Original Article

Searching the literature for information on traumatic spinal cord injury: the usefulness of abstracts

MPJM Dijkers*,¹

¹Department of Rehabilitation Medicine, Mount Sinai School of Medicine, New York, NY, USA

Study design: Systematic review of abstracts of published papers presumed to contain information on chronic pain in persons with spinal cord injury (SCI).

Objectives: To determine to what degree papers on SCI are abstracted in such a way that they can be retrieved, and evaluated as to the paper's applicability to a reader's questions.

Setting: US - academic department of rehabilitation medicine.

Methods: 868 abstracts published in Medline were independently examined by two out of 13 screeners, who answered four questions on the subjects and nature of the paper with 'Yes', 'No' or 'insufficient information'. Frequency of ratings 'insufficient information', and screener agreement were evaluated as affected by screener and abstract/paper characteristics.

Results: Screeners could not determine whether the paper dealt with persons with traumatic SCI for 37% of abstracts; whether chronic pain was a topic could not be determined in 18%. Physicians were less willing than other disciplines to assign 'insufficient information'. Screener agreement was better than chance, but not at the level suggested for quality measurement. Screener discipline and task experience did not make a difference, nor did abstract length, structure, or decade of publication of the paper.

Conclusion: Authors need to improve the quality of abstracts to make retrieval and screening of relevant papers more effective and efficient.

Sponsorship: National Institute on Disability and Rehabilitation Research.

Spinal Cord (2003) 41, 76–84. doi:10.1038/sj.sc.3101414

Keywords: abstracting and indexing; publications; bibliometrics; spinal cord injury; pain

Introduction

Medical journals started printing abstracts to accompany research and other papers in the 1950s. JAMA published the first ones in 1956¹ and others followed suit. These abstracts replaced or supplemented the 'summary' that prior to that time ended many papers. The purpose of the abstract, formatted as one or more free-text paragraphs and printed at the very front, was to give the reader some sense of direction as to what was to follow. Abstracts should (*and sometimes do*) give readers enough information to assess if the paper is relevant to their interests, offers information that might affect the way they practice medicine, and (if a report of empirical research) uses methods that meet basic standards of validity.²

These traditional narrative abstracts have been criticized for a number of reasons, including conclusions that do not follow from the study

findings and an overemphasis on positive conclusions.^{3–5} They often did not convey the information readers would expect.⁶ In the late 1980s proposals for 'more informative abstracts' were made,⁷ which eventually resulted in what are now known as structured abstracts, which first appeared in the *Annals of Internal Medicine* in 1987. Many prestigious journals followed this lead, including *Spinal Cord*.⁸ It was expected that structured abstracts would provide more and more relevant information, and allow readers to quickly assess reliability and applicability of clinical reports. In addition, structured abstracts were expected to facilitate peer review of manuscripts, and help in the accurate indexing and retrieval of reports.^{7,9,10} Structured abstracts have indeed been found to be more informative,^{11–13} to a degree this is due to the fact that they are longer (on average) than unstructured ones.^{11,14} However, they also may be more readable,^{11,15} and produce more indexing terms, which makes it easier to retrieve them from electronic data bases such as Medline and Embase.¹⁴ Only a few journals have not hopped on

*Correspondence: M Dijkers, Department of Rehabilitation Medicine, Box 1240, Mount Sinai School of Medicine, One Gustave Levy Place, New York, NY 10029-6574, USA;
E-mail: marcel.dijkers@mssm.edu

the bandwagon, citing issues of restricting creativity of authors.¹⁶

Structured abstracts are not without shortcomings, however.^{12,17} While studies that compared them to non-structured abstracts declared them improved,^{12,14} many lack a description of the study population,^{10,17,18} or are missing other information considered essential.¹³ Others do not evidence correspondence between the abstract and the paper, in that the data reported are different, or there are data in the abstract entirely lacking from the paper.^{2,19}

However, informative and correct abstracts are of extreme importance: they are the component of papers most likely to be read,^{1,19,20} and sometimes they are the only part read.^{21,22} Not infrequently, clinicians make decisions based on nothing more than a perusal of the abstract.²³ The purpose of this paper is to evaluate to what degree papers on persons with spinal cord injury (SCI) contain informative abstracts, that is abstracts that can be successfully retrieved and scanned as to their applicability for particular questions. The effect of certain abstract and abstract reader characteristics on judgments of informativeness is also evaluated.

