

Table 2 Brief summary of articles based on structured abstracts and additional elements.

Reference no.	17	18	19	20
Author	Jarrott SE, et al.	Kam MCY, et al.	Tse MMY.	Kim SY, et al.
Citation	Am J Alzheimers Dis Other Demen 2010; 25: 657–65.	Hong Kong J Occup Therapy 2010; 20: 80–86.	J Clin Nurs 2010; 19: 949–958	J Korean Soc Hortic Sci 2003; 44: 780–785
Title	Comparing responses to horticultural-based and traditional activities in dementia care programmes.	Evaluation of a horticultural activity programme for persons with psychiatric illness	Therapeutic effects of an indoor gardening programme for older people living in nursing homes	Effect of horticultural therapy on functional rehabilitation in hemiplegic patients after stroke
Aim/objective	To identify whether participants in horticultural-based activities (HT-based activities) exhibit higher levels of adaptive behavior and experience more positive affect than participants in traditional activities (TA) during the presented activities.	To investigate the effect of applying horticulture activity on stress, work performance and quality of life in persons with psychiatric illness.	To explore the activities of daily living and psychological well-being of older people living in nursing homes and also to examine the effectiveness of a gardening programme in enhancing socialisation and life satisfaction, reducing, loneliness and promoting activities of daily living for older people living in nursing homes.	To evaluate the influence of horticultural therapy (HT) activities on the improvement in rehabilitation of individuals with stroke patients
Setting/place	8 care programmes (facilities) in rural south west Virginia	All the horticulture activities took place in the five outdoor theme gardens of the Farm, namely the Sensory garden, Activity garden, Farm garden, Display garden and Practical garden in Hong Kong.	All gardening activities were carried out in the multiple function room of the nursing homes in Hong Kong.	National rehabilitation center in Korea
Participants	129 persons with a diagnosis of dementia	Twenty-four participants with severe mental illness such as schizophrenia spectrum disorder (schizophrenia, schizoaffective disorder, schizophreniform disorder, psychosis not otherwise specified), bipolar disorder, or major depression were recruited to participate in this study.	Fifty-three participants in nursing homes were recruited to participate in this study.	Forty-two individuals among poststroke hemiplegic patients

Table 2 (Continued)

Reference no.	17	18	19	20
Intervention	HT-based activities were implemented twice weekly at 4 treatment sites for 6 weeks, while regular TA were observed at comparison sites. Two facilitators developed and facilitated HT-based activities at the treatment sites. Activities were selected for anticipated therapeutic benefits within social, physical, psychological, and cognitive domains. In addition, activities were simple, cost-effective, and versatile. Plant materials were researched to ensure safety. Group sizes at HT-based activities ranged from 4 to 20 participants; the average number of participants varied at each site. When group size exceeded 8 participants, the facilitators divided participants into 2 groups to provide step-by-step instructions, physical and verbal assistance, and a steady supply of materials to each participant. Participation was always voluntary and elders were informed that the focus of the activities would be gardening. Horticultural therapy-based activities ranged from sowing seeds and training topiaries to craft activities that incorporated horticultural materials or themes.	The horticulture activity programme was a standardized group that was conducted in 10 consecutive days within 2 weeks. The group objectives are: (a) to teach basic horticultural knowledge and skills, and to develop interests in working with plants, (b) to share relaxing experience and coping strategies through working with plants, (c) to promote sharing and social support among participations. Participants in the experimental group attended a 1-h horticultural activity session for 10 days. Once they finished the 1-h horticulture session, they joined the regular sheltered workshop training. Participants in the central group remained in the conventional workshop training throughout the study period. Based on the work of previous studies, each session of the standardized horticultural programme have a specific theme and objectives, and all main activities were related to work with living plants. In each session, there is an introduction and warm-up, followed by horticulture activities and group sharing.	Older people from the experimental groups were invited to join the eight week indoor gardening programme, while older people in the other two nursing homes were treated as the control groups: they received regular care and visits each week, but not the eight week indoor gardening programme. Posttest measures were collected on the completion of the eight week indoor gardening programme for all participants and participant of the experience of the experimental groups were invited for an interview regarding their experience of the programme. Participants in the experimental group were invited to attend an interview conducted by the researcher at the end of the eight week indoor gardening programme. The interviews were conducted in the function rooms of the nursing homes. Interview lasted from 10 to 15 min. To created before proceeding to ask the participants about the gardening experience.	Treatment group attended a 1 h horticultural therapy programme 5 times per week for 6 weeks. The programme carried out during horticultural therapy included various activities such as seeding, growing, and utilization of postharvest products to make them to experience a whole life cycle of plants. For evaluating the effect of programme several scales including self-esteem scale, powerlessness scale, beck depression inventory, neurobehavioral cognitive status examination, motor-free visual perception test, and functional independence measure were applied. Measurements were conducted at 2 times periods; before and after 6 weeks horticultural therapy programme.

Main and secondary outcomes	Affect (the Apparent Affect Rating Scale) and engagement (Menorah Park Engagement Scale)	The Chinese version of Depression Anxiety Stress Scale 21 (DASS21), Work Behavior Assessment (WBA), and Chinese version Personal Wellbeing Index (PWI-C).	The Life Satisfaction Index-A Form, The Revised UCLA Loneliness Scale, and The Lubben Social Network Scale.	Self-esteem scale, powerlessness, beck depression inventory (BDI), neurobehavioral cognitive status examination (NCSE), motor-free visual perception test (MVPT), and functional independence measure (FIM).
Randomization	No description	They were assigned a number and randomly allocated by an independent research assistant (voluntary helper) who was blinded to the hypothesis and the intervention programme.	No description	No description
Blinding/masking	No description	A blinded independent assessor (voluntary helper) conducted all the outcome measurements before and after the intervention programme.	No description	No description
Numbers randomized	HT-based activities group ($n = 75$) and control group ($n = 54$). These were allocated each facility unit.	Experimental group ($n = 12$) and control group ($n = 12$)	Experimental group ($n = 26$) and control group ($n = 27$). They were allocated each facility unit.	Treatment group ($n = 21$) and control group ($n = 21$)
Recruitment	8 care programmes (facilities) in rural southwest Virginia	Twenty-four participants with severe mental illness such as schizophrenia spectrum disorder (schizophrenia, schizoaffective disorder, schizophreniform disorder, psychosis not otherwise specified), bipolar disorder, or major depression were recruited to participate in this study.	An organization operating resident care homes for older people in Hong Kong was approached and invited to participate in the study.	Forty-two individuals with poststroke hemiplegic patients (right or left hemiplegia, cerebral infarct, hemorrhage) participated in patient stroke rehabilitation programme of national rehabilitation center were recruited to participate in this study.
Numbers analyzed	HT-based activities group ($n = 75$) and control group ($n = 54$).	Experiment group ($n = 10$) and control group ($n = 12$)	Experimental group ($n = 26$) and control group ($n = 27$). They were allocated each facility unit.	Treatment group ($n = 21$) and control group ($n = 21$)

Table 2 (Continued)

Reference no.	17	18	19	20
Outcome	Results revealed no differences between groups on affective domains. Levels of adaptive behavior differed between the groups, with the treatment group demonstrating higher levels of active, passive, and other engagement and the comparison group demonstrating higher levels of self-engagement.	There was a significant difference in change scores of the DASS21 ($p=0.01$) between experimental and control group. There were no significant differences in change scores of the PWI-C between the two groups.	There were significant improvements in life satisfaction and social network and a significant decrease in perception of loneliness for older people in the experimental group after the eight week indoor gardening programme, while the activities of daily living were unchanged for both groups after programme.	In treatment group, the self-esteem scale and the powerlessness were significantly improved, and beck depression inventory score, motor-free visual perception test raw score, and functional independence measure were also significantly decreased. Among the items of neurobehavioral cognitive status examination, similarity score was significantly increased.
Harm	Nothing	No description	No description	No description
Conclusion	Our results highlight the value of HT-based programmes and the importance of simultaneously capturing participants' affective and behavioral responses.	Horticultural therapy is effective in decreasing the levels of anxiety, depression and stress among participants in this pilot study, but the impact of the programme on work behavior and quality of life will need further exploration.	The present study demonstrates the therapeutic effects of an indoor gardening programme. Findings provide evidence for the inclusion of an indoor gardening programme in nursing home care practice with older people.	Horticultural therapy was very effective as one of the alternative medicines and played an important role in rehabilitation of the stroke patients, especially in neuropsychological aspects.
Trial registration	No description	No description	No description	No description
Fund	Alzheimer's and Related Diseases Research Award Fund from the Virginia Center on Aging (VCOA), Fund # 04-3.	No description	The Hong Kong Polytechnic University supported the study.	No description
Withdrawals	Nothing	There were two drop-outs in the experiment group. These reasons were not described.	Nothing	Nothing
Cost of intervention	No description	No description	No description	No description

Table 3 International Classification of target diseases in each article.

