

Achieving High Research Reporting Quality Through the Use of Computational Ontologies

Amrapali Zaveri · Luciana Cofiel · Jatin Shah ·
Shreyasee Pradhan · Edwin Chan · Olivier Dameron ·
Ricardo Pietrobon · Beng Ti Ang

Published online: 16 October 2010
© Springer Science+Business Media, LLC 2010

Abstract Systematic reviews and meta-analyses constitute one of the central pillars of evidence-based medicine. However, clinical trials are poorly reported which delays meta-analyses and consequently the translation of clinical research findings to clinical practice. We propose a Center of Excellence in Research Reporting in Neurosurgery (CERR-N) and the creation of a clinically significant computational ontology to encode Randomized Controlled Trials (RCT) studies in neurosurgery. A 128 element strong computational ontology was derived from the Trial Bank ontology by omitting classes which were not required to perform meta-analysis. Three researchers from our team tagged five randomly selected RCT's each, published in the last 5 years (2004–2008), in the Journal of Neurosurgery (JoN), Neurosurgery Journal (NJ) and Journal of Neurotrauma (JoNT). We evaluated inter and intra observer reliability for the ontology using percent agreement and kappa coefficient. The inter-observer agreement was

76.4%, 75.97% and 74.9% and intra-observer agreement was 89.8%, 80.8% and 86.56% for JoN, NJ and JoNT respectively. The inter-observer kappa coefficient was 0.60, 0.54 and 0.53 and the intra-observer kappa coefficient was 0.79, 0.82 and 0.79 for JoN, NJ and JoNT journals respectively. The high degree of inter and intra-observer agreement confirms tagging consistency in sections of a given scientific manuscript. Standardizing reporting for neurosurgery articles can be reliably achieved through the integration of a computational ontology within the context of a CERR-N. This approach holds potential for the overall improvement in the quality of reporting of RCTs in neurosurgery, ultimately streamlining the translation of clinical research findings to improvement in patient care.

Keywords Systematic review · Meta-analyses · Evidence-based medicine · Reporting · RCT · Neurosurgery · Standardized · Ontology · Kappa coefficient

A. Zaveri (✉) · B. T. Ang
National Neuroscience Institute,
Singapore, Singapore
e-mail: Amrapali_Jyotindra_ZAVERI@ttsh.com.sg

A. Zaveri · J. Shah · S. Pradhan · R. Pietrobon
Duke-NUS Graduate Medical School,
Singapore, Singapore

E. Chan
Australasian Cochrane Centre,
Singapore, Singapore

O. Dameron
INSERM U936, Université de Rennes1,
Rennes, France

L. Cofiel
Research on Research Group,
Sao Paulo, Brazil

Introduction

Systematic reviews and meta-analyses (SR-MAs) constitute one of the central pillars of evidence-based medicine (Manchikanti et al. 2008; Sackett 1996), summarizing the best available evidence to answer specific clinical questions. SR-MAs also constitute an important component of clinical practice guidelines, thus establishing the connection between clinical research and clinical practice (Manser and Walters 2001).

Although SR-MAs have an important role in evidence-based practice, their creation and maintenance is a complex and demanding process. First, the information contained in original articles used to generate SR-MAs can often be poorly reported, ultimately causing data to be left out of

SR-MA (Moher et al. 2008). The inclusion of reports of low-quality Randomized Controlled Trials (RCTs) in meta-analysis is likely to affect the estimates of intervention efficacy (Moher et al. 1998; Biesen et al. 2008). Also, without complete, clear and transparent reports, readers cannot judge the reliability and usefulness of health research (Moher et al. 2008). Second, in the event of poor quality reporting/incomplete articles, meta-analysts frequently need to communicate with the authors of these papers. It adds significant delay to the already time consuming meta-analysis process. Consequently the translation of clinical research findings to clinical practice is delayed.

Over the last decade, there have been numerous efforts to improve the quality of reporting for randomized-trials, meta-analyses, diagnostic studies, non-randomized designs and observational biomedical research studies by establishing standards, checklists and guidelines (Moher et al. 2000; Stroup et al. 2000; Moher et al. 2001a; Bossuyt 2004; Jarlais et al. 2004). As a consequence reporting quality has significantly improved in many areas of biomedical research (Moher et al. 2001a; Kane et al. 2007; Fung et al. 2009; Han et al. 2009), though it is still not standardized or optimal (Bossuyt 2004; Wilczynski 2008; Xu et al. 2008). Importantly none of the reporting guidelines fully incorporate the full set of trial details required for the conduct of SR-MA's (Meinert et al. 1984; Pocock et al. 1987; Clarke and Stewart 1994; Meinert 1998; Moher et al. 2001b) For example the CONSORT (Consolidated Standards of Reporting Trials) statement, which is a check-list developed to alleviate the problems arising from inadequate reporting of RCTs, includes only 22 of the 100 items needed for successfully conducting SR-MA's (Altman et al. 2001; Hopewell et al. 2008).

Ontologies are alternative methods to creating reporting standards of research articles. By definition, ontology is a specification of a conceptualization (Gruber 2008). They provide structured vocabularies, definitions and reasoning that enable standardized and semantically interconnected machine readable sections in a research article. The search process is very time consuming and there are chances of missing important articles. With machine-readable articles, the ontologies can assist and thus increase speed in searching for relevant articles as well as ensure that all related articles are included. Also, ontologies not only improve research reporting but also provide a method to simultaneously extract qualitative and quantitative information from constituent research articles thus enabling semi-automation of the meta-analysis process, as well as facilitating real time meta-analysis (Cook et al. 2007). In this context, The Trial Bank ontology (Sim and Detmer 2005) was one of the first frame-based ontologies developed for RCT's. Frame-based ontologies are those based on frames, which are equivalent to classes in object-oriented systems and represent collection of

instances, and slots, which are equivalent to attributes thus providing greater structure to the ontology representation. It proposed the co-publication of RCT articles in prose and machine readable formats (Sim et al. 2000, 2001). However, the Trial Bank ontology is not being widely used for SR-MA as it is highly comprehensive and lacks terms that are needed for the conduct of SR-MAs. Thus there is a lack of an RCT ontology focused on SR-MAs that can semi-automate the SR MA process and update it in real time.

