Journal of Affective Disorders 246 (2019) 506–514

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Review article

Gatekeeper training for suicidal behaviors: A systematic review

Naohiro Yonemoto[a,b,*], Yoshitaka Kawashima[a], Kaori Endo[a], Mitsuhiko Yamada[a]

[a] Department of Psychopharmacology, National Center of Neurology and Psychiatry, 4-1-1 Ogawahigashi, Kodaira, Tokyo, Japan
[b] Department of Biostatistics, Kyoto University School of Public Health, Japan

ARTICLE INFO

Keywords:
Gatekeeper training
Suicide
Suicidal behavior
Suicide prevention
Systematic review

ABSTRACT

Background: Gatekeeper training (GKT) is a common intervention aiming to prevent suicidal behavior. We investigated updated evidence for the effectiveness of GKT in suicide prevention using data from randomized controlled trials (RCTs) and intervention studies, and we also describe variations in existing GKT programs.
Methods: We performed a systematic review. The literature search was conducted using PubMed, PsycINFO, CINAHL, the Cochrane databases, and reference lists from previous reviews. RCTs or intervention studies utilizing prospective or quasi-experimental designs were included.
Results: The search terms identified 343 articles. Ten randomized clinical trials and six intervention studies were identified as eligible for inclusion. Among the eligible studies, a number of different types of GKT were identified, including Question, Persuade, and Refer, Applied Suicide Intervention Skills Training, OSPI, Youth Aware of Mental Health, and approaches based on e-learning. For the RCTs, the effects of GKT remained unclear in relation to knowledge, appraisals, and self-efficacy after training, though some supportive evidence was found in the uncontrolled pre-post studies. The overall quality for each RCT was rated as either low or unclear.
Limitations: We could not perform a meta-analysis because comparable outcomes could not be identified across studies.
Conclusions: The effects of GKT remain unclear. There are many variations in GKT and there is a need to replicate studies in target populations. Future research should examine the effectiveness of a standardized GKT program using high-quality RCTs which include the evaluation of pre-specified primary outcomes in comparison with appropriate control groups.

1. Introduction

Suicide is a tragic yet prevalent cause of death, with over 800,000 suicides occurring worldwide each year (World Health Organization, 2014). However, consistent evidence in support of effective strategies to prevent suicide behaviors is lacking (Mann et al., 2005; Zalsman et al., 2016), because of the complexity of factors involved.

Gatekeeper training (GKT) aims to enable non-specialists to identify and respond to those at risk of suicidal behavior and is one of the most common interventions in suicide prevention (Mann et al., 2005). To date, the definition of GKT has slightly varied in the literature. Centers for Disease Control and Prevention (CDC) was defined as training designed to train teachers, coaches, clergy, emergency responders, primary and urgent care providers, and others in the community, to identify people who may be at risk of suicide and to respond effectively, including facilitating treatment seeking and support services. As such, the guidance could be implemented in a variety of settings to identify and support people at risk (Stone et al., 2017). According to the Surgeon General's National Strategy for Suicide Prevention (2001), a gatekeeper was identified as anyone in a position to recognize a crisis and the warning signs that someone may be contemplating suicide. This includes the individuals who are strategically positioned to recognize and refer someone at risk of suicide. In the Research Agenda in Suicide Research (National Strategy for Suicide Prevention, 2014), a gatekeeper was defined as an individual in a community who has face-to-face contact with large numbers of community members as part of their usual routine. They may be trained to identify persons at risk of suicide and refer them to treatment or supporting services as appropriate. GKT

Abbreviations: ASSIST, Applied Suicide Intervention Skills Training; CDC, Centers for Disease Control and Prevention; GKT, Gatekeeper training; RCT, Randomized control trial; QPR, Question, Persuade, and Refer; SC, Suicidal crises; YAM, Youth Aware of Mental Health

* Corresponding author at: Department of Psychopharmacology, National Center of Neurology and Psychiatry, 4-1-1 Ogawahigashi, Kodaira, Tokyo, 187-8553, Japan.
E-mail addresses: nyonemoto@gmail.com (N. Yonemoto), kawashima@ncnp.go.jp (Y. Kawashima), kaosd_sea@yahoo.co.jp (K. Endo), mitsu@ncnp.go.jp (M. Yamada).

https://doi.org/10.1016/j.jad.2018.12.052
Received 19 August 2018; Received in revised form 26 October 2018; Accepted 16 December 2018
Available online 19 December 2018
0165-0327/

also more generally refers to programs that seek to develop individuals' knowledge, attitudes, and skills to identify those at risk, determine levels of risk, and make referrals when necessary according to a report by the SPRC Registry for Evidence Based Practices (Suicide Prevention Resource Center (SPRC), 2013).

A comprehensive review of GKT evaluations concluded that "Gatekeeper training holds promise as a part of a multi-faceted strategy to combat suicide" (Isaac et al., 2009, p.266). The authors identified three ancillary strengths of gatekeeper training. First, the training can be tailored to specific regional (or other) needs. Second, the training is usually undertaken by individuals familiar to a community, which takes advantage of their knowledge of existing pathways to care. Third, the training can strengthen environments (Isaac et al., 2009). A comprehensive review of the literature has shown that there are four important factors impacting upon GKT: knowledge, perceptions about suicide prevention, reluctance, and self-efficacy. Although the previous research is scant, we can draw three general conclusions, as follows. First, there is some evidence that gatekeeper training can improve the knowledge, beliefs, attitudes, self-efficacy, and reluctance to the intervention. Second, the transfer of knowledge, beliefs, and skills learned in the training to actual intervention behavior is almost unstudied. Third, individual and contextual factors relating to the participant in training are associated with how effective the gatekeeping programs will be (Burnette et al., 2015). A systematic review in indigenous peoples reported that GKT may be a promising suicide intervention on the basis of uncontrolled research studies, but it needs to be tailored to the target population (Nasir et al., 2016). More research is required to confirm the transferability of these findings and to understand the roles of gatekeepers. A recent systematic review in adolescent participants reported that the eligible school-based GKT studies appeared to have the potential to change participants' knowledge and skills in suicide prevention. However, it was noted that a larger number of better quality studies is needed to determine its effectiveness in changing gatekeepers' attitudes. Finally, methodological issues, such as the lack of RCTs and the inconsistent use of validated measures, jeopardize the conclusions that can be drawn from the studies conducted to date (Mo et al., 2018). To this end, randomized controlled trials (RCTs) are required.

