SPECIAL REPORT

WHEN IS A RESEARCH QUESTION NOT A RESEARCH QUESTION?

Nancy E. Mayo, BSc(PT), MSc, PhD[1,2,3], Miho Asano, BSc, MSc, PhD[4] and Skye Pamela Barbic, BSc, BScOT, MSc[1,2,3]

From the [1]Division of Division of Clinical Epidemiology, [2]Research Institute, McGill University Health Center (MUHC), [3]School of Physical and Occupational Therapy and [4]Faculty of Applied Health Sciences, School of Rehabilitation Therapy, Queen's University, Kingston, Canada

Background: Research is undertaken to answer important questions yet often the question is poorly expressed and lacks information on the population, the exposure or intervention, the comparison, and the outcome. An optimal research question sets out what the investigator wants to know, not what the investigator might do, nor what the results of the study might ultimately contribute.
Objective: The purpose of this paper is to estimate the extent to which rehabilitation scientists optimally define their research questions.
Methods: A cross-sectional survey of the rehabilitation research articles published during 2008. Two raters independently rated each question according to pre-specified criteria; a third rater adjudicated all discrepant ratings.
Results: The proportion of the 258 articles with a question formulated as methods or expected contribution and not as what knowledge was being sought was 65%; 30% of questions required reworking. The designs which most often had poorly formulated research questions were randomized trials, cross-sectional and measurement studies.
Conclusion: Formulating the research question is not purely a semantic concern. When the question is poorly formulated, the design, analysis, sample size calculations, and presentation of results may not be optimal. The gap between research and clinical practice could be bridged by a clear, complete, and informative research question.

Key words: research methods; research; rehabilitation.

INTRODUCTION

Research is undertaken to answer important questions, without bias and with precision. The research question sets out what the investigator wants to *know*, not what the investigator might *do* or what the results of the study might ultimately *contribute* to that particular field of science. Thus, two common errors are seen in posing research questions: the question is posed as a method or the question is posed as the expected contribution.

Without a clear question which is specific as to the knowledge the investigator wants to gain, it is impossible to identify who should be included, what the outcomes should be, and when the outcomes need to be measured with respect to the study initiation. Most research aims to answer questions about how two or more variables are related to each other. In most instances, the researcher wants to know if one variable is causally related to a second variable, in other words the research questions is about cause and effect. These variables have different labels depending on the field of study but for this article we will use the epidemiological framework of exposure and outcome to refer to these variables.

To get the question right, the research needs to be clear on the following elements: (*i*) the population; (*ii*) the exposure with its specific levels to be compared; (*iii*) the outcome with its time frame; and (*iv*) the parameter that links the exposure and the outcome. Let us be more specific. The population is that group of persons for whom the knowledge is required. This is often a very specific subset of the total population with the condition under study. The exposure is what the investigator hypothesizes is related to the outcome. The exposure is a variable and must have different levels which could be two (treated/untreated; men/women) or more (no depression/mild depression/severe depression), or could be continuous such as age or a score on a test. A common misconception occurs when a study involves subjects receiving an intervention and the investigators want to know if persons changed on an outcome after participating in the intervention. As all subjects got the intervention, it does not vary and could not be the exposure variable. In fact, the only thing that varied across subjects was time and in this type of pre-post design, time is the exposure. In fact, a unique characteristic of the study sample is that they all participated in the intervention and this needs to be expressed in the population part of the research question. The outcome is what the investigator believes is the most important aspect of a person's physical, psychological or social health that needs to be improved or understood, is expected to change owing to the exposure variable, and is measurable.

Probably the most common error in formulating research questions in being unclear about the parameter that is being used to link the exposure and outcome. The exposure and

outcome can be linked in several ways. The exposure and outcome may be associated but the cause and effect nature cannot be discerned in the study because of the design; this is a feature of a cross-sectional design. The exposure may predict the outcome which can be determined in a longitudinal study. Researchers can ask questions about parameters in which case these questions usually start with "the extent to which" or researchers can ask questions that support hypotheses, i.e., is the outcome for exposure level A better than for exposure level B. The researcher must use a very strong operational verb in the question and avoid verbs such as explore, examine, assess, investigate, understand, describe, collect, gather, etc., as they cannot be used to identify when the question has been answered. When, for example, would the researcher know that there has been enough exploration or describing; when does understanding begin and end; when has enough data been collected or gathered? If the researcher uses phrasing such as "estimate the extent to which high level of depressive symptomatology predicts recovery of function post stroke", it is possible to design a study to determine within a specified degree of certainty the strength of this prediction.

