# Examining the evidence in anesthesia literature: a survey and evaluation of obstetrical postdural puncture headache reports

[Examen de la preuve dans la documentation sur l'anesthésie : enquête et évaluation des articles sur les céphalées obstétricales post-ponction durale]

Peter T.-L. Choi MD FRCPC MSc,\*† Saramin E. Galinski MHSc MD,† Stefan Lucas MD, Lawrence Takeuchi MSc MD FRCPC,\* Alejandro R. Jadad MD DPhil FRCPC‡

**Purpose:** To describe a bibliographic database on the literature of postdural puncture headache (PDPH) in the obstetrical population, to describe the research architecture in this field, and to evaluate the quality of case-control studies, cohort studies, and controlled clinical trials on PDPH.

Methods: Computerized bibliographic searches, citation review, and hand searches were conducted to find all relevant citations on incidence, clinical course, prevention, or treatment of PDPH in parturients. The study design and topic(s) covered by each study were evaluated. Case-control studies and cohort studies were evaluated using the Quality Index; clinical trials were evaluated using the Jadad scale.

**Results:** One hundred ninety-six relevant citations were published between 1949 and 1999. Research on PDPH has been increasing rapidly with the majority of studies published in the 1990's. Incidence and prevention were the focus of over half of all citations. Optimal study designs were infrequently utilized. The methodological quality was poor for observational studies (Quality Index 10/29) and clinical trials (Jadad scale 2/5).

**Conclusion:** Although the amount of research on PDPH in parturients is increasing, use of optimal study designs and improvement in methodology is needed.

**Objectif**: Décrire une base de données bibliographiques des articles sur les céphalées post-ponction durale (CPPD) en obstétrique, présenter la méthodologie des études et évaluer la qualité des études rétrospectives, prospectives et des essais cliniques contrôlés sur les CPPD. **Méthode :** Des recherches bibliographiques électroniques, une revue des citations et des recherches manuelles ont été menées pour trouver toutes les citations pertinentes sur l'incidence, l'évolution clinique, la prévention ou le traitement des CPPD chez les parturientes. Le devis de recherche et les sujets traités dans chacune des études ont été évalués. Les études rétrospectives et prospectives ont été évaluées au moyen de l'Indice de qualité et, les essais cliniques avec l'échelle de Jadad.

**Résultats :** Cent quatre-vingt-seize citations utiles ont été publiées entre 1949 et 1999. La recherche sur les CPPD s'est accrue rapidement et la majorité des études a été publiée dans les années 1990. L'incidence et la prévention ont prédominé dans plus de la moitié de toutes les citations. Les devis de recherche optimaux n'ont pas été utilisé souvent. La qualité méthodologique était pauvre pour les études par observation (Index de qualité de 10/29) et les essais cliniques (échelle de Jadad de 2/5).

**Conclusion :** Malgré le nombre croissant de recherches sur les CPPD chez les parturientes, l'usage de devis de recherche optimal et l'amélioration de la méthodologie sont nécessaires.

VIDENCE-BASED health care can be defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."<sup>1</sup> The process involves "conscious choice on the part of the clinician and patient, explicit and

Revision accepted September 26, 2001.

From the Department of Anesthesia, St. Joseph's Healthcare\* and McMaster University,† Hamilton, Ontario, Canada; the Departments of Anesthesia and Health Administration, University Health Network, University of Toronto,‡ Toronto, Ontario, Canada; and the Department of Anesthesia, University of Rochester, S Rochester, New York, USA.