In this review, we investigate updated evidence for the effectiveness of GKT for suicide prevention and describe the existing variations in GKT. This is important because most countries are required to develop suicide prevention strategies on the basis of currently available high-quality evidence.

2. Methods

This systematic review was conducted and reported in accordance with the published protocol and Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (Liberati et al., 2009). The study protocol was registered with PROSPERO (CRD: 42016053861).

2.1. Search strategies

We searched the following electronic bibliographic databases: PubMed (from 1949), PsycINFO (from 1806), CINAHL (from 1981), and the Cochrane databases (from 1939). We also examined the reference lists of the identified studies and three previous systematic reviews (Issac et al., 2009; Burnette et al., 2015; Nashir et al., 2016). The following were used as search terms: (suicid* OR self-harm* OR selfharm* OR self-poison* OR selfpoison* OR overdose* OR over-dose* OR self-injur* OR selfinjur* OR self-mutilation* OR selfmutilation* OR auto-mutilation* OR auto-mutilation* OR self-destructive* OR self-destructive*) AND ((gatekeeper*) OR (gate-keeper*)). Keywords were identified in the title, abstract, or both. There were no language restrictions. The search was performed in February 2017.

2.2. Inclusion and exclusion criteria

The focus of our review was comprehensive interventions in GKT, consistent with that of a previously published review (Isaac et al., 2009). There were no restrictions in relation to participants or study populations, because GKT is implemented in a variety of settings and populations (Stone et al., 2017).

RCTs or intervention studies (prospective single-arm trials, non-randomized studies, and quasi-experimental studies) on articles with "gatekeeper training" as keywords were included in this review. We defined intervention trials as those with "prospective", "controlled", or "quasi-experimental" designs stated in the article. Observational studies such as cohort studies, case-control studies, retrospective pre-post studies, case series, and case reports were excluded.

2.3. Data extraction and management

All records identified from the electronic databases and hand-searches were loaded into ENDNOTE version X6 (Thomson Reuters, USA) and duplicate records removed.

Titles and/or abstracts of studies retrieved using the search strategy and those from additional sources were screened independently by two review authors to identify studies that potentially met the inclusion criteria outlined above. The full texts of these potentially eligible articles were retrieved and independently assessed for eligibility by two reviewers. Any disagreement over the eligibility of particular articles was resolved through discussion with a third review author. Two review authors extracted the data independently, any discrepancies were identified, and resolved through discussion (with a third author where necessary) (Fig. 1).

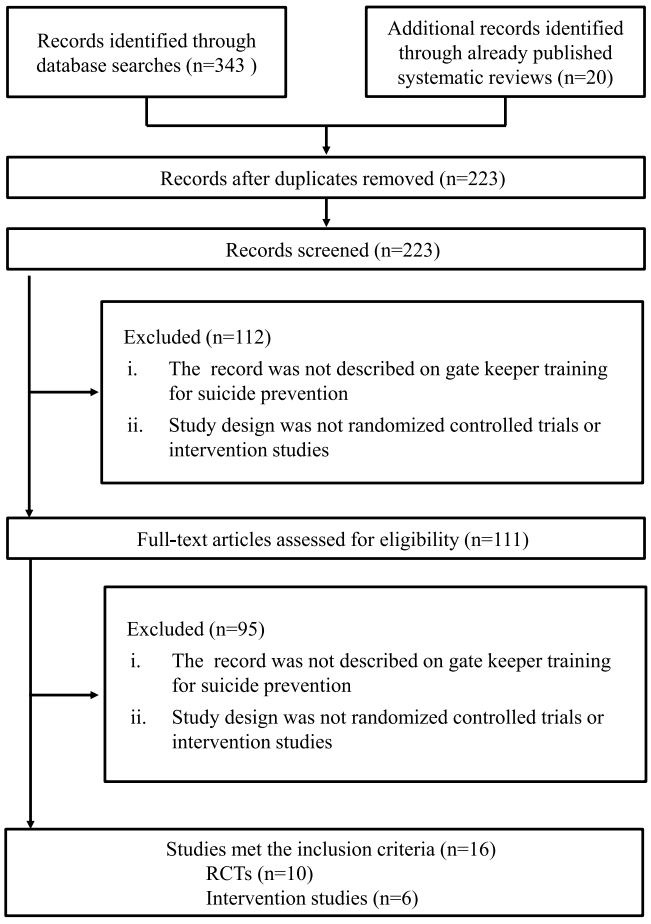

Fig. 1. Flow chart of screening process.

Data extraction was conducted independently by at least two authors. We extracted information and basic bibliographic data regarding the country of implementation, study design, characteristics of study participants, number of participants, type of intervention, follow-up period, primary and secondary outcomes, and effects of the intervention.

We present summary tables according to study design, that is, whether studies were an intervention study or RCT. The categorization of the study depended upon the intervention arm based on a standardized package program (such as QPR and ASSISTS). This categorization has been used by previous reports (Suicide Prevention Resource Center (SPRC), 2013; Stone et al., 2017). Summary tables were created on the eligible studies but meta-analysis was not performed because of the large variation in outcomes across all studies.

2.4. Assessment of methodological quality

The risk of bias was assessed for RCTs according to the Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (Higgins et al., 2011). Judgments were conducted independently by at least two reviewers in relation to random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective reporting (reporting bias), and other biases. In the case of disagreement, the issue was discussed with a third reviewer and resolved through consensus.

3. Results

3.1. Search results

As presented Fig. 1, the search terms identified 343 articles as follows: PubMed (n = 143), PsycINFO (n = 159), CINAHL (n = 40), and Cochrane databases (n = 1). Twenty articles were identified from published systematic reviews. After removing duplicates and studies that did not meet the selection criteria, 223 articles remained. These 223 articles were screened on the basis of titles and abstracts, leaving 111 potentially relevant articles that were reviewed for inclusion. The 111 articles were assessed on the basis of the full-text for eligibility. Of these, 16 articles, comprising 10 randomized clinical trials and six intervention studies, were identified.

3.2. Study characteristics

3.2.1. Intervention studies

As presented in Table 1, we categorized the six intervention studies into four categories by type as follows: Question, Persuade, and Refer, ASSIST/RESPONSE, OSPI, and “others”.