Some types of research questions, particularly those that intend to answer questions about evidence to guide practice are well suited to the format which has become known under the acronym PICO or PICOT to include time (1–3): "*Among people with specific characteristics defining the target Population, does a particular Intervention, in Comparison to a specified alternative intervention, usual care, or placebo, result in altered Outcomes at a specified Time.*" This format can be adapted to exposures that are not interventions (PECOT) such as our example above: "*Among people in the sub-acute phase of stroke (P), does a moderate or high degree of depressive symptomatology (E) in comparison to a low degree (C) predict community participation (O) at 6 months post-stroke (T)?*" The key elements are that the population is specified, the intervention or exposure and the comparison group or levels are given, the outcomes must be measurable, and the time frame specified.

The importance of getting the question right has not been missed in the field of literature as illustrated by this comment by the computer in Douglas Adams' epic work, *The Hitch Hiker's Guide to the Galaxy* (4):

"*'Forty-two!' yelled Loonquawl.*

'*Is that all you've got to show for seven and a half million years' work?*'

'*I checked it very thoroughly*', said the computer, '*and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never really known what the question is… Once you know what the question is, you'll know what the answer means.'*"

This paper is the third in a series dealing with methodological rigour in rehabilitation science. The first two papers (5, 6), dealt with the mislabelling and misunderstanding of the case-control study. A structured literature review identified that of 86 rehabilitation articles labelled as a case-control study by the authors, 83 (97%) were incorrectly classified. One of the reasons why there was such a high rate of misuse

of the case-control design in the rehabilitation literature was that in many instances the authors had not declared what they wanted to know rather they focused on what data they wished to collect. As designs are tools to answer questions (7), it is essential that the research question is correctly specified in order that the researcher chooses the most optimal design. The purpose of this paper is to estimate the extent to which rehabilitation scientists optimally define their research questions in scientific publications.

METHODS

A cross-sectional survey of the published literature in general rehabilitation journals was carried out. All rehabilitation journals listed on the Web of Science for the year 2008 with an impact factor ≥1.0 were identified (10 did not meet this criteria and 6 dealt with specific clinical conditions). Issues for review were selected at random; the number selected depended on the number of issues published per year. Only the research articles were selected for review. Management was done through Reference Manager version 12.

The criteria for evaluating the research question came from the course notes of the senior author (NM) that were based on the PICOT or PECOT framework. These criteria are given in Appendix I. To establish rating consistency, each of the 3 raters reviewed one article and discrepancies in rating were discussed. A second article was then reviewed in the same manner. On the third article, all 3 raters used the criteria consistently.

The two raters (MA, SB) independently rated each question in each article. The senior author (NM) rated 22 articles and agreed with at least one of the other raters on 21 of these articles. Based on these results, rather than having all 2 authors review all articles, the senior author adjudicated all discrepant ratings.

The association between journal and type of study and overall question quality was tested using logistic regression with the reference category being the mean of all categories.

RESULTS

Table I lists the journals selected with the number of articles reviewed which totalled 258. Of the 258 articles, the senior author adjudicated 183 (71%) on one or more element; 92 (36%) of the articles had discrepant ratings on two or more elements. There were 25 (10%) discrepant ratings on the population, 14 (5%) on the exposure or intervention, 77 (30%) on

Table I. *Selection of journals, issues and articles*

Journal	Issues per year	Issues selected	Articles
Am J Phys Med Rehabil	12	3	20
Arch Phys Med Rehabil	12	3	70
Aust J Physiother	4	1	6
Clin Rehabil	6	2	23
Disabil Rehabil	24	4	26
J Orthop Sports Phys Ther	12	3	13
J Rehabil Med	10	2	38
J Rehabil Res Dev	4	1	15
Man Ther	12	1	16
Neurorehabil Neural Repair	4	1	13
Phys Ther	12	3	18
		Total	258

Table II. *Proportion (%) of articles with optimal rating on key question criteria according to journal of publication*

