Development of Complex Interventions in Stroke Care
A Systematic Review

Judith Redfern, MSc; Christopher McKeivitt, PhD; Charles D.A. Wolfe, MD, FFPHM, FRCOG

Background and Purpose—Stroke care is complex, requiring input from professionals, patients and carers. Identifying and developing appropriate intervention components to meet these complex needs is difficult. The Medical Research Council (MRC) Framework for developing and evaluating ‘complex’ (nonpharmacological) interventions aims to improve intervention development. This study uses the Framework to review complex interventions in stroke care.

Methods—Systematic review with multiple search strategies (electronic databases, recent journals, gray literature) was used. The MRC Framework was used to guide the search strategy and assess study quality. ‘Complex interventions’ were defined as educational/psychosocial interventions to change knowledge, beliefs or behaviors.

Results—Sixty-seven studies were included: 39 randomized controlled trials (RCT) and 28 other designs. Complex interventions targeted healthcare professionals (17), and patients, carers and the general population (21 targeting primary or secondary prevention; 30 targeting adjustment and recovery after stroke). Compared with recovery studies, primary and secondary prevention studies were significantly less likely to have been evaluated in RCTs. Interventions evaluated in RCTs were significantly less likely to influence primary outcomes (26%) compared with other designs (44%). Theoretical grounding to support intervention choice was reported in 40 studies but only 14 were theoretically ‘well developed’; 21 RCTs listed multiple primary outcome measures, with 10 listing 5 or more. Of these only 3 reported considering statistical power before recruitment and none was sufficiently powered.

Conclusion—Few complex interventions in stroke care have been adequately developed or evaluated. This may explain failures to demonstrate efficacy. In future, greater attention is needed to theoretical development and methodological quality. (Stroke. 2006;37:2410-2419.)

Key Words: methodology ■ prevention ■ randomized controlled trials ■ stroke

Stroke prevention and management are complex processes requiring intervention at multiple levels (physiological, psychological, social, environmental). Although there is considerable evidence to support the use of pharmacological interventions in aspects of stroke care, the best way to intervene to improve health outcomes using nonpharmacological ‘complex interventions’ is often unclear.¹ Complex interventions often have particular characteristics making them less likely to succeed when evaluated using randomized controlled trial (RCT) methods, such as having multiple components, aiming to target multiple outcomes, being difficult to implement or evaluate, or aiming to achieve outcomes which are notoriously difficult to influence.¹ Even when RCTs of complex interventions are successful it is difficult to fully understand how and why the interventions work.² For example, it is now widely accepted that stroke units save lives and that these units should feature particular characteristics (assessment and monitoring, early management, multidisciplinary team rehabilitation) but we still have relatively little understanding of how these components interact to improve outcomes.³,⁴

A number of models exist to assist in developing and evaluating of complex interventions with the aim of improving intervention quality.⁵-⁷ The UK Medical Research Council (MRC) Framework outlines 5 key phases for intervention development, paralleling those used in pharmaceutical trials¹ (Figure). The Framework does not prescribe specific research methods to be used in intervention development but does recommend specific questions to be addressed in each phase in order to ensure that studies are well designed. The Framework also states that efficacy must be tested in a RCT. Recent studies have started to use this Framework to guide the development of complex interventions including those in stroke care.⁸,⁹

Using the MRC Framework to define quality in intervention development, this review investigates the theoretical and methodological development of complex interventions in stroke care.

Methods
The review included published research articles of evaluations of complex interventions in stroke prevention and management. Be-
Iterative phases of the MRC Framework for developing complex interventions

cause there is little clarity over the definition of a ‘complex intervention’, we defined complex interventions as educational or psychosocial interventions aimed at changing knowledge, beliefs or behaviors. Specific rehabilitation or therapy interventions and those targeted at service organization were excluded. These included: diagnostic tools; nonpharmacological therapies (including complementary medicine, psychotherapy, occupational therapy, speech and language therapy, cognitive-behavioral therapy); stroke units; early discharge interventions; integrated stroke care-pathways.

The review was not restricted by language or study design but interventions evaluated only in preliminary analyses or pilot studies were excluded.