Question, Persuade, and Refer: Question, Persuade, and Refer (QPR) was developed by Quinnet (QPR Institute, 2006). The GKT program teaches individuals to recognize warning signs, question suicidal intent, listen to problems, and refer or help others using a format of frequent 2-h group sessions. The quasi-experimental, non-equivalent, control group design consisted of one group who received training, three groups who only completed assessments (control groups), and a wait-list control group (who received training after follow-up assessment). Participants were 240 college student residence advisers from six private institutions. The outcomes were QPR quizzes on knowledge, appraisals, and gatekeeper behaviors (Tompkins and Witt, 2009).

ASSIST and RESPONSE: The Applied Suicide Intervention Skills Training (ASSIST) and RESPONSE program is an extension of QPR for use with young people. The RESPONSE component is a brief high school-based program involving school-wide awareness and establishment of policies, protocols, a curriculum for youth, advanced suicide intervention program for youth, suicide prevention for specific staff

Table 1
Characteristics of the evaluations of gatekeeper training programs: non-randomized intervention study.

Authors	Study design	Training details	Country	Study participants	Outcomes
Tompkins, T. L. and Witt, J.	Quasi-experimental non-equivalent control group design.	QPR	USA	240 college student residence advisers (RAs) from six private institutions administrators at two schools self-selecting their RAs to receive training. Three opted to only complete surveys.	QPR quiz, Appraisals, Gatekeeper behaviors
Arensman E, et al.	Prospective pre-post. (Part of the OSPI study)	4-hour course on OSPI program	Germany, Ireland, Portugal	828 police officers	Depression Stigma Scale, (DSS) Suicide Intervention Knowledge Test (IKT), Morris confidence Scale (CS)
Coleman D and Del Quest A	Quasi-experimental three-group repeated measures design (pre-, post-, and 6-months after training).	Two brief sessions: (Question, Persuade, Refer: QPR and RESPONSE) and one longer (Applied Suicide Skill Training: ASSIST) training session	USA	126 trainees (QPR n = 28, RESPONSE, n = 75, ASSIST n = 23) on social service agencies, schools, churches, and other civic organization	Preparedness, attitude, efficacy, behavior, asking about suicide and distress, referral behaviors based on the QPR clinical trial by Wyman et al. (2008)
Coppens E, et al.	Prospective pre–post (pre-, post-, and 6-months after the training). (Part of OSPI study)	4-hour course on the OSPI program (with variations)	Germany, Hungary, Ireland, Portugal	1276 community facilitators (teachers, pharmacists, nurses, clergy, social workers, counselors, managers, carers for the elderly)	DSS, IKT, CS
Chauliac N, et al.	Controlled quasi-experimental study over 1-year. Trained vs untrained control.	2-days training, (14 hours in total)	France	Nursing home staff and support staff in 12 trained and 12 untrained nursing homes (no individual-level data available)	A sense of competence, detection of suicidal crises (SC), management of SC, eight institution-level measures
Teo AR, et al.	Quasi-experimental design. Trained vs attention control using mixed methods.	2 hours, multi-modal and interactive event	USA	Japanese Americans, Expatriates, and related stakeholders, 85 participants in the intervention group and 48 in the control	Suicide-specific skills, self-efficacy, social norms, personal impact and qualitative data

Table 2
Characteristics of the evaluations of gatekeeper training: randomized controlled trials.

Study	Study design	Training details	Country	Study Participant	Outcomes
Wyman, P.A, et al.	Cluster RCT (pre-, post-, 1-year average).	QPR program	USA	Secondary school staff from 10 middle and 6 high schools. There were 166 participants (of whom 122 were analyzed) in the experimental group, 176 staff (127 analyzed) in the control group.	Staff suicide prevention survey, self-reported knowledge, appraisals of efficacy, gatekeeper behaviors, communication with students, service access
Jacobson, J.M., et al.	RCT(pre, post, 4 months)	QPR program	USA	Advanced masters in social work students. 35 participants in the experimental group, 38 in the control group	Knowledge, Attitudes suicide prevention, Self-efficacy (Perceived preparedness), Working with clients at risk for suicide, Suicide prevention behaviors, Satisfaction
Cross, W. et al.	RCT (pre-, post-, 3- months)	Brief gatekeeper training (QPR) with training plus brief behavioral rehearsal (i.e., role play practice)	USA	School personnel (mental health professionals, teachers, bus drivers) in a university medical center and a rural–suburban school. 72 were analyzed in the experimental group, and 75 were analyzed in the control group (Not shown allocation number)	Declarative knowledge, self-perceived knowledge about suicide, self-efficacy for intervening, Observational Rating Scale of Gatekeeper Skills, actor adherence, gatekeeper behavior and diffusion
Sareen J, et al.	RCT (assessments were conducted pre-, post-, and 6-months post intervention), Clinical.gov NCT01287416	2 days of ASSIST program	Canada	First Nations community members, 31 participants in the ASSIST training group and 24 participants in the control group consisting of a resilience retreat	Primary outcome; Suicide Intervention Response Inventory (SIRI) Secondary outcomes; self-reported preparedness measures, gatekeeper behaviors, suicide ideation and attempts by participants.
Wasserman D, et al.	Multi-center cluster RCT (assessments at baseline, 3-months, and 12-months). German Trials Registry DRKS00000214	QPR program, The Youth Aware of Mental Health Programme (YAM), 3-hour role-play sessions with interactive workshops combined with a 32-page booklet, six educational posters displayed in classroom and two 1-hour interactive lectures about mental health at the beginning and end of the intervention, Screening by Professionals (Prof Screen), selective or indicated intervention based on responses to questionnaire on pre-intervention.	10 EU countries (Austria, Estonia, France, Germany, Hungary, Ireland, Italy, Romania, Slovenia, Spain)	40 schools [2692 pupils] QPR, 45 [2721] YAM, 43 [2764] Prof Screen, 40 [2933] control With six educational posters	Primary outcome: Number of suicide attempt(s) by 3-month and 12-month follow-up. Other outcomes: Suicide, suicide attempts, severe suicidal ideation, Strengths and Difficulties Questionnaire
Ghoncheh R et al.	RCT with web-based assessments (baseline, after the program and at 3-months). Netherlands Trial Register NTR3625	Metal Health Organization (MHO) Program, 8 e-learning modules alongside additional information regarding adolescent suicide prevention	Netherlands	Workers in health care and education working with adolescents aged 12–20 years old. 94 participants in the experimental group and 96 in the waitlist control group	Actual knowledge, perceived knowledge, perceived self-confidence
Tsai, WP, et al.	RCT, (pre-, post-training),	Regular monthly continuing education class and a 90-min Gatekeeper Suicide-Awareness Program	Taiwan	Nurses in a general hospital. 98 participants were in the experimental group, and 97 were in the control group	Awareness of Suicide Warning Signs Questionnaire
Wu CY, et al.	RCT (assessments pre- and post-training), Clinical.gov NCT02033915	2-hour suicide gatekeeper lecture and an additional 5-hour group discussion about suicide risk assessment skills	Taiwan	Nurses in a general hospital in internal medicine, surgical and emergency care sectors. Allocated number was 112 (of whom 54 were analyzed) in the experimental group, 90 (57 analyzed) the control group.	Case vignette, Chinese Sad Persons Scale, Brief Symptom Rating Scale