Chapter	ICD code	Classification	Reference no. (Detail ICD code)
1	A00-B99	Certain infectious and parasitic diseases	
2	C00-D48	Neoplasms	
3	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	
4	E00-E90	Endocrine, nutritional and metabolic diseases	
5	F00-F99	Mental and behavioral disorders	17(F00-01), 18(F20,30-33)
6	G00-G99	Diseases of the nervous system	20(G81)
7	H00-H59	Diseases of the eye and adnexa	
8	H60-H95	Diseases of the ear and mastoid process	
9	I00-I99	Diseases of the circulatory system	
10	J00-J99	Diseases of the respiratory system	
11	K00-K93	Diseases of the digestive system	
12	L00-L99	Diseases of the skin and subcutaneous tissue	
13	M00-M99	Diseases of the musculoskeletal system and connective tissue	
14	N00-N99	Diseases of the genitourinary system	
15	O00-O99	Pregnancy, childbirth and the puerperium	
16	P00-P96	Certain conditions originating in the perinatal period	
17	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities	
18	R00-R99	Symptoms, signs and abnormal clinical and laboratory finding not elsewhere classified	
19	S00-T98	Injury, poisoning and certain other consequences of external causes	
20	V00-Y98	External causes of morbidity and mortality	
21	Z00-Z99	Factors influencing health status and contact with health services	
22	U00-U99	Code for special purpose	
—	Unidentified	Because many illnesses were mixed, we could not identify it.	19

quality assessment indicated substantial agreement for all 44 items (percentage agreement 86% and $k=0.794$).

This assessment evaluated the quality of how the main findings of the study were summarized in the written report. There was a remarkable lack of execution and/or description in the randomization, concealment, blinding, and intention-to-treatment (ITT) analysis, in general. The items for which the description was lacking (very poor; <50%) in many studies were as follows: "Was the method of randomization adequate?" (25%); "Was the treatment allocation concealed?" (25%); "Was the patient blinded to the intervention?" (0%); "Was the care provider blinded to the intervention?" (0%); "Was the outcome assessor blinded to the intervention?" (25%); and "Did the analysis include an ITT analysis?" (25%).

Meta-analysis

We could not perform meta-analysis because of heterogeneity of all outcome measurements and intervention methods.

Withdrawals and adverse events

Three studies^{18–20} did not describe adverse events or harm (Table 2). Three studies^{17,19,20} reported no withdrawals (dropouts), and one study showed two dropouts, but reasons were not described.

Costs of intervention

All studies^{17–20} provided no information about the costs of intervention (Table 2).

Trial registration

There were no descriptions about this information in all studies.

Discussion

This is the first SR of the effectiveness of HT based on RCTs. Our study is unique because it summarized the evidence for

Table 4 Evaluation of the quality of methodology for each article.

No	Criteria list	17	18	19	20	Present description**	
						No/4	Rate (%)
1	Was the method of randomization adequate?	?	y	?	n	1	25%
2	Was the treatment allocation concealed?	?	y	?	n	1	25%
3	Were the groups similar at baseline regarding the most important prognostic indicators?	y	y	y	y	4	100%
4	Was the patient blinded to the intervention?	n	n	?	n	0	0%
5	Was the care provider blinded to the intervention?	n	n	n	n	0	0%
6	Was the outcome assessor blinded to the intervention?	n	y	n	n	1	25%
7	Were cointerventions avoided or similar?	y	y	y	y	4	100%
8	Was the compliance acceptable in all group?	?	y	y	n	2	50%
9	was the drop-out rate described and acceptable?	n	y	y	n	2	50%
10	Was the timing of the outcome assessment in all groups similar?	y	y	y	y	4	100%
11	Did the analysis include an intention-to-treat analysis?	n	n	y	n	1	25%
Present description no/11		3	8	6	3	—	
Rate (%)		27%	73%	55%	27%		

Yes: y, no: n, do not know or unclear: ?, not applicable: n/a.

each target disease according to ICD-10 classification. We assume that this study will be helpful to researchers who want to understand the effects of HT comprehensively, and it could provide indispensable information for the organization that is going to make guidelines according to each disease.

Among four RCTs that were identified, target diseases and/or symptoms included dementia, severe mental illness such as schizophrenia, bipolar disorder, and major depression, frail elderly in nursing home, and hemiplegic patients after stroke.

Tendency of target disease and outcome

The most reported target diseases were “Mental and behavioral disorders”,^{17,18} and the effect of HT on these diseases was improved mental health (e.g., anxiety and depression) and adaptive behavior. The main reason given in these articles for improved mental health might be that participants could relax from work stress and felt joyful in the horticultural environment during the program.¹⁸ This trend was also seen in a research study for hemiplegic patients after stroke, with the target disease category “Disease of the nervous system (G31)”.²⁰ The improvement on adaptive behavior for patients with dementia might also be that HT supported participants’ attainment of the adaptation level, which was characterized by optimal person-environment fit.¹⁷ Another study reported an unidentified targeted disease,¹⁹ and indoor HT for older people living in a nursing home improved mental-psychosocial outcomes. These common results were

congruent with previous studies that used other research designs.^{21,22}

Horticultural environment may play a significant role in the improvement of mental health in patients of various types, but there is some suggestion that the influence of facilitators is very important for outcome.^{17,23,24} An analysis of facilitators’ interaction styles and their knowledge of dementia-care practices would also supplement the evidence and guide subsequent practice.¹⁷

Overall evidence and quality assessment

The Cochrane’s list is the most important tool related to the internal validity of trials. In this SR, there were serious problems with the conduct and reporting of the target studies. Our review especially detected omissions of the following descriptions: method used to generate randomization, concealment, blinding, and ITT analysis. Descriptions of these items were lacking (very poor; <50%) in many studies. In addition, the results of this study suggested that the RCTs conducted have been of relatively low quality.

In the Cochrane Review,²⁵ the eligibility criteria for a meta-analysis are strict, and for each article, heterogeneity and low quality of reporting must first be excluded. We could not perform a meta-analysis. Due to poor methodological and reporting quality and heterogeneity, there was insufficient evidence in the studies of HT, and we are therefore unable to offer any clear conclusions about the effects of HT based on RCTs. Moreover, the CONSORT 2010²⁶ and the CONSORT for nonpharmacological trials checklists²⁷ are relatively new, but it was shown that the study protocol

Table 5 Overall evidence and future research agenda to build evidence of horticulture therapy.

Overall evidence in the present	Research agenda
Due to poor methodological and reporting quality and heterogeneity, although there was insufficient evidence in the studies of HT, it may be an effective treatment for mental and behavioral disorders such as dementia, schizophrenia, depression, terminal-care of cancer.	1. Satisfactory description and methodology including the CONSORT 2010 and the CONSORT for nonpharmacological trials 2. Description of adverse effects (e.g. allergy) and withdrawals 3. Description of intervention dose (if pragmatic intervention) 4. Description of cost 5. Power analysis 6. Development of the original check-list for HT 7. Effect on symptomatic relief of PTSD* caused by disaster or war

* Post-traumatic stress disorder (PTSD).

description and implementation for HT studies should be subjected to these checklists.

Although further accumulation of strict RCT data is necessary, HT may be effective treatment for the following diseases and symptoms: mental and behavioral disorders such as dementia, depression, and schizophrenia, and various other clinical conditions.

Future research agenda to build evidence

Table 5 shows the future research agenda for studies on the treatment effect of HT. In a recent review for gardening as a mental health intervention, Clatworthy et al.²⁸ also suggested that there is a need further high-quality research such as adequate outcome measures in this field. Researchers should use the appropriate checklists for research design, appropriate analysis methods such as power analysis, and intervention methods, which would lead to improvement in the quality of a study, and would contribute to the accumulation of evidence. Researchers should also present not only efficacy data, but also a description of any adverse events or harmful phenomena and the reasons for withdrawals and non-participation.