Method

The data presented here are a 'by-product' of an ongoing project to conduct meta-analyses of the literature on chronic pain in persons with traumatic SCI. The purpose of the meta-analyses is to quantitatively summarize information on the incidence, prevalence and risk factors for chronic pain; assessment and diagnosis of chronic pain; impact of pain on functioning and quality of life; and chronic pain treatment outcomes, side effects, and costs. In order to perform these analyses, we needed to identify published and unpublished papers on any of these topics, based on original data reported for individuals with traumatic spinal cord injury. We sought potential papers in various ways, including systematic searches of Medline, CINAHL, Embase, PsychInfo, the Cochrane collaboration and other databases. The search criteria for Medline were the Medical Subject Headings (MeSH) terms spinal cord injury, paraplegia, quadriplegia, tetraplegia, pain (all exploded), and the parallel text words, as well as such words as ache, spinal lesion, etc. The search terms for other databases were similar. In addition, we performed hand searches of journals most likely to contain pertinent information (*Spinal Cord*, *Journal of Spinal Cord Medicine*, etc.), and of various co-investigators' own files. The reference lists of all papers, books and other documents identified in this way were scrutinized (ancestry search), and likely publications added to the list of references to be reviewed. We also performed a descendant search, searching the Web of Science database for any 'offspring' of (key) papers: later articles that quoted one of more of the papers identified as appropriate.

This method was abandoned fairly quickly as not cost-efficient.

The number of titles mushroomed quickly to over 2000. The main reason was that the Medline MeSH term 'spinal cord injury' retrieves papers on low back pain, other diseases affecting the spinal column or cord without causing paralysis, and non-traumatic causes of paralysis such as cerebral palsy, metastatic cancer, etc. In fact, the term 'spinal cord injury' is found at three places in the MeSH tree: as a subcategory of 'spinal cord diseases'; as a subcategory of 'trauma, nervous system'; and once more, as a grouping under 'wounds and injuries'. A second reason is that often papers are referenced in an article without it being clear what specific information they contain; in order not to miss any relevant information, we included rather than excluded these references. Obtaining copies of more than 2000 papers and books would have been excessively time-consuming and expensive. Because for many papers (especially the more recent ones) abstracts are available in the databases, we decided to use abstract-based screening to identify those papers that were potentially appropriate for the meta-analyses. Abstracts were printed, and provided to two screeners who completed a screening form guided by a 3-page syllabus.

The screeners were 13 individuals, including four rehabilitation physicians, three physical therapists, four occupational therapists, and two others. All had at least several years of clinical and/or research experience with SCI, and were familiar with the literature in their field of expertise. The screeners were to answer four key questions:

- Does the paper abstracted deal with humans (rather than animals, SCI units, etc.)?
- Are these individuals with traumatic SCI (rather than spinal injury from other origin, or other, non-spinal, disorders)?
- Is information provided on chronic pain experienced by these persons?
- Does the paper report original data (rather than constituting a review, theoretical treatise, etc.)?

For each question, the reviewers could answer 'Yes', 'No' or 'Insufficient information in the abstract'. 'Insufficient information' indicates lack of informativeness of abstracts, and possible lack of perceptiveness of the screener. Screeners were instructed to err on the conservative side, and mark '?' (insufficient information) rather than 'No' if there was doubt in their mind. Once the four questions had been answered, assignment of the paper was automatic: four times 'Yes' meant it was to be included in the meta-analysis; one 'No' excluded it, and a combination of 'yes' and '?' answers meant that a decision could not be based on the abstract alone.

We decided to have each abstract screened by two independent reviewers, in order to minimize changes that a paper would be selected out even though it was appropriate, or at least should be classified as such

based on the abstract information. Only if *both* reviewers declared a paper not relevant, and for the *same* reason, was it discarded. Disagreement between the two independent screeners reflects lack of precision of the abstract language, or possibly screener shortcomings or different tendencies to interpret terms in a certain way. An answer of 'Yes' on all four questions by *both* screeners was required for a paper to be a candidate for the meta-analyses. An answer of 'No' by both for the *same* question meant that the paper was no longer considered. Any other combination of answers necessitated obtaining the full text of the paper in order to answer the questions based on complete information.

In addition to the two screeners' judgments on the four key items, the following information was collected on each paper or abstract:

- Journal in which published
- Year of publication
- Original language of paper (classified as English or any other; all *abstracts* were in English, whatever the language of the paper)
- Whether the abstract was structured or not
- Whether the abstract as published in the database was truncated. (Medline truncates non-structured abstracts at 250 (if the paper is less than 10 pages) or 400 words, but never truncates structured abstracts)
- Abstract length in lines

Because of the latter, this analysis is limited to abstracts that were available in Medline, whatever the method or source that originally generated the paper as a possible data source for meta-analysis. The standard Medline Text printout, which provides the entire abstract as a single continuous paragraph (whether or not the original published in the journal had a different format), allows for easy counting of the number of lines. Partial last lines were counted as one. The average Medline Text line has 10.7 words (as counted by the 'Word count' utility in Microsoft Word), so a short 4-line abstract has about 42 words, while a lengthy 50-line one has in excess of 500.