Search Strategy

Multiple search strategies were used: electronic online database searches; hand searching of individual journals; and a ‘gray literature’ search. Five online databases were searched: MEDLINE 1966 to 2005; EMBASE 1980 to 2005; PsychInfo 1967 to 2005; Science Citation Index (SCI) and Social Science Citation Index (SSCI) 1900 to 2005. Recent journals (2000–2005) were hand searched to crosscheck for articles not identified by the electronic search (Stroke, Cerebrovascular Diseases, Lancet, New England Journal of Medicine, JAMA, BMJ, Health Education Research, Health Psychology, British Journal of Health Psychology, Social Science and Medicine, Patient Education and Counseling, Journal of Advanced Nursing, Health Promotion International). Existing reviews, UK health education and health policy documents were searched, as were the reference lists of included articles. Both a MeSH subject heading search and a key word search were conducted (last searched November 2005). The choice of search terms was guided by the MRC Framework and included multiple key words and phrases.


Articles describing the theoretical or methodological development of included studies were retrieved in addition to those documenting the evaluation itself.

One investigator (J.R.) conducted the search and rated studies for inclusion. Where it was unclear whether or not a study should be included, relevant articles were retrieved and discussed with a second investigator (C.M.) until consensus was achieved. In one case where the article could not be retrieved, the authors were contacted to request the full article. Decisions on 4 studies published in languages other than English were made based on translations of the abstract. The key article of 1 study judged to be relevant was retrieved in Chinese and translated in full.

Interventions were categorized by study aim and target group, and data extracted into predesigned tables to ensure standardization across studies.

Analysis

Because the review focused on the influence of theoretical and methodological development rather than on intervention efficacy, a meta-analysis was not conducted.

A number of criteria exist for evaluating study quality but most are not relevant in the context of this review (for example ‘blinding’ is a less relevant criterion to assess most complex interventions where it is impossible to conceal from intervention recipients or providers whether or not they have received the intervention). Instead, we used the MRC Framework recommendations for study development to define quality criteria. Studies were first classified according to the reported development process and whether the intervention was theoretically grounded (Table 1). Secondly, they were classified according to the chosen evaluation methods: RCT versus other designs; outcome measures (primary and secondary outcomes and use of standardized assessment measures) and consideration of statistical power. We considered it unlikely that studies would report explicit ‘phases’ of intervention development because the MRC Framework is not routinely used for this purpose. Studies might also predate the Framework.

The impact of interventions on outcomes (study ‘success’) was defined using three categories: interventions demonstrating a significant beneficial impact on all primary outcomes were defined as ‘successful’, those having a beneficial impact on at least one but not all primary outcomes were defined as ‘partially successful’, those which either failed to demonstrate an impact on any primary outcomes, or demonstrated a detrimental impact were defined as ‘failed’. Simple frequencies were used to describe study characteristics and χ² tests used to explore associations between aspects of study design and outcomes.

Results

Over 12 000 references were identified and 733 retrieved yielding 95 articles covering 67 complex intervention studies. Six hundred and thirty-eight articles were excluded: 293 were not evaluation studies; 217 referred to specific interventions excluded from the review; 98 were not stroke specific; 21 reported pilot or preliminary findings only; 9 were not evaluated studies; 217 referred to specific interventions excluded from the review; 98 were not stroke specific; 21 reported pilot or preliminary findings only; 9 were not evaluated studies; 217 referred to specific interventions excluded from the review; 98 were not stroke specific; 21 reported pilot or preliminary findings only; 9 were not evaluated studies; 217 referred to specific interventions excluded from the review; 98 were not stroke specific; 21 reported pilot or preliminary findings only; 9 were
There were significant differences in the success of studies in different areas ($\chi^2=12.64$, df=4, $P=0.013$). Interventions to improve recovery, in particular social support or information interventions, were less likely to succeed, 12 (41.3%) failed compared with 3 (15.0%) primary/secondary prevention interventions and 3 (17.7%) professional interventions. However, prevention interventions and professional interventions were also less likely to have been evaluated using RCT methods (9, 45% and 6, 35% respectively compared with 24, 80% of interventions to improve recovery).

