA	N	PN	NA	PN	NA	PN	Y	NI
	4.5	If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	NI	NA	NA	NA	NA	NA	NA	NA	Y	PN
D5	5.1	Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	NI	PY	NI	PY	NI	PY	PN	PN	NI	PY
	5.2	... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	PN	PY	PN	PN	NI	NI	PN	PN	NI	PN
	5.3	... multiple eligible analyses of the data?	PN	PN	NI	NI	NI	NI	PY	PN	NI	PN

Table A.2 (caption): Responses to the signalling questions from the Risk of Bias 2 (RoB 2) tool, categorized by domain (D1–D5) and reported for each included RCT. Ratings were independently assigned by two reviewers (R1 and R2). The answers to signalling questions guide domain-level risk-of-bias judgments in Table

A.1. Domains: D1: Bias arising from the randomization process, D2: Bias due to deviations from intended interventions, D3: Bias due to missing outcome data, D4: Bias in measurement of the outcome, D5: Bias in selection of the reported result. Response codes: Y = Yes, PY = Probably Yes, N = No, PN = Probably No, NI = No Information, NA = Not Applicable.

Supplementary Material B

Table B.1 Risk of bias assessment using the ROBINS-I tool for non-randomized studies

Study	D1		D2		D3		D4		D5		D6		D7		Overall	
	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
To et al. (2021)	S	S	L	L	L	L	M	M	S	S	L	M	L	M	S	S
Dhinagaran et al. (2021)	S	S	S	S	M	M	M	M	C	S	S	M	M	S	C	C
Maher et al. (2020)	C	S	L	L	L	M	M	M	S	M	M	M	S	L	C	S

Table B.1 (caption): Risk of bias (RoB) judgments for each domain (D1–D7) and the overall assessment for included non-randomized studies, evaluated using the ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool. Ratings were independently provided by two reviewers (R1 and R2). Domains: D1: Bias due to confounding, D2: Bias in classification of interventions, D3: Bias in selection of participants into the study (or into the analysis), D4: Bias due to deviations from intended interventions, D5: Bias due to missing data, D6: Bias in measurement of the outcome, D7: Bias in selection of the reported result. Judgement levels: Low (L), Moderate (M), Serious (S), Critical (C).

Table B.2 Responses to signalling questions for ROBINS-I risk of bias assessment.

Domain		Question	R1	R2	R1	R2	R1	R2
			To et al. (2021)		Dhinagaran et al. (2021)		Maher et al. (2020)	
D1	1.1	Did the authors control for all the important confounding factors for which this was necessary?	WN	WN	NI	NI	SN	SN
	1.2	If Y/PY/WN to 1.1: Were confounding factors that were controlled for (and for which control was necessary) measured validly and reliably by the variables available in this study?	NI	NI	NA	NA	NA	NA

	1.3	If Y/PY/WN to 1.1: Did the authors control for any post-intervention variables that could have been affected by the intervention?	PN	N	NA	NA	NA	NA
	1.4	Did the use of negative controls, quantitative bias analysis, or other considerations, suggest serious unmeasured confounding?	Y	PY	N	PN	Y	PN
D2	2.1	Did assignment of participants to the intervention group or the comparator group rely on events or measurements that occurred after the start of follow up?	PN	N	NI	NI	N	N
	2.2	If Y/PY to 2.1: Were participants included in the comparator group until they fulfilled the definition of the intervention (or vice versa)?	NA	NA	NI	NI	NA	NA
	2.3	If N/PN to 2.1: Was all information used to classify intervention and comparator groups recorded at or before the time the interventions started?	Y	PY	NA	NA	Y	PY
	2.4	Was classification of intervention status influenced by knowledge of the outcome or risk of the outcome?	N	PN	NI	NI	N	PN
	2.5	If N/PN to 2.1 and WY/N/PN/NI 2.4: Was intervention status classified correctly for all, or nearly all, participants?	Y	Y	NA	NA	Y	PY
D3	3.1	Did assignment of participants to the intervention group or the comparator group rely on events	PN	N	NI	NI	N	PN

		or measurements that occurred after the start of follow up?						
	3.2	If Y/PY to 3.1: Were participants excluded after the start of follow-up because they did not meet the definition of either the intervention or the comparator?	NA	NA	NA	NA	NA	NA
	3.3	Were start of follow up and start of intervention the same for most participants?	Y	PY	NI	NI	Y	NI
	3.4	If N/PN to 3.3: Is the effect of intervention expected to be constant over the time period studied?	NA	NA	NA	NA	NA	NA
	3.5	Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention (additional to the situations addressed in 3.1 and 3.3)?	PN	N	NI	NI	N	PN
	3.6	If Y/PY to 3.5: Were the post-intervention variables that influenced selection likely to be associated with intervention?	NA	NA	NA	NA	NA	NA
	3.7	If Y/PY to 3.6: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	NA	NA	NA	NA	NA	NA
	3.8	If Y/PY to 3.2, N/PN 3.4 or Y/PY to 3.7: Is it likely that the analysis corrected for all of the potential selection biases identified in 3.1-3.2, 3.3-3.4 or 3.5-3.7 above?	NA	NA	NA	NA	NA	NA