As a gradual increase of intervention is necessary in cure and rehabilitation programs, it is easy to assign settings like "Stage" for the intervention, such as first stage and second stage. Therefore, we also expect to understand the results and detailed descriptions of "pragmatic trials"²⁹ as well as "explanatory trials" for the treatment effect of HT. Jarrott et al. emphasized that systematically assessing the dosage of therapeutic programming needed to achieve desired outcomes would help program planners maximize benefit.¹⁷

Bowen et al.³⁰ suggested that public health is moving toward the goal of implementing evidence based intervention. But the feasibility of possible interventions, and whether comprehensive and multilevel evaluations are needed to justify them, must be determined. It is at least necessary to show the cost of such interventions. Introduction of an interventional method must be based on its cost–benefit, cost–effectiveness, and cost–utility. A recent preliminary study reported on the feasibility and acceptability of a mentored gardening intervention, and suggested that it may offer a novel and promising strategy to improve fruit

and vegetable consumption, physical activity, and physical function in cancer survivors.³¹

In addition, HT as an intervention is unique and completely different than pharmacological or traditional rehabilitation methods. Therefore, it may be necessary to add some original items like herbal intervention,³² acupuncture,³³ traditional Chinese medicine³⁴ and balneotherapy³⁵ to the CONSORT 2010 checklist as alternative and/or complementary medicines.

Wang et al.³⁶ suggested that gardening may be an activity that promotes overall health and quality of life, physical strength, fitness and flexibility, cognitive ability, and socialization. Furthermore we expect new research on victims of the East Japan Great Earthquake Disaster and the Collapse of the Fukushima Nuclear Power Plant that occurred on March 11, 2011.³⁷ A research protocol has been developed to investigate whether intervention using HT provides symptomatic relief of mild post-traumatic stress disorder (PTSD) caused by the disaster.

Strength and limitations

This review had several strengths: (1) the methods and implementation registered high on the PROSPERO database; (2) it was a comprehensive search strategy across multiple databases with no data restrictions; and (3) it involved detailed data extraction to allow for collecting all articles' content into a recommended structured abstract. The conduct and reporting of this review also aligned with the PRISMA statement³⁸ for transparent reporting of SRs and meta-analyses.

This review also had several limitations that should be acknowledged. Firstly, some selection criteria were common across studies, as described above; however, bias remained due to differences in eligibility for participation in each study. Secondly, publication bias was a limitation. Although there was no linguistic restriction in the eligibility criteria, we searched studies with only English and Japanese key words. In addition, this review reported on a relatively small and heterogeneous sample of studies. Moreover, we could not follow standard procedures for estimating the effects of moderating variables.

Conclusion

Although there was insufficient evidence in the studies of HT due to poor methodological and reporting quality and heterogeneity, HT may be an effective treatment for mental and behavioral disorders such as dementia, schizophrenia, depression, and terminal-care for cancer.

To most effectively assess the potential benefits of HT, it will be important for future research to utilize and describe (1) RCT methodology when appropriate, (2) adverse effects and withdrawals, (3) intervention dose, (4) intervention cost, and (5) development of the original check-list for HT.

Conflict of interest statement

None declared.

Author contribution

HO, MY, SO, SP, TH, SH, and HK conceived the study and take responsibility for the quality assessment and summary

of included studies and data extraction. KT, MY, and YM are the guarantors. HK, SO, HO, and SH designed the study. TH conducted the statistical analyses. SJP, TH, and HK assessed the quality of articles. All authors critically revised the manuscript for important intellectual content.

Funding

Supported by Grant-in-Aid for Scientific Research (C) number 23500817 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, 2012.

Acknowledgement

We would like to express our appreciation to Ms. Rie Higashino and Ms. Rinako Kai (paperwork) and Ms. Satoko Sayama (all searches of studies) for their assistance in this study.

Appendix.

References to studies excluded in this review

Exclusion no.	Author. Journal (year)	Title	Reason of exclusion
1	Verra ML, et al. Altern Ther Health Med (2012)	Horticultural therapy for patients with chronic musculoskeletal pain: results of a pilot study	Not randomized controlled trial
2	Annerstedt M, et al. Scand J Public Health (2011)	Nature-assisted therapy: systematic review of controlled and observational studies.	Not randomized controlled trial
3	Luk KY, et al. Int J Geriatr Psychiatry (2011)	The effect of horticultural activities on agitation in nursing home residents with dementia.	Not randomized controlled trial
4	Garland K, et al. Am J Geriatr Psychiatry (2007)	A comparison of two treatments of agitated behavior in nursing home residents with dementia: simulated family presence and preferred music	Not horticultural therapy
5	Verra ML, et al. Alternative Therapies Health Med (2012)	Horticultural therapy for patients with chronic musculoskeletal pain: results of a pilot study	Not randomized controlled trial
6	Rice JS, et al. J Offender Rehabil (1998)	Impact of horticultural therapy on psychosocial functioning among urban jail inmates	Not curative effect
7	Lee YA, et al. Korean J Horticultural Sci Technol (2007)	Effect of horticultural therapy using pressed flower based upon logotherapy on the improvement of the purpose in life and ego identity of middle-aged women	Not randomized controlled trial