Address correspondence to: Dr. Peter T.-L. Choi, Department of Anesthesia, McMaster University, 1200 Main Street West, Room HSC-2U5, Hamilton, Ontario L8N 3Z5, Canada. Phone: 905-521-2100 ext. 75174; Fax: 905-523-1224; E-mail: choip@mcmaster.ca

This work was presented in part at the American Society of Anesthesiologists 1999 Annual Meeting (October 1999, Dallas, USA) and was completed in partial fulfilment of the requirements for a Masters of Science (Health Research Methodology) degree to Dr. Choi. *Accepted for publication August 27, 2001.* 

TABLE I Search strategy used computerized literature searches

E I Search strategy used
exp headache/
headache\$.tw.
cephalalg\$.tw.
or/1-3
exp spinal puncture/
dural.tw.
epidural.tw.
postdural.tw.
lumbar puncture.tw.
or/5-9
exp punctures/
exp rupture/
punctur\$.tw.
ruptur\$.tw.
perforat\$.tw.
or/11-15
and/4,10,16
exp obstetrics/
exp labour/
exp delivery/
obstetric\$.tw.
labour\$.tw.
labour\$.tw.
delivery.tw.
postpartum.tw.
or/18-25

27 and/17,26

exact decisions being made and carried out, and always wisdom, experience, and judgment used to evaluate and apply this evidence."<sup>2</sup> To carry out this process, five steps are necessary:<sup>2</sup> definition of the question, acquisition of the evidence to be used in answering the question, evaluation of the validity and relevance of the evidence, integration of the evidence with patient factors to arrive at the decision, and evaluation of the decision-making process.

Inherent in the framework of evidence-based health care is the ability to obtain high quality information upon which decisions may be based. A comprehensive bibliographic database is important for several reasons. First, comprehensiveness minimizes the random error and various biases.<sup>3</sup> Second, a comprehensive database permits assessment of patterns and trends in publication (type of study, reporting quality, methodological quality), clinical practice (types of intervention, changes in incidence or prevalence, perspectives of patients and clinicians), and research (direction of research efforts). Third, a comprehensive bibliographic database can be a valuable information resource to patients, clinicians, researchers, and policy makers by providing information in a single source and by minimizing duplication of search efforts. Fourth, such

databases are the foundation for systematic reviews and meta-analysis, which are heavily dependent on comprehensive acquisition of all relevant data to generate valid conclusions.

Anesthesia-specific bibliographic databases are relatively infrequent. For example, over 100 years of literature on postdural puncture headache (PDPH) exists. Tourtellotte *et al.*<sup>4</sup> systematically compiled a bibliography of the literature up to the 1960's; since then, no comprehensive compilation has been undertaken. In this paper, we report the development of a bibliographic database of the literature on obstetrical PDPH, describe the research architecture in this field, and evaluate the quality of observational and experimental studies of PDPH.

### Methods

## Search strategy

To identify all citations on PDPH in the obstetrical population, the bibliographic databases MEDLINE (1966 to January 1999), CINAHL® (1982 to December 1998), HealthSTAR (1975 to December 1998), and the Cochrane Library (1998 issue 4) were searched. The search strategy (Table I) was initially designed for the MEDLINE database and then tailored to the others. The reference list of each citation was reviewed for potential additional citations. Further citations were also sought from abstracts and conference proceedings of major anesthesia meetings from the year of their first publication to 1998 (Table II). We included citations that were not published in English. All citations were recorded in a Reference Manager® (Research Information Systems, Carlsbad, CA, USA) version 9 bibliographic database. The search was completed on February 19, 1999 and the database was closed nine days later.

## Selection of eligible citations

For a citation to be included in the database, it had to discuss at least one of four topics relating to PDPH: incidence, clinical course, prevention, or treatment. Inclusion criteria were developed for each topic based on the population, intervention or triggering event, outcome(s), and study design (Table III). Definitions for each criteria were developed by consensus by four of the authors (PTC, SEG, SL, LT) and field tested over three sessions using randomly selected citations from the literature search. A citation was included if it met all four criteria for any of the four topics; it was possible for a citation to meet all the inclusion criteria for more than one topic. Once the inclusion criteria were established, two of us (PTC, SEG) reviewed all the citations identified by the literature search. For