**TABLE 1. Criteria to Assess Theoretical Quality**

<table>
<thead>
<tr>
<th>Quality Assessment</th>
<th>Theoretical Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) Well grounded</td>
<td>The intervention itself was informed by some form of literature review and in addition by either an established theoretical framework or empirical investigation conducted by the authors.</td>
</tr>
<tr>
<td>B) Moderately grounded</td>
<td>A detailed analysis of at least one of the following was reported: an established theoretical framework; evidence from a published systematic review supporting the intervention; empirical investigation conducted by the authors; or review of empirical studies conducted by the authors supporting the intervention.</td>
</tr>
<tr>
<td>C) Minimally grounded</td>
<td>A brief overview of literature was reported in support of the intervention or a detailed review of the study area but not specifically in support of the intervention.</td>
</tr>
<tr>
<td>D) No theoretical grounding</td>
<td>Interventions did not report any research/cite any literature to support the use of the intervention.</td>
</tr>
</tbody>
</table>

**TABLE 2. Theoretical Quality: Summary of Studies With Theoretical Support for Interventions**

<table>
<thead>
<tr>
<th>Evidence from:</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic review</td>
<td>N=8</td>
</tr>
<tr>
<td>Published theoretical frameworks:</td>
<td>N=19</td>
</tr>
<tr>
<td>Family systems theory</td>
<td>54,59,60,63,78</td>
</tr>
<tr>
<td>Lazarus &amp; Folkman/stress-buffer theories</td>
<td>73,83,85</td>
</tr>
<tr>
<td>Counselling-enabling/social problem solving theory</td>
<td>64,79</td>
</tr>
<tr>
<td>Self efficacy/health locus of control</td>
<td>61</td>
</tr>
<tr>
<td>Prochaska &amp; DiClemente – 'Stages of Change'</td>
<td>52</td>
</tr>
<tr>
<td>Weinstein – Bias in health perception</td>
<td>32</td>
</tr>
<tr>
<td>Orem/Bobath - self care nursing (patient empowerment)</td>
<td>15,22,30</td>
</tr>
<tr>
<td>Chronic care model (patient-centred care)</td>
<td>50</td>
</tr>
<tr>
<td>Shared decision-making models</td>
<td>39</td>
</tr>
<tr>
<td>Rogers - Diffusion-innovation theory</td>
<td>29</td>
</tr>
<tr>
<td>Empirical Research:</td>
<td>N=12</td>
</tr>
<tr>
<td>Testing applicability of theoretical model</td>
<td>61,64</td>
</tr>
<tr>
<td>Testing intervention components</td>
<td>15,45,58,67,87</td>
</tr>
<tr>
<td>Canvas user views/ service needs</td>
<td>9,26,30,37,71,79</td>
</tr>
</tbody>
</table>

**Theoretical Grounding**

Theoretical grounding was difficult to establish from published reports because most studies presented only a brief introduction to the study area and an overview of the methods. Forty-two of the 67 studies reported some form of theoretical grounding for their intervention (supplemental Tables I to III) but only 14 were judged to be theoretically well developed (an ‘A’ rating).

Table 2 presents a list of theoretical research in support of the interventions. Although most studies included some form of literature review, this was not necessarily in support of the chosen intervention. Only 8 studies reported attempts to systematically review the literature or cited an existing systematic review. Nineteen studies referenced published theoretical frameworks to support their intervention choice (mostly from the psychological literature) and 13 conducted empirical research to inform intervention development.

**Outcome Assessment**

Study outcome measures are presented in Table 3. A range of single item questions, standardized/published scales (53 in total), or author-developed scales (15) were used. Author designed outcome measures (with or without validation) were mostly used to assess knowledge or satisfaction with care but also to assess self-rated quality of life, health expectations, ‘recovery-efficacy’, illness perceptions, assertiveness, decision-making and confidence. There was no association between type of measure (single item or scale), measure development (published or author developed) and study outcomes. Most studies used multiple measures to evaluate impact. Twenty-one RCTs and 16 quasi-experimental/obser-
vational studies listed 3 or more primary outcomes of interest with 10 RCTs and 12 non-RCTs listing 5 or more primary outcomes (supplemental Tables I to III). Studies with large numbers of primary outcomes (5 or more) were statistically less likely to be completely successful than those with fewer outcomes (5, 22.7% compared with 17, 38.6%) but were also less likely to fail (2, 9.1% compared with 16, 36.4% respectively, $\chi^2 = 12.1$, $df = 2$, $P = 0.002$).