The analysis focuses on the informativeness of abstracts, specifically the ability of screeners to determine the answer to the four questions, and the factors (year of publication, abstract length) that affect informativeness. Agreement between screeners is also analyzed, and factors affecting agreement. The percentage of time screeners agreed is reported, as well as weighted and unweighted kappa. Kappa quantifies agreement adjusted for chance agreement between judges. Weighted kappa does the same, but gives partial credit for similar (but not identical) answers. For instance, 'Yes' and '??' are not the same, but are somewhat similar, compared to 'Yes' and 'No' answers. Kappa and weighted kappa range from 0.00 (agreement is not better than chance) to 1.00 (perfect agreement).

Statistical testing is used, but readers should be aware of the limitations of the generalizations that seem possible. For instance, it is not known whether any conclusions can be generalized to non-Medline journals, or diagnostic groups other than traumatic SCI. However, it is likely that our findings would be applicable to a search for literature on eg SCI and spasticity.

Results

The total number of abstracts included in this analysis is 868, covering the years 1975 (the first year for which Medline started printing abstracts, rather than just titles) through 2001. The number increased from 60 for the 1975–79 5-year period to 322 for the 1995–99 period. The year 2000 produced 69, and 2001 (an incomplete year) 10. These abstracts had been published in 286 different journals, indicating the range of journals publishing on SCI or closely related topics (Table 1). Most journals produced only one or two papers on (presumed) SCI pain during the entire period (or during the time they were published or covered in Medline), and *Paraplegia/Spinal Cord* was the only one with over 60 papers. Papers were concentrated, in that 11 journals (3.8% of the total) published 40.8% of the abstracts; Table 1 provides further information on the concentration in a limited number of journals of papers presumed to be on chronic pain in SCI.

Most papers (741 or 85.4%) were in English. Only a few (21 or 2.4%) abstracts had been truncated because of excessive length. A total of 117 (13.5%) had a structured format, all published in 1991 or later years. While a few papers had three-line abstracts (5 or 0.6%), most had more, with the longest having 51 lines. The average was 15.9, with a standard deviation of 7.5. The average for truncated abstracts was 25.7 (std dev 7.9), and the average for structured abstracts was 23.2 (std dev 7.3). English-language papers had abstracts that were marginally longer (16.1, with std dev of 7.5) than foreign-language papers (14.6 ± 7.2).

The screeners each reviewed between 36 and 311 abstracts. Assignments came in batches of about 50, and there was no selection to match abstract topics or characteristics to screener. The pairing of screeners assigned to the same abstract was more or less random. Table 2 reflects the judgments of individual reviewers, and those of the disciplines and the group as a whole. (Because each abstract was screened independently by two screeners, the number of judgments is twice as large as the number of abstracts). It is clear that all screeners had trouble making a final decision on the four characteristics to be judged, resulting in a large number of 'insufficient information' judgments. A final assignment of a paper depended on a 'Yes' answer on all four issues, or a 'No' on at least one of them; all other abstracts were put in the category 'indeterminate' (??) for the overall decision.

Table 1 Number of papers and number of journals included in the analysis

Journal name and category	Journals			Papers		
	Number	%	Cum. %	Number	%	Cum. %
Paraplegia/Spinal Cord	1	0.3	0.3	85	9.8	9.8
Arch Phys Med Rehabil	1	0.3	0.7	55	6.3	16.1
Spine	1	0.3	1.0	44	5.1	21.2
J Neurosurg	1	0.3	1.4	35	4.0	25.2
Neurosurgery	1	0.3	1.7	32	3.7	28.9
Pain	1	0.3	2.1	22	2.5	31.5
Am J Phys Med Rehabil/Am J Phys Med	1	0.3	2.4	19	2.2	33.6
Acta Neurochir (Wien) (incl. supplement)	1	0.3	2.8	18	2.1	35.7
No Shinkei Geka	1	0.3	3.1	16	1.8	37.6
Clin Orthop	1	0.3	3.5	15	1.7	39.3
J Spinal Disorders	1	0.3	3.8	13	1.5	40.8
Subtotal: Journals with 10 or more papers	11	3.8	3.8	354	40.8	40.8
Journals with 9 papers	2	0.7	4.5	18	2.1	42.9
Journals with 8 papers	5	1.7	6.3	40	4.6	47.5
Journals with 7 papers	4	1.4	7.7	28	3.2	50.7
Journals with 6 papers	3	1.0	8.7	18	2.1	52.8
Journals with 5 papers	7	2.4	11.2	35	4.0	56.8
Journals with 4 papers	13	4.5	15.7	52	6.0	62.8
Journals with 3 papers	16	5.6	21.3	48	5.5	68.3
Journals with 2 papers	50	17.5	38.8	100	11.5	79.8
Journals with 1 paper	175	61.2	100.0	175	20.2	100.0
Total	286	100.0	—	868	100.0	—

