Generating the evidence base for the National Service Framework for Long Term Conditions: a new research typology

Lynne Turner-Stokes, Richard Harding, Judith Sergeant, Carol Lupton and Kath McPherson

ABSTRACT – The UK National Service Framework (NSF) for Long Term Conditions was published in May 2005. This article describes the challenges and some proffered solutions towards development of the evidence base to support best practice in the management of life-long neurological conditions, which are the principal focus of the NSF. The inherent limits in current systems for appraisal of evidence and their lack of applicability to these conditions are discussed. A new typology of evidence is proposed, which acknowledges the importance of expert opinion from users, carers and professionals as well as encompassing a broad range of research designs. To apply the typology, a brief evaluation tool is presented, which provides simple assessment of both qualitative and quantitative research evidence in terms of design, quality and applicability, and is practical for use by clinicians. Preliminary testing and application in development of the evidence base for the NSF are described.

KEY WORDS: evidence appraisal, long-term conditions, National Service Framework, research

Background

Evidence-based healthcare provision has become the guiding principle for clinical practice on the grounds of both optimising patient care[1] and efficient allocation and utilisation of health resources.[2] Clinical standards and guidelines have become the tools by which evidence-based practice is determined and delivered. In the past five years, standards for the National Health Service (NHS) in England have been set out in a series of National Service Frameworks (NSFs). The latest in this series – the NSF for Long Term Conditions – was released in March 2005 and presented some new challenges requiring novel solutions which have relevance for developing evidence-based practice both in the UK and in other countries.

Evidence-based practice requires the integration of individual clinical expertise with the best available external evidence from rigorous clinical research.[1] Randomised controlled trials (RCTs) – long considered to be the most reliable means of obtaining unbiased and robust information[3] – have become the accepted 'gold standard' for providing best quality evidence, with alternative quasi-experimental research designs offering acceptable but less highly ranked evidence. However, it is increasingly recognised that experimental quantitative research methods cannot answer all the questions that need to be answered.[4] In particular:

- they do not provide certainties, but only statistical probabilities
- they only deliver evidence for areas that can be controlled for, measured, counted and analysed in quantifiable terms, and experimentally manipulated[5]
- strict inclusion criteria and enhanced quality of care within the trial setting may limit the generalisability of their findings to real-life practice.[6]

Even where fit for purpose, RCTs can be undertaken well or badly and are only as good as the quality of theoretical work by which they are underpinned. If a trial asks the wrong questions, it will find the wrong answers.

Life-long neurological conditions are the principal focus of this latest NSF, and development of the evidence-base posed a number of challenges for traditional research methods, namely:

1 The effects of a long-term condition unfold over many years – a timescale beyond the scope of most clinical trials.
2 Interventions are complex and played out over a long period, changing progressively in the light of the individual's response to what has gone before. They are frequently multidisciplinary in nature, and any one intervention will often overlap and interact with others.
3 The effect of any clinical intervention must be assessed not only in the context of other clinical care, but also against a complex background

Lynne Turner-Stokes[2] DM FRCP, Herbert Dunhill Professor of Medicine

Richard Harding[1] PhD, Lecturer, Department of Palliative Care Medicine

Judth Sergeant,[3] NSF Development Team

Carol Lupton[3] PhD, Research and Development

Kath McPherson[4] PhD, Professor of Rehabilitation

[1]Department of Palliative Care and Policy, King's College London, UK
[2]Academic Rehabilitation, Northwick Park Hospital and King's College London, UK
[3]Department of Health, UK
[4]Auckland University of Technology, Auckland, New Zealand

Professor Turner-Stokes was Deputy Chair of the External Reference Group and Chair of the Research and Evidence Sub-group for the NSF in Long Term Conditions

Clin Med 2006;**6**:91–97

array of social and environmental factors, which are not adequately described by existing quantitative techniques. To control properly for these as 'unseen confounders' potentially requires a much larger sample even than the total affected population. This may be particularly so for the less common conditions.