Methodological Quality

There were significant differences in the success of interventions using different evaluation methods ($\chi^2 = 17.2$, $df = 2$, $P = 0.001$). Only 4 non-RCT studies distinguished between primary and secondary outcomes,22,35,42,60 but all achieved at least partial success and 12 (44.4%) achieved all of their aims. One study conducted a qualitative evaluation of a family nursing intervention concluding that all aspects of the intervention were useful, but did not specifically test a hypothesis.78 Ten (25.6%) trials were successful, 11 (28.2%) were partially successful but nearly half (18, 46%) were unsuccessful. Sixteen (41.0%) RCT studies included a power calculation; 6/21 RCT studies with multiple (>2) primary outcomes reported considering statistical power (Table 4). One study considered power for 4/7 primary outcomes and although it was still slightly underpowered (it did not reach the target recruitment figure) it was successful in influencing all of the outcomes for which power was estimated.68 In the only study to consider power for all measures, the authors anticipated that the intervention would lead to a 25% difference between groups in all risk factor control outcomes.49 The intervention did not have the anticipated impact, the authors concluding that risk factor control rates in the control group were better than expected, and consequently that the study was underpowered.

Only one RCT study reported conducting any form of exploratory trial of the intervention before the main study. It was used to test the feasibility of implementing a telephone intervention to provide nursing follow-up to improve adjustment and

<table>
<thead>
<tr>
<th>Measure:</th>
<th>Professional Interventions</th>
<th>Primary/Secondary Prevention</th>
<th>Recovery &amp; Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>N=7 (Refs. 13, 18, 23, 25, 26, 28, 30)</td>
<td>N=7 (Refs. 33, 34, 36, 37, 45-47)</td>
<td>N=9 (Refs. 55-60, 69, 72, 86)</td>
</tr>
<tr>
<td>Psychological and social adjustment, perceived health-status, self-rated quality of life, social resources, confidence and self-esteem</td>
<td>N=1 (Ref. 17)</td>
<td>–</td>
<td>N=17 (Ref. 17)</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>N=1 (Ref. 11)</td>
<td>N=2 (Refs. 36, 37)</td>
<td>N=12 (Refs. 9, 53, 56, 58, 59, 64, 69, 75, 77, 78, 84, 86, 87)</td>
</tr>
<tr>
<td>Physical functioning, independence, disability in activities of daily living</td>
<td>N=1 (Ref. 17)</td>
<td>–</td>
<td>N=8</td>
</tr>
<tr>
<td>Mood: anxiety and depression</td>
<td>N=1 (Ref. 17)</td>
<td>–</td>
<td>N=8</td>
</tr>
<tr>
<td>Mortality, morbidity and institutionalization</td>
<td>N=2 (Refs. 12, 24)</td>
<td>N=6 (Refs. 40, 42-45, 51)</td>
<td>N=1 (Ref. 68)</td>
</tr>
<tr>
<td>Changes in clinical practice</td>
<td>N=9 (Refs. 11, 13-16, 18, 22, 23, 29)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Health behavior/medication use/adherence</td>
<td>N=3 (Refs. 17, 21, 30)</td>
<td>N=7 (Refs. 37, 39-41, 48, 49, 52)</td>
<td>–</td>
</tr>
<tr>
<td>Physiological risk factor measures</td>
<td>–</td>
<td>N=6 (Refs. 40, 41, 44, 45, 49, 52)</td>
<td>N=2 (Refs. 66, 69)</td>
</tr>
<tr>
<td>Use/receipt of health and social services</td>
<td>–</td>
<td>N=3 (Refs. 35, 36, 44)</td>
<td>N=2 (Refs. 82, 90)</td>
</tr>
<tr>
<td>Health perceptions, attitudes, behavioural intentions</td>
<td>N=3 (Refs. 17, 26, 30)</td>
<td>N=3 (Refs. 32, 36, 48)</td>
<td>N=1 (Ref. 61)</td>
</tr>
<tr>
<td>Family functioning, carer burden/stress/competence</td>
<td>–</td>
<td>–</td>
<td>N=6 (Refs. 54, 59, 64, 68, 85, 86)</td>
</tr>
<tr>
<td>Global health outcomes</td>
<td>N=1 (Ref. 30)</td>
<td>N=2 (Refs. 48, 50)</td>
<td>N=1 (Ref. 77)</td>
</tr>
<tr>
<td>Patient choice/decision making</td>
<td>N=1 (Ref. 30)</td>
<td>N=1 (Ref. 37)</td>
<td>N=1 (Ref. 53)</td>
</tr>
</tbody>
</table>
recovery and to test the validity of outcome measures. In one of the social support intervention studies, the authors had already conducted an RCT of 1 of the intervention components (the information pack) which could be described as a pilot evaluation. The authors also conducted research to inform their power calculation before the main trial and this work could be defined as the “modeling phase” of development. However, because this was the only study that did so, the relationship between study quality and success is difficult to assess. Studies reporting a power calculation to inform trial methods were no more likely to have been successful than those which did not report considering statistical power.