Journal	Articles n	PICOT	Worded as "knowledge"[a] n (%)	Needs no or only minor rewording/revising n (%)
Am J Phys Med Rehabil	20	19/20/9/15/4	10 (50)	17 (85)
Arch Phys Med Rehabil	70	60/70/36/59/15	16 (33)	43 (61)
Aust J Physiother	6	6/6/1/6/1	4 (67)	5 (83)
Clin Rehabil	23	20/20/3/21/2	7 (30)	15 (65)
Disabil Rehabil	26	24/25/8/20/4	11 (52)	20 (77)
J Orthop Sports Phys Ther	13	12/13/5/12/0	5 (38)	10 (77)
J Rehabil Med	38	37/35/19/33/12	18 (47)	26 (68)
J Rehabil Res Dev	15	10/15/8/9/3	12 (13)	9 (60)
Man Ther	16	12/15/16/12/3	6 (37)	13 (81)
Neurorehabil Neural Repair	13	12/11/8/11/1	1 (8)	7 (54)
Phys Ther	18	16/17/11/16/5	10 (56)	15 (83)
ALL	258		90 (35)	180 (70)

[a]Indicates that the question was stated as what the investigators wanted to know ("knowledge") as opposed to what they wanted to do or the ultimate aim of the knowledge gathered. PICOT: Population/Exposure or Intervention/Outcome/Comparison/Time.

the comparison condition, 39 (15%) on the outcome; 42 (16%) on the time; 51 (20%) on the type of question (knowledge, methods, or expected contribution), and 73 (20%) on whether the question needed to be reworked or could stand either as is or with some rewording.

Table II shows the rating of the crucial elements (PICOT) according to journal. The proportion of articles which clearly identified the population ranged from 67% to 100% (mean 88%). Identifying the exposure or intervention was rarely omitted but identifying the comparison occurred only in 24% to 58%. The outcome was not always clear with a range across journals of 60% to 100%. In contrast, the time frame was rarely specified in the question (range 0% to 32%). The proportion of articles that worded the question by stating what the investigators wanted to know ("knowledge") as opposed to stating what they wanted to do ("methods") or the ultimate use for the knowledge gained ("expected contribution") was 35%, with a range across journals from 8% to 77%. We identified that 70% of questions, would have benefited from some minor rewording (e.g. examine to estimate) or revising (include missing PICOT elements) but this meant that 30% of ques-

tions required reworking, with a range across journals from 15% to 46%. There was no association between journal and quality of the questions, however, statistical power was low for these comparisons. The magnitude of the odds of having a question flaw for each journal relative to the odds over all journals (odds ratio, OR) ranged from 0.4 to 2.0 but all of the 95% confidence intervals (CI) included the null value of 1.0.

Table III presents the same results on question quality according to the type of study. The most common study type was a randomized controlled trial (RCT) and the element most often missing from these questions was an indication of the comparison group and time; 63% needed no or only minor rewording or revising and, hence, the remainder (37%) needed to be reworking. Pre-post designs also lacked clarity on the intervention and control elements as in these designs the time (pre to post) represents the control and intervention situations. There were no statistical differences across designs on whether the question was worded as "knowledge" and not as methods or expected contribution but statistical power was low for these comparisons. RCTs (OR: 13.9; 95% CI:1.9–104.3), cross-sectional designs (OR: 9.6; 95% CI: 1.3–73.4), and measurement (psychometric) studies (OR: 23.6; 95% CI: 3.2–173.9) were

Table III. *Proportion of articles with optimal rating on key question criteria according to study design*

Design	Articles n	PICOT	Worded as "knowledge"[a] n (%)	Needs no or only minor rewording/revising n (%)
Deliberate Interventions				
RCT	27	24/27/11/20/4	1 (4)	17 (63)
Cross-over	19	16/19/11/16/3	1 (5)	14 (74)
Pre-post	20	16/20/4/14/2	3 (15)	10 (50)
Controlled/SS	7	5/7/3/3/2/	2 (39)	4 (57)
Longitudinal	61	54/59/29/52/23	33 (52)	44 (72)
Cross-sectional	65	64/65/34/58/10	38 (58)	49 (75)
Measurement	45	36/41/23/39/3	1 (2)	31 (69)
Qualitative	14	13/9/2/12/3	11 (79)	3 (21)
ALL	258		90 (35)	180 (70)

[a]Indicates that the question was stated as what the investigators wanted to know ("knowledge") as opposed to what they wanted to do or the ultimate aim of the knowledge gathered. RCT: randomized controlled trial; SS: single subject. PICOT: Population/Exposure or Intervention/Outcome/Comparison/Time.