4 By definition, in the context of a long-term condition, 'cure' or reversal of pathology is rarely a goal for treatment. Instead, the intended outcomes focus on reducing the impact of the disease – for example, on quality of life or societal participation. Whilst a variety of standardised 'measures' have been developed to evaluate these issues, many appear to provide a less than satisfactory reflection of real-life experiences in chronic conditions.[7,8]

5 Response shift is likely to occur throughout the trajectory of a long-term condition and may confound evaluative efforts.[9]

We do not suggest that the typical or 'pure' RCT has no place in the context of life-long neurological conditions. Undoubtedly there are questions that lend themselves to these experimental approaches, and for which they may remain the design of choice. However, there are other areas of investigation for which alternative quantitative designs or qualitative techniques will be more appropriate. Some would argue that observational studies and other uncontrolled studies are open to bias and may therefore overestimate the magnitude of treatment effects. In fact, two detailed reviews comparing the findings of observational studies and RCTs found no evidence for this.[10,11] Moreover, in an analysis of 'survival of truth' – that is time for which the study conclusions remained 'true' to current belief – Poynard and colleagues[12] identified that the 20-year survival of conclusions from meta analysis (57% SD = 10) was lower than that from individual non-randomised studies (87% SD = 2), or randomised trials (85% SD = 3) in the area of hepatitis and cirrhosis. These reviews highlight the fact that a re-think of 'best evidence' is required, based more on the quality of research than the design itself.

Key Points

Current systems for appraisal of evidence overemphasise quantitative experimental research and have limited applicability in the context of long-term conditions

A new typology of evidence was developed for the National Service Framework (NSF) for Long Term Conditions, which encompasses expert opinion from users, carers and professionals as well as a broad range of research designs

The research evaluation tool provides simple assessment of both qualitative and quantitative research evidence, which is practical for use by clinicians

Evaluation of the tool and its application in the appraisal of research evidence for the NSF are described

The new typology offers a practical way forward to more inclusive approaches to the gathering and appraisal of evidence to support best clinical practice

In the light of these methodological challenges to evidence generation, the NSF protocol proposed the adoption of a much wider definition of 'evidence' encompassing a range of robust research methods. This article describes the development and evaluation of a new 'typology of evidence' which is designed to underpin the recommendations of the NSF for Long Term Conditions, and which offers much wider application in other areas of clinical practice that demand a holistic approach to evidence gathering and appraisal.

Methods

Policy framework

Preparation of an NSF involves an initial period (usually about 18 months) during which the appointed External Reference Group (ERG) assembles evidence-based 'advice' for the Minister on the nature and content of the standards to be contained within the NSF. A second phase of approximately 6–9 months follows. During this time the Minister, supported by the Department of Health (DH) NSF development team, considers in depth the ERG's proposals, their likely impact and the expected cost of meeting them. The final NSF is then drafted accordingly, laying out a dozen or so standards (or 'Quality Requirements' as they were called in this case), together with a set of more detailed recommendations ('Markers of Good Practice') and the evidence base to support them.

External Reference Group

The External Reference Group for this NSF was appointed in 2002. Developing the evidence base was identified as an early priority, and a 'Research and Evidence' (R&E) subgroup was set up to lead this task. Its role included (a) defining a system for the evaluation and presentation of evidence to underpin the NSF recommendations, and (b) overseeing the assembly and evaluation of evidence in accordance with that system. The ERG handed its advice to the Minister in May 2004, but the R&E subgroup continued to support the generation of the evidence base up until February 2005 to ensure the quality and accuracy of citations during drafting of the final document. The process of development of the NSF and generation of the evidence base is outlined in Fig 1.

Pre-existing research appraisal techniques

An initial review of existing research classifications produced the following conclusions:

1 Hierarchical classifications such as those used by the National Institute for Health and Clinical Excellence (NICE)[13] and the Scottish Intercollegiate Guidelines Network (SIGN)[14] form a useful basis for research evaluation, but tend to overemphasise quantitative experimental designs, and make little or no mention of well-conducted qualitative or mixed methods.

2 The NSF for Older People used a research 'typology' which attempted to elevate 'well-designed qualitative research' to

level B3, and also to identify separately expert opinion from users (U), carers (C) and professionals (P). However, its hierarchical classification was still based solely on experimental design, with no systematic attempt to assess research quality. It also failed to encompass the growing body of social scientific research that collects both qualitative and quantifiable data.