Journal	Years*	Conference
Anesthesia and Analgesia	1986 - 1998	International Anesthesia Research Society
Anesthesiology	1979 - 1998	American Society of Anesthesiologists Society for Obstetrical Anesthesia and Perinatology
British Journal of Anaesthesia	1993 - 1998	European Society of Anaesthesiologists
Canadian Journal of Anesthesia Regional Anesthesia and Pain Medicine†	1981 - 1998 1982 - 1998	Canadian Anesthesiologists' Society American Society of Regional Anesthesia

TABLE II List of abstract and conference proceedings searched by hand

\*All journals were searched from 1998 back to the first year in which conference proceedings were published. †This journal also contained occasional proceedings from meetings of the Asian and Oceanic, European, and Latin American Societies of Regional Anesthesia.

TABLE III Inclusion criteria for citations on postdural headache in parturients

Topic	Inclusion criteria Population	Intervention/event	Outcomes(s)	Study design
Incidence	Obstetrical patients	Lumbar epidural or spinal analgesia or anesthesia or combined spinal epidural	Posdural puncture headache (including mention of "no PDPH")	Case series Chart reviews Surveys Prospective studies Systematic reviews
Clinical course	Obstetrical patients	Postdural puncture headach	e Spontaneous resolution rate of PDPH, and/or duration of PDPH	Case series Chart reviews Surveys Prospective studies Systematic reviews
Prevention	Obstetrical patients with intentional or accidental dural punctures	Any intervention performed to prevent PDPH (study must state purpose to reduce incidence of PDPH or to prevent (PDPH)	l Postdural puncture headache e	Case reports Case series Chart reviews Prospective studies Systematic reviews
Treatment	Obstetrical patients with PDPH	Any intervention performed to treat PDPH	Resolution of PDPH	Case reports Case series Chart reviews Surveys Prospective studies Systematic reviews

non-English citations, translation was obtained when needed. Disagreements were resolved by consensus. The extent of agreement was recorded and the interrater reliability (expressed as kappa  $\pm$  standard error)<sup>5</sup> was calculated for the four topics.

The bibliographic data from all selected citations were copied to a separate Reference Manager® file, which was called the McMaster Obstetrical PDPH Evidence Database (MOPED). Each citation was coded using the US National Library of Medicine's Medical Subject Headings (MeSH). Additionally, each citation was classified by topic ("incidence", "clinical course", "prevention", or "treatment") and by study design.

## Evaluation of quality

Two reviewers (PTC, SEG) independently evaluated

the quality of observational and experimental studies listed in MOPED. In the event of duplicate publication, the publication with the most methodological information was used. For case-control studies and cohort studies, the Quality Index<sup>6</sup> was used. The Quality Index is a 27-item, partially validated checklist that assesses the reporting quality, generalizability, internal validity, and power of non-randomized and randomized studies.<sup>6</sup> For non-randomized studies, scores may range from 0 (poor) to 29 (excellent). Controlled clinical trials were evaluated with the Jadad scale,7 which is a validated three-item scale that evaluates randomization, blinding, and description of withdrawals. Possible scores range from 0 (not randomized, not double-blind, no description of withdrawals) to 5 (randomized appropriately, double-blinded appropriately, withdrawals described).

Reviewers were not blinded to the author(s), journal, or year of publication. Disagreements between the two reviewers were resolved by consensus. The extent of agreement was recorded and the inter-rater reliability was calculated using generalizability theory<sup>5</sup> for the two instruments used for rating quality. As the distributions of the quality scores were not expected to be normal, the median and range were chosen as the measures of central tendency and dispersion. We compared the quality scores of abstracts and full papers using the Mann-Whitney U nonparametric test. All statistics were performed using SPSS© 8.0 for Windows (SPSS Inc, Chicago, IL, USA).

