Effectiveness of brief suicide management training programme for medical residents in Japan: a cluster randomized controlled trial

Y. Suzuki[1]*, T. A. Kato[2], R. Sato[3], D. Fujisawa[4], K. Aoyama-Uehara[3], N. Hashimoto[5], N. Yonemoto[6], M. Fukasawa[1] and K. Otsuka[7]

[1] *Department of Adult Mental Health, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan*

[2] *Department of Neuropsychiatry, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan*

[3] *Department of Psychiatry, Yokohama City University School of Medicine, Yokohama, Japan*

[4] *Psycho-oncology Division, National Cancer Center East, and Department of Neuropsychiatry, Keio University School of Medicine, Tokyo, Japan*

[5] *Department of Psychiatry, Hokkaido University Graduate School of Medicine, Sapporo, Hokkaido, Japan*

[6] *Department of Epidemiology and Biostatistics, Translational Medical Center, National Center of Neurology and Psychiatry, Tokyo, Japan*

[7] *Department of Disaster Psychiatry and Community Psychiatry, Iwate Medical University, Iwate, Japan*

Aims. To evaluate the effectiveness of a brief suicide management training programme for Japanese medical residents compared with the usual lecture on suicidality.

Methods. In this multi-center, clustered randomized controlled trial, the intervention group attended a structured suicide management programme and the control group, the usual lecture on depression and suicidality. The primary outcome was the difference in residents' cumulative competency score to manage suicidal persons from baseline (T0) to 1 month after the intervention (T2), determined using the Suicide Intervention Response Inventory (SIRI-1) score, at individual level.

Results. Analysis of 114 residents (intervention group $n = 65$, control group $n = 49$) assigned to two clusters in each group revealed no change in SIRI-1 score from T0 to T2 or immediately after the intervention (T1) between the two groups. As a secondary analysis, discrepancy in judgement between the participants and Japanese suicidologists was examined immediately after the intervention in the adjusted model, with a mean difference in score of 9.98 (95% confidence interval: 4.39–15.56; $p = 0.001$).

Conclusions. The structured programme was not proven to improve competency in suicide management when measured by the SIRI-1 score. Further elaboration of the programme and valid measurement of its outcome would be needed to show the program's effectiveness.

Introduction

Suicide has been a national concern for Japan ever since numbers first exceeded 30 000 per year in 1998 (Shiho *et al.* 2005). However, psychiatric services have not been readily accessible to those with depression in Japan, and more than half of individuals with new onset depression consult their general medical practitioners first, often delaying their receipt of psychiatric care (Fujisawa *et al.* 2008). Health

* Address for correspondence: Dr Yuriko Suzuki, Department of Adult Mental Health, National Institute of Mental Health, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8553, Japan.

(Email: yrsuzuki@ncnp.go.jp)

professionals across a range of fields are therefore in a position to have important first contacts to detect and intervene with patients with mental health problems, but they require appropriate knowledge and skills to do this (Sato *et al.* 2006).

In Japan, sufficiently detailed training for health professionals such as physicians on suicide management of individuals in mental crisis has been lacking. Yet such training would seem vital for them in their role as gatekeepers for persons with health problems, particularly as they are expected to assume important roles in suicide prevention (Mann *et al.* 2005). We hypothesized, therefore, that suicide management training for medical residents could help them increase their competency in managing persons with mental health problems, and ultimately contribute to suicide prevention.

This intervention study aimed to improve the competency of medical residents, at individual level, in managing suicidal persons they encounter in medical settings through their attendance at a brief structured training programme for suicide management. Its effectiveness was compared with the conventional training that medical residents receive on suicidality.

Method

Study design

This study employed a cluster randomized controlled trial design to avoid contamination of the programme among individual medical residents and to facilitate group-level participatory intervention. The trial involved five institutions in Japan (Hokkaido University, Iwate Medical University, Keio University, Yokohama City University and Kyushu University). Each institution offers a training programme for medical residents on suicide prevention and the programmes were provided during their psychiatric rotation (July 2009–March 2010). This study was registered in the national clinical trial registry, the UMIN-CTR (registration number: UMIN000002245), on 24 July, 2009.

Participants

Participants were second-year medical residents on their psychiatry rotation. In the Japanese medical education system, newly certified physicians must go through 2 years of postgraduate training as general medical residents before they chose a clinical specialty (Morriss *et al.* 2005). During these 2 years, psychiatry was one of the required areas for training for at least 1 month in the second year, the time when this research was conducted (Kozu, 2006; Teo, 2007).

Randomization

Lists of medical residents in each institution in 2009 were obtained. The unit of randomization was the group of medical residents on their psychiatric rotation during the study period. A cluster constituted a mix of groups completing a psychiatry rotation in the participating institutions. The rotation groups comprised 1–9 medical residents and the groups were equivalent in terms of content of training and individual factors such as preference for future speciality. Since there was variation in the size of the groups, we operationally created a cluster containing several rotation groups in order to make a cluster with 9–18 members. After randomization to clusters, a training session was assigned to each cluster and all medical residents in the rotation group were then invited to participate in the study.