We propose to create a Center of Excellence in Research Reporting (CERR) that will not only set standards for RCT reporting using ontology but will also serve as a repository that will support immediate update of meta-analysis and facilitate real time meta-analysis. In this endeavor rather than preparing ontology from scratch we will modify the Trial Bank ontology to make it more purpose specific. In the context of a CERR ontology, accuracy and observer annotation consistency will not only determine the accuracy of reporting but will also significantly influence the meta-analysis report. Despite its importance, the degree of agreement in the same observer as well as among different observers annotating the same article using an ontology is rarely evaluated in the literature. Yet, agreement should be taken into consideration during ontology evaluation because using an ontology involves applying a categorization and the usefulness of that categorization depends on consensus among "categorizers". Although some authors are aware of the importance of agreement for testing reliability of standards and have investigated alternatives for measuring agreement (Hripcsak and Heitjan 2002; Petersson et al. 2002), surprisingly, this gap is present even in areas where the issues associated with lack of observer agreement have been well documented, such as in the case of ontology-assisted meta-analyses (Tendal et al. 2008). For example, although the Trial Bank ontology adheres to strict ontology engineering principles, to our knowledge it's inter and intra-observer reliability has not been evaluated up to this point. The most thorough examination of the issue of reliability for meta-analysis is a study of inter-rater reliability. Since the distinction between what is intended to be measured and what one can reliably measure may decrease the quality of information collected and interfere in its usefulness (Petersson et al. 2002), agreement between ontology users is of great importance. If there is little agreement among the raters, then their responses are unreliable and the quality of the reference standard is suspected.

Rather than applying CERR to a broad biomedical research domain we decided to focus on the improvement of trial reporting in neurosurgery. Previous studies have noted that reporting of neurosurgical randomized trials is frequently of poor quality (Haines 1983) and needs higher accuracy (Vranos et al. 2004). Further, being a diverse discipline

traversing many scales and employing many data types, ontologies provide a suitable means to integrate data across neurosurgery. Such integration will help facilitate SR MA's and ultimately improve quality of patient care. In the context of CERR, it will initially help validate the concept, which can later be scaled up and applied to clinical trial reporting in different fields of biomedical research. Ontologies, when applied to field of neurosurgery will not only help standardize clinical data reporting but also facilitate integrating the sparse data and subsequently stream-lining the meta-analysis process, ultimately achieving answers to important questions with regards to neurosurgery clinical practice.

This article addresses the issues outlined earlier in three phases. First, it describes the concept of Center for Excellence in Research Reporting in Neurosurgery (CERR-N), which will encode manuscripts leading to improved and standardized research reporting, data extraction and semi-automated incorporation in SR-MAs. Second, it describes the structured classifications used to encode these articles in the form of computational ontologies. Specifically, it describes the development of a computational ontology for Randomized Controlled Trials (RCTs). Lastly, it describes the inter and intra observer agreement to check for the ontology's accuracy and consistency in tagging a manuscript.

Methods

Computational Ontology Development

The RCT ontology we created was derived as a sub-set of the RCT ontology created for the Trial Bank Project, with additional input from Cochrane experts (EC and FS). Specifically, we created a list of classes from the Trial Bank ontology and compared them with the elements currently collected in RevMan (Review Manager), which is a software used to support the production of systematic reviews to ensure a standard format for inclusion in the Cochrane Database of Systematic Reviews. It can perform meta-analysis of the data entered, and present the results graphically. We obtained input from the two Cochrane members about which classes should be included or excluded. Based on a series of meetings and discussions, we modified the ontology and included only those classes specific to MA. We attempted to reproduce both the qualitative tables used to compare studies as well as the individual point estimates and confidence intervals related to tables for the meta-analysis. The class and sub-class structure was maintained in the derived ontology. An ontology expert (OD) helped create the relationships between the various classes in the ontology.

We developed the ontology by creating classes and their definitions, which when instantiated would assist in the

meta-analysis process. We implemented the following steps mentioned in Noy and McGuinness's (Noy and McGuinness 2001) guide for creating a new ontology: (1) Determine the domain and scope: Since our purpose was to be able to perform meta-analysis, we aligned the classes in the ontology accordingly; (2) Consider reusing existing ontology: Rather than building an ontology from scratch we decided to reuse portions of the Trial Bank ontology that were relevant to our endeavour. We omitted classes that were outside our scope; (3) Enumerate important terms: We wrote down all relevant terms expected to appear in the ontology and then created definitions for each of the classes which were included in the final version of our ontology; (4) Define classes and class hierarchy: The classes and class hierarchy was maintained as present in the Trial Bank ontology; (5) Define properties of classes: Properties such as objective and data type properties as well as subclass relationship, domain, range, disjointness amongst sub-classes was defined while creating the ontology; (6) Define facets: Different facets such as cardinality, values, relationships, restrictions, inconsistencies was defined for the classes in the ontology; (7) Create instances: Instantiation of the ontology was specific to a single clinical trial; (8) Check for anomalies: We checked for inconsistencies after creating the ontology by using the Racer OWL tool (Racer Systems GmbH & Co. KG, 2004). The validation of the ontology was done by using the SPARQL query language, that would allow us to retrieve instances of the classes based on the information we require.

Observer Agreement

Initially, we tagged five randomly selected RCT's from the Journal of Neurotrauma, published within the last 5 years (2004–2008). We chose only one journal initially so that we could measure the ontology's accuracy based on the small sample, make changes if required and then apply it to all 15 articles from three journals. Three observers (JS, AZ, SP) tagged the five articles using the ontology. The tagging took an average of one hour per paper. We calculated inter-observer kappa-coefficients and overall percent agreement for all five articles for the three observers. Based on the analysis and further discussion, we modified the ontology to make its classes and definitions clearer and aligned with the goal of facilitating the process of meta-analysis. The three observers, two with a background of ontologies (AZ and SP) and third with a clinical background (JS), were then e-trained, that is over the internet by using video conferencing and Skype, on the ontology class and sub-class meanings. This was followed by a repeated evaluation of inter and intra observer reliability, for the articles from the Journal of Neurotrauma.

Thereafter, we tagged five randomly selected RCT's from both the Journal of Neurosurgery and Neurosurgery Journal, published within the last 5 years (2004–2008). We calculated the inter-observer agreement, overall percent agreement and intra-observer agreement for all the articles for each of the journals. Also, we analyzed the results and identified those elements which were used by two or all observers more frequently than the ones not used at all. Those which were used by only one of the observers were not considered to be frequent.