## Results

## Search results

One hundred thirteen citations were identified by MED-LINE search. Of these, 25 citations were completely unrelated to PDPH and were excluded prior to import of citations to Reference Manager®. The Cochrane Library identified 13 additional citations; CINAHL® and HealthSTAR did not identify any additional citations. Citation review and manual searching added 65 and 105 citations respectively. A total of 269 eligible citations were identified and entered into Reference Manager®. Attempts were made to obtain full text, paper copies of each citation: three articles could not be found in any North American library; therefore, only 266 citations were evaluated for selection.

#### Research architecture

One hundred ninety-six citations (73.7% of retrieved citations) met selection criteria for at least one topic.<sup>a</sup> Information on incidence, clinical course, prevention, and treatment was present in 75.5%, 5.3%, 52.0%, and 15.3% of the citations meeting selection criteria respectively. Inter-rater reliability was high for all topics (incidence k 0.82  $\pm$  0.04; prevention k 0.91  $\pm$  0.03; treatment k 0.98  $\pm$  0.02) except clinical course (k 0.24  $\pm$  0.18).

The classification of the selected citations by topic and study type is outlined in Table IV. The classification has been delineated by categories similar to the "levels of evidence" described by Sackett.<sup>8</sup> Only onethird of citations (34.7%) were systematic reviews or randomized controlled trials (RCTs). The majority of citations (53.1%) were observational reports.

Citations with information on the incidence of PDPH were mostly of observational reports (52.7%) and RCTs (33.8%). A similar distribution was seen for information on the clinical course of PDPH but only

14 of the 196 citations covered this topic. Information on prevention of PDPH was seen primarily in RCTs (46.1%) and case series (29.4%). The 30 citations with information on treatment of PDPH were mostly case series and case reports (83.3%); the remaining 16.6% were RCTs.

## Publication details

Citations were published in 27 journals and seven languages (English, Finnish, French, German, Norwegian, Spanish, and Chinese). The majority of citations (176 of 196; 89.8%) were published in anesthesia journals. Over three-quarters of all selected citations (151 of 196; 77.0%) were published in the six anesthesia journals with the highest citation impact factors (CIF): Anesthesiology, Anesthesia and Analgesia, Regional Anesthesia and Pain Medicine, British Journal of Anaesthesia, Anaesthesia, and Canadian Journal of Anesthesia (in descending order of CIF). Most citations (187 of 196; 95.4%) were published in English.

Table V classifies the citations by the decade in which they were published and by the category of evidence. The number of citations in each category increased over each decade. The majority of citations (140 of 196; 71.4%) were published in the 1990's. Experimental studies (RCTs and non-RCTs) have increased more than observational studies although the latter still constituted more than 50% of all selected citations in the 1990's. MEDLINE and the Cochrane Library indexed selected citations from 1972 to 1999 and 1950 to 1998 respectively. Citation review and manual searches found selected citations from 1949 to 1995 and 1975 to 1999 respectively. Thirty-eight (29.4%) of the 129 selected citations found by citation review and manual searches were present in MEDLINE but unidentified by the computerized search strategy.

Sixty-seven of the 196 citations (34.1%) were meeting abstracts. Eighteen pairs of citations (36 of 196; 18.4%) contained duplicate data. Three pairs were duplicate publication of abstracts, one was an abstract subsequently published as a letter, one was a letter subsequently published as a full paper, and 12 were abstracts subsequently published as full papers. Two papers contained duplicate data from each other.

# Quality assessment of case-control and cohort studies

There were 24 case-control or cohort studies; three duplicate publications were excluded. Twenty-one casecontrol or cohort studies (14 full papers, seven abstracts) were assessed using the Quality Index. The scores ranged from 2 to 19 with a median score of 10. Table VI enumerates the median and the range for each section of the Quality Index. The inter-rater reliability

<sup>&</sup>lt;sup>a</sup> The references and the details on the study design and the topics covered by each citation are available from the corresponding author.