(*continued on next page*)

Table 2 (*continued*)

Study	Study design	Training details	Country	Study Participant	Outcomes
Klingman & Hochdorf	RCT (pre-, post-training [2 weeks])	Three-phase model as educational and conceptual skill acquisition, rehearsal, and application. Seven units were completed during 12 weekly 1-hour sessions	Israel	Students in grade 8, junior high-school with primarily low- to middle-class backgrounds. 116 participants were in the experimental group, and 121 in the control group	Israeli Index of Potential Suicide, UCLA Loneliness Scale, story completion, semantic differential, knowledge, program/feedback evaluation
Chagnon F, et al.	RCT (pre-, post-, and 6-months [experimental group only])	A program over 3 days (1 day a week for 3 consecutive weeks) by senior staff from the suicide prevention center. Used a directive, problem-solving approach and commonly used counseling methods	Canada	Helpers from ten educational establishments and community or institutional organizations serving the youth clientele in the metropolitan area. 43 participants in the experimental group, 28 in the control group	Suicide Intervention Questionnaire, knowledge, skills

(e.g., councilors and social workers), and brief gatekeeper training for all staff. The ASSIST component applies suicide interventions skills training (Living Works, 2010) over a period of two days. There were 126 participants across three groups (28 in QPR, 75 in RESPONSE, 23 in ASSIST) in the quasi-experimental and repeated measures design. Outcomes were preparedness, attitude, efficacy, behavior, questioning about suicide and distress, and referral behaviors. Outcomes were based on Wyman et al. (2008) for the QPR program (Coleman and Del Quest, 2015).

OSPI: The OPSI program is a short 4-h course (with variations) which has been conducted as a part of a multimodal community model for suicide prevention. A prospective, pre-post study design has been used in previous research. In this, participants were 1276 community facilitators (teachers, pharmacists, nurses, clergy, social workers, counselors, managers, or carers for the elderly), and 828 police officers. Outcomes were a depression stigma scale, suicide intervention knowledge test, and the Morris Confidence Scale (Arensman et al., 2016; Coppens et al., 2014).

Others: Another type of GKT consisting of two days training (14 h) was tested in a controlled quasi-experimental study over one year. Participants were nursing home and support stuff in 12 trained and 12 untrained nursing homes. Outcomes were the sense of competence, detection of suicidal crises (SC), and management of SC by eight institution-level measures. The study only had data by institution unit, but no individual data (Chauliac et al., 2016).

A further program was two hours in duration and was a multi-modal and interactive event. This was investigated using a quasi-experimental design with an attention control group using mixed methods quantitative and qualitative data. The 85 participants were Japanese Americans, expatriates, and related stakeholders. Outcomes were suicide-specific skills, self-efficacy, social norms, personal impacts, and qualitative data (Teo et al., 2016).

3.2.2. Randomized controlled trials

As presented in Table 2, we categorized the ten RCTs into five groups according to the type of intervention as follows: QPR, ASSIST, extensions to these programs, e-learning, and “others”.

Question, Persuade, and Refer: A cluster RCT with repeated measurements at pre-intervention, post-intervention, and 1-year average was conducted with secondary school staff. There were 10 middle and six high schools in both the experimental (166 staff) and control group (176 staff). The outcomes were a staff suicide prevention survey, self-reported knowledge, appraisals of efficacy, gatekeeper behaviors, communication with students, and service access (Wyman et al., 2008).

Another RCT used repeated measurements at pre-intervention, post-intervention, and four months with students on an advanced Masters in social work program. There were 35 trainees in the experimental group and 38 in the control group. The outcomes were knowledge and attitudes on suicide prevention, self-efficacy (perceived preparedness), working with clients at risk for suicide, suicide prevention behaviors, and satisfaction (Jacobson et al., 2012).

A program with QPR with the addition of brief behavioral rehearsal has also been developed. An RCT included measurements at pre-intervention, post-intervention, and 3 months on school personnel (72 participants in the experimental group and 75 in the control group). Outcomes were declarative knowledge, self-perceived knowledge about suicide, self-efficacy in intervening, observational rating scale of gatekeeper skills, actor adherence, gatekeeper behavior, and diffusion (Cross et al., 2011).

ASSIST: An RCT with assessments at pre-intervention, post-intervention, and 6 months post-intervention was conducted with First Nations community members. There were 31 participants in the ASSIST training group and 24 in the control group with a resilience retreat. The primary outcome was the Suicide Intervention Response Inventory. Secondary outcomes were self-reported preparedness measures, gatekeeper behaviors, suicide ideation and attempts by participants (Sareen et al., 2013).

Youth Aware of Mental Health Program, screening by Professionals, and QPR: The Youth Aware of Mental Health Program (YAM) consisted of 3-h role-play sessions with interactive workshops combined with a 32-page booklet, six educational posters displayed in classrooms, and two 1-h interactive lectures about mental health at the beginning and end of the intervention. Screening by professionals (Prof Screen) was a selective or indicated intervention based on responses to the questionnaire at pre-intervention. The study consisted of four arms: YAM, Prof Screen, QPR, and control. This cluster RCT involved measurements at baseline, 3 months, and 12 months in a multicenter study including 10 EU countries and 40 schools [2692 pupils] QPR, 45 [2721] YAM, 43 [2764] Prof Screen, 40 [2933] control with six educational posters. The primary outcome was the number of suicide attempt(s) by the 3-month and 12-month follow-ups. Other outcomes were suicide, suicide attempts, severe suicidal ideation, and the Strengths and Difficulties Questionnaire (Wasserman et al., 2015).