References

- Ohtsuka Y, Yabunaka N, Takayama S. Shinrin-yoku (forest-air bathing and walking) effectively decreases blood glucose levels in diabetic patients. *Int J Biometeorol* 1998;41:125–7.
- Frumkim K. Beyond toxicity human health and the natural environment. *Am J Prev Med* 2001;20:234–40.
- Ulrich RS. Natural versus urban scenes: some psycho-physiological effects. *Environ Behav* 1981;13:523–56.
- Li Q. Effect of forest bathing trips on human immune function. *Environ Health Prev Med* 2010;15:18–26.
- Söderback I, Söderström M, Schäländer E. Horticultural therapy: the 'healing garden' and gardening in rehabilitation measures at Danderyd Hospital Rehabilitation Clinic, Sweden. *Pediatr Rehabil* 2005;8:235–7.
- Verra ML, Angst F, Beck T, Lehmann S, Brioschi R, Schneiter R, et al. Horticultural therapy for patients with chronic musculoskeletal pain: results of a pilot study. *Altern Ther Health Med* 2012;18:44–50.
- Annerstedt M, Wahrborg P. Nature-assisted therapy: systematic review of controlled and observational studies. *Scand J Public Health* 2011;39:371–88, <http://dx.doi.org/10.1177/1403494810396400>.
- Roberts S, Bradley AJ. Horticultural therapy for schizophrenia. *Cochrane Database Syst Rev* 2011, <http://dx.doi.org/10.1002/14651858.CD009413>.
- American Horticultural Therapy Association. <http://ahta.org/sites/default/files/DefinitionsandPositions.pdf> [accessed 22.08.13].
- Gonzalez MT, Hartig T, Patil GC, Martinsen EW, Kirkevold M. Therapeutic horticulture in clinical depression: a prospective study. *Res Theory Nurs Pract* 2009;23:312–28.
- van Tulder M, Furlan A, Bombardier C, Bouter L, The Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine* 2003;28:1290–9.
- Hopwell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomized trials in journal and conference abstracts. *Lancet* 2008;371:281–3.
- Hopewell S, Ravaud P, Baron G, Boutron I. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ* 2012;344:e4178.
- Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490–7.
- International Prospective Register of Systematic Reviews; August 9, 2013. http://www.crd.york.ac.uk/NIHR_PROSPERO/display_record.asp?ID=CRD42013005340 [accessed 28.08.13].
- Booth A, Clarke M, Ghersi D, Peticrew M, Stewart L. An international registry of systematic review protocols. *Lancet* 2011;377:108–9.
- Jarrott SE, Gigliotti CM. Comparing responses to horticultural-based and traditional activities in dementia care programs. *Am J Alzheimers Dis Other Dement* 2010;25:657–65, <http://dx.doi.org/10.1177/1533317510385810>.
- Kam MCY, Siu AMH. Evaluation of a horticultural activity programme for persons with psychiatric illness. *Hong Kong J Occup Therapy* 2010;20:80–6.
- Tse MM. Therapeutic effects of an indoor gardening programme for older people living in nursing homes. *J Clin Nurs* 2010;19:949–58.
- Kim SY, Son KC, Jung HJ, Yoo JH, Kim BS, Park SW. Effect of horticultural therapy on functional rehabilitation in hemiplegic patients after stroke. *Korean J Hortic Sci Technol* 2003;44:780–5 [in Korean with English abstract].
- Dijkstra K, Pieterse ME, Pruyn A. Stress-reducing effects in the built healthcare environment: the mediating role of perceived attractiveness. *Prev Med* 2008;47:279–83.
- Wichrowski M, Whiteson J, Haas F, Mola A, Rey MJ. Effects of horticultural therapy on mood and heart rate in patients participating in an inpatient cardiopulmonary rehabilitation programme. *J Cardiopulm Rehabil* 2005;25:270–4.
- Burgio LD, Allen-Burge R, Roth DL, Bourgeois MS, Dijkstra K, Gerstle J, et al. Come talk with me: improving communication between nursing assistants and nursing home residents during care routines. *Gerontologist* 2001;41:449–60.
- Jarrott SE, Gigliotti CM. Layers of influence: important contextual factors in directing dementia care programs. *J Appl Gerontol* 2011;30:113–22.
- Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration; 2011. <http://www.cochrane.org/training/cochrane-handbook> [accessed 05.06.13].
- Moher D, Hopwell S, Schulz KF, Montori V, Gotzsche PC, Dev-ereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
- Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT group. Methods and processes of the CONSORT group: example of an extension for trials assessing nonpharmacologic treatments. *Ann Intern Med* 2008;148:W60–6.
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Clin Epidemiol* 2009;62:499–505.
- Clatworthy J, Hinds J, Camic PM. Gardening as a mental health intervention: a review. *Mental Health Rev J* 2013;18:214–25.
- Bowen DJ, Kreuter M, Spring B, Cofta-Werpel L, Linnan L, Weiner D, et al. How we design feasibility studies. *Am J Prev Med* 2009;36:452–7.
- Blair CK, Madan-Swain A, Locher JL, Desmond RA, Loss Santos JD, Affuso O, et al. Harvest for health gardening intervention feasibility study in cancer survivors. *Acta Oncol* 2013;52:1110–8.
- Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C, et al. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med* 2006;144:364–7.
- MacPherson H, White A, Cummings M, Jobst K, Rose K, Niemtzow R. Standards for reporting interventions in controlled trials of acupuncture: the STRICTA recommendations. *Complement Ther Med* 2001;9:246–9.
- Wu T, Shang H, Bian Z, Zhang J, Li T, Li Y, et al. Recommendations for reporting adverse drug reactions and adverse events of traditional Chinese medicine. *J Evid-Based Med* 2010;3:11–7.
- Kamioka H, Kawamura Y, Tsutani K, Maeda M, Hayasaka Y, Okuizumi H, et al. A checklist to assess the quality of reports on spa therapy and balneotherapy trials was developed using the Delphi consensus method: the SPAC checklist. *Complement Ther Med* 2013;21:324–32.
- Wang D, Macmillan T. The benefits of gardening for older adults: a systematic review of the literature. *Act Adapt Aging* 2013;37:153–81.
- Kotzaki Y. *Intervention study of psychotherapy on elderly women who live in disaster area*. University hospital Medical Information Network-Clinical Trial Registration (UMIN-CTR 00008936). <https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr.cgi?function=brows&action=brows&type=summary&recptno=R000010226&language=J> [accessed 05.09.13].
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009;151:W65–94.

診療ガイドラインとシステマティック・レビュー

Clinical Practice Guidelines and Systematic Review

唐 文涛¹⁾ 小島原典子²⁾
河合富士美³⁾ 津谷喜一郎¹⁾

ABSTRACT

The definition of clinical practice guidelines (CPGs) by Institute of Medicine (IOM, USA) was revised in 2011, and the term of “systematic review” (SR) was incorporated into this definition as the basis for CPGs development for the first time. In this review, we investigated the definitions and number of published SRs in the world. We also analyzed database records and the evidence used for answering clinical questions in Japanese CPGs that are selected by Minds (Medical Information Network Distribution Service), which works mainly as Japanese CPGs clearinghouse. We found 1) differences among the definitions or descriptions of SR published by different organizations (IOM, The Cochrane Library, PubMed) that may lead to differences in standards of CPGs development; 2) publication by 31 Dec 2013 of approximately 5,500 SRs in the CDSR (Cochrane Database of Systematic Reviews) and 13,000 SRs in DARE (Database of Abstracts of Reviews of Effects), respectively; 3) comparing the CPGs selected by Minds before and after 2011, an increase in the proportion of CPGs without any record of database searches from 21.5% to 29.1%; 4) use of not only SRs, but also observational studies and case studies to answer clinical questions. We recommend that CPG development groups consider the number, quality, and freshness of existing SRs as well as the reviews of SRs for developing high-quality CPGs in the future.

(Jpn Pharmacol Ther 2014 ; 42 : 189-97)

KEY WORDS Clinical practice guidelines, Systematic review, Minds, Clinical question, Review of systematic reviews

はじめに

1990年代より世界に登場した「エビデンスに基づく医療」(evidence-based medicine: EBM)とシステ

マティック・レビュー(systematic review: SR)の発展に伴い、診療ガイドライン(clinical practice guidelines: CPGs)の概念とその作成はさまざまな臨床領域に普及してきた。

¹⁾ 東京大学大学院薬学系研究科 医薬政策学
教育・研究センター医学図書館

²⁾ 東京女子医科大学 衛生学公衆衛生学第二講座

³⁾ 聖路加国際メディカルセンター

Wentao Tang and Kiichiro Tsutani : Department of Drug Policy & Management, Graduate School of Pharmaceutical Sciences, The University of Tokyo ; Noriko Kojimahara : Department of Public Health, Tokyo Women's Medical University ; Fujimi Kawai : Medical Library, Center for Education and Research, St. Luke's International Medical Center

CPGs の定義としては、米国医学研究所 (Institute of Medicine: IOM) の 1990 年のもの¹⁾が世界的に使われてきたが 2011 年に改訂された²⁾。表 1 に示す。

従来から「システマティックに作成された文書」という文言は入っていたが、今回の改定で「エビデンスのシステマティック・レビューに基づき、……」とより明示的となった。

この新しい定義によって、SR の CPGs 作成における重要性は、日本を含めて世界中で徐々に認識されていくであろう。それを契機として、日本の CPGs 関連の事業のなかで、これまでの CPGs 作成状況を見直し、質の良い SR を CPGs の作成により活用できるようにする時期が来ていると考えられる。

そこで本稿では、第 1 に SR の定義、第 2 に SR の数、第 3 に日本の CPGs 作成における検索対象データベース、第 4 に日本の CPGs 作成における SR の使用の現状、第 5 に CPGs 作成にあたって SR を「つかう」か「つくる」か、また CPGs における使用したデータベースの記載法について述べる。

I システマティック・レビューの定義

SR の統一された世界的な定義はない。広く解釈を含むと医学領域では、以下の三つが代表的である。

1 コクラン共同計画 (The Cochrane Collaboration: CC) による 1992 年の定義³⁾

システマティック・レビューは、ある特定のリサーチクエスションに答えるために、すべての経験的 (empirical) エビデンスをあらかじめ定めた基準で網羅的に収集し、評価し、統合する方法である (A systematic review attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a given research question.)。

2 IOM による 2011 年の定義⁴⁾

システマティック・レビューは、類似するが別々の研究の知見を収集し、選択し、評価し、統合する方法である。医薬品、医療機器、そのほかのヘルスサービスにおいて考えられる益と害に関する既知と未知の事項を明らかにするのに役立つことができる (Systematic reviews identify, select, assess, and synthesize the findings of similar but separate studies.

