Can English-speaking reviewers correctly identify foreign-language articles that meet eligibility criteria for a systematic review of management for fibromyalgia?

Busse JW[1], Bruno P[2], Malik K[1], Connell G[3], Torrance D[3], Ngo T[3], Kirmayr K[4], Avrahami D[3], Riva JJ[1], Ebrahim S[1], Struijs P[5], Brunarski D[6], Burnie S[3], LeBlanc F[3], Coomes EA[1], Steenstra IA[7], Montori V[8], Guyatt GH[1]

[1]McMaster University, Canada; [2]University of Regina, Canada; [3]Canadian Memorial Chiropractic College, Canada; [4] German Hospital, Argentina; [5]Academic Medical Center, The Netherlands; [6]Ontario Chiropractic Association, Canada; [7]The Institute for Work & Health, Canada; [8]Mayo Clinic, USA

Introduction

Systematic reviews endeavour to capture all publications that meet pre-defined eligibility criteria. Non-English studies may present resource challenges in meeting this goal. If English-speaking reviewers could differentiate eligible from ineligible foreign language publications, it would limit demands for participation in the review by those speaking other languages.

The purpose of our study was to explore whether English-speaking reviewers can differentiate eligible from ineligible foreign-language studies in a systematic review of all treatments for fibromyalgia.

Methods

Eligibility criteria
Eligible studies randomly assigned patients with fibromyalgia to any form of therapy or a control group.
Information sources
Relevant randomized controlled trials (RCTs), in any language, by a systematic search of AMED, CINAHL,EMBASE, MEDLINE, HealthSTAR, PsycINFO, Papers First, Proceedings First and CENTRAL, from inception of each database to April, 2011.
Study Appraisal
Two reviewers, independently and in duplicate, completed title and abstract screening, and full text review. All non-English were screened by pairs of reviewers fluent in the language of publication, and again by pairs of English-speaking reviewers.

Review of foreign language articles by English-speaking reviewers was guided by a 10-question form designed to detect features associated with RCTs (*Table 1*).
Synthesis of results
We calculated the agreement between native-language and English-speaking reviewers for eligibility of foreign-language studies using the following statistics: raw agreement and chance-corrected agreement (Kappa). We interpreted the kappa statistic results using the guidelines proposed by Landis and Koch: values of 0 to 0.20 represent slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and greater than 0.80 almost perfect agreement.

We suspected that clearly reported CONSORT diagrams, tables comparing baseline characteristics, or forest plots and search strategies would be rare among foreign-language papers. For the remaining screener questions with the potential for a directional effect, we hypothesized, *a priori*, that mention of the word "random", greater than 2 authors, length of more than 2 pages, and

reporting of statistical analysis would be associated with a greater likelihood of a study being eligible for our review. We entered these variables into a generalized linear model.

Table 1. Screening questions for English reviewers

Question	Anticipated effect
1. Does the title suggest eligibility?	n/a
2. Does the abstract suggest eligibility?	n/a
3. Do the methods suggest eligibility?	n/a
4. Does the word "random" appear?	If so, more likely to be an RCT
5. How many authors?	More than 2 authors more likely to be an RCT
6. What is the page length?	Greater than 2 pages more likely to be an RCT
7. CONSORT diagram reported?	If so, more likely to be an RCT
8. Table of baseline data reported?	If so, more likely to be an RCT
9. Statistical analysis reported?	If so, more likely to be an RCT
10. Forest plots or literature search results reported?	If so, less likely to be an RCT

Results

Figure 1. Study eligibility flow
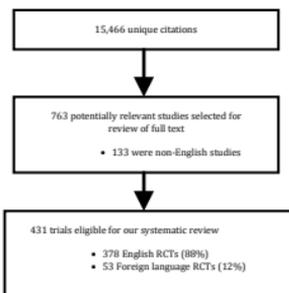
Raw agreement between English and native-language reviewers for assessment of eligibility of the 133 foreign language articles was 89%. The change-corrected agreement was substantial (kappa=0.77).

63 of 133 (47%) of study titles were in English, but the majority (87%) failed to convey sufficient information to establish eligibility. 74 articles provided an English abstract, which allowed for determination of eligibility in 62% of cases. 95% of studies either lacked a methods section (n=33), or English-speaking reviewers were unable to determine eligibility due to a language barrier (n=93).

The word "random" clearly appeared in 56 articles (42%), 72 of the 133 articles were authored by more than 2 investigators (54%), and 32 (24%) were ≤2 pages in length.

Only 6 studies (5%) presented a CONSORT flow diagram, and only 16 (12%) clearly reported a table of baseline characteristics between 2 groups. 79 articles (59%) clearly presented a statistical analysis.

The 53 foreign-language articles that were eligible for our review

represented 19 different languages (Table 2).

Table 2. Language of publication for foreign trials

Language of publication	English-speaking screeners Not Used		Use of English-speaking screeners	
	No. of studies requiring native speakers	No. of eligible studies identified	No. of studies requiring native speakers	No. of eligible studies identified
German	56	16	15	14
Italian	12	5	6	5
French	11	4	3	3
Turkish	10	9	9	8
Chinese	8	7	7	6
Spanish	7	5	5	5
Dutch	6	1	1	1
Portuguese	5	3	4	3
Swedish	4	0	0	n/a*
Russian	3	1	2	1
Danish	2	0	1	0
Norwegian	2	0	0	n/a*
Croatian	1	0	1	0
Hebrew	1	0	0	n/a*
Hungarian	1	0	0	n/a*
Korean	1	1	0	0
Polish	1	1	1	1
Romanian	1	0	0	n/a*
Serbian	1	0	1	0
TOTAL	**133**	**53**	**55**	**47**

* No articles in this language required review by native speakers as they were ruled out by English-speaking reviewers

Of the 19 languages represented in the 133 foreign papers reviewed, 8 were not among eligible trials. However, our English reviewers incorrectly identified 3 of the 8 as eligible (Danish, Croatian and Serbian). As such, use of English-language reviewers to screen foreign language papers would have resulted in 5 fewer teams of foreign-language reviewers (Swedish, Norwegian, Hebrew, Hungarian and Romanian) required for data abstraction. Our English-reviewers did incorrectly exclude 6 of 53 eligible articles (11%). Alternately, If we rejected all languages with ≤2 potential articles, we would eliminate 9 teams at a cost of 2 missed eligible articles.

In our adjusted regression analysis, reporting the word "random" (odds ratio=5.4, 95% confidence interval= 2.7 to 10.8) and clear reporting of statistical analyses (odds ratio=4.6, 95% confidence interval= 2.1 to 10.0) were independently associated with eligibility for our review.

Discussion

Use of English language screeners appears to be inferior to a more simple rule excluding languages with only one or two articles.