Table 2 Decision on each of four criteria and on assignment of paper, by screener^a

Discipline and screener	Papers screened		Human?			Traumatic SCI?			Chronic pain?			Original data?			Assignment ^b		
	Number	%	Yes %	No %	? ^c %	Yes %	No %	? ^c %	Yes %	No %	? ^c %	Yes %	No %	? ^c %	Incl. %	Excl. %	?? %
Rehab. medicine 1	112	6.5	98	1	1	21	55	24	53	33	14	85	11	5	9	76	15
Rehab. medicine 2	93	5.4	82	18	0	31	43	26	41	51	5	77	20	2	15	65	20
Rehab. medicine 3	100	5.8	88	3	9	50	25	25	86	9	5	73	17	10	34	47	19
Rehab. medicine 4	99	5.7	100	0	0	54	33	13	42	47	11	79	14	7	30	59	11
Physical therapy 1	201	11.6	95	4	1	20	56	24	62	27	11	58	10	32	6	77	17
Physical therapy 2	36	2.1	100	0	0	28	6	67	78	3	19	75	14	11	22	22	56
Physical therapy 3	89	5.1	99	0	1	11	34	55	62	12	26	88	6	7	8	46	46
Occup. therapy 1	131	7.5	89	6	5	33	29	38	25	44	31	41	10	50	5	60	34
Occup. therapy 2	148	8.5	96	1	3	22	30	49	36	35	30	48	24	28	2	73	25
Occup. therapy 3	148	8.5	95	2	3	14	46	40	47	41	12	75	18	7	3	81	15
Occup. therapy 4	133	7.7	95	2	3	11	38	51	19	29	52	69	26	5	2	74	23
Other 1	135	7.8	80	8	12	17	45	38	30	60	10	39	50	11	1	81	19
Other 2	311	17.9	94	3	3	21	37	43	67	20	14	81	16	4	10	60	30
Rehab. medicine	404	23.3	92	5	3	39	39	22	56	34	9	79	15	6	22	62	16
Physical therapy	326	18.8	97	3	1	18	44	37	64	20	16	68	9	23	8	62	29
Occup. therapy	560	32.3	94	3	3	20	36	45	32	37	30	58	20	22	3	73	24
Other	446	25.7	90	5	6	20	39	41	55	32	13	68	26	6	7	66	27
Total	1376 ^d	100.0	93	4	3	24	39	37	50	32	18	68	18	14	9	67	24

^aEntries in bold indicate a statistically significant difference in percentage 'insufficient information' (versus 'Yes' and 'No' combined) between the four disciplines (chi-square test, alpha level 0.05). ^bInsufficient information to make a decision. ^cIncl.: include in meta analysis. Excl.: exclude from meta analysis. ??: full paper is needed to make a decision. ^dThe number of papers is twice that in Table 1, because each abstract was screened by two persons. See text

The discipline subtotals indicate that physicians were more likely to assign a 'Yes' or a 'No' answer. Within each disciplinary group, there were screeners

whose decisions seemed to differ from those of the others (eg physician 3 for 'human or not'; physical therapist 2 for 'traumatic SCI or not'); however, it

should be remembered that all screeners did not receive the same package of abstracts, and that the ones assigned were not randomly sampled from the pool. Especially for screeners reviewing only a small number of abstracts (eg physical therapist 2), 'aberrant' profiles are quite possible. Experience in the task did not seem to make much of a difference in the ability of willingness to make a decision whether an abstract satisfied a particular criterion: for all screeners combined, the correlation between number of abstracts screened previously and assignment of a 'not enough information' was very weak and positive for the issues of yes/no traumatic SCI, yes/no chronic pain, and final assignment (0.08 or 0.09), but negative for yes/no human (-0.08), and only 0.03 for yes/no original. Two screeners improved in making all four decisions, and one deteriorated in all four. Logistic regression analysis confirmed these findings.

Several characteristics of the papers included and their abstracts do not appear to systematically affect informativeness: year of publication and language of the paper, the length of the abstract, and whether the abstract was structured or not (Table 3). For every finding indicating that a particular characteristic makes decisions more feasible (eg length of abstract and yes/no original data decision), there is an opposing trend (eg length and yes/no traumatic SCI). These characteristics are not independent from one another: over time, abstracts have grown in length; structured abstracts are longer (and never truncated),

etc. Logistic regression suggested no findings other than the bivariate relationships shown in Table 3.