The inter and intra observer agreement was calculated using kappa statistic or coefficient. A kappa of one indicates perfect agreement, whereas a kappa of zero indicates agreement equivalent to chance. Agreement among observers was evaluated using kappa coefficient calculated with 95% confidence intervals and also by percent agreement. The R language (<http://www.r-project.org/>, version 2.8.1) was used for the inter & intra observer agreement calculations. The inter-observer agreement was calculated using Fleiss's kappa statistic as it is usually used when there are more than two observers. The intra-observer agreement was calculated using Cohen's kappa statistics, as it is usually used when there are two observers, in this case two events were considered of tagging the first time and then after one week.

Article Selection

Journals Included

The Journal of Neurosurgery (JoN), Neurosurgery journal (NJ) and the Journal of Neurotrauma (JoNT) were hand-searched to retrieve all RCT's published within the last 5 years. The operational definitions served as guidelines to choose or reject certain studies.

Design Definition and Inclusion Criteria

Randomized Controlled Trials which included detailed information about the study were selected in our analysis. Details such as a) trial design mentioning few or all of the following: prospective, randomized, double-blinded, placebo-controlled, multi-center trials; b) Hypothesis, objective and outcomes of the trials; c) Particulars of the experimental arm, comparison arm, drug and placebo used; d) Number of population enrolled, screened, randomized, excluded post randomization, calculation of the sample size as well as details of the randomization and the blinding involved; e) Explicit mention of the inclusion and exclusion rule for the population; f) Primary recruitment flowchart of the population; g) Mention of follow up activity, duration, adverse effects, patient safety; and h) Statistical analysis along with the results obtained.

Sample Selection

A sample of five articles from each journal was selected. The article details such as year of publication, journal volume, title and link to download the full text, were populated in a spreadsheet. Using the R language (<http://www.r-project.org/>, version 2.8.1), a random number sequence was generated and five articles from each journal were selected to form the sample. The articles were numbered starting from serial number one and the random number sequence was corresponded with the serial numbers in the spreadsheet.

Article Encoding

Three observers (JS, AZ, SP) tagged five articles, from respective journals, using the corresponding computational ontology, tagging being repeated after 1 week for evaluation of intra-observer agreement. A figure demonstrating how this tagging was performed using OntoClassTag is displayed in Fig. 2.

Center for Excellence in Research Reporting in Neurosurgery (CERR-N)

The goal of CERR-N is to provide a resource for clinical researchers in Neurosurgery so that the overall quality of the reporting of their manuscripts is improved at the time of the publication of their articles. This standardization is aimed at facilitating the subsequent extraction of information from each of these papers so that they can be included in future SR-MAs. Specifically, the CERR-N focuses on articles using the research design category for Randomized Controlled Trials (RCTs). We envision the services of CERR-N to be incorporated on a world-wide basis. Research projects, pertaining to neurosurgery, turn into research papers, which will be accepted by CERR-N. The articles will be reviewed at the point of submission to ensure consistent quality of articles present in CERR-N. The reporting standards, as specified by our ontology, will be applied to the papers and feedback will be provided to ensure that each manuscript contains the required main elements listed by the ontology. All articles collected at CERR-N will then be categorized, using the ontology, and the similar papers can be included in a potential meta-analysis. Consequently, this can help to construct clinical practice guidelines. The results can be used to apply for grants, for that research project.

For articles demonstrating a potential to be integrated into existing SR-MA, investigators will be invited to participate in the preparation of an updated SR-MA. For example, if a randomized controlled trial on low-back pain can be mapped to an existing SR-MA with the same focus, the CERR-N will initiate the review of that SR-MA while incorporating the new study once it is accepted for

publication. Since that publication has already been revised according to the ontology, and the ontology contains all the elements required for integration of the manuscript into a SR-MA, the productivity and quality of the updated process will increase. Researchers will also be invited to contribute with individual anonymized subject data, for a subject-level SR-MA. As a long-term strategy, our goal is to link the CERR-N to peer-reviewed journals in Neurosurgery so that the review process can be incorporated as an ancillary resource during their internal peer-review process.

Results

Ontology Structure

We identified 128 elements, which are required for SR-MA, from the Trial Bank ontology and included them in ours which are displayed in Table 1. The ones marked with an asterisk are the classes that we defined and the rest of the classes are the ones present in the Trial Bank ontology. We executed a number of queries in SPARQL to test whether they could accurately describe the point estimates, confidence intervals or measures of dispersion for each of the elements in a quantitative table for a meta-analysis.

A few examples of the query are as follows:

1. # Retrieve all the instances of the Trial class with their start and end dates

```
SELECT ?trial ?startDate ?endDate
WHERE { ?trial rdf:type rct:Trial .
        ?trial rct:hasStartDate ?startDate .
        ?trial rct:hasEndDate ?endDate .
}
```

2. # Find impossible trials (ending before they have begun)

```
SELECT ?trial ?startDate ?endDate
WHERE { ?trial rdf:type rct:Trial .
        ?trial rct:hasStartDate ?startDate .
        ?trial rct:hasEndDate ?endDate .
}
```

After analyzing the results, we identified 53 elements out of the 128 which were used more frequently, that is in about 70% of all 15 articles. Therefore we recommend that these articles should be included for the meta-analysis of RCTs. These elements are present in Table 1 and are marked in bold text. Figure 1 represents the schema of the ontology.

Observer Agreement

We found a high degree of agreement between the three observers. This is associated with a fair to poor level of agreement according to the Landis and Koch scale (Landis and Koch 1977). This confirms consistency amongst them of tagging a section of the manuscript with the same element from the ontology (Table 2, Fig. 2).