E-learning on the MHO program: The MHO program consisted of eight e-learning modules alongside additional information regarding adolescent suicide prevention. An RCT investigated performance on web-based assessments at baseline, after the program, and at 3 months after the program for workers in health care and education working with adolescents aged 12–20 years old. There were 94 participants in the experimental group and 96 in the waitlist control group. The outcomes were actual knowledge, perceived knowledge, and perceived self-confidence (Ghoncheh et al., 2016).

Others: Four studies were categorized into "others". The first program was a regular monthly continuing education class and a 90-min Gatekeeper Suicide-Awareness Program for general nurses. An RCT on the program was conducted in a general hospital in Taiwan. There were 98 participants in the experimental group and 97 in the control group. Outcomes were assessed using Awareness of Suicide and Warning Signs Questionnaire (Tsai et al., 2011).

The second consisted of a 2-h suicide gatekeeper lecture and a 5-h group discussion about suicide risk assessment skills for nurses. An RCT on the program was conducted with nurses in a general hospital. There were 112 participants in the experimental group and 90 in the control group. Outcomes were assessed using a case vignette, the Chinese Sad Persons Scale, and the Brief Symptom Rating Scale (Wu et al., 2014).

The third was a program using a three-phase model comprising educational/conceptual skill acquisition, rehearsal, and application, built upon seven units covered in 12 weekly 1-h sessions with students. Participants were grade 8 students in junior high school from primarily low- to middle-class backgrounds. There were 116 participants in the experimental group and 121 participants in the control group. Outcomes were the Israeli Index of Potential Suicide, UCLA Loneliness Scale, story completion, semantic differentiation, knowledge, and program/feedback evaluation (Klingman and Hochdorf, 1993).

The last was a program conducted over three days (one day a week for three consecutive weeks) by senior staff from a suicide prevention center and used a directive, problem-solving approach, and commonly used counseling methods. Participants were helpers from 10 educational establishments and community or institutional organizations serving the clientele of youth. There were 43 participants in the experimental group and 28 participants in the control group. Outcomes were attitudes on the Suicide Intervention Questionnaire, knowledge, and skills (Chagnon et al., 2007).

3.3. Study findings

3.3.1. Intervention studies

When considering the nonrandomized designs, most studies demonstrated at least one positive effect of GKT, but for the core outcomes, only unchanged or unclear changes were found.

In the quasi-experimental, non-equivalent control study, there were no significant gains in the QPR Quiz Scores or QPR Behaviors, although there were gains for sustained prevention training in the residence advisor's appraisals of their preparation, efficacy, knowledge of resources, and intention to intervene with suicidal students. Although there were increases in appraisals over time for those who received no training, they tended to be less pronounced than changes in the QPR group. However, statistical support was limited, and only four of the nine outcomes in total being reported as showing statistical significance (Tompkins and Witt, 2009).

In the three-arm quasi-experimental study in which the QPR, RESPONSE and ASSIST programs were assessed pre-post and at 6-month follow-up, all five outcome measures statistically significantly improved from training ($p < 0.01$). The QPR was a reference group. Cohen's d on differences pre-post was as follows: Preparedness; QPR 1.10 RESPONSE 0.71 ASSIST 0.73, Efficacy; QPR 0.90 RESPONSE, 1.00 ASSIST 0.90, Referred; QPR 0.16 RESPONSE, 0.22, ASSIST 0.45, Ask; QPR 0.16 RESPONSE, 0.17 ASSIST 1.10, Referral; QPR 0.50 RESPONSE, 0.40, ASSIST 0.50 (Coleman and Del Quest, 2015)

In an OSPI program for police officers using a pre-post design, the effect size of training differed across countries for the Depression Stigma Scale (DSS), Suicide Intervention Knowledge Test (IKT), and Morris Confidence Scale (CS). The overall effect of training for each country was not provided. For community facilitators, an increase in mean scores on the DSS occurred after training (Mean [SD], 32.99 [4.6] vs. 35.18 [4.8], $p < 0.001$), on the IKT (3.88 [1.3], vs. 4.35 [1.3], $p < 0.001$), and confidence scale (4.48 [1.8] vs. 5.85 [1.8], $p < 0.001$). An increase in mean scores on the personal depression stigma subscale of the DSS was observed (31.57 [5.1] vs. 33.01 [5.6] at baseline and at 35.15 [4.8] at post-training), and for scores on the Suicide IKT (3.95 [1.3] vs. 4.27 [1.3] at post-training) but not at baseline (3.82 [1.3]). All of the three outcomes were reported as statistically significant in relation to pre-post assessments. There were three outcomes in total, with heterogeneity by country (Arensman et al., 2016; Coppens et al., 2014).

In a study with a quasi-experimental controlled design over one year for elderly people in nursing homes, the mean score for sense of competence was 6.0 in trained nursing homes, and 3.0 in untrained nursing homes, $p < 0.001$. During the monitoring phase, across all 24 nursing homes, 36 SC in trained nursing homes and 33 in untrained nursing homes ($p = 0.14$) were found. In the untrained nursing homes, in over 20% of cases of SC no specific measures were taken, while at least one measure was taken in all cases in the trained nursing homes. The trained nursing homes also took significantly more measures on average (2.7 vs. 1.9, $p = 0.03$). Most of the institution-level measures that played a part in suicide prevention were significantly changed and the number of institution-level measures by visit number increased between trained and untrained nursing homes ($p = 0.03$) (Chauliac et al., 2016).

In a controlled intervention study with pre-post assessments, intervention groups showed improvement in their likelihood of using overall behavior and all three specific behaviors (questioning, persuading, and referring), while control participants showed no change. The intervention group showed improvement overall and in all four individual components of self-efficacy related to assisting someone at risk for suicide. The control group showed unclear improvement, except in their self-perceived ability to tell if someone close to them was at risk for suicide. Intervention participants showed increases in overall treatment-promoting social norms, but not control participants. In adjusted models, the mean improvement in gatekeeper behavior (0.52, 95% CI 0.28–0.75) and self-efficacy (0.55, 0.32–0.79) was greater in the intervention than the control group, but the impact on social norms was not clear. Thus, two of the three outcomes in total were reported as statistically significant (Teo et al., 2016).