The agreement between the screeners on the four decisions is displayed in Table 4. (The assignment of the paper, as determined by the four decisions, is shown too, but it should be remembered that two screeners for a particular abstract may both have decided on a 'not suitable for the meta analysis', but for *different* reasons. For instance, screener 1 may have decided that no chronic pain data were reported, [while screener 2 thought there were such data, or was not sure], and screener 2 decided that there were no original data [while screener 1 was the opposite opinion]). Unweighted kappas are around 0.30; weighted kappas (which give partial credit in cases where one screener decides 'Yes' or 'No', the other '?') is slightly higher. However, because of the implications of the decisions (our goal was to select papers, not to rate them on an ordinal scale), the unweighted kappas are presumably the more useful indicators of the agreement between screeners. The data in Table 5 indicate that characteristics of the papers or abstracts do not appear to affect the ability of screeners to agree in their judgments. Again, whatever trends there appear to be (eg number of lines and 'human?') are contradicted by findings for other decisions. This is not surprising given the results in Table 3, which indicated that paper/abstract attributes do not affect the ability of individual screeners to exclude 'insufficient information to make a decision'.

Table 3 Decision on each of four criteria and on assignment of paper, by paper/abstract characteristics^a

Screener	Papers screened		Human?			Traumatic SCI?			Chronic pain?			Original data?			Assignment ^b		
	Number	%	Yes %	No %	? ^c %	Yes %	No %	? ^c %	Yes %	No %	? ^c %	Yes %	No %	? ^c %	Incl. %	Excl. %	?? %
<i>Abstract structured</i>																	
Yes	1496	86.2	95	3	2	26	37	37	45	40	15	78	11	10	10	68	22
No	240	13.8	93	4	4	23	39	37	50	31	19	66	20	15	9	66	24
<i>Language of paper</i>																	
English	1482	85.4	93	4	3	25	37	38	50	33	17	69	18	13	10	66	24
Other	254	14.6	94	2	4	17	54	29	48	27	25	61	19	20	4	73	23
<i>Year of publication</i>																	
1975–79	120	6.9	93	1	6	23	40	38	56	24	20	60	17	23	8	65	27
1980–84	174	10.0	91	4	5	23	40	37	45	34	21	60	18	22	6	69	25
1985–89	244	14.1	98	1	1	25	37	38	52	31	17	76	11	13	12	59	29
1990–94	396	22.8	93	4	3	22	40	39	46	37	17	65	19	16	7	69	24
1995–99	644	37.1	92	4	4	25	38	37	51	31	18	70	20	10	11	66	23
2000–02	158	9.1	90	7	3	26	42	32	51	30	19	65	23	12	8	77	15
<i>Length of abstract</i>																	
<10 lines	370	21.3	92	3	5	18	46	37	50	31	19	55	24	21	5	75	21
10–19 lines	876	50.5	94	3	3	26	37	37	50	32	18	69	17	15	11	64	26
20–29 lines	388	22.4	90	6	4	26	37	38	49	31	20	72	19	9	13	63	24
30+ lines	102	5.9	96	3	1	21	40	39	47	42	11	84	10	6	5	75	21
Total	1736 ^d	100.0	93	4	3	24	39	37	50	32	18	68	18	14	9	67	24

^aEntries in bold indicate a statistically significant difference in percentage 'insufficient information' (versus 'Yes' and 'No' combined) between categories of the characteristic shown (chi-square test, alpha level 0.05). ^bIncl.: include in meta analysis. Excl.: exclude from meta analysis. ??: full paper is needed to make a decision. ^cInsufficient information to make a decision. ^dThe number of papers is twice that in Table 1, because each abstract was screened by two persons. See text

Table 4 Agreement between screeners on each of four criteria and on assignment of paper

<i>Screener 1 decision</i>	<i>Screener 2 decision</i>				<i>Consensus measure</i>		
	<i>Yes</i>	<i>?</i>	<i>No</i>	<i>Total</i>	<i>Agreement</i>	<i>Kappa unweighted</i>	<i>Kappa weighted</i>
<i>Human?</i>							
Yes	773	22	19	814	90%	0.28	0.35
?	15	3	2	20			
No	12	13	9	34			
Total	800	38	30	868			
<i>Traumatic SCI?</i>							
Yes	101	78	19	198	56%	0.33	0.43
?	95	152	79	326			
No	19	88	237	344			
Total	215	318	335	868			
<i>Chronic pain?</i>							
Yes	293	46	68	407	55%	0.27	0.33
?	77	39	70	186			
No	86	45	144	275			
Total	456	130	282	868			
<i>Original data?</i>							
Yes	465	92	63	620	67%	0.33	0.41
?	55	27	23	105			
No	31	23	89	143			
Total	551	142	175	868			
<i>Assignment^a</i>	Incl.	??	Excl.	Total			
Include	28	37	15	80	69%	0.37	0.45
Need full paper (??)	43	88	82	213			
Exclude	13	78	484	575			
Total	84	203	581	868			