The pre and post modification results using overall percent agreement and kappa statistics, for the five articles from the Journal of Neurotrauma, amongst the three observers, are reported in Table 3. The average kappa coefficient before training was 0.173 and after training was 0.602. The average overall percent agreement before training was 37% and after training was 76%. These results indicated that after training the overall agreement amongst

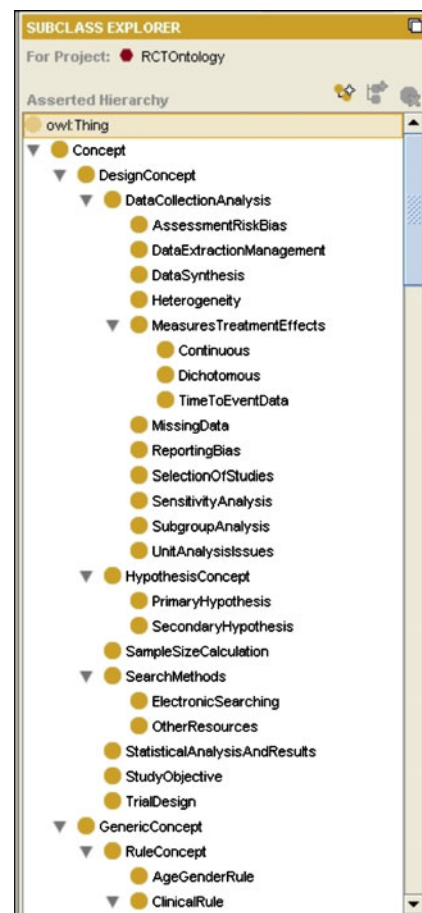


Fig. 1 Schema of the RCT ontology

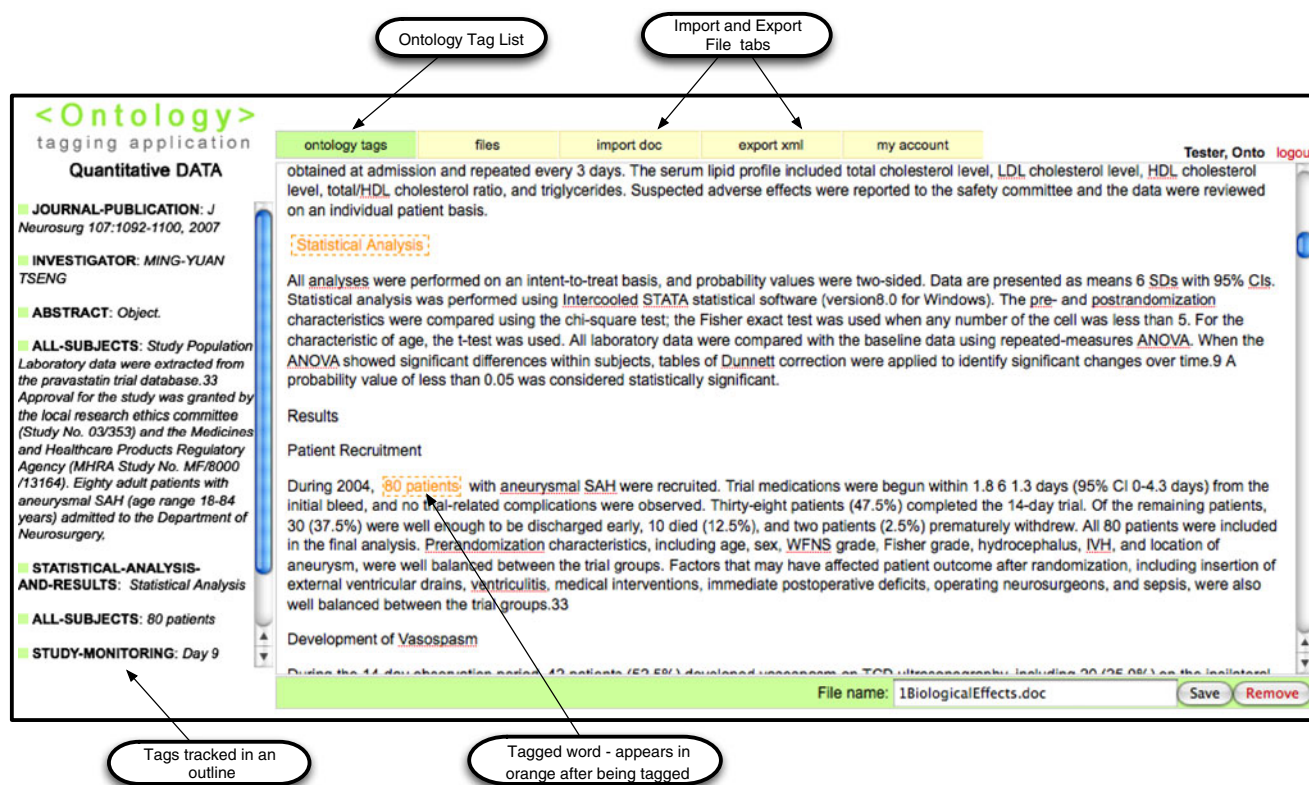


Fig. 2 Demonstrating the use of OntoClassTag to perform the tagging

observers increased significantly which suggests that training amongst observers is essential to obtain consistency in tagging an article.

Table 4 shows the inter-observer agreement and overall percent agreement calculated for all five articles of all three journals amongst the three observers. These results demonstrated high agreement amongst the observers which also confirms consistency, indicating the ontology's accuracy. Table 5 shows the intra-observer agreement, tagging repeated after 1 week interval, and overall percent agreement calculated for all five articles of all three journals for the three observers. This also indicates high degree of agreement within each individual observer and also confirms coding consistency over a period of time.

Discussion

In this study, we developed and evaluated an ontology specifically for Randomized Controlled Trials in order to facilitate the production of systematic reviews and meta-analysis. To our knowledge, this is the first study analyzing the agreement among different observers for the ontology classes. We found a high degree (Landis and Koch 1977) of inter and intra observer agreement which indicates the

ontology's accuracy and confirms consistency between observers to tag the same element for different sections of a given scientific manuscript.

Ontologies enable information sharing and reuse by making implied concepts and knowledge explicit and machine readable (Gruber 1993). The main reason why we chose to develop an ontology and use it to standardize reporting of clinical trials, is because they support the sharing and reuse of formally represented knowledge and common understanding of the structure of information, among people or software agents.