### TABLE 4. Methodological Quality: Consideration of Statistical Power

<table>
<thead>
<tr>
<th>Study &amp; Date</th>
<th>Outcome Measures</th>
<th>Time Point</th>
<th>Power Calculation</th>
<th>No. Recruited</th>
<th>Impact on Primary Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allen 2004,50 USA</td>
<td>Global “well-being” (5 domain score developed by authors)</td>
<td>3 months postintervention</td>
<td>96 patients gives 90% power to detect small to moderate effect sizes across the five domains ( \alpha = 0.05 ).</td>
<td>96</td>
<td>There was a significant effect size of 0.53 on the global test score. Effect sizes were significant on all domains</td>
</tr>
<tr>
<td>Boter 2004,75 Netherlands</td>
<td>Dissatisfaction with stroke care (Satisfaction with Stroke Care questionnaire - SASC-19)</td>
<td>6 months postdischarge</td>
<td>524 participants gives 80% power to detect a 50% reduction in dissatisfaction (( \alpha ) not reported)</td>
<td>536</td>
<td>No statistical differences in dissatisfaction scores between groups except intervention patients had higher scores on role limitation.</td>
</tr>
<tr>
<td>Clark 2003,54 Australia</td>
<td>1) Family functioning (McMaster Family Assessment Device–FAD) 2) Functional status (Barthel Index – BI) 3) Social recovery (Adelaide Activities Profile – AAP) Proportion of patients whose risk factors were ‘on target’:</td>
<td>6 months postdischarge</td>
<td>30 families give 80% power to detect a 0.14 decline in FAD from premorbid state to 6 months (( \alpha ) not reported).</td>
<td>62</td>
<td>FAD scores remained stable in the intervention arm but declined in the control arm. The BI was slightly higher for the intervention group than controls. The AAP was better in the intervention group than control group for domestic chores, household maintenance and social activities.</td>
</tr>
<tr>
<td>Ellis 2005,49 UK</td>
<td>1) All relevant risk factors controlled 2) Blood pressure &lt; 140/85 3) Smoking cessation 4) Blood glucose &lt; 8.0 mmol/L 5) HbA1c &lt; 7.5% 6) Cholesterol &lt; 5.0 mmol/L</td>
<td>5 months postintervention</td>
<td>89 patients per group gives power to detect a change from 25% risk factor ‘on target’ to 50% (( \alpha ) not reported)</td>
<td>205</td>
<td>After adjusting for baseline differences the intervention had no significant impact for any risk factors.</td>
</tr>
<tr>
<td>Forster 1996,79 UK</td>
<td>Improvement in social activities (Frenchay Activities Index–FAI)</td>
<td>12 months postrecruitment</td>
<td>160–220 patients gives 90% power to detect a 4-point improvement in the FAI (( \alpha ) not reported)</td>
<td>240</td>
<td>No significant differences in FAI between groups. Qualitative analysis revealed that aspects of the nursing role (concern, attention, empathy and interest) were valued by the patients and careers.</td>
</tr>
<tr>
<td>Harari 2004,68 UK</td>
<td>Self-reported bowel movements (BM) per week</td>
<td>1, 3, 6 and 12 months postintervention</td>
<td>120 patients would give 90% power to detect a 54% increment in BMs/week (( \alpha ) not reported)</td>
<td>146</td>
<td>Mean No. of BMs per week was higher in the intervention group compared to the controls and persisted at 12 months. Self-rated normal BMs were also higher in the intervention group.</td>
</tr>
<tr>
<td>Kalra 2004,68 UK</td>
<td>1) Death and institutionalization</td>
<td>3 &amp; 12 months poststroke</td>
<td>300 patients would give the study 80% power to detect:</td>
<td>268</td>
<td>No significant differences were found in mortality/ institutionalization. Patients in the intervention arm had lower median HAD scores and higher median Euroqol scores. A higher proportion were less disabled and had low Rankin scores at 3-months.</td>
</tr>
</tbody>
</table>
Impact of Study Quality on Outcomes