3.3.2. Randomized controlled trials

Most RCTs identified some positive outcomes but not all studies reported clear effects. Unfortunately, all trials had issues arising from being a low study quality and/or having unclear reporting.

QPR: The authors suggested that training positively impacts on self-reported knowledge (effect size [ES] 0.41), appraisals of efficacy (ES

Table 3
Risk of bias in randomized controlled trials.

	Random sequence generation	Allocation sequence concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective outcome reporting	Other potential sources of bias
Wyman, P.A, et al.	Unclear	Unclear	Unclear	Unclear	High	Low	High
Jacobson, J.M., et al.	Unclear	Unclear	High	Unclear	Low	High	High
Sareen J, et al.	Low	Unclear	Unclear	Unclear	High	Low	High
Cross, W. et al.	Unclear	Unclear	Unclear	Unclear	Unclear	High	High
Wasserman D, et al.	Low	High	Unclear	Unclear	High	Low	High
Ghoncheh R et al.	Unclear	Unclear	High	High	High	High	High
Tsai, WP, et al.	High	High	Unclear	Unclear	Low	High	Unclear
Wu CY, et al.	High	High	Low	Low	High	Low	High
Klingman & Hochdorf	Unclear	Unclear	Unclear	Unclear	Unclear	High	High
Chagnon F, et al.	Unclear	Unclear	Unclear	Unclear	High	High	High

1.22), and service access (ES 1.07) whereas the impact on behaviors was concentrated within staff in certain job classes and in those communicating with distressed youths about suicide at baseline. Statistical evidence for differences between the groups was not shown and the analysis did not include cluster variations. Two of the 10 outcomes in total were reported as statistically significant (Wyman et al., 2008).

In another study, the effect of GKT was not clear. There was no difference in mean scores from pre- to 4 months, between the intervention and control groups, or in the interaction for time and group on knowledge. Interaction effects between group and time suggested an improvement among the intervention group with regard to knowledge, efficacy to perform the gatekeeper role, and skills. No changes in attitudes were observed. Two of the 12 outcomes in total were reported as statistically significant. Interaction terms between time and group on three outcomes were reported as statistically significant (Jacobson et al., 2012).

For the third study, the effect of the GKT was positive for skills only, but not on other measures. For knowledge, attitudes as self-perceptions and self-efficacy, gatekeeper behavior and diffusion, there was no effect of training. Scores in the intervention group were significantly higher than those in the control for the total score of gatekeeper skills ($p < 0.05$). Seven of the 30 outcomes in total were reported as statistically significant (Cross et al., 2011).

ASSIST: In this study, the effect of GKT was negative for the primary outcome. No significant impact was shown on three outcomes of the study based on the intention-to-treat analysis. Knowledge of risk as a secondary outcome was significant, but marginal ($p < 0.03$). There was a trend toward an increase in suicidal ideation among those who participated in the ASSIST group compared with those who were in the resilience retreat group (Sareen et al., 2013).

YAM, Prof Screen, and QPR: In this study, the effect of GKT was negative at 3 months but positive at 12 months. No clear differences between the experimental groups and control group were recorded at the 3-month follow-up. At the 12-month follow-up, YAM was associated with a significant reduction of suicide attempts (odds ratios [OR] 0•45, 95% CI 0•24–0•85; $p = 0•014$) and severe suicidal ideation (0•50, 0•27–0•92; $p = 0•025$), compared with the control group. In this study, 14 pupils (0•70%) reported suicide attempts at the 12-month follow-up in the YAM group vs. 34 (1•51%) in the control group, and 15 pupils (0•75%) reported severe suicidal ideation in YAM group vs. 31 (1•37%) in the control group. No participants completed suicide during the study period. Two of the 12 outcomes at 12 months were reported as statistically significant (Wasserman et al., 2015).

E-learning on MHO program: In this study, the effect of the GKT was positive and all of the three outcomes were reported as statistically significant. The actual knowledge, perceived knowledge, and perceived self-confidence of gatekeepers in the experimental group improved significantly compared with those in the waitlist control group at post-test, and the effects remained significant at 3-month follow-up. The overall effect sizes were 0.76, 1.20, and 1.02, respectively, across assessments (Ghoncheh et al., 2016).

Others: In a study among general nurses, the effect of GKT was not clear. The authors suggested nurses who participated in the program were much more aware of suicide warning signs and more willing to refer patients for professional counseling. Statistical evidence was not shown for differences between groups for pre-post comparisons. The only evidence was for pre and post differences (Tsai et al., 2011).

In another study among general nurses, the effect of GKT was not clear. One of the two outcomes in total was reported as statistically significant. The authors suggested the intervention was an effective and practical strategy for enhancing general nurses' abilities in relation to suicide risk identification and assessment. Statistical evidence for differences between groups on pre-post was not presented. The only evidence supplied was for pre and post differences (Wu et al., 2014).

In a study with adolescents, the effect of GKT was found to be positive. Four of the 10 outcomes in total were reported as statistically significant. The authors suggested the program had a positive effect on an index of potential suicide ($p < 0.05$), awareness of distress coping skills ($p < 0.05$), potency dimension ($p < 0.001$) and activity dimension ($p < 0.04$) at semantic difference. However, overall effect sizes were not presented, only p-values (Klingman and Hochdorf, 1993).

In a study among youth helpers, the effects of GKT were positive. All three outcomes were reported as statistically significant. The training group improved in knowledge, attitudes, and intervention skills following the training, compared with the control group across measures at pre-post ($p < 0.0001$). Scores in the training group were maintained at 6-month follow-up (Chagnon et al., 2007).

As shown in Table 3, the quality scores for the trials were either low or unclear. This suggests that the data available are vulnerable to severe bias.

4. Discussion

4.1. Main findings

A variety of types of GKT, including QPR, ASSIST, OSPI, YAM, e-learning, and others, have been implemented to date. From the evidence obtained from RCTs, the effects of GKP are still unclear, even for knowledge, appraisals, and self-efficacy after training, although stronger evidence was obtained from uncontrolled studies. Only one report with suicidal behavior in the target population as an outcome was reported. Importantly, most of the RCTs had issues relating to quality.