3 Earlier classifications focused solely on design regardless of quality,[15] but current procedures for research evaluation including SIGN,[16] Consort Quorum[17] and the Centre for Reviews and Dissemination,[18] do now include an assessment of quality based on the proportion of criteria met from a predefined checklist for each design category. Checklists for evaluation of other designs including cohort and case-control designs have also been developed.[19,20]

4 Many of these quality checklists, however, tend to be long (the SIGN checklists are 23–26 items) and thus were not considered practical in this instance. They also focus predominantly on quantitative methods. The development of techniques to synthesise qualitative findings is latterly underway,[21-23] and integrative techniques to synthesise research from both quantitative and qualitative methods are also being explored through the Health Development Agency[24] and the Economic and Social Research Council (ESRC)[25] and other sources.[26] However, at the time of starting this development, no simple practical tools were immediately to hand.

Development of a new evidence typology

From the above review of grading tools, we were unable to identify a suitably practical classification system which satisfactorily recognised the breadth of research design required in health and social care for these patient populations. Recognising that greater strength of evidence that is provided when both expert opinion and research of different types reinforce one another,

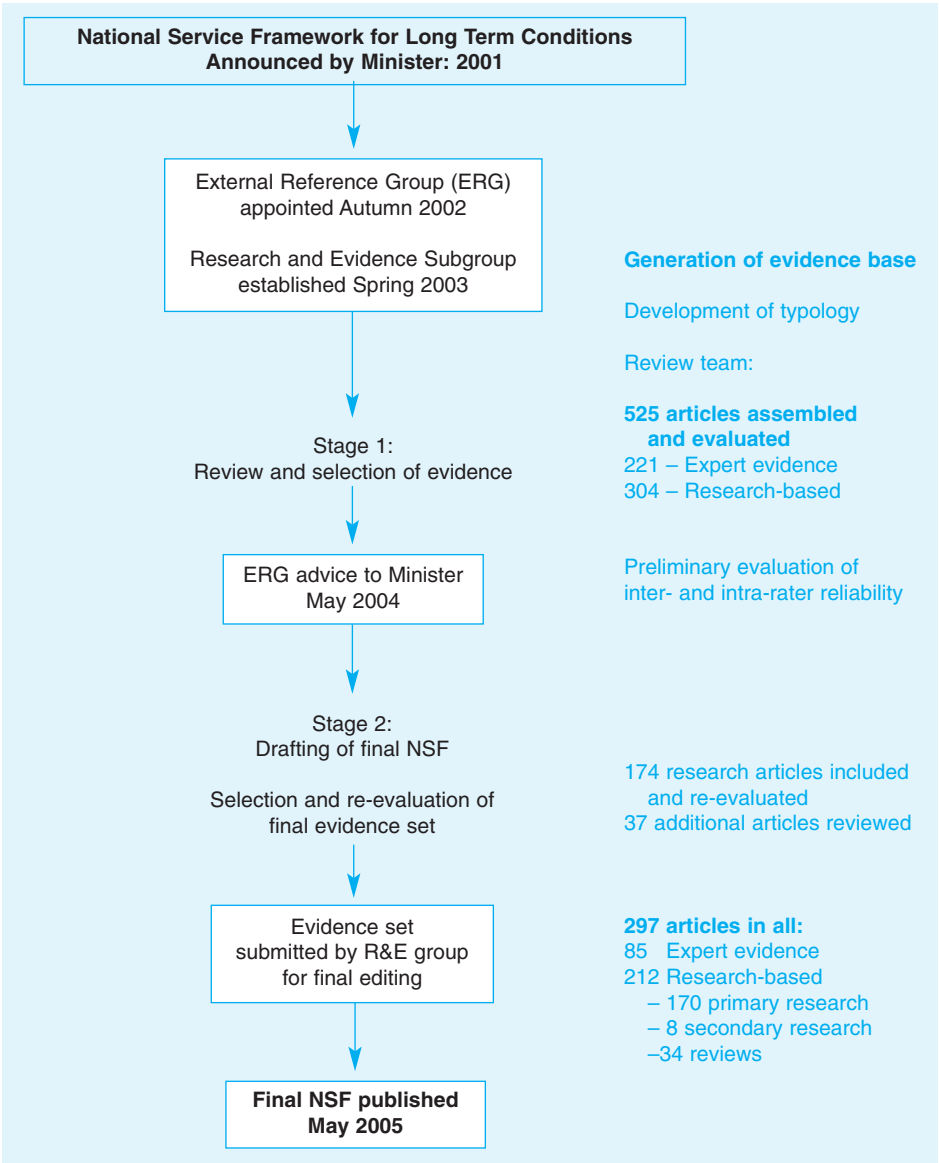

Fig 1. Process of NSF development and generation of the evidence base.