After adjusting for aspects of methodological quality (study design and statistical power) there was no evidence that theoretical development was associated with success in primary outcomes. Six RCT studies were judged to be theoretically well developed and included a power calculation; 2 of these were successful,\textsuperscript{50,54} 2 were partially successful,\textsuperscript{64,86,88} and 2 unsuccessful.\textsuperscript{9,71} Of the 2 successful interventions, the first had a significant impact on family functioning, which remained stable in the intervention group but declined over

### TABLE 4. Continued

<table>
<thead>
<tr>
<th>Study &amp; Date</th>
<th>Outcome Measures</th>
<th>Time Point</th>
<th>Power Calculation</th>
<th>No. Recruited</th>
<th>Impact on Primary Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mant 2000,\textsuperscript{67} UK</td>
<td>1) Quality of life (Dartmouth Coop charts; SF-36) 2) Disability (Barthel Index &amp; Rivermead mobility index) 3) Handicap (London Handicap Scale - LHS) 4) Emotional health (Hospital anxiety and depression scale - HADS)</td>
<td>6 &amp; 12 months poststroke</td>
<td>300 participants gives 80% power to detect: 7-point difference in the LHS; 2.5-point difference in the HADS; 7-point difference in secondary outcome: SF-36 (α not reported).</td>
<td>323</td>
<td>Significant differences in carer FA, and some quality of life domains in favor of the intervention (5 parts of SF-36 and 1 part Dartmouth Coop chart). No significant differences in patient outcomes.</td>
</tr>
<tr>
<td>Pennington 2005,\textsuperscript{29} UK</td>
<td>Adherence to clinical guidelines (10-item tool developed by the authors).</td>
<td>8–12 months postintervention</td>
<td>24 departments recruiting 50 patients each gives 80% power to detect a 0.24 difference in adherence at the α = 0.05.</td>
<td>717</td>
<td>No difference between groups in adherence to practice guidelines.</td>
</tr>
<tr>
<td>PRISM 2003,\textsuperscript{21} UK</td>
<td>Relative risk reduction (RRR) in ischaemic and haemorrhagic vascular events achieved by antithrombotic therapy Vs no therapy.</td>
<td>6 months postintervention</td>
<td>27 centres recruiting 40 patients each (1080 patients in total) gives 80% power to detect a 5% change in RRR, assuming an ICC of 0.15 (α not reported).</td>
<td>1952</td>
<td>No difference in RRR between groups, or the proportion of patients receiving optimal therapy.</td>
</tr>
<tr>
<td>Rodgers 1999,\textsuperscript{71} UK</td>
<td>Patients’ and carers’ perceived health status (SF-36)</td>
<td>6 months poststroke</td>
<td>160 patients and 216 carers gives 80% power to detect differences ranging from 5–10 points in the subscales of the SF-36 (α not reported).</td>
<td>204</td>
<td>No differences in SF36 outcome between groups.</td>
</tr>
<tr>
<td>SAFIRE 2004,\textsuperscript{24} Australia</td>
<td>Fall rate</td>
<td>3 months postintervention</td>
<td>348 patients gives power to detect a 10% reduction in fall rate in the intervention group (α and β not reported).</td>
<td>715</td>
<td>Fall rates increased in both trial arms over the study period. No significant differences between groups.</td>
</tr>
<tr>
<td>Smith 2004,\textsuperscript{22} UK</td>
<td>Knowledge score (developed by the authors)</td>
<td>3 &amp; 6 months poststroke</td>
<td>Sample size was calculated for secondary outcomes only; 65 patients per group were needed to achieve 90% power to detect a difference of 8 points in the London Handicap Scale (α not reported).</td>
<td>170</td>
<td>No difference in knowledge score between groups at either time point.</td>
</tr>
</tbody>
</table>
time in the control group; the second had a significant effect on an author-defined (and validated) global well-being score (0.53 effect size), but the implications of this for clinical practice are difficult to interpret. Five non-RCT studies were also considered well developed; 1 of these was successful, and 3 were partially successful and 1 did not test a specific hypothesis. If the theoretically moderately well developed studies (‘B’ rated) and well developed studies (‘A’ rated) are combined, the findings remain the same with no clear evidence that theoretical or methodological quality was associated with improved outcome.