Table IV. *Examples of reworked research questions*

Original wording	Reworked question
Randomized Controlled Trial (8): To examine whether the addition of either electroacupuncture (EA) or interferential electrotherapy (IE) to a standard shoulder exercise programme would lead to better clinical outcomes in the management of frozen shoulder.	Among people receiving a standard exercise program for frozen shoulder, to what extent does 4 weeks of EA in comparison to 4 weeks of IE alter the time course of recovery of functional performance and pain.
Pre-post design (9): The primary objective of this study was to examine the impact of a multifaceted falls prevention program including balance exercise and educational components on 2 psychologic factors related to balance: balance confidence and perceived balance (9).	To what extent does balance confidence and perceived balance change following 12 weeks of participation in a multifaceted falls prevention program?
Longitudinal design (10): Abstract – matched pairs of patients with stroke, with and without lateropulsion, were compared for functional outcomes and discharge destination following inpatient rehabilitation. Text – our study was designed to test the null hypothesis that lateropulsion does not affect Functional Independence Measure efficiency or discharge destination (living situation) following inpatient rehabilitation.	Among people with stroke undergoing inpatient rehabilitation, to what extent does deficit in lateropulsion impact on rapidity of gain in lower extremity mobility and ambulation (outcomes known to contribute to global functional recovery and length of stay in rehabilitation)?
Cross-sectional study (11): The aims of this study were: (*i*) to classify subgroups according to the degree of pain intensity, depression, and catastrophizing, and investigate distribution in a group of patients with chronic whiplash-associated disorders; and (*ii*) to investigate how these subgroups were distributed and inter-related multivariately with respect to consequences such as health and quality of life outcome measures.	Among people with chronic whiplash injury, to what extent does differing pain, depression and pain catastrophizing symptom-clusters impact on life satisfaction and health-related quality of life?

significantly more likely to need reworking in comparison to the overall probability; cross-over (OR: 0.4; 95% CI: 0.2–0.7), and qualitative designs (OR: 0.1; 95% CI: 0.04–0.5) were less likely.

Table IV presents examples of questions from different types of study designs that were judged in need of reworking. The first question is from a RCT (8). The key reason why this question needed reworking is that the phrasing of the question did not match the data presented introducing doubt into what the authors really wanted to know. There were also a number of elements that were suboptimal. The word examine is not an operational (when has the investigator done enough examination; what result is expected from examining); the outcomes are not specified, neither is the time frame. While the intervention and comparison contrasts are given, the wording implies either a 2-group comparison: EA+exercise and IE+exercise or a 3-group comparison with the third group being exercise alone. However, the design has 3 groups but the third group has no intervention. In addition, the data presented covers 5 time points (baseline to 6 months), but only for the two treated groups, as the un-treated control group has data only at baseline and at 4 weeks. If the question was reworked based on the data presented including the contrast with no treatment and then the longitudinal contrast with multimodal treatment, it would read something like: "The extent to which a multimodal intervention for frozen shoulder involving exercises and either EA or IE results in greater improvement in functional performance and pain in comparison to no treatment and, secondarily, the extent to which the time course of recovery in functional performance and pain differed between people treated with EA or IE". However, the emphasis in the study and the data analysis seems to point to the longitudinal comparison of EA and IE and hence our research team interpreted this as the author's question. The implication of the reworking of this question is that parameters quantifying recovery are estimated rather than performing a

series of hypotheses tests at each time point. Reworking of the question would lead the analysis towards a regression-based mixed model of group, time, time post-treatment, group × time and group × time post-treatment (12). This would allow for a comparison of both the treatment and maintenance effects between groups, because clearly the authors did not expect the gains made during treatment to continue at the same linear rate of increase post-treatment. The study would also be powered for this type of analysis. This regression approach with two linear slopes is preferable to a mixed model analysis of variance which had to be followed up by 5 separate *t*-tests in order to interpret when difference occurred. The estimates of the effect of group over the two time periods (treatment and maintenance) and their error terms are more useful parameters as it is possible to make treatment and prognostic decisions based on the magnitudes and confidence intervals of the parameters, whereas the *p*-values are uninformative.

The second question is an example of a pre-post design (9). This question has two major flaws and needs reworking. First, it is not possible to examine an impact but it is possible to estimate impact. However, as there is no control group, the impact of the program cannot be discerned. The intent of the study is to contribute evidence for the impact of this type of intervention on the outcomes of interest, albeit evidence from an observational study. The rewording would help focus the analysis to an estimate of change with a 95% confidence interval, which could be used to estimate sample size for a randomized controlled trial.