The ontology used in this study is a modified version of the Trial Bank ontology developed by Ida Sim which was previously validated by our group for the conduction of HIV meta-analysis (Cook et al. 2007). Sims ontology was designed to capture RCT information into an RCT Bank, a knowledge base designed to support systematic reviews and evidence-based practice. The Trial Bank ontology is the most complete trial ontology for trial interpretation and application to clinical care (Sim et al. 2004). It has an extensive range covering terms related to RCTs. Despite its great utility, it has many terms that are not necessary for conducting meta-analysis. Our goal was to develop a more focused ontology that could be used specifically for meta-analysis purposes. Therefore, during the adaptation of the ontology in this study, the elements

Table 1 Classes of the ontology demonstrating 128 different elements used to define RCT's

Secondary Study	Primary Recruitment Flowchart	Withdrawal Reason	Cost	Selective Reporting*
Trial	Randomized Population	Statistical Analysis and Results	Outcome Assessment	Blinding
Erratum	Enrolled Population	Sample Size Calculation	Category	Blinding Methods
Publication Details	Eligible Population	Trial Design	Range	Blinding Efficacy
Trial Entry Details	Screened Population	Primary Hypothesis	Kaplan Meier Timepoint	Drug Step
Administrative Details	Excluded Postrand Population	Secondary Hypothesis	All Comparisons at Time X	Non Drug Intervention Step
Conclusion Details	Not Randomized Population	Study Objective	Single Time X Comparison	Comparison Arm
Background Details	Not Enrolled Population	Study Monitoring	Cointervention Data Point	Experimental Arm
Stopping Details	Not Eligible Population	Electronic Searching*	Cross over population	Crossover
Retraction Details	All Subjects	Other Resources*	Data Point Population	Run In
Correction Details	Crossover Population	Assessment Risk Bias*	Kaplan Meier Datapoint	Washout
Fraud Details	Study Arm Population	Data Extraction Management*	Blinding Efficacy Datapoint	Other Intervention
Duration	Subgroup Population	Data Synthesis*	Baseline Datapoint	Procedure
Double Anchored	Site Enrollment	Heterogeneity*	Summary Datapoint	Device
Single Anchored	Follow Up Compliance	Continuous*	Regression Analysis and Results	Drug
Date	Follow Up Activity	Dichotomous*	Subgroup Regression	Usual Care
Time Point	Follow Up	Time To Event Data*	Outcome Value Entity	No Treatment
Exclusion Rule	Protocol Change	Missing Data *	Baseline	Placebo or Sham
Inclusion Rule	Treatment Assignment	Reporting Bias*	Primary Outcome	Cointervention
Age Gender Rule	Cluster Individual Protocol	Selecting of Studies *	Secondary Outcome	Compliance Result
Single or Irregular Event	Reason Outcome Not Assessed	Subgroup Analysis*	Ancilliary Outcome	Chain Logic
Regularly Recurring Event	Reason Excluded Postrand	Unit Analysis Issues*	Side Effect	Option Logic
Concurrent Situations	Reason Not Randomized	Rate	Allocation Risk*	Blocking Logic
Non Concurrent Situations	Reason Not Enrolled	Life Year	Blinding Risk*	Conditional Logic
State	Reason Not Eligible	Scored Instrument	Incomplete Outcome Data *	
Subgroup	Reason Off Assigned Intervention	Categorical	Potential Source Bias*	

Asterisk = *

not required for carrying out a meta-analysis were excluded. Specifically, terms related to administrative aspects were excluded. After the Trial Bank ontology was adjusted for this new purpose, it was initially used for tagging the content of five neurosurgery articles. Based on the analysis and further discussion, the ontology content was once more modified to make its classes and definitions more aligned for purpose of use in meta-analysis. This procedure was carried out in order to minimize inconsistencies on the ontology content, such as incomplete content and contradictions, problems commonly observed in ontologies (Rogers 2006).

The Epoch is a suite of clinical trial ontologies that define the vocabulary and semantics necessary to represent information on clinical trials (Shankar et al. 2007) and is composed by a group of seven ontologies (Clinical Trial Ontology, Protocol Ontology, Organization Ontology, Assay Ontology, Labware Ontology, Virtual Trial Data Ontology, Constraint Expression Ontology and Measurement Ontology)(Shankar et al. 2007). Similar to Sims ontology, the Epoch was specially developed for the use in RCTs, and therefore is not useful for the development of SR-MAs.

While the increasing standardization of results reporting for RCTs is contributing to the development

Table 2 List of 15 RCT's derived from the three journals

Journal	Article Title	Year of Publication
Journal of Neurotrauma	Effect of Rosuvastatin on Amnesia and Disorientation after Traumatic Brain Injury (NCT003229758) (Tapia-Perez et al. 2008)	2008
	A Single Dose, Three-Arm, Placebo-Controlled, Phase I Study of the Bradykinin B2 Receptor Antagonist Anatibant (LF16-0687Ms) in Patients with Severe Traumatic Brain Injury (Marmarou et al. 2005)	2005
	Efficacy of Standard Trauma Craniectomy for Refractory Intracranial Hypertension with Severe Traumatic Brain Injury: A Multicenter, Prospective, Randomized Controlled Study (Jiang et al. 2005)	2005
	Multiplex Assessment of Cytokine and Chemokine Levels in Cerebrospinal Fluid following Severe Pediatric Traumatic Brain Injury: Effects of Moderate Hypothermia (Buttram et al. 2007)	2007
	Severe Human Traumatic Brain Injury, but Not Cyclosporin A Treatment, Depresses Activated T Lymphocytes Early after Injury (Mazzeo et al. 2006)	2006
Journal of Neurosurgery	Dosing and safety of cyclosporine in patients with severe brain injury (Hatton et al. 2008)	2008
	Biological effects of acute pravastatin treatment in patients after aneurysmal subarachnoid hemorrhage: a double-blind, placebo-controlled trial (Tseng et al. 2007)	2007
	Effectiveness of neuronavigation in resecting solitary intracerebral contrast-enhancing tumors: a randomized controlled trial (Willems et al. 2006)	2006
	Clinical and radiographic analysis of cervical disc arthroplasty compared with allograft fusion: a randomized controlled clinical trial (Mummaneni et al. 2007)	2007
	Clazosentan (AXV-034343), a selective endothelin A receptor antagonist, in the prevention of cerebral vasospasm following severe aneurysmal subarachnoid hemorrhage: results of a randomized, double-blind, placebo-controlled, multicenter Phase IIa study (Vajkoczy et al. 2005)	2005
Neurosurgery Journal	Intraventricular thrombolysis speeds blood clot resolution: results of a pilot, prospective, randomized, double-blind, controlled trial (Naff et al. 2004)	2004
	Spinal cord stimulation versus repeated lumbosacral spine surgery for chronic pain: a randomized, controlled trial (North et al. 2005)	2005
	The effect of early isolated lumbar extension exercise program for patients with herniated disc undergoing lumbar discectomy (Choi et al. 2005)	2005
	Predictors of cerebral infarction in patients with aneurysmal subarachnoid hemorrhage (Ferguson and Macdonald 2007)	2007
	Comparison of open discectomy with microendoscopic discectomy in lumbar disc herniations: results of a randomized controlled trial (Righesso et al. 2007)	2007

of high quality ontologies, there are a number of problems that continue to plague their reliability. Studies evaluating ontologies reveal several areas from which errors tend to emerge including, philosophical rigor, ontological commitment, content correctness and fit for purpose (Rogers 2006), which can all impact the interoperability of ontologies. While these problem areas are increasingly addressed through quality assurance systems (Schulz et al. 1998), a surprising issue that has been largely overlooked is whether ontologies actually fulfill their intended purpose of reducing the number of conflicting definitions among different users (Fonseca and Martin 2007).