**Discussion**

This review aimed to investigate theoretical and methodological quality in the development and evaluation of complex interventions in stroke care and the impact of this on study outcomes. The 67 intervention studies included in the review were targeted at a range of health outcomes. The majority of interventions identified were targeted at an individual level (general public, stroke patients, care-givers and health professionals) rather than at a social or environmental level. This may reflect the chosen search criteria because environmental and social interventions are more likely to be targeted at the general population and aim to prevent a range of diseases. Such studies might not have been found unless they included stroke specific outcomes.

Observational or quasi-experimental studies tended to yield better outcome success than RCT designs. Differences may reflect difficulties in demonstrating success using RCT methods but may also be influenced by publication bias, because journals are more likely to publish studies if there is a significant treatment effect. If journals are less likely to publish nonsignificant results for observational or quasi-experimental studies than for RCTs, then interventions evaluated in RCTs would appear to be less successful than those evaluated using other methods.

Study design may also explain the apparent lack of success of interventions focusing on patient and carer adjustment and recovery because these studies were more likely to be evaluated in RCTs. There were few RCT studies evaluating interventions in primary or secondary prevention, possibly reflecting problems of recruitment, randomization and follow-up, for relevant community interventions. Such interventions may also present ethical difficulties attributable to the requirement that researchers withhold the intervention from some groups/patients. For the studies included in this review, a number of novel quasi-experimental methods were used in an attempt to overcome such difficulties, including matched controls, time-lag controls, or ‘before and after’ studies, with or without qualitative evaluation. However, the MRC framework suggests that these methods are inadequate for evaluating efficacy because nonrandom allocation of participants may lead to bias. Using the MRC Framework (with its focus on RCTs) as a marker of quality in complex intervention evaluation may lead to exclusion of those interventions which best serve the needs of disadvantaged groups (because these groups are more difficult to identify, recruit and retain in research studies). It may also lead to exclusion of interventions that have the most impact on longer-term population health goals (such as mortality rates or behavior change). Although theoretically longer-term follow up is not a problem of RCTs, in practice RCTs are more difficult and expensive to conduct over long periods of time.

One of the potential strengths of the MRC Framework is that it sets standards for theoretical and methodological development within an RCT design such as highlighting the importance of considering statistical power. However, a substantial proportion of the RCTs in this review did not include a power calculation to justify sample size. No relationship was found between reporting a power calculation and study success, possibly because many of the studies that did not conduct a power calculation were still adequately powered; conversely, not all of the studies that reported a power calculation did achieve statistical power. Consideration of statistical power should be more than an