the R&E group set about designing a new typology for the NSF. The following criteria were identified by consensus discussion within the group:

1 The typology must take account of the opinions and experience of service users, their families and carers, as well as the views of professionals.
2 Emphasis should be placed on the quality of the study design, the integrity of its conclusions, and their relevance to the population served by this NSF – accepting the principle that well designed and conducted qualitative, quantitative and mixed studies can have equal validity when used in the appropriate context.
3 It should include an evaluation tool which could be applied to the full range of research designs, to provide an assessment of research quality that is both *simple* and *quick* to apply, so that it is practical for use by professionals in clinical settings.

The new typology was developed through an iterative process, successive drafts being piloted by members of the group until a satisfactory design was reached. The final structure of the typology is outlined briefly in Table 1, and details of the process for grading and synthesis of the literature are given in Appendix 1.

Application and evaluation of the typology

During the initial phase of advice preparation, ERG members and other contributors provided a large bank of reference material to generate the proposed 'evidence-based recommendations'. This evidence was gathered either from their direct knowledge of the supporting literature, or from specific searches undertaken by the ERG members and research staff. Biomedical databases were searched for each domain of the NSF and for the range of long-term neurological conditions. After evaluation of individual studies in any given area, a synthesis of best research evidence was undertaken to provide an overall grade (A, B or C) based on the quality category and applicability of the studies to support each recommendation.

In all, a team of 11 different reviewers were involved in evaluating and checking the evidence base in two stages.

Stage 1: During the initial assembly of advice for the Minister, five reviewers working in parallel undertook between them the initial sifting and evaluation of evidence for the proposed NSF. These were two members of the R&E group and three paid research assistants

Table 1. Summary of evidence typology for the NSF for Long Term Conditions.

Expert evidence	
Opinion/experience	Of users and/or carers (E1) or professionals (E2)
Research-based evidence	
Design	
Primary research	Quantitative (P1), qualitative (P2) or mixed methods (P3)
Secondary research	Meta-analysis (S1) or other secondary analysis (S2)
Reviews	Systematic (R1), or other descriptive reviews (R2)
Quality assessment	Rated on five parameters (scored out of 10) and categorised into 'high', 'medium' and 'low' quality ratings.
Applicability	*Direct* (within long-term neurological conditions) or *Indirect* (extrapolated evidence from other conditions)

Table 2. Inter- and intra-reliability agreement between independent ratings (Stage 1), and comparison of agreement with initial rating on re-evaluation of final quality ratings (Stage 2).

	Inter-rater reliability 96 paired ratings		Intra-rater reliability 19 paired ratings		Re-evaluation 174 paired ratings	
Criterion	**Kappa**	**Significance (*p*)**	**Kappa**	**Significance (*p*)**	**Kappa**	**Significance (*p*)**
Design category	0.80	<0.001	0.47	<0.001	0.80	<0.001
Quality criterion						
Aims are clearly stated	0.43	<0.001	0.48	0.03	0.85	<0.001
Design appropriate to research question	0.11	N/S	0.11	N/S	0.86	<0.001
Methods are clearly described	0.42	<0.001	0.88	<0.001	0.87	<0.001
Adequate data presented to support authors' conclusions	0.23	0.004	0.62	0.004	0.82	<0.001
Results are generalisable	0.28	<0.001	0.88	<0.001	0.88	<0.001
Quality category (high (>6), medium (4–6) or low (<4))	0.32	<0.001	0.60	<0.002	0.82	<0.001
Applicability (direct or indirect)	0.31	<0.001	1.0	<0.001	0.73	<0.001
	Rho	**Significance**	**Rho**	**Significance**	**Rho**	**Significance**
Total quality score (Spearman rank correlation)	0.34	*p* <0.001	0.95	*p* <0.001	0.84	*p* <0.001

(one full time, two part time for four months). An identified individual from within the review team led for each draft 'Quality Requirement (QR)', to coordinate the process of searching, selection, review and collation of articles to form the set of supporting evidence.