表 1 IOM による CPGs の定義

1990 特定の臨床状況のもとで、臨床家と患者の意思決定を支援する目的で、システマティックに作成された文書。
(Clinical practice guidelines are systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.¹⁾)

2011 エビデンスのシステマティック・レビューに基づき、患者ケアの最適化を目的とする推奨を含む文書。
(Clinical practice guidelines are statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options.²⁾)

They can help clarify what is known and not known about the potential benefits and harms of drugs, devices, and other healthcare services.)。

3 PubMed の MeSH にはシステマティック・レビューの annotation (解説) は存在しないが、PubMed の Clinical Queries には SR の検索方法について subset (sb) としての記述がある⁵⁾

この検索戦略は、システマティック・レビュー、メタアナリシス、臨床試験のレビュー、エビデンスに基づいた医療、コンセンサス会議、ガイドライン、臨床医にとって価値のある研究を専門的にレビューする雑誌からの文献を抽出することを意図するものである。このサブセットは systematic [sb] の検索によって抽出できる (This strategy is intended to retrieve citations identified as systematic reviews, meta-analyses, reviews of clinical trials, evidence-based medicine, consensus development conferences, guidelines, and citations to articles from journals specializing in review studies of value to clinicians. This subset can be used in a search as systematic [sb].)。

上記の三つの SR の定義や記述では、若干の違いがある。文献の選択基準は、CC では「あらかじめ定めた基準で網羅的に収集する」とより明確に述べられている。IOM は「類似するが別々の研究」(findings of similar but separate studies) のみとなっている。PubMed では SR を検索するための検索対象の文献種類のみを述べているが、SR を行う際の文献の選択基準は記載されていない。すなわち、それぞれの研究組織とデータベースによって SR の定義などの

不一致は存在する。これが実際の CPGs 作成時に異なる SR の選択基準をもたらす可能性がある。

共通点として、CC と IOM には SR の定義のなかで “synthesize” (統合) の用語が使われている。「統合」プロセスを定量的 (統計学的) に行ったものは「定量的 SR」(quantitative systematic review, 量的 SR と称する) と称され、それに対して、後述するようにメタアナリシス (meta-analysis: MA) という名称が用いられる⁶⁾。すなわち、MA は SR の一部として理解される。

一方、そのプロセスを定行的に行ったものを「定性的 SR」(qualitative systematic review, 質的 SR と称する) と称する。そこでは、アウトカムとして構造化抄録集やエビデンス・テーブルを提示し、エビデンスのプロファイルを明示することが重要である。それをもとに判断 (judgement) し一定の結論をつける。双方ともに、明確なリサーチクエスションの設定, systematic search, critical review は同じである。以後本稿ではこれらの用語を用いる。

なお、SR は医学以外の分野でも使われる用語である。Chalmers ら⁷⁾による SR と MA の定義は以下のとおりである。

システマティック・レビューは、あるテーマに関するすべての研究の収集、批判的吟味、統合におけるバイアスを制限するための策略の応用。メタアナリシスは、そのプロセスのなかの一部にもなり得るが、必ずしも用いられるとは限らない (Systematic Review: the application of strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic. Meta-analysis may be, but is not necessarily, used as part of this process.)。

メタアナリシスは、類似する別々の研究、すなわち比較できる研究の結果のまとめを定量的に導出するために、それらの研究データを統計的手法で統合するものである (meta-analysis: the statistical synthesis of the data from separate but similar, i. e. comparable studies, leading to a quantitative summary of the pooled results.)。

上記の SR と MA の定義は、Carole Torgerson の教育分野における “Systematic Reviews” と題した本でも引用されている⁸⁾。

表 2 各種データベースにおける SR の数

データ ベース	検索用語 (検索領域)	検索数
CDSR	(—)	5,549*
DARE	(—)	13,032
PubMed	¹⁾ “systematic review” (All Fields)	44,589
	²⁾ meta-analysis (Publication Type)	42,751
	³⁾ Systematic [Subset]	218,955
	(¹⁾ OR ²⁾ AND ³⁾	74,631

*256 件の取り下げられた (withdrawn) SR を除いた数
(検索日: 2013.12.31)

II 各種データベースにおけるシステマティック・レビューの数

各種データベースにおける SR の数を表 2 に示す。

SR というと、The Cochrane Library に収載される Cochrane Database of Systematic Reviews (CDSR, 各レビューは Cochrane Reviews と略される) が想起されることが多い。約 5500 件である。日本ではあまり知られていないが、Database of Review of Effectiveness (DARE) には約 13,000 件と、その 2 倍以上の SR が含まれる。この DARE (<http://www.crd.york.ac.uk/CRDWeb/>) は、英国のヨーク大学の NHS-CDS によって作成・運営されているデータベースの一つであり、Cochrane Review 以外の質の高い SR を systematic に探して第三者のコメント付きの構造化抄録が作成されている。

PubMed で単に “systematic review” と検索した場合、その検索結果 44,589 件には SR の方法論などの文献も入っている。また、SR の書誌情報には “systematic review” と書かれていないものもあり、検索結果に入らない可能性がある。検索の感度、特異度双方ともに問題がある。

一方、PubMed の MeSH には meta-analysis [Publication Type] の annotation (解説) がある (その annotation は以下で検索できる: All Mesh Categories⇒Publication Type Category⇒Study Characteristics⇒meta-analysis)。その annotation では、以下の記述がある。

定量的な手法で独立した研究の結果を組み合わせ、治療の有効性の評価、新たな研究の計画等のために、統合した要約と結論を出すすべての文献を含

表 3 Minds ガイドラインセンターの CPGs 評価選定方法の変遷

2004 年選定方法 ¹⁾ (AGREE (2001 年) に基づき、九つの基本的考え方から診療 ガイドライン選定委員会で選定)	2011 年選定方法 (AGREE II (2009 年) の六つの観点と全体評価から診療 ガイドライン選定委員会で選定)
- わが国における Disease Burden (発症率, 罹病率, 致命率, 死亡者数) が大きい疾病・病態に関するものであること。 - テーマに関わる多分野から作成委員が選出されていること。 - 作成時期(および改定の予定時期)が明示されていること。 - 対象となるテーマ(健康問題, 医療技術など)が明確に定義されているか。 - エビデンスの検索方法が明示されていること。 - エビデンスを抽出した文献を列挙していること。 - エビデンスがレベルづけされていること。 - 推奨がグレードづけされていること。 - 適用にあたって患者の意向に配慮するようにと述べられている。	網羅的に検索収集した文献から、2 段階のスクリーニングを経て、診療ガイドラインの作成方法に焦点を当てた評価ツールである。 AGREE II (Appraisal of Guidelines for Research & Evaluation II, http://www.agreetrust.org/) の評価対象となるガイドラインを絞り込む。 ガイドライン選定ワーキンググループ 4 名以上の結果をもとに、診療ガイドライン選定委員会で選定。 AGREE II の評価項目は、以下の六つの観点 - 対象と目的 - 利害関係者の参加 - 作成の厳密さ - 提示の明確さ - 適用可能性 - 編集の独立性と全体評価

¹⁾原表記は「診療ガイドラインの評価選定の基本的考え方」

めるものである (Works consisting of studies using a quantitative method of combining the results of independent studies (usually drawn from the published literature) and synthesizing summaries and conclusions which may be used to evaluate therapeutic effectiveness, plan new studies, etc.)。

前述したように PubMed には SR の subset があり、それはかなり広義の SR が収録されているため、検索結果は約 22 万件である。Publication Type で meta-analysis での検索結果と、All Fields で “systematic review” の phrase での検索結果を subset の systematic と掛け合わせた検索結果は、約 75,000 件であり、この数は狭義の SR の件数に近いものではないかと推察されるが、今後さらなる研究が必要である。

III 日本の診療ガイドライン作成と 検索対象データベース

日本では、厚生省 (2001 年 1 月から厚生労働省へ再編) が 1999 年度から、根拠に基づく医療 (Evidence-based medicine: EBM) の手法を用いた CPGs 作成支援を開始した⁹⁾。それ以降、多くの専門学会

や研究会が、それぞれの作成方法と手順をふまえて CPGs を作成している。「診療ガイドライン」という名称がつけられ、出版などがなされたものは 2003 年に 600 種類以上と報告された¹⁰⁾。この数は CPGs の質には配慮しない数であり、エビデンス (evidence-based) ではなく専門家の意見に基づく (consensus-based) もので、かつそのことが明示されていないものも含まれていた。

公益財団法人・日本医療機能評価機構 Minds ガイドラインセンター (<http://minds.jcqhc.or.jp/n/>、Minds は Medical Information Network Distribution Service の略称) では、厚生労働省からの補助金を受け、2004 年から評価選定された質の高い CPGs をその website 上に公開している。2013 年 7 月末で 107 件の CPGs が掲載されている。評価選定の方法は表 3 のように 2011 年に変更され、徐々に新しい選択方法に基づくものに置き換えられつつある。

2010 年までに Minds ガイドラインセンターの 2004 年選定方法によって選定され、かつ 2013 年 7 月末の時点で Minds の website に本文が掲載されている CPGs は 65 件であった。この 65 件の CPGs の作成プロセスの文献検索データベースの記載はす