In each institution, two clusters (first half and second half groups) were created and a random number of medical students were assigned to the first group of each institution, yielding two intervention groups and control groups. Consequently, alternate intervention/control conditions were assigned to the second group in each institution to avoid an institutional effect as a whole. Registration of the clusters was conducted by YS, and the randomization procedure was carried out by MF, an independent researcher who was not involved in the intervention. Random numbers were produced by a computer program and four clusters were randomly assigned to the intervention or control group (two clusters each group) (Fig. 1). The nature of the intervention meant that allocation was not concealed.

Interventions

The training programme for suicide prevention implemented in this study for Japanese medical residents (hereafter, Gatekeeper Training Program) was developed by the research team based on the Mental Health First Aid (MHFA) programme implemented in Australia, which was originally developed for the general public. The MHFA programme is a training
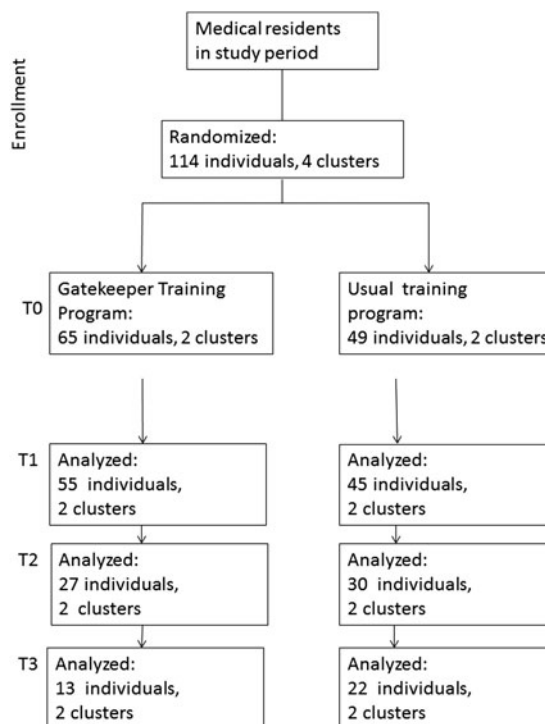

Fig. 1. Flow of the participants in the trial. T0: baseline response before intervention; T1: response immediately after the intervention; T2: response 1 month after the intervention; T3: response 3 month after the intervention.

programme designed to help non-health professionals handle contacts with persons in mental health crisis and this programme has been well researched (Kitchener & Jorm, 2006) and its effectiveness demonstrated in Australia (Jorm *et al.* 2004, 2005).

This programme was a 2-hour structured educational programme on managing people with depression and suicidal thoughts. Since the participants in this study were medical residents, the programme was condensed by minimizing the knowledge-based content of the lecture and including more experiential learning, an essential feature of the MHFA programme. To fit the Gatekeeper Training Program into the tightly arranged residency curriculum for medical residents, the only topic taught in the programme was depression and suicide, including current needs to respond to suicide prevention. With reference to the existing mental health first aid guidelines for Japan (Colucci *et al.* 2011), the training covered the approaches which are universally endorsed, rather than taking a specific approach relevant only to Japanese culture. The first part of the programme presented, in lecture format, factual information on depression and suicide in Japan. Action plans for management of people with depression and suicidal thoughts were provided. In the second part, a clinical scenario was shown in a demonstration DVD in which a hospitalized patient with a physical problem and suicidal thoughts was referred to a physician at the bedside. The facilitator then prompted the participants to role-play the scene by taking on different roles of the patient, family member and medical staff. An interactive discussion that followed was facilitated by certified facilitators to address the questions and comments the participants raised. The details of the programme and results of a pilot study are described elsewhere (Kato *et al.* 2010).

For comparative purposes, one of the researchers presented the usual training session, a didactic lecture on depression and suicidality, the contents of which were based on a standard psychiatric textbook (Nishijima, 2008).

Measures

A self-administered questionnaire comprised of the following inventories and items were distributed to the study participants.

The Suicide Intervention Response Inventory (SIRI)

To assess the participants' competency in managing suicide intervention, we used the SIRI, a 25-item self-rated questionnaire comprised of forced-choice responses to brief clinical interview excerpts with suicidal persons. The SIRI is a well-validated and reliable instrument that is sensitive to change in suicide interview skills as a result of training (Neimeyer & Bonnelle, 1997). On the SIRI, respondents are asked to rate the appropriateness of each response alternative on a 7-point Likert scale ranging from -3 (very inappropriate) to 3 (very appropriate).