Stage 2: During preparation of the final NSF document, the reference pool was refined and reduced, to include only the most salient articles. Some further reference articles were added at this stage to fill identified gaps in the evidence base. Seven reviewers were involved in re-evaluation, selecting and checking the subset of evidence to be included (details of this process are given later). These were three members of the DH NSF team, two members of the R&E group and two part-time paid research assistants.

During the course of this process, repeatability and reliability of the classification and research quality assessments were evaluated. The results are presented in Table 2 and discussed below.

• **Stage 1**

Inter-rater reliability: During Stage 1, 96 research articles had two independent ratings undertaken by different members of the research team. Taking these 96 together, the paired sets of ratings were compared using a computerised statistical package (SPSS version 11.5). Agreement was tested using non-weighted Cohen's Kappa statistics for each of the five quality criteria scores, applicability, category of research and overall quality category (high, medium or low). Correlation of the total quality scores was tested using a Spearman rank correlation. Although agreement for four of the five individual quality criteria reached levels of statistical significance, this could at best be described as 'moderate' according to the classification of Fleiss *et al* (1981).[27] Relatively poor agreement for 'Applicability' reflected differences in the applied definition of 'long-term neurological conditions' which changed slightly during the evolution of the NSF. This definition was tightened for subsequent rounds.

Intra-rater reliability was not tested formally, but was 'put to the test' in 19 instances where a reviewer rated the same article for a different QR – evidently having no recollection of having rated it previously, since they created a new entry on the database rather than copying over a previous record. Here the associations are relative high, indicating a good level of consistency in interpretation of the quality rating tool, at least for the two reviewers who undertook these repeated ratings.

The evaluations of both inter- and intra-rater reliability demonstrated relatively poor agreement in rating the individual quality criterion of 'Design' (whether the study design was appropriate to the research question). This particular criterion relies to a considerable degree on subjective judgement. Further training and some additional instructions were provided for raters in the re-evaluation stage, and these are now offered for new reviewers in an attempt to improve consistency of scoring (details available from the authors).

• **Stage 2**

Re-evaluation: The NSF document includes 11 QRs, each with its own cited reference set – some reference articles were quoted in two or more QRs. In the last draft submitted from the R&E group for final editing by the DH NSF development team, the total number of references in the cited reference set had been reduced, and 37 new research articles were added to strengthen evidence in weaker areas. A designated member of the research evaluation team was responsible for final checking and re-evaluation to assure the quality assessment and the relevance of each cited reference to the specific QR context. During this final re-evaluation, the reviewer was not blinded to the initial rating. However, the citations were also screened by an independent assessor who had not been involved in the first round, and any identified inconsistencies or queries were referred back to the reviewer responsible for that QR to re-assess the rating.

In all, 174 research articles which had already been rated in the earlier round were re-evaluated and the quality rating amended where appropriate. This adjustment process led to a change in total quality score for 22 articles (12.6%) – 16 were adjusted downwards leading to 12 reductions in the quality category (high to medium, medium to low etc). Six were adjusted upwards, leading to three elevations of quality category. Table 2 shows the agreement between the ratings given in the 'draft advice' and in the final NSF document, for this set of 174 references. Agreement is generally satisfactory.

The final cited reference set

The breakdown of the final reference set (n = 297) is shown in Table 3. Of the 212 research based references, 80% were in the

Table 3. Distribution of design categories among the research articles cited in the final NSF document (n = 297).

Expert opinion	n	%
E1: Users/carers	7	2
E2: Professionals	62	22
E1+2: Both	16	5
Total Expert opinion	**85**	**29%**
Research-based evidence	**n**	**%**
Primary research		
P1: Quantitative methods	144	49
P2: Qualitative methods	12	4
P3: Mixed methods	14	5
Secondary research		
S1: Meta analysis	3	1
S2: Other secondary analysis	5	2
Reviews		
R1: Systematic reviews	13	4
R2: Other descriptive reviews	21	7
Total Research-based evidence	**212**	**71%**

primary research category, of which 15% were either qualitative or mixed methods research and 85% were quantitative studies. The large majority (183 (87%)) were 'directly applicable' – studies from within the group of people with long-term neurological conditions; 143 (68%) met the criteria for high quality research, 55 (26%) for medium quality and only 13 (6%) were scored as low quality.

