Appendix 1 Summary Protocol

Study Inclusion criteria

Studies were eligible for inclusion if they met the following criteria:

i) Any intervention delivered directly to individuals where the primary aim is to alleviate loneliness (if there are multiple primary aims, it must be one of three or fewer);
ii) Intervention must be based in an OECD country;
iii) Loneliness must be measured using a validated and standardised measure;
iv) Loneliness must be measured before and after intervention (pre/post design);
v) English language;

The exclusion criteria included:

i) Intervention where alleviating loneliness is a secondary aim or one of 4+ primary aims;
ii) Intervention is based in a non-OECD country;
iii) Do not use validated and a standardised measure of loneliness (including qualitative ascertainment);
iv) Loneliness not measure before and after the intervention;
v) Non-English language;
vi) Medical-based intervention (i.e. pharmaceutical) or delivered within a hospital setting.
vii) Inappropriate record type (i.e. news article, book, dissertation)

Table 1. Overview of study inclusion criteria using the PICOS framework

Element	Inclusion criteria
Population	Intervention takes place in an OECD country.
Intervention	Any intervention that is delivered directly to people where the primary aim is to alleviate loneliness.
Comparator/ Control	Present comparison data from a control group (i.e. no intervention or usual care), or historical time-based comparators (i.e. pre-post test data).
Outcome	Must report loneliness outcomes using a standardised/validated quantitative measure.
Study design	Experimental; Quasi-Experimental; Comparative Before and after study; Mixed-method

Search strategy

The search strategy consisted of two arms: traditional academic databases and grey literature. Both arms combined two key constructs 'loneliness' and 'intervention' using Boolean operators, truncation symbols and MeSH terms/mapped subject headings (see Figure 1). Truncation symbols enabled various spellings of a given phrase to be included; for example, 'program*' would capture 'programme", 'program', 'programe', 'programs', etc.

Searches were conducted across three academic databases (Ovid Medline, ERIC, PsychInfo) using keywords and MeSH terms/mapped subject headings and were restricted to 2008 onwards. Grey literature searches were conducted across the electronic databases and websites/online repositories listed in Table 1 below.

Table 2 Grey literature sources

Electronic databases
Social Science Research Network
SCIE online
OAIster/Worldcat.org
PsychEXTRA
Google (Advanced; Scholar)
Websites and online repositories
Local and central government outputs; Evidence centre websites/repositories
Charity sector funder websites/repositories
Evaluation repositories
Organisational websites

Data Extraction

For all articles that met the inclusion criteria, a single reviewer independently extracted all study data. All extracted data was confirmed by a second member of the review team against the original record. Both members of the review team independently conducted the 10-item critical appraisal (more detail below), with a third member resolving any discrepancies. Where separate records provided data for the same study sample and intervention, both records were used to extract a single set of data for that intervention; if sample size differed, results for the larger sample size were extracted. Where records reported multiple interventions or multiple sample groups (i.e. mentees, mentors), results were extracted separately. The data extraction table consisted of the following components:

Table 3 Data extraction template

Sample characteristics		
Short sample description	[free text]	
Country	[free text]	
Age	Mean, standard deviation, minimum, maximum	1 or more of the following categorical age groups were selected: children (0-10), adolescents (11-18), young adults (19-25), adults (26-49), older adults (50+), older adults (60+), older adults (75+), other
Control group	Yes/no	
Randomisation	Individual randomisation, cluster randomisation, wait-list control group, no randomisation or no wait-list, other:	
Intervention characteristics		
Short intervention description	duration, frequency, time scale, description [free text]	1 or more of the following categorical age groups were selected: children (0-10), adolescents (11-18), young adults (19-25), adults (26-49), older adults (50+), older adults (60+), older adults (75+), other
Overarching theme	Yes/no	
Sub-theme	[developed inductively]	
Setting	Community-based, Healthcare (clinical	

	and/or Social care); Education	
Results	N, mean, SD pre and post intervention (in intervention and control group) for continuous measures N(%) for single-item measures [free text summary of other results]	
Loneliness scale used	UCLA Scales; De Jong Gierveld Scale; Self-report single item, Other	
Primary results of interest	Mean, standard deviation and sample size pre-intervention and post-intervention in intervention group and control group (where applicable)	
Subgroup information	SDs were extracted by subgroup (e.g. gender, age, etc.) and loneliness subscale (e.g. emotional, social, romantic)	
Qualitative data	All available data extracted	

Quality Assessment

Reviewers used the 'What Works Centre for Wellbeing (WWCW) Quality Checklist: quantitative evidence of intervention effectiveness' to appraise the quality of all included studies. The framework and scoring system were developed by WWCW academics and the Office for National Statistics (ONS) based on the Early Intervention Foundation (EIF) Standards of Evidence. The checklist consists of 10 elements: fidelity, measurement, counterfactual, representativeness, sample size, attrition, equivalence, measures, analysis, and interpretation of findings (see Appendix 2 for further details). Each element of the checklist is scored as either 1 (yes) or 0 (no, can't tell or N/A). Points are summed across the 10 elements and the total is used to assign each study an overall level of

confidence of low (0-2), moderate (3-6) or high (7-10). The checklist can be used to appraise the quality of quantitative studies only.

Synthesis and Reporting

As the majority of studies reported mean (SD) loneliness scores before and after the intervention, the primary synthesis of results was a random-effects meta-analyses of standardised mean differences (SMD) to examine the impact of the intervention on loneliness. By using SMDs instead of raw mean differences (i.e. post-score – pre-score), any continuous measure of loneliness could be included in a single meta-analysis. SMDs, also known as Hedge's g, are calculated for each intervention group by dividing the mean change in score (post score – pre score) by the SD of the change score (pooled SD). SMDs are then pooled across all eligible studies to give an aggregate SMD effect size. Briefly, 0.20, 0.50 and 0.80 correspond to small, medium and large effect sizes, respectively. Due to inconsistencies in sample sizes pre- and post-intervention, group differences instead of individual differences were estimated.

In studies that reported mean (SD) for both an intervention and a control group before and after the intervention or a mean change (SD) score for each group, we conducted a second meta-analysis of difference in change scores following the same considerations as above. Where two intervention arms were presented against a single control group, we pooled the two intervention groups into one before including in the meta-analysis. Note that for the primary synthesis, these arms were separately included in the meta-analysis. If means and standard deviations were reported by the study, we utilised approaches recommended by the Cochrane Collaboration for dealing with missing data in meta-analyses (e.g. SD imputation, medians, ranges, interquartile ranges, etc.). WebPlotDigitizer was used to obtain data that were presented in graphs and not tables. If data was still missing, we contacted authors to request missing means, SD and sample sizes. Where only subgroup or subscales were reported, we calculated an overall group mean and SD.

To capture statistical heterogeneity (e.g. high variance or poor overlap of the SMD confidence intervals between studies), we calculated the I2 statistic for each meta-analysis. A high I2, often considered >75%, indicates substantial statistical heterogeneity, and is often a result of clinical heterogeneity (e.g. variability in participants, interventions and outcomes) or methodological variability (variability in study design and risk of bias) that contribute to differing effect sizes between studies.

A link to the full study protocol registered with PROSPERO can be found here.