The SIRI was chosen to evaluate competency in suicide management skills quantitatively because, to the best of the authors' knowledge, this is the only assessment tool validated to measure the competency level in suicide intervention in Japan (Kawashima *et al.* 2010). Also, when physicians encounter a suicidal patient in a general medical setting, they are expected to show an empathetic, non-judgemental attitude, and this is one of the issues that the SIRI aims to measure.

Originally, the SIRI score was calculated by summing the number of appropriate responses made, with a score ranging from 0 to 25 (SIRI-1 score). However, since this method of scoring demonstrated a ceiling effect, the use of a revised scoring system (hereafter, SIRI-2 score) was recommended (Neimeyer & Pfeiffer, 1994). The SIRI-2 score is calculated by summing the discrepancy between participants' ratings of the appropriateness of the responses made and a set of criterion ratings by a panel of American suicidology experts (hereafter, SIRI-2 Original score). The SIRI-2 Original score ranges from 12.90 to 247.28, with lower scores indicating better management of suicidal persons. The SIRI has been translated into Japanese and validated among Japanese suicidologists (Kawashima *et al.* 2010), and thus the discrepancy between the participants' ratings and Japanese experts' ratings can also be calculated (hereafter, the SIRI-2 Japanese score). The primary outcome was chosen as the SIRI-1 score in this study, and the SIRI-2 Original and SIRI-2 Japanese scores were calculated as secondary measures.

The SIRI was administered before the intervention (T0), immediately after the intervention (T1), 1 month after the intervention (T2) and 3 months after the intervention (T3).

Item for assessing confidence in managing suicidal persons

The participants were asked to indicate their confidence in managing suicidal persons on a 5-point Likert scale, which was originally developed for this study, ranging from 1 (not confident at all) to 5 (very confident).

Medical Outcome Study (MOS) Quality of Life (QOL) Questionnaire-Short Form

The MOS QOL Questionnaire-Short Form (SF-8) was used to assess the participants' health status

(Fukuhara & Suzukamo, 2004). The SF-8 is an 8-item self-report questionnaire to assess health-related QOL. It yields two summary measures of physical and mental health: the Physical Component Summary and Mental Component Summary. QOL was measured in this trial because QOL among medical residents generally tends to be poor and is one of the targets for improvement (Drolet *et al.* 2012), and they commonly encounter cases of suicide, which brings significant psychological stress that would affect their QOL (Fang *et al.* 2007). Moreover, a previous study on the MHFA programme demonstrated positive effects on non-professional participants' QOL (Kitchener & Jorm, 2006).

Background characteristics of participants

Information was collected on the demographic and educational characteristics of the participants, as well as their own and close relatives' experience of mental health problems.

Sample size calculation

Regarding the primary outcome of SIRI, the Gatekeeper Training Program group was expected to show a greater increase in SIRI-1 score between T0 and T2 than the control group. Based on our pilot study with a pre-post design, the mean (standard deviation, S.D.) of the SIRI-1 score was 18.31 (2.0) at T0 (unpublished data). The mean difference between SIRI-1 score at T2 and T0 was 1.05 (1.87). From the results of this pilot study, we expected the Gatekeeper Training Program group to increase their SIRI-1 score by 1.5 (2.0) at a power of 0.8 and alpha level of 0.05, with the 4 clusters. With intraclass correlates of 0.15 (Jorm *et al.* 2004), the number of participants needed for one cluster was 22, giving 88 participants for 4 clusters. The number of participants needed for recruitment was estimated as 104, based on 10% attrition at T2.

Analysis

Analysis was by the intention-to-treat principle. In the adjusted analysis, we applied a random (mixed) effects model under the missing-at-random assumption to the intention-to-treat population. We included variation of the allocated clusters in the model in order to allow for the possibility that outcome measures from individuals within the allocated cluster might not be independent in the analysis of the cluster randomized trial (Campbell & Grimshaw, 1998). Mean differences in the two groups from T0 to T1, T2 and T3 and 95%

confidence intervals (CI) were calculated. SAS version 9.1 and JMP version 6.0J were used for all analyses.

Results

From among a total of 120 medical residents on their psychiatric rotation during the study period, six declined to participate and data from the remaining 114 residents (95%) were analysed. The characteristics of the Gatekeeper Training Program and the control programme participants are presented in Table 1. No notable difference was observed in the baseline characteristics of the two groups.

First, with regard to the primary outcome, no group difference was observed in increase in SIRI-1 score at any time point (Table 2). Next, a comparison of the SIRI-2 Original score between the Gatekeeper Training Program and the control groups revealed that the scores decreased in both groups, indicating that the rating gap between Western suicidologists and study participants had narrowed after the training. However, there was no difference between the Gatekeeper Training Program group and the control group. At T3, the difference in the score between T3 and T0 was −8.32 (S.D. = 6.14) in the Gatekeeper Training Program group and 1.00 (12.88) in the control group, showing a group difference, although the proportion of responses was very low at this final time point. Last, a comparison of the SIRI-2 Japanese version score between the two groups revealed a difference between them: score at T1 was 61.28 (18.06) for the Gatekeeper Training Program group and 53.05 (15.79) for the control group ($p = 0.017$), and the difference in the score between T1 and T0 was 9.98 (95% CI; 4.39–15.56) between the two groups in the adjusted model, with a significant group difference ($p = 0.001$).