The utility and reliability of the ontology described here has some limitations. It has been shown that the inter-rater reliability can vary according to the level of the training of observers in a specific scale, with less experienced observers showing lower indexes because their ratings are less consistent (Andersen et al. 1989). Moreover, the use of a structured interview guide can increase the scale reliability in raters with low clinical experience (Crippa et al. 2002). In the present study, analysis of agreement between observers was carried out on two different moments: before and after the training of the ontology class and sub-class meanings. It was observed that the agreement rate before the training was lower. Despite the fact that the observers were experienced researchers, they

Table 3 Pre and post modification percent agreement and kappa statistics for five randomly selected RCT's from Journal of Neurotrauma

Article	Kappa coefficients and percent agreement before training	Kappa coefficients and percent agreement after training
1	0.044 (37.93%)	0.634 (79.31%)
2	0.242 (45.45%)	0.73 (86.67%)
3	0	0.663 (83.33%)
4	0.46 (64.10%)	0.515 (68.89%)
5	0.119 (40.74%)	0.472 (63.63)

Table 4 Overall percent agreement and Inter-Observer agreement for five randomly selected RCT's from Journal of Neurosurgery, Neurosurgery Journal and Journal of Neurotrauma

Journal	Overall Percent Agreement	Inter-observer Agreement
Journal of Neurotrauma	76.4% (63.6%–86.7%)	0.60 (0.47–0.73)
Journal of Neurosurgery	75.97% (75.31%–81.33%)	0.54 (0.50–0.58)
Neurosurgery Journal	74.9% (70%–79.71%)	0.53 (0.50–0.57)

had different research background: two had ontological background and the third had a clinical background, which could be contributing for the lower agreement rate. As expected, the agreement rate improved after the training, where the observers where instructed on the ontology terms, the meanings of the terms they did not understand were explained in detail. Therefore, in order to maintain the reliability, it is advisable that it be used by trained coders and it might not be generalizable to other researchers, unless they participate of a specific training for the use of this ontology, which would be advisable for most CERRs.

Table 5 Overall percent agreement and intra-observer agreement for five randomly selected RCT's from Journal of Neurosurgery, Neurosurgery Journal and Journal of Neurotrauma

Journal	Percent Agreement (%)	Kappa Coefficient
Neurosurgery Journal		
Article 1	76.5	0.8
Article 2	82.4	0.71
Article 3	85.7	0.8
Article 4	91.7	0.82
Article 5	78.6	0.83
Overall	89.8	0.79
Journal of Neurosurgery		
Article 1	77.3	0.81
Article 2	80	0.86
Article 3	82.4	0.81
Article 4	81	0.8
Article 5	83.3	0.83
Overall	80.8	0.82
Journal of Neurotrauma		
Article 1	88	0.8
Article 2	82.4	0.82
Article 3	88.9	0.76
Article 4	88.9	0.78
Article 5	84.6	0.81
Overall	86.56	0.79

However, the ontology could potentially be used in conjunction with text mining, which could expand its use without the need for trained observers.

Through the CERR-N we envision to raise the reporting quality of RCT articles, offering support to the researchers prior to the submission of the article in peer-reviewed journals. Manuscripts will be analyzed according our ontology, to ensure that each manuscript contains the required main elements. Since this support is offered in an advanced phase of the research project, it is possible that studies to be analyzed in the CERR-N do not meet the necessary criteria to be included in a SR-MA. This problem was already faced by Sim (Sim et al. 2004) on the development of the Trial Bank Project. During the first phase of the project, from the 108 RCTs published in JAMA (Journal of the American Medical Association), 54 were excluded due to modeling limitations of RCT Schema (Sim et al. 2004). Therefore, despite the positive impact the CERR-N might have on RCT reporting, it does not prevent RCT to be developed with methodological inconsistencies. This limitation must be addressed in future studies.

At the present moment, we do not have long-term outcomes of what the CERR-N can do in terms of improving the quality of RCTs on a broader scale. However, the present results indicate that it is a reliable ontology for the development of SR-MA. In order to make this ontology available and ensure its utilization, the CERR-N should be linked to peer-reviewed journals so that the review process can be incorporated as an ancillary resource during their internal peer-review process. Future studies should expand this ontology to articles involving other research designs, such as cohort and diagnostic studies. The expansion of CERR-N to other centers and establishing partnerships with peer reviewed journals, to increase overall reporting quality, and expedited translation of articles into clinical practice guidelines, could have a positive impact on the clinical research field as a whole.

Standardizing reporting for neurosurgery articles can be reliably achieved through the integration of a computational ontology within the context of a CERR-N. This approach holds potential for the overall improvement in the quality of reporting of Randomized Controlled Trials in neurosurgery, ultimately streamlining the translation of clinical research findings to improvement in patient care.

Information Sharing Statement

One of the resources used in our work is OntoClassTag, which is an online tagging/annotation application developed by the Research On Research group (www.researchonresearch.org), We used Ontotag for tagging the RCT's. It is accessible online by following the link: <http://www.ceso.duke.edu/tagging/servlet/Controller?cmd=default>. It requires a user-

name and password, which can be obtained by sending an email to the RoR group: contact@researchonresearch.org

Acknowledgments We would like to thank the Research on Research group (www.researchonresearch.org), Duke University, USA for manuscript writing templates. (Phadtare et al. 2009). We would also like to thank FS from the Cochrane group who helped us modify the ontology, by giving us input as to which classes would be required for a meta-analysis.