**TABLE 4. Continued**

<table>
<thead>
<tr>
<th>Study &amp; Date</th>
<th>Outcome Measures</th>
<th>Time Point</th>
<th>Power Calculation</th>
<th>No. Recruited</th>
<th>Impact on Primary Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tilling 2005, UK</td>
<td>Satisfaction with services (Pound Scale)</td>
<td>3 &amp; 12 months postrandomisation</td>
<td>336 patients give 80% power to detect a 15% difference in satisfaction between groups (α not reported)</td>
<td>340</td>
<td>No difference in overall satisfaction score between groups for either patients or carers. Satisfaction with some aspects of care was higher in the intervention group than the control group for patients and carers at particular time points but no consistent patterns.</td>
</tr>
<tr>
<td>Van Den Hueil 2000, Netherlands</td>
<td>1) Confidence in knowledge about relevant themes related to being a caregiver (a new instrument developed by the authors) 2) Use of active coping strategies (Utrecht Coping List)</td>
<td>14 weeks from baseline</td>
<td>76 respondents per group gives 80% power to detect a medium effect in outcomes; 547 respondents per group would give 80% power to detect small effects (α not reported)</td>
<td>130 group 1 78 group 2 49 group 3</td>
<td>Both interventions achieved small increases in confidence in knowledge of patient care and use of active coping strategies. The group program achieved a small increase in social support seeking. No differences between home and group support.</td>
</tr>
</tbody>
</table>
administrative task and needs to be theoretically justified. Exploratory evaluations before the main study could have been used to make more realistic estimates of intervention impact. Authors could then have adjusted sampling or outcome measures at the outset.

For most of the studies there was little evidence that authors considered the theoretical development, or the mechanisms by which the intervention was expected to influence outcomes. However, this was not necessarily the fault of the investigators. For example, in 2 social-support intervention studies the interventions were existing services rather than services designed specifically for the study.77,84

Lack of clarity about the mechanisms by which interventions work makes it difficult to pool the results of similar studies in systematic reviews or meta-analyses, because although intervention components appear similar, we do not know whether interventions are delivering similar ‘packages’ of care. In one systematic review of problem solving interventions for carers after stroke, the authors concluded that poor theoretical and methodological quality coupled with a diverse range of intervention components made it impossible to draw conclusions about efficacy.98 Similar problems have been reported in systematic reviews of other complex stroke interventions such as information and education.99

Requirements for study design and presentation of results inevitably change over time and our review is limited in that inferences about study development are based on published or reported information only. It is possible that some studies did actually conduct empirical work or pilot their interventions but did not report it. However, it is of concern that so few studies reported an appropriate literature review or considered established theory. It would not be acceptable within a standard clinical trial to test out a pharmaceutical intervention in a phase III study without sufficient understanding of the mechanisms by which the drug influences bodily processes. Yet, complex health service interventions seem to be developed based on ad-hoc assumptions, and evaluated using methods where at the end of the study it is impossible to understand the reasons for success or failure.

Complex interventions often aim to influence a number of outcomes, or outcomes that are difficult to encapsulate in a single measure. In just over half of the studies multiple outcome measures were used. In some cases, authors may have incorporated multiple measures in the hope that by ‘hedging their bets’ it would enhance the chance of at least one of the measures showing a significant result. Non-RCT studies were particularly poor at specifying the main outcomes of interest and this may explain why they were less likely to have failed. However, not identifying the key outcomes of a study in which multiple outcomes are assessed suggests poor theoretical development and that the study question itself is insufficiently clear.

Whether using the recommendations of the MRC Framework to assist development will improve our understanding of interventions in the future is unclear. There did not appear to be any convincing evidence that being theoretically or methodologically well developed improved outcomes. The one study that explicitly reported using the MRC framework in development was unsuccessful.9 The Framework does not guarantee success, but in the case of this intervention, despite apparently following recommended procedures we are still no clearer about why the intervention failed. An evaluation of intervention implementation (process evaluation) could have enlightened our understanding.

Conclusion

This literature review has highlighted inadequacies in the development and evaluation of complex interventions in stroke care. Using the MRC Framework as a measure of quality, we found studies to be lacking in terms of theoretical and methodological rigour. This may explain the negative results commonly reported in RCTs. In the future, researchers should pay greater attention to theoretical and methodological rigour. However, new methods also need to be developed to help unpack intervention complexity. Qualitative methods and process evaluation could be used to help interpret trial findings but more research to investigate their role alongside RCTs and their potential impact on study outcomes is required.

Sources of Funding

The authors would like to acknowledge the Medical Research Council for funding J.R. as a Special Training Fellow in Health of the Public Research and the Department of Health Research Capacity Development Programme for funding C.M. as a Public Health Career Scientist.

Disclosures

None.

References