With respect to the residents' interest in mental health problems and confidence in their care of suicidal persons, the interest score lowered immediately after the training, although it did so in both groups, and the confidence score increased by about 1 unit in both groups (Table 3). There was no difference in change in QOL between the two groups at any time point in either the physical or mental domain on the SF-8 (Table 4).

Discussion

This study evaluated the effectiveness of the Gatekeeper Training Program, a brief suicide management training programme, in comparison with the usual programme involving a didactic lecture on depression and suicidality. When measuring competency in managing suicidal persons using the SIRI-1

Table 1. *Participant characteristics (N = 114)*

	Gatekeeper Training Program N = 65	Usual training programme N = 49
Institution, *n* (%)		
Hokkaido University	15 (23.1)	13 (26.5)
Iwate Medical University	3 (4.6)	0 (0.0)
Keio University	13 (20.0)	13 (26.5)
Yokohama City University	11 (16.9)	8 (16.3)
Kyushu University	23 (35.4)	15 (30.6)
Sex, *n* (%)		
Male	36 (55.4)	29 (59.2)
Female	29 (44.6)	20 (40.8)
Age		
Mean (S.D.)	27.5 (2.3)	28.3 (4.2)
Median (25–75%)	27 (26–28)	27 (26–28)
Intended specialty, *n* (%)		
Psychiatry, psychosomatic medicine	5 (7.7)	2 (4.1)
Medicine	20 (30.8)	18 (36.7)
Surgery	7 (10.8)	12 (24.5)
Paediatrics	6 (9.2)	3 (6.1)
OB/GYN	3 (4.6)	2 (4.1)
Family medicine	2 (3.1)	0 (0.0)
Unknown	7 (10.8)	1 (2.0)
Other	15 (23.1)	11 (22.4)
Personal experience of mental health problem, *n* (%)		
Own problem	14 (21.5)	13 (26.5)
Relatives' problem	19 (29.2)	14 (28.6)
Competency in suicide management		
SIRI-1	19.06 (2.49)	18.47 (3.64)
SIRI-2 Original version	64.31 (14.04)	65.96 (15.47)
SIRI-2 Japanese version	52.05 (15.30)	53.97 (16.58)
Interest in mental health problems		
Mean (S.D.)	3.1 (1.0)	2.9 (0.9)
Median (25–75%)	3 (2–4)	3 (2–3)
Confidence in care of suicidal persons		
Mean (S.D.)	1.6 (0.7)	1.7 (0.5)
Median (25–75%)	2 (1–2)	2 (1–2)
Confidence in comparison with previous month		
Mean (S.D.)	3.23 (0.50)	3.21 (0.62)
Median (25–75%)	3 (3–4)	3 (3–4)
SF-8		
Physical component summary	52.55 (4.76)	51.76 (4.48)
Mental component summary	48.17 (5.52)	49.20 (4.81)

SIRI-1: sum of the number of appropriate responses made on SIRI. Higher score indicates appropriate competency in managing suicidal persons.

SIRI-2 Original version: sum of the discrepancy between participants' ratings of appropriateness of the responses made and a set of criterion ratings by a panel of American suicidology experts in SIRI. Higher score indicates fewer competencies in managing suicidal persons compared with the competencies of American suicidology experts.

SIRI-2 Japanese version: sum of the discrepancy between participants' ratings of appropriateness of the responses made and a set of criterion ratings by a panel of Japanese health workers on SIRI. Higher score indicates fewer competencies in managing suicidal persons compared with the competencies of Japanese health workers.

SF8: MOS QOL questionnaire-short form.

score, there was no difference between the two groups at individual level. As a cluster was a combination of rotation groups, we cannot generalize the results to the medical residents at group level; however, it may be applicable at individual level. Below we take a closer look at different scoring methods used in this

Table 2. *SIRI score for Gatekeeper Training Program group and usual training programme group*