4.2. Comparisons with previous systematic reviews and other GKT reports

Our findings on RCTs were based on evidence associated with the assessed quality of the trial, because evidence from previous reviews are from uncontrolled, controlled interventions and observational studies (Isaac et al., 2009; Burnette et al., 2015; Nasir et al., 2016; Mo et al., 2018). Previous review papers have suggested issues related to longevity outcomes, referral patterns, and the effects of suicidal behaviors. These issues are also apparent in the current review. The Garret Lee Smith Suicide Prevention Program on young people assessed rates of nonfatal suicidal behavior and suicide with a multi-site evaluation in 50 states and 50 tribes. However, this report was observational, and not an intervention study (Godoy Garraza et al., 2015).

4.3. Variations in training programs and study populations

There were variations in the type of training and target populations of each study. Some programs were developed with on the basis of theoretical backgrounds and standardized (such as the QPR program), but others were developed in original contexts or were unclear in their theoretical background. Thus, proof of concept and core outcomes such as knowledge, skills, and efficacy were not clear in many reports. In these cases, it is difficult to identify a primary outcome. We need to create a standardized program and develop training with clear concepts based on the theoretical background, conceptualized training content and plans with rationales, with well-considered training manuals including trainers and core outcomes with reliable and validated evaluations. The populations targeted varied, including, for example, members of the community, nurses, and supporters. In the development of policy relating to GKT, we need to consider the transferability of potentially beneficial effects for different populations. Even when effectiveness has been found in one specific group, this might not be replicable in others. There might be differences in the duration of training across studies although our findings from RCTs do not support this. We now need to collect data on the quality of adherence and core conceptual outcomes as knowledge, attitudes, appraisals, self-efficacy, and skills, during and after the training on the evaluation of the dose of training.

4.4. Methodological issues

Inconsistency in the comparisons made created difficulties for the interpretation of findings. Superiority across programs was not clear, though there was one trial which compared active arms and control with no intervention (Wasserman et al., 2015). We cannot identify an appropriate strategy for programs on the basis of evidence to date. The issue of how to determine an appropriate control or active arm in GKT research requires consideration.

We identified 10 RCTs, but most were unclear in their reporting of randomization methods (e.g., random sequence generation and allocation concealment). This is an important and critical issue because, without this information, we cannot evaluate the validity of the trial.

Most of the RCTs had set many outcomes but not identified a primary outcome. The rationale and structure for identifying outcomes were often unclear. In future trials, primary outcomes and secondary outcomes should be clearly identified. All studies set more than five outcomes, meaning the levels for alpha error could be violated. Core outcomes in GKT research should be standardized. Some studies reported analyses on repeated measures. For some trials, interaction terms as time and intervention were significant. Heterogeneity is an issue in the development of GKT programs.

There remain difficulties in measuring the effectiveness on suicidal behaviors of identified and referred persons in the community by trained members. It may be technically difficult to assess this considering the nature of suicidal behaviors in the community. Evidence on the effectiveness of GKT is a key issue. There is a report of an intervention study in elderly people in nursing homes that measured the number of detections of suicidal persons and those referred to a psychologist at an institution-level (Chauliac et al., 2016).

Overall, the studies to date had many reporting issues. Sometimes, essential statistical information was not reported, including the provision of effect sizes with *p*-values. In addition, selective reporting and publication biases might impact upon the estimates provided in some of the studies.

4.5. Limitations

We could not perform a meta-analysis because comparable outcomes could not be identified across studies. Some studies had long-term outcomes after training, but the reporting of these was limited. Cost-effectiveness of GKT was not reported (Ahern et al., 2018). Cost and quality are issues to be addressed in future research.

We excluded some multi-component community intervention studies with GKT programs (Rutz et al., 1992; Knox et al., 2003) because we could not isolate the specific effects of the GKT programs from the other interventions in the published articles.

4.6. Implications

GKT programs have had some positive effects but the supportive evidence is limited. Implementations of GKT need to monitor core conceptual outcomes as knowledge, attitudes, appraisals, self-efficacy, and skills. Policymakers may therefore be encouraged to evaluate and monitor the conceptual outcomes and long-term outcome as suicide behaviors as part of practice. We should also develop programs aiming to identify the referral of supported people at suicidal risk to appropriate medical resources.

5. Conclusions

The effects of GKP remain unclear. There are many variations in the application of GKT and there is a need to replicate findings across target populations. We should develop a standardized GKT program and assess its superiority compared with an appropriate control group on a pre-specified primary outcome to achieve completion of a high-quality RCT.

Acknowledgment

We thank for Ms. Hiroko Kobayashi and Ms. Hiromi Muramatsu for her helpful assistance.

Author statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, and revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the Journal of Affective disorders.

Funding

This review was supported by funding from the Grant-in-Aid for Scientific Research, Kaken (C), 16K08899, Japan.

Availability of data and materials

The data supporting this review can be found in the databases and journals mentioned throughout the manuscript text.

Author contributions

NY was responsible for the study concept and design. NY also performed the review, analysis and drafted the manuscript. YK and KE performed the review, contributed to writing the manuscript and provided critical review. MY contributed to writing the manuscript and provided critical review.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2018.12.052.

References

Ahern, S., Burke, L.A., McElroy, B., Corcoran, P., McMahon, E.M., Keeley, H., et al., 2018. A cost-effectiveness analysis of school-based suicide prevention programmes. Eur. Child Adolesc. Psychiatry 27 (10), 1295–1304.

Arensman, E., Coffey, C., Griffin, E., Van Audenhove, C., Scheerder, G., Gusmao, R., et al., 2016. Effectiveness of depression–suicidal behaviour gatekeeper training among police officers in three European regions: outcomes of the optimising suicide prevention programmes and their implementation in Europe (OSPI-Europe) study. Int. J. Soc. Psychiatry 62 (7), 651–660.

Burnette, C., Ramchand, R., Ayer, L., 2015. Gatekeeper training for suicide prevention a theoretical model and review of the empirical literature. Rand Health Q. 5, 16 Jul 15.

Chagnon, F., Houle, J., Marcoux, I., Renaud, J., 2007. Control group study of an intervention training program for youth suicide prevention. Suicide Life Threat Behav. 37, 135–144.

Chauliac, N., Brochard, N., Payet, C., EGEE (Étude Gatekeepers en EHPAD) study group,Duclos, A., Terra, J.L., 2016. How does gatekeeper training improve suicide prevention for elderly people in nursing homes? A controlled study in 24 centres. Eur. Psychiatry 37, 56–62.

Coleman, D., Del Quest, A., 2015. Science from evaluation: testing hypotheses about differential effects of three youth-focused suicide prevention trainings. Soc. Work Public Health 30, 117–128.