Discussion

As with any evidence classification, the new typology designed for this NSF inevitably involves a degree of subjective evaluation, and the authors advocate a larger scale formal validation protocol. The urgent policy context and associated brief timescale and lack of pre-allocated resource for this NSF precluded a full-scale validation of the typology prior to its application, and we recognise this limitation. Nevertheless, the dissemination of this process provides a more rigorous and transparent evaluation than many other NSFs have adopted to date. Further formal evaluation of the tool is required, and we anticipate that it would be refined and the design reconsidered as a result. In the meantime we hope that we have at least promoted consideration of the inherent weaknesses in current systems for appraisal of evidence and their lack of applicability to these conditions. We hope also that this new typology will offer a practical way forward to develop new, progressive and inclusive approaches to the future gathering and appraisal of evidence to support best clinical practice.

Acknowledgements

The authors would like to acknowledge the help of all those involved in collating and reviewing the articles including Barbara Sullivan, Norman Keen, Debbie Romney-Alexander, Tariq Saleem, Valerie Hansen, and Mike Davies. We are grateful also to members of the R&E group who supported the development of the new typology, including Professor Chris Ward, Dr Keith Andrews and Diana Whitworth. Financial support for part of the evaluation was provided by a project grant (No 030-0059) from the Department of Health R&D Funding Programme, and support for preparation of this manuscript was kindly provided by the Luff Foundation.

The views expressed in this paper are those of the authors and do not necessarily represent those of the Department of Health.

References

1 Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; **312**:71–2.
2 Maynard A. Evidence-based medicine: an incomplete method for informing treatment choices. *Lancet* 1997;**349**:126–8.
3 Cochrane AL. *Effectiveness and efficiency. Random reflections on health services.* London: Nuffield Provincial Hospitals Trust (reprinted in 1989 in association with the *BMJ*), 1972.
4 Whyte J. Traumatic brain injury rehabilitation: are there alternatives to randomised controlled clinical trials? *Arch Phys Med Rehabil* 2002; **83**:1320–22.
5 Malterud K. The art and science of clinical knowledge: evidence beyond measures and numbers. *Lancet* 2001;**358**:397–400.
6 Britton A, McKee M, Black N, McPherson K *et al.* Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* 1998;**2**:1–124.
7 McPherson KM, Myers J, Taylor WJ, McNaughton HK, Weatherall M. Self-valuation and societal valuations of health state differ with disease severity in chronic and disabling conditions. *Med Care* 2004;**42**: 1143–51.
8 Carr AJ, Higginson IJ. Measuring quality of life: Are quality of life measures patient centred? *BMJ* 2001;**322**:1357–60.
9 Ahmed S, Mayo NE, Wood-Dauphinee S, Hanley JA, Cohen SR. Response shift influenced estimates of change in health-related quality of life poststroke. *J Clin Epidemiol* 2004;**57**:561–70.
10 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *New Engl J Med* 2000;**342**:1878–86.
11 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New Engl J Med* 2000;**342**:1887–92.
12 Poynard T, Munteanu M, Ratziu V, Benhamou Y *et al.* Truth survival in clinical research: an evidence-based requiem?[see Comment]. *Ann Intern Med* 2002;**136**:888–95.
13 National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal.* London: NICE, 2004.
14 Scottish Intercollegiate Guidelines Network 50. *A guideline developer's handbook.* Edinburgh: SIGN, 2002.
15 Royal College of Physicians. *National clinical guidelines for stroke.* London: RCP, 2002.
16 Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;**323**:334–6.
17 Consort-Group. The consort statement: revised recommendations for improving the quality of reports of parallel group randomised trials. *Lancet* 2001;**357**:1191–94.
18 Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness.* York: Centre for Reviews and Dissemination, 2001.
19 Steultjens EMJ, Dekker J, Bouter LM, van Schaardenburg D *et al. Occupational therapy for rheumatoid arthritis.* The Cochrane Database of Systematic Reviews. Oxford: Update software, 2003.
20 Wells GA, Shea B, O'Connell D, Peterson J *et al.* Newcastle-Ottawa Scale. www.ohri.ca/programs/clinical_epidemiology/nosgen.doc
21 Public Health Resource Unit. *The Critical Appraisal Skills Programme (CASP) and evidence-based practice.* Milton Keynes: Public Health Resource Unit, 2003.
22 Britten N, Campbell R, Pope C, Donovan J *et al.* Using meta ethnography to synthesise qualitative research: a worked example. *J Health Services Res Policy* 2002;**7**:209–15.
23 Campbell R, Pound P, Pope C, Britten N *et al.* Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Soc Sci Med* 2003;**56**:671–84.
24 Dixon-Woods M, Agarwal S, Young B, Jones D, Sutton A. *Integrative approaches to qualitative and quantitative evidence.* London: Health Development Agency, 2004.
25 Pawson R, Boaz A, Grayson L, Long A, Barnes C. *Types and quality of social care knowledge. Stage two: towards the quality assessment of social care knowledge.* London: Social Care Institute for Excellence, 2003.
26 Hawker S, Payne S, Kerr C, Hardy M, Powell J. Appraising the evidence: reviewing disparate data systematically. *Qual Health Res* 2002;**12**: 1284–99.
27 Fleiss JL. *Statistical methods for rates and proportions.* New York: John Wiley, 1981:218.