	Gatekeeper Training Program $N = 65$ Mean (S.D.)	Usual training programme $N = 49$ Mean (S.D.)	Mean difference (95% CI)	*p*-Value
SIRI-1				
T1–T0 (observed cases)	(*n* = 60)	(*n* = 47)		
Unadjusted	0.40 (2.63)	0.81 (2.20)	−0.41 (−1.35–0.54)	0.394
Adjusted			−0.26 (−1.13–0.62)	0.560
T2–T0 (observed cases)	(*n* = 28)	(*n* = 30)		
Unadjusted	0.50 (2.53)	0.47 (2.17)	0.03 (−1.21–1.27)	0.957
Adjusted			0.39 (−0.73–1.52)	0.487
T3–T0 (observed cases)	(*n* = 16)	(*n* = 22)		
Unadjusted	0.94 (2.89)	−0.09 (2.91)	1.03 (−0.90–2.96)	0.288
Adjusted			1.25 (−2.12–4.63)	0.375
SIRI-2 Original version				
T1–T0 (observed cases)	(*n* = 55)	(*n* = 45)		
Unadjusted	−3.82 (12.56)	−4.20 (9.20)	0.39 (−4.07–4.84)	0.864
Adjusted			0.32 (−2.34–2.97)	0.814
T2–T0 (observed cases)	(*n* = 27)	(*n* = 30)		
Unadjusted	−1.45 (14.47)	−1.00 (10.09)	−0.44 (−7.01–6.12)	0.893
Adjusted			−1.22 (−7.70–5.26)	0.708
T3–T0 (observed cases)	(*n* = 13)	(*n* = 22)		
Unadjusted	−8.32 (6.14)	1.00 (12.88)	−9.33 (−17.10–1.55)	0.020
Adjusted			−10.36 (−23.17–2.44)	0.094
SIRI-2 Japanese version				
T1–T0 (observed cases)	(*n* = 55)	(*n* = 45)		
Unadjusted	9.47 (14.83)	−0.65 (14.22)	10.12 (4.31–15.93)	0.001
Adjusted			9.98 (4.39–15.56)	0.001
T2–T0 (observed cases)	(*n* = 27)	(*n* = 30)		
Unadjusted	7.15 (12.29)	−0.50 (16.54)	7.64 (−0.16–15.45)	0.055
Adjusted			6.82 (−0.93–14.57)	0.083
T3–T0 (observed cases)	(*n* = 13)	(*n* = 22)		
Unadjusted	10.48 (13.41)	6.47 (18.6)	4.01 (−8.02–16.04)	0.502
Adjusted			3.51 (−15.09–22.11)	0.658

Adjusted by score at T0 and cluster effects.
SIRI-1: sum of the number of appropriate responses on SIRI.
SIRI-2 Original version: sum of the discrepancy between participants' ratings of appropriateness of intervention and a set of criterion ratings by a panel of American suicidology experts on SIRI.
SIRI-2 Japanese version: sum of the discrepancy between participants' ratings of appropriateness of intervention and a set of criterion ratings by a panel of Japanese health workers on SIRI.

study, discuss further the programme contents and explore issues of cultural sensitivity.

Evaluation of the Gatekeeper Training Programme according to different SIRI scoring methods

First, with respect to the primary outcome measure, SIRI-1, which involves simple arithmetic addition of 25 items, we failed to find improved competency in suicide management. One possible reason for this is that the medical residents already had basic medical knowledge on depression and suicide-related behaviour, therefore, the degree of improvement was limited (a ceiling effect). In addition, the problem associated with the SIRI-1 score is that zero is assumed for any participant that did not respond. As there was a high proportion of such cases in our dataset, the distribution of data became skewed to zero, indicating wrongly that the score decreased regardless of the programme's effect.

Second, the Gatekeeper Training Program was found to be marginally effective when measured using the SIRI-2 Original score at T3. Given the medical residents' already higher level of knowledge than non-health professionals', our programme intended to further improve the residents' competency through

Table 3. *Scores for interest in mental health problems and confidence in care of suicidal persons for Gatekeeper Training Program group and usual training programme group and mean differences*

	Gatekeeper Training Program N = 65 Mean (S.D.)	Usual training programme N = 49 Mean (S.D.)	Mean difference (95% CI)	p-Value
Interest, mean (S.D.)				
T1–T0 (observed cases)	(n = 60)	(n = 46)		
Unadjusted	−0.95 (1.23)	−0.94 (1.18)	−0.02 (−0.48–0.45)	0.949
Adjusted			−0.04 (−0.58–0.65)	0.885
T2–T0 (observed cases)	(n = 28)	(n = 30)		
Unadjusted	−0.21 (0.92)	0.03 (1.07)	−0.25 (−0.77–0.27)	0.349
Adjusted			−0.12 (−0.61 − 0.38)	0.634
T3–T0 (observed cases)	(n = 16)	(n = 22)		
Unadjusted	−0.13 (1.15)	−0.27 (0.77)	0.15 (−0.48–0.78)	0.637
Adjusted			0.16 (−0.49–0.81)	0.622
Confidence, mean (S.D.)				
T1–T0 (observed cases)	(n = 60)	(n = 46)		
Unadjusted	1.62 (1.29)	1.13 (1.07)	0.49 (0.02–0.95)	0.041
Adjusted			0.45 (−0.23–1.10)	0.164
T2–T0 (observed cases)	(n = 28)	(n = 30)		
Unadjusted	0.50 (0.51)	0.23 (0.73)	0.27 (−0.07–0.60)	0.114
Adjusted			0.25 (−0.18–0.68)	0.207
T3–T0 (observed cases)	(n = 16)	(n = 22)		
Unadjusted	0.56 (0.63)	0.27 (0.83)	0.29 (−0.21–0.79)	0.248
Adjusted			0.28 (−0.59–1.15)	0.451