The third entry is from a longitudinal study (10). Here we present both how the question was presented in the abstract and in the text. The original phrasing in the abstract identifies the methods to be used but not what the authors want to know. The population is identified, people with stroke. The exposure is the presence or absence of lateropulsion. The outcomes

are very general and may or may not be causally linked to the exposure. For example, discharge destination is unlikely to be linked directly to lateropulsion but only through motor and functional outcomes. Outcomes could imply a value at discharge or a change from admission to discharge. The results section of the abstract indicates a change per unit time outcome as in efficiency. In the text, the study aim is worded as: *"Our study was designed to test the null hypothesis that lateropulsion does not affect FIM efficiency or discharge destination (living situation) following inpatient rehabilitation."* This wording has more flaws than the one presented in the abstract, as it is not possible to test a null hypothesis. Research aims to reject the null hypothesis. This wording could imply that the authors wanted to know whether lateropulsion has no impact on outcomes. However, equivalency studies are extremely difficult to conduct because of the issue that low power will render even strong effects, non-significant, or null. In looking at the data presented, it would appear that the authors wanted to know: Among people with stroke undergoing inpatient rehabilitation, to what extent does deficit in lateropulsion impact on rapidity of gain in lower extremity mobility and ambulation (outcomes known to contribute to global functional recovery and length of stay in rehabilitation)? The latter link to more distal outcomes is not needed but would indicate that the main focus would be on those outcomes related to the deficit while at the same time indicating that relationships to other outcomes would also be estimated, to show the relevance of the study question. The danger here is that without a clear statement as to the knowledge desired, the analysis could be perceived as a *post hoc*, searching for outcomes that differed between the two groups of people rather than presenting the pre-specified contrasts.

The fourth and last example is from a cross-sectional study (11). It is not clear from the original wording what the authors want to know by classifying and investigating. In looking at the results presented, one component is prevalence of different pain and depression profiles; but the more important component, and the one that the analysis relates to, is impact of the different pain and depression clusters on health outcomes. The poorly worded question does not do justice to the novel approach to untangling these complex relationships. A better phasing would be: Among people with chronic whiplash injury, to what extent does differing pain, depression and pain catastrophizing symptom clusters impact on life satisfaction and health-related quality of life (HRQL)? The focus is now on knowledge, the impact of symptom clusters on outcomes, not on methods, classify and investigate, and not on the measures, but rather the important outcomes of life-satisfaction and HRQL.

DISCUSSION

The main finding from this review was that a sizeable proportion of the research questions presented in the rehabilitation literature were poorly formulated: 65% did not indicate what the researcher wanted to know and 30% needed to be reworked. The importance of getting the research question right is not just a matter of semantics. It indicates that the researcher is focused on generating

knowledge, and the research is not an exercise in data collection. Many studies were worded to imply that the study's focus was on the data by using such verbs as assess, examine, etc.

When what the researcher want to know is specified in the question, the arguments that need to be made in the background are clear and makes for a more focussed introduction. From a clear question, it is possible to predict all subsequent parts of the research. The measurement strategy, including primary and secondary outcomes, or confirmatory, explanatory and exploratory outcomes (13), are clear and the time frame for these assessments determined. The analysis can be predicted from a clear question as the outcome is specified (and the number of outcomes), the measurement scale of the exposure is specified, and the time frame is specified. This would lead to an appropriate sample size calculation based on the desired effect size and the analysis that matched the question. The latter point is particularly important as with a poorly formulated question, the analyses could be perceived as being post hoc. A clearly formulated question would also lead to a very specific data presentation matching the pre-specified analysis.

Poorly formulated research questions were common across the journals reviewed. However, some research designs were more likely to have poor questions. For example, evaluative and measurement studies were more likely to focus on the methods they wished to use rather than what knowledge was to be gained from the research. Of the RCTs included for review, the objectives were to: assess, determine, describe, compare, establish, evaluate, examine, investigate, quantify, and test the effects of various interventions. RCTs also rarely indicated the comparison group (11/27) or the time frame (4/27) in the question (Table III), both key elements of the PICOT format. Given that the PICOT format was developed for RCTs, it is concerning that rehabilitation scientists have not yet adopted this format.

Questions in qualitative studies were most likely to clearly indicate the knowledge they wished to gain from the research and indicate the population and the outcome (Table III).