	3.9	If N/PN to 3.8: Did sensitivity analyses demonstrate that the likely impact of the potential selection biases identified in 3.1-3.2, 3.3-3.4 or 3.5-3.7 above was minimal?	NA	NA	NA	NA	NA	NA
	3.10	If N/PN to 3.9: Were potential selection biases identified in 3.1-3.2, 3.3-3.4 or 3.5-3.7 above sufficiently severe that the result should not be included in a quantitative synthesis?	NA	NA	NA	NA	NA	NA
D4	4.1	Was the study undertaken in an experimental context?	PY	Y	PY	Y	N	PN
	4.2	If Y/PY to 4.1: Did participants deviate from the intended intervention as a result of the processes of recruiting and engaging them in the study?	N	PN	PN	NI	NA	NA
	4.3	If Y/PY to 4.1: Did study personnel consciously or unconsciously undermine implementation of the intended interventions?	N	PN	PN	NI	NA	NA
	4.4	If Y/PY/NI to 4.2 or 4.3: Were these deviations from intended intervention likely to have affected the outcome?	NA	NA	NA	NA	NA	NA
	4.5	Was an appropriate analysis used to estimate the effect of assignment to intervention?	NI	NI	WN	NI	WN	WN
D5	5.1	Were complete data on intervention status available for all, or nearly all, participants?	NI	NI	PN	PN	Y	PY
	5.2	Were complete data on the outcome available for all, or nearly all, participants?	NI	NI	N	PN	N	PN

	5.3	Were complete data on important confounding variables available for all, or nearly all, participants?	NI	PY	PN	NI	Y	PY
	5.4	If N/PN/NI to 5.1, 5.2 or 5.3: Is the result based on a complete case analysis?	PY	NI	PY	PY	N	PN
	5.5	If Y/PY/NI to 5.4: Was exclusion from the analysis because of missing data (in intervention, confounders or the outcome) likely to be related to the true value of the outcome?	PY	Y	PY	PY	NA	NA
	5.6	If Y/PY/NI to 5.5: Is the relationship between the outcome and missingness likely to be explained by the variables in the analysis model?	WN	NI	SN	NI	NA	NA
	5.7	If N/PN to 5.4: Was the analysis based on imputing missing values?	NA	NA	NA	NA	Y	PY
	5.8	If Y/PY to 5.7: Is it reasonable to assume that data were 'missing at random' (MAR) or 'missing completely at random' (MCAR)?	NA	NA	NA	NA	N	PY
	5.9	If Y/PY to 5.8: Was imputation performed appropriately?	NA	NA	NA	NA	NA	NI
	5.10	If N/PN/NI to 5.7: Was an appropriate alternative method used to correct for bias due to missing data?	NA	NA	NA	NA	NA	NA
	5.11	If PN/N/NI to 5.1, 5.2 or 5.3 AND (Y/PY/NI to 5.5 OR (Y/PY to 5.8 AND WN/SN/NI to 5.9) OR WN/SN/NI to 5.10):	N	PN	N	PN	NA	PY

		Is there evidence that the result was not biased by missing data?						
D6	6.1	Could measurement or ascertainment of the outcome have differed between intervention groups?	N	PN	N	PN	N	PN
	6.2	Were outcome assessors aware of the intervention received by study participants?	N	NI	Y	PY	Y	PY
	6.3	If Y/PY/NI to 6.2: Could assessment of the outcome have been influenced by knowledge of the intervention received?	NA	NI	SY	NI	WY	NI
D7	7.1	Was the result reported in accordance with an available, pre-determined analysis plan?	NI	NI	NI	NI	NI	NI
	7.2	Is the numerical result being assessed likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	PN	NI	NI	NI	NI	PN
	7.3	Is the numerical result being assessed likely to have been selected, on the basis of the results from multiple analyses of the data?	PN	N	NI	NI	PY	PN
	7.4	Is the numerical result being assessed likely to have been selected, on the basis of the results from multiple subgroups?	PN	PN	N	NI	N	PN

Table B.2 (caption): Responses to signalling questions for each domain (D1–D7) of the ROBINS-I tool used to evaluate risk of bias in non-randomized studies. Responses were independently provided by two reviewers (R1 and R2). These questions support domain-level judgments presented in Table B.1. Domains: D1:

Bias due to confounding, D2: Bias in classification of interventions, D3: Bias in selection of participants into the study (or into the analysis), D4: Bias due to deviations from intended interventions, D5: Bias due to missing data, D6: Bias in measurement of the outcome, D7: Bias in selection of the reported result. Response codes: Y: Yes, PY: Probably Yes, N: No, PN: Probably No, NI: No Information, NA: Not Applicable, WN: No, but uncontrolled confounding was probably not substantial, SN: No, and uncontrolled confounding was probably substantial, WY: Yes, but the impact was not substantial, SY: Yes, and the impact was substantial.