Adjusted by score at T0 and cluster effects.

participatory learning. It may take longer and more training to effect a change in competency than a change in knowledge because the participants must take back what they have learned and examine whether it really works or not within their own clinical practice, although evidence suggests that those who work in suicidality on a regular basis early in their career benefit most from training (Saunders *et al.* 2012). In the present study, the actual number of valid returns was low at T3, hence there is the possibility that only self-selected medical residents who were clearly interested and motivated in suicide

Table 4. *QOL score determined by the MOS QOL questionnaire-short form (SF-8) for Gatekeeper Training Program group and usual training programme group and mean differences*

	Gatekeeper Training Program N = 65 Mean (S.D.)	Usual training programme N = 49 Mean (S.D.)	Mean difference (95% CI)	p-Value
Physical component summary				
T2–T0 (observed cases)	(n = 27)	(n = 30)		
	0.62 (5.62)	1.41 (5.33)	−2.03 (−4.93–0.88)	0.168
T3–T0 (observed cases)	(n = 16)	(n = 22)		
	−2.79 (6.33)	−1.00 (1.34)	−1.80 (−5.99–2.39)	0.390
Mental component summary				
T2–T0 (observed cases)	(n = 27)	(n = 30)		
	0.18 (4.68)	−1.75 (8.60)	1.94 (−1.80–5.67)	0.303
T3–T0 (observed cases)	(n = 16)	(n = 22)		
	1.01 (5.09)	−0.14 (7.59)	1.15 (−3.30–5.59)	0.604

management answered the questionnaire at T3. Bearing this in mind, a modest interpretation must be made of the programme's effectiveness.

Third, the SIRI-2 Japanese score increased at T1 (immediately after the intervention), indicating a discrepancy in judgement between the participants and Japanese suicidologists that we had not expected, although a previous study had similarly reported a remarkable difference in SIRI-2 Japanese score between intervention programme participants and Japanese suicidologists (Kawashima *et al.* 2010). In examining the correlation of the three scores, Pearson's correlation coefficient was 0.5907 ($p < 0.01$) between the SIRI-2 Original and SIRI-2 Japanese, suggesting a moderate correlation. In addition, the correlation coefficient was -0.8347 ($p < 0.01$) for baseline scores on the SIRI-1 and SIRI-2 Original, and -0.4251 ($p < 0.01$) for baseline scores on the SIRI-1 and SIRI-2 Japanese, suggesting a strong correlation between SIRI-1 and SIRI-2 Original and a moderate correlation between SIRI-1 and SIRI-2 Japanese. It should be noted that the Japanese experts who provided the ratings from which the SIRI-2 Japanese score was derived tended to judge the appropriateness of suicide management response modestly compared with Western suicidologists (Kawashima *et al.* 2010), and this might explain why, after receiving the Gatekeeper Training Program which was originally developed based on Western views, participants' judgement was more congruent with those of Western suicidologists.

Ambiguity over the principles of suicide management among first aiders was pointed out as a characteristic of Japanese mental health professionals, as many selected the 'do not know, or it depends' option in the consensus building process (Colucci *et al.* 2011). Furthermore, the role that first aiders, especially non-specialist health professionals, should play in the suicide prevention framework and the way they should approach persons with suicidal ideation was not well recognized by Japanese mental health professionals, as suggested by the consensus building process results not endorsing that persons with suicidal ideation themselves or others close to them can manage their psychological crisis (Colucci *et al.* 2011). Thus, while the standards of care for persons with suicidal behaviours have been established (Kuwabara *et al.* 2009), consensus is yet to be reached on the approach people should take in suicide management in Japan. This evolution of suicide prevention strategy in Japan might have had an effect on the results. In spite of there still being the need to firm up experts' views, from a cultural perspective it is clearly important for participants to bring their suicide management skills more in line with those of Japanese experts.

Evaluation of the Gatekeeper Training Programme in relation to medical residents' confidence and QOL

In this study, we adapted the MHFA programme that was originally developed for lay persons and which was reported by previous evaluation studies to have improved knowledge, competency and QOL among such participants, compared with the usual programmes (Jorm *et al.* 2004). The present study similarly showed that medical residents' confidence in suicide management improved, and the improvement was greater, among participants in the Gatekeeper Training Program group than among the control group receiving the usual programme, at a marginally significant level.