に分析されている¹¹⁾。また、2011 年以降、2011 年選定方法で選択された CPGs は 2013 年 7 月末時点で 116 件ある。そのうち 55 件の本文が Minds の website で公開されている。この 55 件を対象に CPGs 作成時の文献データベースの記載について調査し、2010 年までと 2011 年以降の文献データベースの記載状況を比較し、表 4 にまとめた。

2010 年までの 65 件の CPGs で、Medline/PubMed を検索したと記載があったのは 51 件 (78.5%)、医中誌は 41 件 (63.1%) であった。The Cochrane Library は 20 件 (30.8%) で検索されていた。The Cochrane Library には CDSR, DARE, CENTRAL, その他四つ、全七つのサブセットのデータベースが含まれている。しかし、今回対象となった CPGs で、The Cochrane Library のなかのどのサブセットを検索したかが記載されているものは 0 件であった。また、検索したデータベースについてまったく記載がない CPGs も 14 件 (21.5%) あった。

2011 年以降では AGREE II を用いた CPG の評価選定が行われている。それにもかかわらず、データベースの記載のない CPGs の割合が増えている (21.5%⇒29.1%)。用いられているデータベースは、やはり 2010 年と同じ三つのデータベース (Medline, 医中誌, The Cochrane Library) が大半であった。The Cochrane Library を検索した 15 件の CPGs のなかで、The Cochrane Library のどれが検索されたかを明確に記載したものは 3 件のみであった (CDSR: 1 件, CENTRAL: 1 件, CDSR+CENTRAL: 1 件)。

2014 年 3 月公表予定の「Minds 診療ガイドライン作成の手引き 2014」においては、国際標準に従い、網羅的な文献検索の手順について詳細な記載をするよう強調されている。質の高い CPGs を作成するためには、その文献検索のプロセスにおいて、すべての使用したデータベース名やそのサブセット名、その検索式、検索期間、検索実施日を明確に記載することを必須要素とすべきである。

IV 日本の診療ガイドライン作成における システマティック・レビューの使用の現状

日本の CPGs の作成にあたり、エビデンスの検索対象となるデータベースとして、PubMed と医中誌

表 4 CPGs 作成時に検索したデータベース
(2010 年までと 2011 年以降の比較)

データベース 検索記載有の割合	2010 年まで (65 CPGs)	2011 年以降 (55 CPGs)
Medline/PubMed	51 (78.5%)	39 (70.9%)
医中誌	41 (63.1%)	30 (54.5%)
The Cochrane Library	20 (30.8%)	15 (27.3%)
検索したデータベース についての記載なし	14 (21.5%)	16 (29.1%)

がもっともよく使われてきた。ただし、PubMed であれ医中誌であれ、meta-analysis という publication type に絞り込み、検索が行われた CPGs の割合は不明である。

そこで、実際、CPGs においてクリニカル・クエスション (CQ) から推奨を導くのにどのような研究デザインが用いられているのかを分析した。2013 年 7 月末現在、Minds で本文が参照でき、2010 年の選定方法で選択された 55 CPGs のうち、改訂版、かつ文献検索が CQ ごとに行われ、さらに The Cochrane Library を検索したと記載のある CPGs として「子宮頸癌治療ガイドライン 2011」と「頸椎後縦靱帯骨化症診療ガイドライン 2011 (改訂第 2 版)」を質の高い CPGs の代表として、採用された論文のエビデンスレベルを分析した。

第 1 の「子宮頸癌治療ガイドライン 2011」では、エビデンスの質評価基準は、研究デザインによって表 5 のように分類された。35 の CQs に対する 66 の推奨の根拠としてのべ 512 論文が採用され、そのうち“エビデンスレベル I 複数のランダム化試験 (randomized controlled trials: RCT) のメタナリシス”は 20 論文 (4%)、さらにうち 5 件は Cochrane review であった。これ以外にも、先行する 32 件の CPGs, 5 件の規約が採用されていた。CQ ごとに推奨決定に採用された論文のエビデンスの質は、八つの CQs (23%) で“エビデンスレベル I 複数のランダム化比較試験のメタアナリシス”が採用されており、これらすべての CQs で RCT も同時に採用されていた。質の高いエビデンスがなく、III, IV に属する論文のみで推奨が決定されていた CQs も 12 (34%) あった。このように Cochrane Reviews など、質の高いエビデンスがあっても観察研究なども推奨決定に採用していることが明らかとなった。また、SR の

表 5 「子宮頸癌治療ガイドライン 2011」のエビデンスの質評価基準（レベル）とのべ採用論文数と推奨決定

エビデンス レベル	エビデンスの質の評価基準	採用論文数 (n=512)	各 CQ における推奨決定に採用された 論文のエビデンスの質 (n=35)
I	複数のランダム化比較試験のメタアナリシス	20	8 (23%)
	うち Cochrane Reviews	(5)	(3 (9%))
II	ランダム化比較試験, またはよくデザインされた非ランダム化比較試験	106	15 (43%)
III	よくデザインされた準実験的研究, または比較研究, 相関研究, 症例比較研究など, よくデザインされた非実験的研究	326	12 (34%)
IV	専門委員会の報告や意見, または権威者の臨床経験	61	

表 6 「頸椎後縦靱帯骨化症診療ガイドライン 2011 (改訂第 2 版)」のエビデンスの質評価基準（レベル）とのべ採用論文数と推奨決定

エビデンス レベル	エビデンスの質の評価基準	採用論文数 (n=626)	各 CQ における推奨決定に採用された 論文のエビデンスの質 (n=75)
I	全体で 100 例以上の RCT の MA または SR	0	0 (0%)
	うち Cochrane review	(0)	(0 (0%))
II	全体で 100 例以上の RCT	1	1 (1.3%)
III	全体で 100 例未満の RCT の MA または SR	1	1 (1.3%)
	うち Cochrane review	(0)	(0 (0%))
IV	全体で 100 例未満の RCT	2	0 (0%)
V	CCT および cohort study	37	24 (32.0%)
VI	Case control study	83	21 (28.0%)
VII	Case series	464	26 (34.7%)
VIII	Case report	15	1 (1.3%)
IX	その他	23	1 (1.3%)

み, RCT のみで作成されている CQ はなかった。

第 2 の, 「頸椎後縦靱帯骨化症診療ガイドライン 2011 (改訂第 2 版)」では, エビデンスの質評価基準 (レベル) は, 研究デザインによって表 6 のように分類され, 75 の CQs に対する 90 の推奨の根拠として採用されている論文の研究デザイン別のべ論文数を示した。75 の CQs でのべ 626 論文が採用され, そのうち “エビデンスレベル I 全体で 100 例以上の RCT の MA または SR” はなかった。“エビデンスレベル III 全体で 100 例未満の RCT の MA または SR” は 1 件であったが, Cochrane Review ではなかった。CQ ごとに推奨決定に採用された論文

のエビデンスの質は, “エビデンスレベル II 全体で 100 例以上の RCT” と “エビデンスレベル III 全体で 100 例未満の RCT の MA または SR” が採用された CQ は各 1 件 (1.3%) であった。それらの CQs の推奨決定には cohort study, case-control study, case series も同時に採用された。RCT の MA または SR のみで推奨が決定された CQ はなかった。それ以外の 73 の CQs の推奨は, 質の高いエビデンスがなく, 観察研究や case series で決定されたものであった。

V 診療ガイドライン作成にあたってシステマティック・レビューを「つかう」か「つくる」か CPGs 作成にあたっての SR のあり方には三つの種類がある

第1は「テーラーメイド SR」(tailor-made systematic review)である。作成しようとする CPGs に使える SR が CDSR や DARE に存在しない場合は、SR を新たに「つくる」べきである。実際に「つくる」かどうかの判断にあたっては、CPGs 中の CQ の重要性の軽重、臨床試験登録システムで登録されている臨床試験の進行と結果公表の状況、CPGs 作成に用いることのできる経済的・人材的資源の多寡などを考慮する。その際に、RCT に対する CONSORT 声明と同様に、27 項目からなる SR を作る際の PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)¹²⁾のチェックリストを用いて SR を実施し報告することにより、SR の質向上に貢献でき、それは CPGs の質向上にもつながる。さらに、臨床試験登録システムと同様に、SR の publication bias を防ぐための登録システム PROSPERO (<http://www.crd.york.ac.uk/Prosperto/>) が 2011 年 2 月に設立された。また、2012 年 2 月から日本でも UMIN-CTR (<http://www.umin.ac.jp/ctr/index-j.htm>) で SR の登録ができるようになった。SR の protocol を PROSPERO や UMIN-CTR に登録することが望まれる。