^aIncl.: include in meta analysis. Excl.: exclude from meta analysis. ??: full paper is needed to make a decision

Table 5 Agreement between screeners (unweighted kappa) on each of four criteria and on assignment of paper, by characteristics of the paper/abstract

<i>Decision</i>	<i>Abstract structured</i>		<i>Language of paper</i>		<i>Year of publication</i>							<i>Number of lines of abstract</i>			
	<i>All</i>	<i>Yes</i>	<i>No</i>	<i>English</i>	<i>All other</i>	<i>75–79</i>	<i>80–84</i>	<i>85–89</i>	<i>90–94</i>	<i>95–99</i>	<i>00–01</i>	<i>1–9</i>	<i>10–19</i>	<i>20–29</i>	<i>30+</i>
Human?	0.28	0.34	0.28	0.30	0.15	–0.06	0.30	0.19	0.30	0.23	0.59	0.13	0.20	0.49	0.48
Traumatic SCI?	0.33	0.34	0.33	0.33	0.32	0.29	0.51	0.35	0.33	0.29	0.38	0.33	0.37	0.28	0.23
Chronic pain?	0.27	0.35	0.26	0.29	0.13	0.16	0.29	0.25	0.35	0.24	0.26	0.28	0.26	0.26	0.30
Original data?	0.33	0.27	0.33	0.33	0.32	0.32	0.20	0.31	0.30	0.37	0.41	0.31	0.35	0.30	0.16
Assignment	0.37	0.40	0.37	0.36	0.42	0.34	0.43	0.28	0.38	0.38	0.38	0.39	0.37	0.38	0.17
Number of papers	868	120	748	742	126	60	87	122	198	322	79	185	438	194	51

Discussion

Abstracts are provided in order to allow readers to quickly scan papers for applicability to their own situation. For clinical applications where the clinician has a treatment dilemma, the abstract should offer the following information: nature of the patients in the study reported, the specific problem(s) treated; the treatment provided, the results of that treatment, and recommendations. For other applications (eg preparation of a research project) similar information is needed. Our data suggest that the abstracts published

in the spinal cord injury literature (or the broader literature of papers that potentially deal with spinal cord injury) fail to provide the information necessary to perform such screening efficiently. It seems that the introduction of longer abstracts and of structured abstracts has not made a difference, as the data in Table 3 indicate. All of the improvements in research reporting since the early 1970s taken together appear to have increased the informativeness of abstracts minimally if at all.

As indicated in Table 2, often it is not feasible to determine from the abstract whether the subjects

involved were persons with traumatic SCI, and whether data on chronic pain were reported. Given the fact that all 13 screeners had this problem to more or less the same degree, it is unlikely that this can be blamed on the performance of the screeners. There was a tendency for physicians to be more decisive, which may be related to their better knowledge of the literature involved (most of which was published in medical rather than allied health journals). However, it may just reflect physicians' training to be decisive, rather than wait and see. It is possible that the physicians less often specified '??', but that their decisions as to a 'Yes' or 'No' were wrong more often. A separate analysis would be needed to determine whether physicians agreed with physicians more often than non-physicians agreed with non-physicians; that would give a preliminary indication of the physiatrists' superior knowledge as opposed to a tendency to select 'Yes' or 'No', and avoid 'dithering'.

The data in Tables 4 and 5 indicate that even when the number of 'maybe' answers to the screening questions is small (as it is for 'yes/no human'), the answers of the screeners do not necessarily agree: kappa, which indicates how much better agreement is than chance, is far from the level traditionally used as a cut-off for adequate interrater reliability (0.60 or 0.80). It is known that when the table's marginal totals are substantially unbalanced, kappa is low in spite of high agreement between raters.²⁴ The recommendation has been made to either report proportionate agreement with respect to a positive ('Yes') decision and with respect to a negative ('No') decision,²⁵ or kappa and percentage agreement.²⁶ As is clear from Table 4, except for the decision whether the paper deals with human subjects, the per cent agreement was also low; therefore, kappa was used in Table 5.

The suggestion was made that the blame for the poor consensus lies with the abstracts, providing insufficient or ambiguous information. It is likely that the problem is with the instructions the screeners received, or with the conscientiousness with which they followed these instructions? That is possibly part of the explanation – but one would assume that most searches for literature are done with a similar level of preparation and precision, and most with less. Our study indicates the upper level of performance that can be achieved in routine literature searches; in most situations, especially clinical ones, the level will be less: in searches done in preparation for patient care, education or research, there is no syllabus with unchanging instructions, explicit criteria, and a requirement to state an explicit judgment on four elements. Thus, many searches will bypass papers that are relevant, because the abstract fails to mention the pertinent information, or provides it in a way that is ambiguous. These searches also will lead to perusal of the full text of many papers, which will result in a rejection because the ultimate judgment based on full text information is 'not relevant'.