No differences in participants' QOL were revealed between the two groups. One possible explanation is that overload and burnout was common among the medical residents, as has been commonly reported (Dyrbye *et al.* 2008), and that the intervention was too brief under such circumstances to address their QOL, which is a complex interplay of academic competency, financial status and other work characteristics (West *et al.* 2009; Campbell *et al.* 2010). Improving skills to handle difficult patients, such as those with suicidality, may contribute to disentangling these factors, whereas a stand-alone training programme may have limited effect in improving QOL among medical residents. However, in the long run, receiving a suicide management programme in advance of clinical encounters may enable residents to handle emotional distress when actually faced with patients exhibiting suicidal behaviour in a better manner.

Limitations

This study did not address the assessment of emotional and behavioural change when residents encounter suicidal patients. Our ultimate goal was to change medical residents' competency, and a more behavioural assessment anchored by objective observation would have provided a stronger basis for evaluating the programme's effect. Considering current knowledge on suicide prevention measures and the complex interplay of cause and path of suicide-related factors, suicide prevention strategies should be underpinned by multi-faceted training programmes: any single programme may not be able to contribute enough to effecting a gross change. The limited observation period in the present study, during which encounters with patients with suicidality were presumably few, did not allow us to adopt these objective behavioural measures. However, it is noteworthy that the residents' confidence was increased after the intervention.

With regard to methodological issues, the CONSORT 2010 Statement (Campbell *et al.* 2012) recommended that the minimum number of clusters per arm should be at least four. However, our study had only two clusters per arm owing to practical issues with recruiting the participants. Therefore, our study might be still confounded by cluster effects.

The proportion of residents that participated at the follow-up time points was low and the number of respondents in the two groups was unbalanced. At T1, the proportion lost to follow-up was 7.7% (5/65) and 4.1% (2/49), respectively, but more than a half of the participants did not return the questionnaire at T2 and T3, and therefore some respondent and follow-up bias might be present. In the medical residency system, psychiatric rotation is allocated only for 1 month before the residents move on to training in another department. The follow-up questionnaires were sent out when the medical residents were busy adjusting to training in the new department and they might therefore have been too busy to respond to the mailed questionnaire. Several reminders of the follow-up response were sent, but resulted in very few further returns. We took the view not to send out further additional reminders so as not to give a negative impression about psychiatric training by insistently demanding a response from the participants. This dilemma, resulting in a major limitation of the study, is inherent to this kind of study.

In addition, the proportion lost to follow-up was greater in the Gatekeeper Training Program group than in the control group for reasons that are unknown. We compared the characteristics of the respondents and drop-outs at T2 (data are available upon request), but no differences were found in scores for SIRI, interest in mental health or confidence in managing people with suicidal thoughts. Thus, unknown factors might have affected the respondents' data. The results may well be different if we had a dataset for those who were lost to follow-up.

In conclusion, our study did not demonstrate that the Gatekeeper Training Program improved medical residents' competency in suicide management when measured by the SIRI-1 score, and it showed an unexpected, unfavourable effect in the short term when measured by the SIRI-2 Japanese version score. While this study failed to show that the Gatekeeper Training Program increased medical residents' competency in suicide management skills, this may be due to a mixture of contributing factors. Selection of the SIRI as the primary outcome measure is questionable because of its limited validity, which suggests that the Japanese community must reconsider and build consensus on the principles first aiders should follow when encountering persons with suicidal thoughts.

Further elaboration of the programme and valid measurement of its outcome would be needed to show the programme's effectiveness.

Conflict of Interest

The authors have no conflicts of interest.

Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The study protocol was approved by the institutional review board of each participating institution and the National Center of Neurology and Psychiatry. Written informed consent was obtained from all participating medical residents.

References

Campbell MK, Grimshaw JM (1998). Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored. *British Medical Journal* **317**, 1171–1172.

Campbell J, Prochazka AV, Yamashita T, Gopal R (2010). Predictors of persistent burnout in internal medicine

residents: a prospective cohort study. *Academic Medicine* **85**, 1630–1634.

Campbell MK, Piaggio G, Elbourne DR, Altman DG; CONSORT Group (2012). Consort 2010 statement: extension to cluster randomised trials. *British Medical Journal* **345**, e5661.

Colucci E, Kelly CM, Minas H, Jorm AF, Suzuki Y (2011). Mental Health First Aid guidelines for helping a suicidal person: a Delphi consensus study in Japan. *International Journal of Mental Health System* **19**, 5–12.

Drolet BC, Christopher DA, Fischer SA (2012). Residents' response to duty-hour regulations – a follow-up national survey. *New England Journal of Medicine* **366**, e35. Retrieved 9 April 2013 from http://www.nejm.org/doi/full/10.1056/NEJMp1202848.