Categories of evidence	Incidence	Clinical course	Prevention	Treatment	Total
Structured summaries	(0.7%)*		(2.9%)		(1.5%)
Systematic reviews	1	0	3	0	3
Randomized clinical trials	(33.8%)	(35.7%)	(46.1%)	(16.6%)	(33.2%)
Parallel design	49	5	46	4	63
Crossover design	1	0	1	1	2
Non-randomized clinical trials	(12.8%)	(7.1%)	(21.6%)		(12.2%)
Concurrent cohorts	14	0	15	0	17
Historical cohorts	5	1	7	0	7
Observational reports	(52.7%)	(57.1%)	(29.4%)	(83.3%)	(53.1%)
Patient surveys	6	3	0	0	6
Case series, prospective	35	2	11	9	44
Case series, retrospective	36	3	19	11	48
Case reports	1	0	0	5	6
Totals	148	14	102	30	196

TABLE IV Number of citations by type of study and topic

\*Percentage of all citations within a given topic.

TADIEV	Number of	citations l	w decod	a of pu	ublication	and study	decion
IADLE V	Number of	citations t	by uccau	e or pu	ioncation	and study	uesign

Publication years	Systematic reviews	Randomized clinical trials	Non-randomized clinical trials	Observational	Totals
<1950*	0	0	1	0	1
1950-1959	0	1	0	2	3
1960-1969	0	0	1	2	3
1970-1979	0	3	5	11	19
1980-1989	0	10	1	19	30
1990-1999	3	51	16	70	140
Totals	3	65	24	104	196

\*Earliest citation was published in 1949.

was 0.68, which is similar to the inter-rater reliability obtained by the original developers of the index.<sup>6</sup>

The range of scores was narrower for abstracts than for full papers. No significant differences in quality were seen between abstracts and papers in any of the five categories or in the total score. In general, casecontrol and cohort studies rated low in all categories (reporting, external validity, bias, confounding, and power; Table VI). No study discussed issues of power or adjusted for confounding in data analyses. Less than half of the 21 studies clearly described characteristics of study participants and those lost to follow-up nor reported all important adverse events. A similar proportion did not report estimates of random variability or the actual probability values for the main outcomes. Issues relating to external validity, blinding, and adjustment for differing lengths of time or followup were also poorly addressed. The remaining issues were addressed adequately in over half of the studies: description of the intervention was the item addressed most frequently (95.2%).

#### Quality assessment of randomized clinical trials

There were 65 RCTs; ten duplicates were excluded. Fifty-five RCTs (38 full papers, 17 abstracts) were assessed using the Jadad scale. The scores ranged from 0 to 5 with a median score of 2. Table VII describes the proportion of responses in each category per question. The inter-rater reliability was 0.58, which is consistent with the reliability obtained by Jadad *et al.*<sup>7</sup>

No statistically significant differences in randomization, blinding, withdrawal, or overall score were seen between papers and abstracts. All but one RCT stated that randomization was used. Only 27.8% of RCTs described the method of randomization. Of these, 26.7% used methods that were inappropriate. Less than

TABLE VI Median Quality Index score and ranges for case-control studies and cohort studies in MOPED

Category		Full papers* (n=14)	Meeting abstracts* (n=7)	All citations (n=21)
Reporting	(max 11)	5 (1-9)	5 (2–9)	5 (1–9)
External validity	(max 3)	0 (0-3)	0 (0–1)	0 (0–3)
Bias	(max 7)	3 (0-6)	3 (1–5)	3 (0–6)
Confounding	(max 4)	1 (0-3)	2 (0-3)	2 (0-3)
Power	(max 5)	0 (0-0)	0 (0-0)	0 (0-0)
Total	(max 30)	9 (2-19)	10 (4-15)	10 (2-19)

\*All values are expressed as the median score and the range.