Many studies have addressed the issue of finding published research using bibliographic data bases, especially finding randomized clinical trials (RCTs) needed for clinical applications or systematic reviews.^{27–30} The deficiencies noted (published RCTs not indexed at all; failure by Medline staff or other database indexers to index the fact that randomization was used, etc.) has led to hand searching of journals, and development of trials data bases, eg the Cochrane collection. The problem addressed in this paper was not the difficulty of finding papers, but deciding on applicability once they are found. The issues are related, however, and can be compared to the issues of sensitivity and specificity in diagnostic screening tests. Inability to find the existing studies that are (potentially) applicable to one's question is the problem of false negatives – these studies should have shown up as relevant, but did not. Inclusion of many studies that according to our search strategy had relevant information, but did not upon review of the abstract, is the problem of false positives. The latter problem is compounded by the fact that in many instances, the abstract was not sufficient to decide whether the paper in question was applicable to our research question (a true positive) or not (a true negative). In these cases the abstract was insufficient as a screening test, and a review of the full text of the paper was necessary to decide whether it was a 'case' – comparable to the need for a full battery of tests to diagnose a disease.

With the easy availability of abstracts in electronic data bases, the search for papers to answer clinical or other questions will likely more and more be based on a computer search, based on text words included in the abstract, or the index terms assigned to the paper by the abstracters. Because index term assignments may be based on the abstract only,³¹ completeness and accuracy of abstracts assumes increased importance. Several journals have developed guidelines that are disseminated to potential authors to help them summarize in the abstract all the relevant information in their papers, for instance *JAMA*.¹ The structured abstract instructions offer a template of what is to be included, and where (ie under what heading) – although both set of instructions may be disregarded.¹³ Word limitations on abstracts may make it fairly impossible to satisfy all criteria, while maintaining a minimal level of readability.

Taddio *et al.* mention the paradox of abstracts: the better (longer, more informative) they are, the less likely it is that a reader may want to read the entire paper.¹² Longer, better abstracts may contain so much information that the reader may base clinical and other decisions on the abstract alone. However, doing so has been disapproved of unanimously – if for no other reason than that authors may put less of a gloss on their findings in the details of the 'Discussion' section than they do in the short sentence that summarizes the conclusion in the abstract. As Haynes *et al.* stated: structured abstracts have both been

complimented and criticized for making reading the rest of an article superfluous.⁹

It has been claimed that with the full-text availability of journals (resulting from larger and cheaper storage capability of computers), full-text searching of text, rather than abstract searching, is feasible. It may be attractive to rely on the full text of papers to search for those applicable to a particular clinical or research question. However, more complete retrieval based on full text also brings with it the problem of more false positives – eg papers that to a search engine appear to concern both concept X and clinical population Y, but that only happen to mention X as *not* found in the population Y, which population the authors studied because of problem Z. We had this problem in selecting papers based on abstracts, and it is reasonable to expect it to be worse if the full text of papers is the basis for searching. Abstracts, if well written, have the advantage of containing all the key lexical elements of the full paper, without the ‘accidental’ ones. One study comparing abstract and full text lexical elements for 1103 papers published in four major medical journals concluded that ‘the abstracts do reflect the language of the article and thus are lexical, as well as intellectual, surrogates of the articles they describe’.³² The problem when it comes to expeditiously identifying literature may not be one of easier access to full texts, but abstracts that do a better job of summarizing the crucial substantive and methodological information on the study described.

Three recommendations can be made: Authors of papers should pay attention to the abstract, and make sure they make clear the nature of the subject or patient population studied, and the specific clinical problem or issue evaluated, diagnosed or treated. Journal editors and manuscript reviewers must insist that authors include the information necessary for efficient screening in the abstract. Readers should be aware of the pitfalls involved in identifying literature. They may want to work with a librarian to formulate a search strategy that optimizes the yield of their database searches.

Acknowledgements

Thanks to the abstract screeners: Thomas Bryce M.D., Stacy Glass PT, Allison Gurwitz OT, Ralph Marino MD, Heather Muterspaw OT, Kate Parkin PT, Kristjan Ragnarsson MD, Carrie Schauer PT, Audrey Schmerzler RN, Mary Shea OT, Adam Stein MD and Sarah Zimmerman OT. Venecia Toribio, Adejoke Akimlamilo, Salma Akter and Jessie Gonzales printed abstracts, counted lines and/or entered data with great diligence. Dr Marino provided feedback on an earlier version of this paper. Drs Bryce, Marino, Ragnarsson and Stein are Co-investigators on the meta-analyses which produced some of the data. This work was supported in part by a grant from the National Institute on Disability and Rehabilitation Research (NIDRR), Office of Special Education Services, US

Department of Education to Mount Sinai School of Medicine (H133N000027). The conclusions of this paper are not necessarily endorsed by the Co-investigators or NIDRR.