Dyrbye LN, Thomas MR, Massie FS, Power DV, Eacker A, Harper W, Durning S, Moutier C, Szydlo DW, Novotny PJ, Sloan JA, Shanafelt TD (2008). Burnout and suicidal ideation among U.S. medical students. *Annals of Internal Medicine* **149**, 334–341.

Fang F, Kemp J, Jawandha A, Juros J, Long L, Nanayakkara S, Stepansky C, Thompson LB, Anzia J (2007). Encountering patient suicide: a resident's experience. *Academic Psychiatry* **31**, 340–344.

Fujisawa D, Hashimoto N, Masamune-Koizumi Y, Otsuka K, Tateno M, Okugawa G, Nakagawa A, Sato R, Kikuchi T, Tonai E, Yoshida K, Mori T, Takahashi H, Sato S, Igimi H, Waseda Y, Ueno T, Morokuma I, Takahashi K, Sartorius N (2008). Pathway to psychiatric care in Japan: a multicenter observational study. *International Journal of Mental Health Systems* **2**, 14.

Fukuhara S, Suzukamo Y (2004). *Manual of the SF-8 Japanese Version*. Institute for Health Outcomes & Process Evaluation Research: Kyoto.

Jorm AF, Kitchener BA, O'Kearney R, Dear KB (2004). Mental health first aid training of the public in a rural area: a cluster randomized trial [ISRCTN53887541]. *BioMed Central Psychiatry* **4**, 33.

Jorm AF, Kitchener BA, Mugford SK (2005). Experiences in applying skills learned in a Mental Health First Aid training course: a qualitative study of participants' stories. *BioMed Central Psychiatry* **5**, 43.

Kato TA, Suzuki Y, Sato R, Fujisawa D, Uehara K, Hashimoto N, Sawayama Y, Hayashi J, Kanba S, Otsuka K (2010). Development of 2-hour suicide intervention program among medical residents: first pilot trial. *Psychiatry and Clinical Neurosciences* **64**, 531–540.

Kawashima D, Kawano K, Ito H (2010). Development of the Japanese version of the Suicide Intervention Response Inventory (SIRI). *Seishinigaku* **52**, 543–551.

Kitchener BA, Jorm AF (2006). Mental health first aid training: review of evaluation studies. *Australian and New Zealand Journal of Psychiatry* **40**, 6–8.

Kozu T (2006). Medical education in Japan. *Academic Medicine* **81**, 1069–1075.

Kuwabara H, Kawanishi C, Kawano K, Ito H (2009). To support person with suicidality, guidelines for counsellors [online]. Retrieved 9 April 2013 from http://www.mhlw.go.jp/bunya/shougaihoken/jisatsu/dl/02.pdf.

Mann JJ, Apter A, Bertolote J, Beautrais A, Currier D, Haas A, Hegerl U, Lonnqvist J, Malone K, Marusic A, Mehlum L, Patton G, Phillips M, Rutz W, Rihmer Z, Schmidtke A, Shaffer D, Silverman M, Takahashi Y, Varnik A, Wasserman D, Yip P, Hendin H (2005). Suicide prevention strategies: a systematic review. *Journal of the American Medical Association* **294**, 2064–2074.

Morriss R, Gask L, Webb R, Dixon C, Appleby L (2005). The effects on suicide rates of an educational intervention for front-line health professionals with suicidal patients (the STORM Project). *Psychological Medicine* **35**, 957–960.

Neimeyer RA, Bonnelle K (1997). The Suicide Intervention Response Inventory: a revision and validation. *Death Studies* **21**, 59–81.

Neimeyer RA, Pfeiffer AM (1994). Evaluation of suicide intervention effectiveness. *Death Studies* **18**, 131–166.

Nishijima H (2008). *Suicide Prevention Manual*. Akashi Shoten: Tokyo.

Sato R, Kawanishi C, Yamada T, Hasegawa H, Ikeda H, Kato D, Furuno T, Kishida I, Hirayasu Y (2006). Knowledge and attitude towards suicide among medical students in Japan: preliminary study. *Psychiatry and Clinical Neurosciences* **60**, 558–562.

Saunders KE, Hawton K, Fortune S, Farrell S (2012). Attitudes and knowledge of clinical staff regarding people who self-harm: a systematic review. *Journal of Affective Disorders* **139**, 205–216.

Shiho Y, Tohru T, Shinji S, Manabu T, Yuka T, Eriko T, Ikuko S, Naoki W (2005). Suicide in Japan: present condition and prevention measures. *Crisis* **26**, 12–19.

Teo A (2007). The current state of medical education in Japan: a system under reform. *Medical Education* **41**, 302–308.

West CP, Tan AD, Habermann TM, Sloan JA, Shanafelt TD (2009). Association of resident fatigue and distress with perceived medical errors. *Journal of the American Medical Association* **302**, 1294–1300.