Appendix 1. Description of the research typology for the NSF

Each piece of evidence is reviewed and given either an 'E' or an 'R' rating:

E: reflects 'expert' (user/carer/professional) evidence

R: reflects 'research-based' evidence.

Expert evidence is that expressed through consultation or consensus processes rather than formal research designs. It could be professional opinion, or that of users and/or carers or other stakeholders.

Research evidence is that gathered through formal research processes. Each piece of research-based evidence is awarded a rating based on three categorisations:

- *Design* – category of research design.
- *Quality* rating – high, medium or low.
- *Applicability* – population context of the study.

Design is classified according to the categories shown in Table 4.

Quality is assessed on the basis of five questions to reach a maximum score of 10 (Table 5).

Table 4. Categories used to classify design.

Primary research-based evidence	
P1	Primary research using quantitative approaches
P2	Primary research using qualitative approaches
P3	Primary research using mixed methods (qualitative and quantitative)
Secondary research-based evidence	
S1	Meta-analysis of existing data analysis
S2	Secondary analysis of existing data.
Review-based evidence	
R1	Systematic reviews of existing research
R2	Descriptive or summary reviews of existing research

Table 5. Five questions used to assess quality.
Each quality item is scored as follows: Yes = 2, In part = 1, No = 0.

		Score
1	Are the research question/aims and design clearly stated?	
2	Is the research design appropriate for the aims and objectives of the research?	
3	Are the methods clearly described?	
4	Is the data adequate to support the authors' interpretations/ conclusions?	
5	Are the results generaliseable?	
	Total	**/10**

High quality research studies are those which score at least 7/10.
Medium quality studies score 4–6/10.
Poor quality studies score 3/10 or less.

Applicability is classified on the basis of context into 'direct' and 'indirect' categories:

Direct: Studies that focus on people with long-term neurological conditions

Indirect: Extrapolated evidence from populations with other conditions.

So, for example:

- a well-conducted qualitative study, scoring 8/10 and demonstrating the benefits of a given intervention in people with multiple sclerosis would be classified as: **P2 High Direct.**
- a post-hoc analysis scoring 5/10 on quality assessment, demonstrating the benefits of palliative care in cancer would be classified as: **S2 Medium Indirect.**

Grade of research evidence

Each individual recommendation is then given an overall 'grade of research evidence' rating of A, B or C based on the quality of all the evidence supporting it and how much of it was directly relevant. The grade of research evidence is rated as shown in Table 6.

Overall weight

The overall weight that can be placed on the available evidence is therefore signposted by an indicator that combines a description of the type of evidence with an overall rating of the quality and applicability of any research-based evidence.

For example:
The finding that 'individuals require prompt diagnosis' might carry the following indicator: **(E, RA).**

This indicates that there is expert opinion to support this statement as well as research of high quality, derived directly within the field of study – suggesting that considerable weight could be placed on the findings of this evidence.

Table 6. Rating of research evidence.

Grade	Criteria
Research Grade A:	• More than one study of high quality score (≥7/10) *and* • At least one of these has direct applicability
Research Grade B:	• One high quality study *or* • More than one medium quality study (4–6/10) *and* • At least one of these has direct applicability *Or* • More than one study of high quality score (≥7/10) of indirect applicability
Research Grade C:	• One medium quality study (4–6/10) *Or* • Lower quality (2–3/10) studies *or* • Indirect studies only