TABLE VII Results from the assessment of randomized clinical trials with the Jadad scale

Item	Full papers	Meeting abstracts	All citations
Randomization described and appropriate	8 / 38	3 / 17	11 / 55
and appropriate Withdrawals and dropouts	11 / 38	7 / 17	18 / 55
described	21 / 38	5 / 17	26 / 55

half of all RCTs stated double-blinding (38.2%) or described withdrawals (47.2%). In the 21 RCTs where double-blinding was stated, most (80.9%) described an appropriate method.

#### Discussion

We have compiled a bibliographic database of published information on the incidence, clinical course, prevention, and treatment of PDPH in the obstetrical population. The included citations date from 1949 to 1999. Although the chosen search strategy is similar to ones used for systematic reviews, our database is more inclusive with respect to type of study. A survey of the information on PDPH indicates that literature in this field has been growing exponentially over the past 50 years irrespective of the study design. Similar proliferation of information has been observed in the number of RCTs and systematic reviews in pain medicine and anesthesia.9-12The database does not include editorials, commentaries, unpublished data, or websites; we cannot comment on the extent of information available from these types of information.

The findings of the various strategies employed to seek out relevant information reveal the danger of utilizing only a single strategy. The computerized search of MEDLINE found 113 citations, in which 60 were selected ("true positives"). Thirty-eight citations found by citation review or hand search were indexed in MED-LINE but were missed by the computerized search ("false negatives"). The remaining 91 citations obtained by citation review or hand search were abstracts or articles from journals not cited in any of the four computerized bibliographic databases. The sensitivity of the MEDLINE search strategy for all MEDLINE-indexed citations was only 61.2% (60 / 98). For all selected citations, the sensitivity of the MEDLINE strategy dropped to 30.6% (60 / 196). Most of the citations included in MOPED would have been missed if reliance had been placed on the MEDLINE search alone.

Our computerized searches did not include the EMBASE or LILACS databases. EMBASE and LILACS indexes many European and Latin American journals respectively. At the time of this study, neither EMBASE nor LILACS were freely available to the public and the costs of performing the searches were beyond our allotted budget for this project. (LILACS can be accessed now, free of charge, under "Scientific literature" at the website http://www.bireme.br/ bvs/l/ihome.htm). There is a possibility that we missed relevant European or Latin American literature during the development of MOPED; however, this is unlikely at least for RCTs and systematic reviews as the Cochrane Library routinely searches the two databases for citations through its different Cochrane Centers across the world.

In selecting citations for inclusion in MOPED, individuals with no experience in research methodology were trained to read and evaluate the contents of papers identified by the literature search. Other researchers have found that valid and reliable results can be obtained by training students to screen for methodological content.<sup>13</sup> The inter-rater reliability for selecting articles on incidence, prevention, and treatment all were high (k 0.82, 0.92, and 0.98 respectively), which is consistent with near perfect agreement.<sup>14</sup> Reliability was low (k 0.24; fair agreement)<sup>14</sup> for selecting articles on clinical course. There are likely two reasons for the fair strength of agreement. First, differences existed amongst the two reviewers in the interpretation of the outcome used in the inclusion criteria for clinical course despite extensive field testing. At the consensus meeting, both reviewers were able to resolve disagreements very quickly once the difference in interpretation was brought to light. Second, the k statistic is a function of sensitivity, specificity, and prevalence.<sup>15</sup> For any given sensitivity and specificity, a decrease in prevalence will deflate the k value.<sup>15,16</sup> The low prevalence of citations on clinical course may have contributed to the low k value observed in this topic.

The type of studies published in each topic was also revealing. Optimally, the incidence and clinical course of PDPH would be examined using prospective follow-up of cohorts (with survival-type analysis for the latter topic) and interventions for prevention and treatment would be studied using RCTs. A survey of the literature indicates that the optimal study designs for incidence, clinical course, and treatment were infrequently utilized. Prevention of PDPH was an exception with the majority of studies being RCTs. This observation highlights a problem in the practice of evidence-based medicine with respect to PDPH. Already, the clinician is faced with a hurdle: most of the available information upon which reliance is placed for evidence-based decision making regarding PDPH is not optimal in the study design.