References

- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., et al. (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine*, *134*(8), 663–694.
- Andersen, J., Larsen, J. K., Schultz, V., Nielsen, B. M., Kørner, A., Behnke, K., et al. (1989). The brief psychiatric rating scale. Dimension of schizophrenia—reliability and construct validity. *Psychopathology*, *22*(2–3), 168–176.
- Biesen, W. V., Verbeke, F., & Vanholder, R. (2008). An infallible recipe? A story of cinnamon, soufflé and meta-analysis. *Nephrology, Dialysis, Transplantation*, *23*(9), 2729–2732.
- Bossuyt, P. (2004). The quality of reporting in diagnostic test research: getting better, still not optimal. *Clinical Chemistry*, *50*, 465–466.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., et al. (2004). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Family Practice*, *21*, 4–10.
- Buttram, S., Wisniewski, S., Jackson, E., Adelson, D., Feldman, K., Bayir, H., et al. (2007). Multiplex assessment of cytokine and chemokine levels in cerebrospinal fluid following severe pediatric traumatic brain injury: effects of moderate hypothermia. *Journal of Neurotrauma*, *24*(11), 1707–1718. doi:10.1089/neu.2007.0349.
- Choi, G., Raiturker, P. P., Kim, M. J., Chung, D. J., Chae, Y. S., & Lee, S. H. (2005). The effect of early isolated lumbar extension exercise program for patients with herniated disc undergoing lumbar discectomy. *Neurosurgery*, *57*, 764–772.
- Clarke, M. J., & Stewart, L. A. (1994). Obtaining data from randomised controlled trials: how much do we need for reliable and informative meta-analyses? *BMJ*, *309*(6960), 1007–1010.
- Cook, D. J., Greengold, N. L., Ellrodt, A. G., & Weingarten, S. R. (1997). The relation between systematic reviews and practice guidelines. *Annals of Internal Medicine*, *127*(3), 210–216.
- Cook, C., Hannley, M., Richardson, J. K., Michon, J., Harker, M., & Pietrobon, R. (2007). Real-time updates of meta-analyses of HIV treatments supported by a biomedical ontology. *Accountability in Research*, *14*(1), 1–18.
- Crippa, J. A. S., Sanches, R. F., Hallak, J. E. C., Loureiro, S. R., & Zuardi, A. W. (2002). Factor structure of Bech's version of the brief psychiatric rating scale in Brazilian patients. *Brazilian Journal of Medical and Biological Research*, *35*(10), 1209–1213.
- Ferguson, S., & Macdonald, R. L. (2007). Predictors of cerebral infarction in patients with aneurysmal subarachnoid hemorrhage. *Neurosurgery*, *60*, 658–667.
- Fonseca, F., & Martin, J. (2007). Learning the differences between ontologies and conceptual schemas through ontology-driven information systems. *JAIS - Journal of Association for Information Systems - Special Issue on Ontologies in the Context of IS*, *8*(2), 129–142.
- Fung, A. E., Palanki, R., Bakri, S. J., Depperschmidt, E., & Gibson, A. (2009). Applying the CONSORT and STROBE statements to evaluate the reporting quality of neovascular age-related macular degeneration studies. *Ophthalmology*, *116*(2), 286–296.
- Groves, T. (2008). Enhancing the quality and transparency of health research. *QBMJ*, *337*, a718.
- Gruber, T. (2008). Ontology. Entry in the Encyclopedia of Database Systems. Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, to appear in 2008. Provides a definition of ontology as a technical term for computer science, tracing its historical context from philosophy and AI.
- Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing.
- Haines, S. J. (1983). Randomized clinical trials in neurosurgery. *Neurosurgery*, *12*(3), 259–264.
- Han, C., Kwak, K. P., Marks, D. M., Pae, C. U., Wu, L. T., Bhatia, K. S., Masand, P. S., Patkar, A. A., et al. (2009). The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry. *Contemporary Clinical Trials*, *30*(2), 116–122.
- Hopewell, S., Altman, D. G., Moher, D., & Schulz, K. F. (2008). Endorsement of the CONSORT statement by high impact factor medical journals: a survey of journal editors and journal 'Instructions to Authors'. *Trials*, *9*, 20.
- Hripesak, G., & Heitjan, D. (2002). Measuring agreement in medical informatics reliability studies. *Journal of Biomedical Informatics Archive*, *35*(2), 99–110.
- Jarlais, D. C., Lyles, C., Crepez, N., & TREND Group. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Public Health*, *94*, 361–366.
- Jiang, J. Y., Wei, X. U., Li, W. P., Xu, W. H., Jun, Z., Bao, Y. H., et al. (2005). Efficacy of standard trauma craniectomy for refractory intracranial hypertension with severe traumatic brain injury: a multicenter, prospective, randomized controlled study. *Journal of Neurotrauma*, *22*(6), 623–628.
- Kane, R. L., Wang, J., & Garrard, J. (2007). Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *Journal of Clinical Epidemiology*, *60*(3), 241–249.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.
- Manchikanti, L., Derby, R., Wolfer, L., Singh, V., Datta, S., & Hirsch, J. A. (2008). Evidence based medicine, systematic reviews, and guidelines in interventional pain management, part I: introduction and general considerations. *Pain Physician*, *11*(2), 161–186.
- Manser, R., & Walters, E. H. (2001). What is evidence-based medicine and the role of the systematic review: the revolution coming your way. *Monaldi Archives for Chest Diseases*, *56*(1), 33–38.
- Marmarou, A., Guy, M., Murphey, L., Roy, F., Layani, L., Combal, J. P., et al. (2005). A single dose, three-arm, placebo-controlled, phase I study of the bradykinin B2 receptor antagonist Anatibant (LF16-0687Ms) in patients with severe traumatic brain injury. *Journal of Neurotrauma*, *22*(12), 1444–1455.
- Mazzeo, A. T., Kunene, N. K., Gilman, C. B., Hamm, R. J., Hafez, N., & Bullock, M. R. (2006). Severe human traumatic brain injury, but not cyclosporin a treatment, depresses activated T lymphocytes early after injury. *Journal of Neurotrauma*, *23*(6), 962–975.
- Meinert, C. L. (1998). Beyond CONSORT: need for improved reporting standards for clinical trials. Consolidated standards of reporting trials. *JAMA*, *279*(18), 1487–1489.
- Meinert, C. L., Tonascia, S., & Higgins, K. (1984). Content of reports on clinical trials: a critical review. *Controlled Clinical Trials*, *5*(4), 328–347.
- Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., et al. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*, *352*(9128), 609–613.