How can we improve the situation? First, novice researchers need to be taught the importance of developing a clear question and this will improve over time with an increased focus on evidence-based practice. Second, we think the journals have a responsibility to establish some minimum guidance for what is acceptable as a phrasing of a research question and have editors read each question for conformity. Much like a structured abstract, a structured question could be imposed. This would also greatly facilitate systematic reviews and meta-analyses. Reviewers also have a role in ensuring the research question is acceptable, but if they themselves do not recognize lack of clarity and its consequences for research rigour, it cannot be left solely to the peer review system to ensure excellence. Critiques such as this can only go so far by pointing out areas for improvement. However, rephrasing the question correctly to satisfy a reviewer or an editorial policy may only be window dressing if the study designed and presented does not derive from the question.

In conclusion, getting the research question right is the most important step that a researcher can make in ensuring research rigour. While it may be inferred from reading further on in a

paper what the researcher wanted to know by doing the study, stating this explicitly at the outset will never detract from the science. It is also the most important focus for clinicians reading research reports as they may not be experts in methods or statistics, but they are experts in identifying what they want to know. Perhaps the gap between research and clinical practice could be bridged by a clear, complete, and informative research question.

REFERENCES

1. Birch DW, Eady A, Robertson D, De PS, Tandan V. Users' guide to the surgical literature: how to perform a literature search. Can J Surg 2003; 46: 136–141.
2. Timm DF, Banks DE, McLarty J. Critical appraisal process: step-by-step. South Med J 2012; 105: 144–148.
3. Wilton NK, Slim AM. Application of the principles of evidence-based medicine to patient care. South Med J 2012; 105: 136–143.
4. Adams D. The illustrated hitch hiker's guide to the galaxy. London: Weidenfeld and Nicolson; 1994.
5. Mayo NE, Goldberg MS. When is a case-control study not a case-control study? J Rehabil Med 2009; 41: 209–216.
6. Mayo NE, Goldberg MS. When is a case-control study a case-control study? J Rehabil Med 2009; 41: 217–222.
7. Bailar JC, Mosteller F. Medical Uses of Statistics. 2nd edn. Massachusetts Medical Society; 1992, p. 146.
8. Cheing GL, So EM, Chao CY. Effectiveness of electroacupuncture and interferential electrotherapy in the management of frozen shoulder. J Rehabil Med 2008; 40: 166–170.
9. Filiatrault J, Gauvin L, Richard L, Robitaille Y, Laforest S, Fournier M, et al. Impact of a multifaceted community-based falls prevention program on balance-related psychologic factors. Arch Phys Med Rehabil 2008; 89: 1948–1957.
10. Babyar SR, White H, Shafi N, Reding M. Outcomes with stroke and lateropulsion: a case-matched controlled study. Neurorehabil Neural Repair 2008; 22: 415–423.
11. Borsbo B, Peolsson M, Gerdle B. Catastrophizing, depression, and pain: correlation with and influence on quality of life and health – a study of chronic whiplash-associated disorders. J Rehabil Med 2008; 40: 562–569.
12. Singer JD, Willett JB. Applied Longitudinal Data Analysis Modeling Change and Event Occurrence. New York: Oxford University Press; 2003.
13. Fairclough DL. Summary measures and statistics for comparison of quality of life in a clinical trial of cancer therapy. Stat Med 1997; 16: 1197–1209.

APPENDIX I. Criteria for judging adequacy of the research questions

Criteria	Response options
What is the original question as stated?	1=abstract only
Where was the question found?	1=abstract only
	2=introduction/background only
	3=abstract and introduction/background
	4=elsewhere/5=not found
Is question is stated in one sentence?	yes/no
Is there ONE question?	yes/no (refers to one overaching question; if there are secondary questions, the same design would be used)
Is there a discrepancy between question in abstract and background?	yes/no
Formulation of question	1=knowledge (parameter to be estimated or hypothesis to be tested is stated)
	2=methods (what the author wants to do – explore, assess, examine, correlate, compare, collect data are examples of methods)
	3=expected contribution (understand, elucidate or other terms indicating that after the question has been answered it will contribute to a more global cause)
Population, exposure/intervention, comparison, outcome, time	1=fully defined
	2=partially defined
	3=not defined clearly
Primary/target outcome defined	yes/no/not clear
Question type	1=hypothesis generation
	2=hypothesis testing
	3=parameter estimation
	4–5=mixed
	6=other
Suggestions for improvement	1=adequate as written
	2=needs rewording: one or two words needed to be changed such as investigate to estimate)
	3=needs revising: one or more PICOT elements missing which were crucial for question clarity
	4=needs reworking: the question or objective was absent, or did not reflect the study design or data presented