第2は「既存の SR」(ready-made systematic review)である。すでに存在する SR を CPGs 作成に「つかう」ものである。この場合には、既存の SR と CQ の関連を PICO (Patient, Intervention, Comparison, Outcomes) の形式で吟味し、またその質を評価し、さらに賞味期限以内のものを「つかう」ことが重要である。

SR の質を評価するために、11 項目からなる AMSTAR (assessment of multiple systematic reviews) のチェックリストが 2007 年に開発された¹³⁾。また、SR の質を定量的に評価するために、AMSTAR の修正版である R-AMSTAR (revised assessment of multiple systematic reviews) が 2010 年に開発された¹⁴⁾。AMSTAR または R-AMSTAR で SR の質を評価することが望ましい。

既存の SR が CQ と関連性がない場合、また質が

好ましくない場合、自ら SR を新たに「つくる」ことがより合理的であろう。

また、作成しようとする CPGs にとって SR の賞味期限が切れている場合がある。具体的な何年前の SR であればそのまま採用できるかの判断は疾患領域により異なるが、たとえば SR の最終文献検索から 3 年を過ぎた場合、自ら同じ検索式で、検索時点までに公表された新たな 1 次研究の RCT などについて検索することが推奨される。追加検索で得られた論文については選択基準に従って採用か不採用かを判断し、採用となる研究が CQ と関連しているか、またその質を評価しエビデンス・テーブルを作成し、その結果を更新された SR に反映させるべきである。

第3は「SR のレビュー」(review of systematic reviews)である。CPGs のある特定の CQ に対して複数の SR が存在する場合、SR のレビューを検索し、それを「つかう」ことができる。SR のレビューは、「別々の SR の知見を比較・対照し、臨床の意思決定者に必要とされるエビデンスを提供するための合理的方法である」(Systematic reviews (or overviews) of reviews are a logical and appropriate next step, allowing the findings of separate reviews to be compared and contrasted, providing clinical decision makers with the evidence they need.)¹⁵⁾。すでにいくつかの SR のレビューが公表されたが^{16,17)}、その全体の現存数についてさらなる調査が必要である。

複数の SR が存在するが、既存の SR のレビューが存在しない場合は、自ら SR のレビューを「つくる」こともできる。そのとき、各種のデータベースだけではなく、SR の登録システム PROSPERO も検索することは、publication bias の抑制、また最新の一次研究の知見を組み入れた新たな SR の進行・公表状況の把握に役立つであろう。

複数の SR は、MA の定量的な方法を用いて統合することができる。個々の一次研究間の異質性(heterogeneity)や effect size の報告不足により定量的な統合が不可能な場合には、定性的な統合を考慮する。また、CPGs 中の CQ の重要性の軽重、PROSPERO などの登録システムで登録されている SR の進行と結果公表の状況、CPGs 作成に用いることのできる経済的・人材的資源の多寡などを考慮したう

えて、一つないし複数の SR を選択し、それらをもとに判断するような定性的な統合もできる。

2011 年に改訂された IOM の CPGs の定義で SR が明示的に記され、その重要性が強調された現在、SR を検索したデータベース名や検索式を CPGs 中に具体的に明記すべきである。なお、2013 年度中に、CDSR の abstract に Minds で作成された日本語訳が掲載される予定である。これが日本でさらに広く使われることが望ましい。そのために PubMed から CDSR へのリンク機能について日本において強い広報活動が望まれる。さらに CPGs 作成にあたって、PubMed から CDSR を検索した場合、それを CPGs のなかに明記すべきである。

おわりに

CPGs の社会との関わりは一般に認識されている以上に強く、多様な側面を持ち、各側面の研究が望まれている。第 1 にカバーする医療費の推計¹⁸⁾、第 2 に医師の professional autonomy との関係¹⁹⁾、第 3 に法的問題との関係はすでに論じられてきた²⁰⁾。本研究は CPG 作成における SR 利用の技術的な側面を論じたものである。

SR の定義は各研究組織やデータベースによって異なるところがあり、それが実際の CPGs 作成時に異なる作成基準をもたらす可能性がある。SR の現存数は、CDSR に約 5500 件、DARE に約 13,000 件存在する。CPGs における SR の利用状況は、2010 年までと 2011 年以降の比較で、CPGs における文献検索のデータベースの記載が依然として不十分であることが明らかとなった。CPGs 作成にあたって既存の SR を用いる際、データベース名やサブセット名、検索式を CPGs 中により具体的に明記すべきである。

また実際の CPGs 中での CQs への回答は、「子宮頸癌治療ガイドライン 2011」で 23%、「頸椎後縦靱帯骨化症診療ガイドライン 2011 (改訂第 2 版)」で 1.3% が、SR によってなされていた。このことは、これらの CQ に対する SR がもともと少ない可能性と同時に、CPGs を作成する際に SR の検索が不十分である可能性も示唆する。

また、CPGs の作成にあたって SR を「つくる」か

「つかう」かを判断する際に、既存 SR の有無、その質また賞味期限、さらに「SR のレビュー」の公表状況を総合的に考えることが必要である。SR を「つかう」際の質の吟味も必要であり、SR を「つくる」際にも質の高いものを目指すべきである。さらに、SR のレビューの現状に対する調査と、今後の動向に対する注目が必要である。

【謝辞】 本稿を作成するにあたり協力を得た、東京女子医科大学衛生学公衆衛生学第二講座/日本医療機能評価機構・山口直人、日本医療機能評価機構 EBM 医療情報部・奥村晃子の両氏に謝意を表する。

本研究は平成 25 年度 (2013) 厚生労働科学研究補助金 (地域医療基盤開発推進研究事業) 「システムティックレビューを活用した診療ガイドラインの作成と臨床現場における EBM 普及促進に向けた基盤整備」 (主任研究者: 中山健夫) の一環として実施されたものである。

文 献

- 1) Institute of Medicine. Clinical Practice Guidelines: Direction for a New Program. Washington DC : National Academy Press ; 1990.
- 2) Institute of Medicine. Clinical practice guidelines we can trust. Washington DC : National Academy Press ; 2011.
- 3) About Cochrane Systematic Review And Protocols. <http://www.thecochranelibrary.com/view/0/AboutCochrane-SystematicReviews.html> (accessed 12 Dec 2013)
- 4) Finding What Works in Health Care : Standards for Systematic Reviews. <http://www.iom.edu/Reports/2011/Finding-What-Works-in-Health-Care-Standards-for-systematic-Reviews.aspx> (accessed 15 Dec 2013)
- 5) Search strategy used to creat systematic reviews subset on PubMed. http://www.nlm.nih.gov/bsd/pubmed_subsets/sysreviews_strategy.html (accessed 12 Dec 2013)
- 6) Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-analysis. *J Clin Epidemiol* 1995 ; 48 (1) : 167-71.
- 7) Chalmers I, Hedges LV, Cooper H. A Brief History of Research Synthesis. *Eval Health Prof* 2002 ; 25 : 12-37.
- 8) Torgerson C. (eds). Systematic review. Bodmin, Cornwall : Continuum. 2003.
- 9) 中山健夫. 診療ガイドライン: 現状と今後の展望. 通風と核酸代謝 2009 ; 33 (2) : 137-47
- 10) 吉田雅博. 医療情報サービス Minds 事業報告 (accessed 28 May 2013) [minds4.jcqhc.or.jp/seminar/091024/yoshida.pdf].