Coppens, E., Van Audenhove, C., Iddi, S., Arensman, E., Gottlebe, K., Koburger, N., et al., 2014. Effectiveness of community facilitator training in improving knowledge, attitudes, and confidence in relation to depression and suicidal behavior: results of the OSPI-Europe intervention in four European countries. J. Affect. Disord. 165, 142–150.

Cross, W.F., Seaburn, D., Gibbs, D., Schmeelk-Cone, K., White, A.M., Caine, E.D., 2011. Does practice make perfect? A randomized control trial of behavioral rehearsal on suicide prevention gatekeeper skills. J. Prim. Prev. 32, 195–211.

Ghoncheh, R., Gould, M.S., Twisk, J.W., Kerkhof, A.J., Koot, H.M., 2016. Efficacy of adolescent suicide prevention e-learning modules for gatekeepers: a randomized controlled trial. JMIR Ment. Health 3, e8.

Godoy Garraza, L., Walrath, C., Goldston, D.B., Reid, H., McKeon, R., 2015. Effect of the Garrett Lee Smith memorial suicide prevention program on suicide attempts among youths. JAMA Psychiatry 72, 1143–1149.

Higgins, JPT, Green, S (Eds.), 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Isaac, M., Elias, B., Katz, L.Y., Belik, S.L., Deane, F.P., Enns, M.W., Sareen, J., Swampy Cree Suicide Prevention Team, 2009. Gatekeeper training as a preventative intervention for suicide: a systematic review. Can. J. Psychiatry 54, 260–268.

Knox, K.L., Litts, D.A., Talcott, G.W., Feig, J.C., Caine, E.D., 2003. Risk of suicide and related adverse outcomes after exposure to a suicide prevention programme in the US air force: cohort study. Br. Med. J. 327, 1376–1378.

Jacobson, J.M., Osteen, P.J., Sharpe, T.L., Pastoor, J.B., 2012. Randomized trial of suicide gatekeeper training for social work students. Res. Soc. Work Prac. 22, 270–281.

Klingman, A., Hochdorf, Z., 1993. Coping with distress and self harm: the impact of a primary prevention program among adolescents. J. Adolesc. 16, 121–140.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P., et al., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 339, b2700.

Mann, J.J., Apter, A., Bertolote, J., Beautrais, A., Currier, D., Haas, A., et al., 2005. Suicide prevention strategies: a systematic review. JAMA 294 (16), 2064–2074.

Mo, P.K.H., Ko, T.T., Xin, M.Q., 2018. School-based gatekeeper training programmes in enhancing gatekeepers' cognitions and behaviours for adolescent suicide prevention: a systematic review. Child Adolesc. Psychiatry Ment. Health 12, 29.

Nasir, B.F., Hides, L., Kisely, S., Ranmuthugala, G., Nicholson, G.C., Black, E., et al., 2016. The need for a culturally-tailored gatekeeper training intervention program in preventing suicide among Indigenous peoples: a systematic review. BMC Psychiatry 16, 357.

National Action Alliance for Suicide Prevention, 2014. Rockville, MD. http://actionallianceforsuicideprevention.org/sites/actionallianceforsuicideprevention.org/files/Agenda.pdf.

National Strategy for Suicide Prevention, 2001. Goals and objectives for action. Center for Mental Health Services (US); Office of the Surgeon General (US), Rockville (MD): US Public Health Service. https://www.ncbi.nlm.nih.gov/books/NBK44281/.

Rutz, W., von Knorring, L., Walinder, J., 1992. Long-term effects of an educational program for general practitioners given by the Swedish committee for the prevention and treatment of depression. Acta Psychiatr. Scand. 85, 83–88.

Sareen, J., Isaak, C., Bolton, S.L., Enns, M.W., Elias, B., Deane, F., et al., 2013. Gatekeeper training for suicide prevention in First nations community members: a randomized controlled trial. Depress. Anxiety 30, 1021–1029.

Stone, D.M., Holland, K.M., Bartholow, B., Crosby, A.E., Davis, S., Wilkins, N., 2017. Preventing Suicide: A Technical Package of Policies, Programs, and Practices. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA. https://stacks.cdc.gov/view/cdc/44275.

Suicide Prevention Resource Center (SPRC), 2013. Comparison table of suicide prevention gatekeeper training programs. http://www.sprc.org/sites/default/files/migrate/library/SPRC_Gatekeeper_matrix_Jul2013update.pdf.

Teo, A.R., Andrea, S.B., Sakakibara, R., Motohara, S., Matthieu, M.M., Fetters, M.D., 2016. Brief gatekeeper training for suicide prevention in an ethnic minority population: a controlled intervention. BMC Psychiatry 16, 211.

Tompkins, T.L., Witt, J., 2009. The short-term effectiveness of a suicide prevention gatekeeper training program in a college setting with residence life advisers. J. Prim. Prev. 30, 131–149.

Tsai, W.P., Lin, L.Y., Chang, H.C., Yu, L.S., Chou, M.C., 2011. The effects of the gatekeeper suicide-awareness program for nursing personnel. Perspect. Psychiatr. Care 47, 117–125.

Wasserman, D., Hoven, C.W., Wasserman, C., Wall, M., Eisenberg, R., Hadlaczky, G., et al., 2015. School-based suicide prevention programmes: the SEYLE cluster-randomised, controlled trial. Lancet 385, 1536–1544.

World Health Organisation, 2014. WHO. http://apps.who.int/iris/bitstream/10665/131056/1/9789241564779_eng.pdf?ua=1&ua=1 Accessed 5 July 2018.

Wu, C.Y., Lin, Y.Y., Yeh, M.C., Huang, L.H., Chen, S.J., Liao, S.C., Lee, M.B., 2014. Effectiveness of interactive discussion group in suicide risk assessment among general nurses in Taiwan: a randomized controlled trial. Nurse Educ. Today 34, 1388–1394.

Wyman, P.A., Brown, C.H., Inman, J., Cross, W., Schmeelk-Cone, K., Guo, J., Pena, J.B., 2008. Randomized trial of a gatekeeper program for suicide prevention: 1-year impact on secondary school staff. J. Consult. Clin. Psychol. 76, 104–115.

Zalsman, G., Hawton, K., Wasserman, D., van Heeringen, K., Arensman, E., Sarchiapone, M., et al., 2016. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 3, 646–659.