Attempts to minimize language bias were made by including citations of all languages although it is possible that relevant citations not published in English may have been missed using the described search strategy. Few citations were published in other languages.

Nearly one-fifth of the citations in our database contained duplicate data. Most of the duplicates involved abstracts that were subsequently published as full papers. Of the 67 abstracts, 12 were eventually found as full papers. These observations raise an interesting dilemma. Meeting abstracts are often the earliest reports of studies and can contain important information,<sup>17</sup>especially since only half of all abstracts are published in full.<sup>18,19</sup> However, the peer review process may be less stringent than for full manuscripts and the quality of abstracts may be low.<sup>20</sup> As well, the findings of abstracts may differ drastically from the subsequent papers. In a subset of obstetrical anesthesia abstracts, Halpern et al. found a large number of discrepancies between the abstracts and the full papers.<sup>21</sup> The choice to include or exclude abstracts remains controversial.

Our findings echo previous observations of the validity and statistical concerns in anesthesia evidence. There is room for improvement in the quality of nonrandomized and randomized studies of PDPH in obstetrical patients. Median quality scores were low for both types of studies. There were no differences in quality between full papers and meeting abstracts.

This study is the first to evaluate the quality of observational studies in anesthesia. Both reporting and methodological quality (based on external validity, bias, confounding, and power) were poor with median scores less than half of the maximum possible score per category. No observational study in the examined set evaluated power. This observation reiterates the concerns raised previously by other methodologists who have evaluated anesthesia studies.<sup>22,23</sup> Similarly, the results of these studies were infrequently generalizable. Issues with bias and confounding were also poorly addressed.

Based on these observations, we would suggest the following changes to improve quality of observational studies in this field:

1. Subjects should be representative of the entire population of potential participants.

2. Where possible, patients of different cohorts (for cohort studies) or cases and controls (for case-control studies) should be recruited from the same population over the same time period to minimize confounders relating to population and time.

3. Where possible, subjects and individuals measuring the outcomes should be blinded. The same individuals should measure all groups.

4. Patient characteristics, possible confounders, interventions, and all relevant outcomes and adverse effects should be reported clearly. A flow diagram may assist the reader in tracking the fate of patients.

5. Follow-up should be reported along with a description of patients who withdraw or are lost to follow-up.

6. Analyses need to adjust for confounders and time factors.

7. Power analysis should be performed to ensure an adequate sample size for the effect that one wishes to detect. The analysis should be reported.

The weaknesses observed in the RCTs in MOPED were similar to those reported by Bender *et al.*,<sup>24</sup> whose set of studies was in a similar population (obstetrical patients). Again, the majority of RCTs gave inadequate description of the randomization process (72%), blinding (63%), and withdrawals (54%). Our study, which evaluated RCTs from 1949 to 1999, confirms that further work is needed to improve the quality of RCTs.

As mentioned previously, part of the difficulty in evaluating methodological quality is due to poor reporting. Often for blinding and withdrawal, we could not differentiate between failure to perform blinding and track withdrawals or failure to report them. None of the RCTs were reported using standardized reporting criteria. Given the inadequacies of reports examined in anesthesia so far, we advocate the adoption of standardized criteria such as the CON-SORT statement<sup>25</sup> to improve quality of RCTs.

In summary, we have compiled a bibliographic database that covers the literature on obstetrical PDPH from 1949 to 1999. We hope to update this database on a regular basis. Examination of the architecture of obstetrical PDPH research and the quality of observational and experimental studies indicate that further work is needed to improve the research in this field.

## Acknowledgements

The authors wish to thank Ms. Nalajini Nadarajah and Ms. Theresa Wolf for their secretarial assistance and Dr. Kari Smedstad for translation of the Finnish, German, and Norwegian publications. This work was supported in part by the St. Joseph's Hospital Anesthesiologists' Research Fund (Hamilton, Ontario, Canada) to Dr. Choi.