- 11) 津谷喜一郎. Minds 医療技術評価部会. 第 8 回 CPGs 作成グループ意見交換会 (配布資料, 2010.12.25)
- 12) The PRISMA Statement. <http://www.prisma-statement.org/statement.htm> (accessed 14 Oct. 2013)
- 13) Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR : a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007 ; 7 : 10.
- 14) Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, et al. From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care : Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. *Open Dent J* 2010 ; 4 : 84-91.
- 15) Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol* 2011 ; 11 (1) : 15.
- 16) Smith V, Devane D, Begley CM, Clarke M, Higgins S. A systematic review and quality assessment of systematic reviews of randomised trials of interventions for preventing and treating preterm birth. *Eur J Obstet Gynecol Reprod Biol* 2009 ; 142 : 3-11.
- 17) Singh JA, Christensen R, Wells GA, Suarez-Almazor ME, Buchbinder R, Lopez-Olivo MA, et al. Biologics for rheumatoid arthritis : an overview of Cochrane reviews. *Sao Paulo Med J* 2010 ; 128 (5) : 309-10.
- 18) 津谷喜一郎, 菊田健太郎. 厚生労働省科学研究費補助事業による 20 の診療ガイドラインがカバーする医療費の推計. *臨薬理* 2004 ; 35 (1) : 140S
- 19) 津谷喜一郎, 長澤道行. 医師と診療ガイドライン : “professional autonomy” の視点から. *日医師会誌* 2003 ; 129 : 1793-803.
- 20) 長澤道行, 中山健夫, 津谷喜一郎. 診療ガイドラインの新たな法的課題. *医事新報* 2010 ; 4504 : 54-64.

* * *

Series

EQUATOR Network から得られる、
質の高い研究報告のための国際ルール —— ③

STROBE：疫学における観察研究報告のためのルール

上岡 洋晴^{*1} 中山 健夫^{*2} 津谷喜一郎^{*3}

STROBE 作成の背景と経緯

国際医学雑誌編集者委員会 (International Committee of Medical Journal Editors : ICMJE) が 1979 年に公表した「生物医学雑誌への統一投稿規定」は世界的に多くの雑誌で採択されている。数度の改訂を経て、現在の主たる目的は、正確で明瞭簡潔かつ容易にアクセス可能な論文を作成・配布するという、著者と編集者の共同作業を支援することである。査読者、出版社、メディア、患者とその家族、一般の読者など、その他の多くの関係者に対しても、生物医学研究の原稿執筆および編集過程における有用な情報を提供している^{1,2)}。

そうした生物医学研究の多くは、コホート研究 [cohort study : 以下, コホート研究], 症例対照研究 [case-control study : 以下, ケース・コントロール研究], および横断研究 [cross-sectional study : 以下, 横断研究] からなる観察研究 (observational study) である。観察研究は、疾病の頻度, 原因 (リスク因子), 予後の解明など, 介入の評価を目的とするランダム化比較試験 (randomized controlled trial : RCT) などの介入研究とは異なる問題・課題に対応するものである。また, RCT が限定された対象者で介入の有効性 (efficacy) を評価するのに対して, 観察研究はより real world を反映した効果 (effectiveness) の評価や, 害・副作用に関する検討でも役立てられている³⁾。

しかしながら, 観察研究の報告は, 往々にして不十分なものが少なくなく, 研究の強さと弱さおよび一般化可能性 (外的妥当性) の評価を妨げていることが指摘されてきた。先行研究となる出版済みの観察研究の報告には, 重要な情報がしばしば欠落しているか, または不透明である場合がある。報告の透明性が十分でなければ, さらなる研究によって判断せざるをえなくなるため, 研究による恩恵がもたらされるのが遅れる可能性も指摘されてきた。そうした背景において, 観察研究の報告に関する手引き書の必要性が求められていた。

そこで, 2004 年に行われた 2 日間にわたるワークショップで, 方法論学者, 研究者, および雑誌編集者が集い, チェックリスト項目の草案が作成され, さらにコーディネートグループによる数回の会議と, 大規模な貢献者グループによる E-mail を用いた討議を経て, 22 項目からなる STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) が 2007 年に開発された⁴⁾。さらに, 「STROBE の解説・詳細」の論文も, 同時に発表されている⁵⁾。

STROBE は, EQUATOR (Enhancing the Quality and Transparency of Health Research) Network の web サイトにも, CONSORT 2010 や PRISMA などの各チェックリストとともに掲載され, 世界中に広く知られるようになってきた⁶⁾。実際に STROBE は, 日本語⁷⁾だけでなく, 中国語, スペイン語, ド

^{*1}東京農業大学地域環境科学部
研究科医薬政策学講座

^{*2}京都大学大学院医学系研究科社会健康医学系専攻健康管理学講座

^{*3}東京大学大学院薬学系

イツ語, イタリア語, ペルシャ語, ポルトガル語, ギリシャ語で翻訳されている⁸⁾。「STROBE の解説と詳細」についても, スペイン語⁸⁾と日本語⁹⁾で翻訳されている。2011 年には, 分子遺伝学を含む観察研究のチェックリスト「STROBE-ME (Strengthening the Reporting of Observational Studies in Epidemiology-Molecular Epidemiology)」も開発されている¹⁰⁾。本論では, 前出の STROBE に限定して解説することとする。

観察研究の定義と取扱いの範囲

STROBE は, 観察研究の正確かつ完全なる報告に含まれるべき事項に関する推奨文(チェックリスト)であり, 取扱う範囲は, コホート研究, ケース・コントロール研究, および横断研究という 3 つの主要な試験デザインと定めている。

STROBE が推奨する意図は, 「良好な研究報告の仕方の説明」であることを強調し, 「研究をどのように実施すべきか」あるいは「報告の透明性を保つことは評価の前提条件となるものの, チェックリストが観察研究の質を評価するための手段となるわけではない」と述べている。また, 「STROBE の解説と詳細」と合わせて使用することを強く推奨している。本論では, これらをふまえたうえで, 研究の報告で記載されるべき 22 項目のチェックリストを概説する。

STROBE の構成要素とポイント

STROBE は, 観察研究のより良い報告のために重要と考えられる 22 項目からなるチェックリストである。これらの項目は, 「論文タイトル・抄録 (項目 1)」、「はじめに (項目 2, 3)」、「方法 (項目 4~12)」、「結果 (項目 13~17)」、「考察 (項目 18~21)」、「その他の情報 (項目 22: 研究の財源に関する項目)」に対応している。

18 の項目は 3 つの研究デザインすべてに共通している。一方, 4 つの項目 (項目 6, 12, 14, 15) は, 項目内容のすべてまたは一部が研究デザインごとに異なっている。また, これとは別に 4 つの項目 (*印項目 8, 13, 14, 15) に関して, ケース・コントロール研究ではケース (患者) 群とコントロール群, コホート研究と横断研究においては曝露群と非曝露群で, それぞれ別々に説明すべきであることを

意味している。元来の論文中のチェックリスト^{4,7)}では, これら 3 つの研究デザインをひとまとめにしてあり, やや使いづらい。前述の STROBE のウェブサイト⁸⁾では, これら 3 つの研究デザインごとのチェックリストを閲覧することができる。その日本語訳を表 1~3 に示した。

観察研究の 3 つのデザインのなかで, ケース・コントロール研究に関して, 臨床家のあいだで誤解が少なくない。コホート研究が曝露要因のデータを得て, 追跡によりアウトカム (イベントあり) のデータを把握して, 曝露要因とアウトカムの因果関係を検証する (横断研究も曝露要因からアウトカムの関係をみようとするが, データを得るのが同時点) のに対して, ケース・コントロール研究は, まずアウトカムの有無からサンプリングして, さかのぼって曝露要因の保有状況を調べる。すなわち「アウトカムから曝露を後ろ向きに (retrospective に) みる」ことで, 「曝露とアウトカム」の因果関係を推測するものである。

よくみられる誤解が 2 つある。そのひとつは, 形のうででケース群とコントロール群の同時点の測定結果を比較するだけで, 要因への曝露状況の違いを過去にさかのぼって比較していない研究を, ケースとコントロールという言葉に引かれてケース・コントロール研究としてしまう場合である。これらは, 実際には「横断研究」である場合が多い。

もうひとつは, 介入の有効性評価に際して, 介入群をケース群, 非介入群をコントロール群として, 両群のアウトカム (イベント) 頻度を比較する研究がケース・コントロール研究とよばれる場合である。これらは, 介入の有無という曝露要因のデータを収集し, 時系列に沿ってその結果を測定して, 両群で頻度を比較しているので, これもケースとコントロールという言葉が使われていても疫学的なケース・コントロール研究ではない。両群の決定 (要因の割り付け) が研究者の意図に基づいて行われていれば介入研究 (多くは非ランダム化比較試験) であり, 診療の場で臨床医の判断で行われていればコホート研究となる。

以上のように, ケース・コントロール研究, そしてケース, コントロールという言葉の扱いには注意が必要である。