- Moher, D., Cook, D., Olkin, I., Stroup, D. F., & for the QUOROM Group. (2000). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. QUOROM Group. *The British Journal of Surgery*, 87(11), 1448–1454.
- Moher, D., Schulz, K. F., & Altman, D. G. (2001a). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, 357(9263), 1191–1194.
- Moher, D., Jones, A., Lepage, L., & CONSORT Group (Consolidated Standards for Reporting of Trials). (2001b). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*, 285(15), 1992.
- Moher, D., Simera, I., Schulz, K. F., Hoey, J., & Altman, D. G. (2008). Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. *BMC Medicine*, 6, 13.
- Mummaneni, P. V., Burkus, J. K., Haid, R. W., Traynelis, V. C., & Zdeblick, T. A. (2007). Clinical and radiographic analysis of cervical disc arthroplasty compared with allograft fusion: a randomized controlled clinical trial. *Neurosurgery Spine*, 6, 198–209.
- Naff, N. J., Hanley, D. F., Keyl, P. M., Tuhim, S., Kraut, M., Bederson, J., et al. (2004). Intraventricular thrombolysis speeds blood clot resolution: results of a pilot, prospective, randomized, double-blind, controlled trial. *Neurosurgery*, 54, 577–584.
- North, R. B., Kidd, D. H., Farrokhi, F., & Piantadosi, S. A. (2005). Spinal cord stimulation versus repeated lumbosacral spine surgery for chronic pain: a randomized, controlled trial. *Neurosurgery*, 56, 98–107.
- Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.
- Petersson, H., Gill, H., & Ahlfeldt, H. (2002). A variance-based measure of inter-rater agreement in medical databases. *Journal of Biomedical Informatics*, 35(5–6), 331–342.
- Phadtare, A., Bahmani, A., Shah, A., & Pietrobon, R. (2009). Scientific writing: a randomized controlled trial comparing standard and on-line instruction. *BMC Medical Education*, 9, 27.
- Pocock, S. J., Hughes, M. D., & Lee, R. J. (1987). Statistical problems in the reporting of clinical trials. A survey of three medical journals. *The New England Journal of Medicine*, 317(7), 426–432.
- Racer Pro 2.0 available at: <http://www.racer-systems.com/index.phtml>.
- Righesso, O., Falavigna, A., & Avanzi, O. (2007). Comparison of open discectomy with microendoscopic discectomy in lumbar disc herniations: results of a randomized controlled trial. *Neurosurgery*, 61, 545–549.
- Rogers, J. E. (2006). Quality assurance of medical ontologies. *Methods of Information in Medicine*, 45(3), 267–274.
- Sackett, D. (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, 312, 71–72.
- Schulz, E., Barrett, J. W., & Colin, P. (1998). Read code quality assurance from simple syntax to semantic stability. *Journal of the American Medical Informatics Association*, 5, 337–346.
- Shankar, R. D., Martins, S. B., O'Connor, M., Parrish, D. B., & Das, A. K. (2007). An ontology-based architecture for integration of clinical trials management applications. AMIA Annu Symp Proc. 661–665.
- Sim, I., & Detmer, DE. (2005). Beyond trial registration: a global trial bank for clinical trial reporting. *PLoS Medicine*, 2(11), e365.
- Sim, I., Owens, D. K., Lavori, P. W., & Rennels, G. D. (2000). Electronic trial banks: a complementary method for reporting randomized trials. *Medical Decision Making*, 20(4), 440–450.
- Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., et al. (2001). Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association*, 8(6), 527–534.
- Sim, I., Carini, S., Olasov, B., & Jeng, S. (2004). An ontology of randomized controlled trials for evidence-based practice: content specification and evaluation using the competency decomposition method. *Journal of Biomedical Informatics*, 37(2), 108–119.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., et al. (2000). Meta-analysis of observational studies in epidemiology a proposal for reporting. *JAMA*, 283, 2008–2012.
- Tapia-Perez, J. H., Sanchez-Aguilar, M., Torres-Corzo, J. G., Gordillo-Moscoso, A., Martinez-Perez, P., Madeville, P., et al. (2008). Effect of rosuvastatin on amnesia and disorientation after traumatic brain injury (NCT003229758). *Journal of Neurotrauma*, 25, 1011–1017.
- Tendal, B., Higgins, J. P., Jüni, P., Hrobjartsson, A., Trelle, S., Nüesch, E. et al. (2008). Observer variation when extracting data for the calculation of a standardized mean difference. Poster presentation at 16th Cochrane Colloquium, Freiburg, 3-7 October 2008.
- Tseng, M. Y., Hutchinson, P. J., Turner, C. L., Czornyka, M., Richards, H., Pickard, J. D., et al. (2007). Biological effects of acute pravastatin treatment in patients after aneurysmal subarachnoid hemorrhage: a double-blind, placebo-controlled trial. *Journal of Neurosurgery*, 107, 1092–1100.
- Vajkoczy, P., Meyer, B., Weidauer, S., Raabe, A., Thome, C., Ringel, F., et al. (2005). Clazosentan (AXV-034343), a selective endothelin A receptor antagonist, in the prevention of cerebral vasospasm following severe aneurysmal subarachnoid hemorrhage: results of a randomized, double-blind, placebo-controlled, multicenter Phase IIa study. *Journal of Neurosurgery*, 103, 9–17.
- Vranos, G., Tatsioni, A., Polyzoidis, K., Ioannidis, J. P. A., et al. (2004). Randomized trials of neurosurgical interventions: A systematic appraisal. *Neurosurgery*, 55(1), 18–26.
- Wilczynski, N. L. (2008). Quality of reporting of diagnostic accuracy studies: no change since STARD statement publication—before-and-after study. *Radiology*, 248(3), 817–823.
- Willems, P. W., Taphoorn, M. J., Burger, H., Berkelbach van der Sprenkel, J. W., & Tulleken, C. A. (2006). Effectiveness of neuronavigation in resecting solitary intracerebral contrast-enhancing tumors: a randomized controlled trial. *Journal of Neurosurgery*, 104, 360–368.
- Xu, L., Li, J., Zhang, M., Ai, C., & Wang, L. (2008). Chinese authors do need CONSORT: reporting quality assessment for five leading Chinese medical journals. *Contemporary Clinical Trials*, 29(5), 727–731.