## References

- 1 Sackett DL, Rosenberg WM, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996; 312: 71-2.
- 2 McKibbon A, Eady A, Marks S. PDQ Evidence-Based Principles and Practice. Hamilton: BC Decker Inc., 1999.
- 3 *Counsell C.* Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med 1997; 127: 380–7.
- 4 *Tourtellotte WW*, *Haerer AF*, *Heller GL*, *Somers JE*. Post-Lumbar Puncture Headaches. Springfield, IL: Charles C. Thomas, 1964.
- 5 Streiner DL, Norman GR. Health Measurement Scales.
  A Practical Guide to Their Development and Use, 2<sup>nd</sup>
  ed. Oxford: Oxford University Press, 1995.
- 6 *Downs SH, Black N.* The feasibility of creating a checklist for the assessment of the methodological quality of both randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998; 52: 377–84.
- 7 Jadad AR, Moore A, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996; 17: 1–12.
- 8 Sackett DL. Rules of evidence and clinical recommendations. Can J Cardiol 1993; 9: 487–9.
- 9 Jadad-Bechara AR. Meta-Analysis of Randomised Clinical Trials in Pain Relief [D. Phil. Thesis]. Oxford: University of Oxford, 1994.
- 10 Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. J Clin Epidemiol 1996; 49: 235–43.
- 11 Halpern SH, Jadad AR, Choi PT-L. Evidence based practice in anaesthesia – how good is the evidence? In: Tramèr MR (Ed.). An Evidence Based Resource in Anaesthesia and Analgesia. London: BMJ Books, 2000: 27–44.
- 12 Choi PT-L, Halpern SH, Malik N, Jadad AR, Tramèr MR, Walder B. Examining the evidence in anesthesia literature: a critical appraisal of systematic reviews. Anesth Analg 2001; 92: 700–9.
- 13 Sands ML, Murphy JR. Use of kappa statistic in determining validity of quality filtering for meta-analysis: a

case study of the health effects of electromagnetic radiation. J Clin Epidemiol 1996; 49: 1045–51.

- 14 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–74.
- 15 Spitznagel EL, Helzer JE. A proposed solution to the base rate problem in the kappa statistic. Arch Gen Psychiatry 1985; 42: 725–8.
- 16 Grove WM, Andreason NC, McDonald-Scott P, Keller MB, Shapiro RW. Reliability studies of psychiatric diagnosis. Theory and practice. Arch Gen Psychiatry 1981; 38: 408–13.
- 17 Kelly JA. Scientific meeting abstracts: significance, access, and trends. Bull Med Libr Assoc 1998; 86: 68–76.
- 18 Yentis SM, Campbell FA, Lerman J. Publication of abstracts presented at anaesthesia meetings. Can J Anaesth 1993; 40: 632–4.
- 19 Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta- analysis. JAMA 1994; 272: 158–62.
- 20 Rubin HR, Redelmeier DA, Wu AW, Steinberg EP. How reliable is peer review of scientific abstracts? Looking back at the Meeting of the Society of General Internal Medicine. J Gen Intern Med 1993; 8: 255–8.
- 21 *Halpern SH, Palmer S, Angle P, Tarshis J.* Published abstracts in obstetrical anesthesia: full publication rates and data reliability. Anesthesiology 2001; 94(1A): A69 (abstract).
- 22 Goodman NW, Hughes AO. Statistical awareness of research workers in British anaesthesia. Br J Anaesth 1992; 68: 321–4.
- 23 Goodman NW, Powell CG. Could do better: statistics in anaesthesia research (Editorial). Br J Anaesth 1998; 80: 712–4.
- 24 Bender JS, Halpern SH, Thangaroopan M, Jadad AR, Ohlsson A. Quality and retrieval of obstetrical anaesthesia randomized controlled trials. Can J Anaesth 1997; 44: 14–8.
- 25 Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996; 276: 637–9