References

- 1 Winker MA. The need for concrete improvement in abstract quality. *JAMA* 1999; **281**: 1129–1130.
- 2 Estrada CA *et al.* Reporting and concordance of methodologic criteria between abstracts and articles in diagnostic test studies. *J Gen Intern Med* 2000; **15**: 183–187.
- 3 Evans M, Pollock AV. Trials on trial. Review of trials of antibiotic prophylaxis. *Arch Surg* 1984; **119**: 109–113.
- 4 Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987; **317**: 426–432.
- 5 Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; **10**: 31–56.
- 6 Narine L, Yee DS, Einarson TR, Ilersich AL. Quality of abstracts of original research articles in CMAJ in 1989. *CMAJ* 1991; **144**: 449–453.
- 7 Ad Hoc Working Group for Critical Appraisal of the Medical Literature. A proposal for more informative abstracts of clinical articles. *Ann Intern Med* 1987; **106**: 598–604.
- 8 Harris P. A structured form of abstracts for Spinal Cord. *Spinal Cord* 1997; **35**: 865.
- 9 Haynes RB *et al.* More informative abstracts revisited. *Ann Intern Med* 1990; **113**: 69–76.
- 10 Froom P, Froom J. Deficiencies in structured medical abstracts. *J Clin Epidemiol* 1993; **46**: 591–594.
- 11 Hartley J, Benjamin M. An evaluation of structured abstracts in journals published by the British Psychological Society. *Br J Educ Psychol* 1998; **68**: 443–456.
- 12 Taddio A *et al.* Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ* 1994; **150**: 1611–1615.
- 13 Trakas K *et al.* Quality assessment of pharmacoeconomic abstracts of original research articles in selected journals. *Ann Pharmacother* 1997; **31**: 423–428.
- 14 Harbourt AM, Knecht LS, Humphreys BL. Structured abstracts in MEDLINE, 1989–1991. *Bull Med Lib Assoc* 1995; **83**: 190–195.
- 15 Hartley J, Sydes M. Are structured abstracts easier to read than traditional ones? *J Res Reading* 1997; **20**: 122–136.
- 16 Spitzer WO. The structured sonnet. *J Clin Epidemiol* 1991; **44**: 729.
- 17 Sanchez-Thorin JC, Cortes MC, Montenegro M, Villate N. The quality of reporting of randomized clinical trials published in Ophthalmology. *Ophthalmology* 2001; **108**: 410–5.
- 18 Scherer RW, Crawley B. Reporting of randomized clinical trial descriptors and use of structured abstracts. *JAMA* 1998; **280**: 269–272.
- 19 Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA* 1999; **281**: 1110–1111.

- 20 Pitkin RM, Branagan MA. Can the accuracy of abstracts be improved by providing specific instructions? A randomized controlled trial. *JAMA* 1998; **280**: 267–269.
- 21 Haynes RB *et al.* Online access to MEDLINE in clinical settings. A study of use and usefulness. *Ann Intern Med* 1990; **112**: 78–84.
- 22 Saint S *et al.* Journal reading habits of internists. *J Gen Intern Med* 2000; **15**: 881–884.
- 23 Barry HC *et al.* Family physicians' use of medical abstracts to guide decision making: style or substance? *J Am Board Fam Pract* 2001; **14**: 437–442.
- 24 Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; **43**: 543–549.
- 25 Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; **43**: 551–558.
- 26 Steijnmans VW *et al.* Interobserver agreement: Cohen's kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. *Int J Clin Pharmacol Ther* 1997; **35**: 93–95.
- 27 Derry S, Kong Loke Y, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. *BMC Med Res Methodol* 2001; **1**: 7.
- 28 Watson RJ, Richardson PH. Identifying randomized controlled trials of cognitive therapy for depression: comparing the efficiency of Embase, Medline and PsycINFO bibliographic databases. *Br J Med Psychol* 1999; **72**: 535–542.
- 29 Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; **309**: 1286–1291.
- 30 Avenell A, Handoll HH, Grant AM. Lessons for search strategies from a systematic review, in The Cochrane Library, of nutritional supplementation trials in patients after hip fracture. *Am J Clin Nutr* 2001; **73**: 505–510.
- 31 Lodge H. Improving the accuracy of abstracts in scientific articles. *JAMA* 1998; **280**: 2071.
- 32 Ries JE *et al.* Comparing frequency of content-bearing words in abstracts and texts in articles from four medical journals: an exploratory study. *Medinfo* 2001; **10**: 381–384.