Epilepsy Research 154 (2019) 116–123

Contents lists available at ScienceDirect

Epilepsy Research

journal homepage: www.elsevier.com/locate/epilepsyres

Review article

Assessing evidence quality in research reporting neurocognitive outcomes following paediatric temporal lobe surgery for epilepsy

Alice E. Flint[a], Mitch G. Waterman[b], Poppy Siddell[c], Alexandra L. Houston[c], Gayatri Vadlamani[d], Paul Chumas[e], Matthew C.H.J. Morrall[c,*]

[a] Physical Health Psychology, Brunswick Road, Cannock Chase Hospital, Cannock, Staffordshire, WS11 5XY, UK
[b] School of Psychology, University of Leeds, Leeds, LS2 9JT, UK
[c] Paediatric Neuropsychology, The Leeds Teaching Hospitals NHS Trust, E Floor, Martin Wing, Leeds General Infirmary, Great George Street, Leeds, LS1 3EX, UK
[d] Paediatric Neurology, The Leeds Teaching Hospitals NHS Trust, Leeds General Infirmary, Great George Street, Leeds, LS1 3EX, UK
[e] Department of Neurosurgery, G Floor, Jubilee Wing, Leeds General Infirmary, Great George Street, Leeds, LS1 3EX, UK

ARTICLE INFO

Keywords:
Research quality
Bias
Pediatric epilepsy surgery
Cognitive

ABSTRACT

Purpose: RCTs are the gold standard in determining intervention efficacy with journal impact factor assumed to index research quality. Flint et al's (2017) systematic review examined neurocognitive outcomes following paediatric temporal lobe epilepsy surgery. Retrieved evidence was restricted to non-RCTs, which pose greater risk of bias and thus diminish research quality. The current study evaluated risk of bias in sources retrieved by Flint et al. and explored whether impact factor related to research quality within this selected field.
Methods: Methodological and reporting bias was evaluated using categories of bias specified by Cochrane. The relationship between the identified number of biases and journal impact factors of retrieved sources was examined.
Results: All studies carried substantial risk for bias. Methodology bias included low sample size (76.71%; 56/73), risk of confounding cognitive outcomes due to failure to report pre-surgery neurocognitive data (21.92%; 16/73) and to determine whether patients were prescribed antiepileptic drugs at follow-up (53.42%; 39/73). Reporting bias included overstating claims based on findings (53.42%; 39/73), failure to report individual patient characteristics (66%; 33/50) and omitting the nature of surgical interventions (15.07%; 11/73). The number of sources of common bias within studies was not associated significantly with journal impact factor ($p = .878$).
Conclusion: This evaluation highlights risk of bias when sources are predominantly uncontrolled non-RCTs and provides evidence that journal impact factor is not a reliable indicator of quality within this field. Authors should limit bias in their methods and reporting of results, to ensure the highest quality evidence possible is used to inform treatment decisions and prognosis.

1. Introduction

A systematic review must assess the methodological quality of studies included, so that the evidence presented may be understood in the context of each study's risk of bias in its results or conclusions and the applicability of its findings (Higgins and Green, 2011). Increasing bias sizeably reduces the validity and reliability of research findings; therefore, conscious efforts must be made by authors to limit this factor (Ioannidis, 2005).

Cochrane reviews generally only use randomised control trials (RCTs), which are typically considered as the gold standard for evaluating intervention outcomes as this research design minimises the risk of bias (Centre for Reviews and Dissemination, 2009). However, emphasis on the use of RCTs becomes problematic when considering areas of research that lack sufficient numbers to accommodate large scale trials, such as temporal lobe surgery for an epilepsy in both adult (Krucoff et al., 2017) and paediatric populations (Flint et al., 2017). A large proportion of studies in this field are of single case experimental design which are limited to the extent they permit control for confounding factors. Single case studies and case reports are considered low in the hierarchy of evidence; although some case study designs, such as high quality n-1 trials, are now recognised by the Oxford Centre for Evidence-Based Medicine (OCEBM) as some of the highest quality evidence (OCEBM Levels of Evidence Working Group, 2011).

A systematic review conducted by Flint et al. (2017) examined evidence reporting neurocognitive outcomes following temporal lobe surgery for an epilepsy in childhood and no RCTs were retrieved. While this review

* Corresponding author.
E-mail address: m.morrall@nhs.net (M.C.H.J. Morrall).

https://doi.org/10.1016/j.eplepsyres.2019.03.013
Received 1 November 2018; Received in revised form 17 December 2018; Accepted 21 March 2019
Available online 30 March 2019
0920-1211/

found evidence to suggest that the majority of study participants remained stable in terms of neuropsychological outcomes following neurosurgical treatment, the authors specified that no quantified assessment could be conducted due to a combination of mixed study designs, incomparability of assessments used and study quality. Quality assessment of studies which determine surgical intervention outcomes in this patient cohort is fundamental to ensuring high quality evidence is used when clinicians and families are considering whether to proceed with neurosurgery. Epidemiological research indicates that 27 in every million children might benefit from resective surgery for a form of epilepsy (Berg et al., 2009). While this equates to approximately 405 children per year in the UK, only about 25% of such epilepsy surgery procedures took place prior to 2013 (NHS Commissioning Board, 2013). The few RCTs in adults and adolescents aged > 12 years reported improved quality of life and seizure reduction with surgical intervention compared to long-term drug treatment (Wiebe et al., 2001; Engel et al., 2012). However, these sources do not present adolescent participant outcomes separately from adult data, with Engel et al. (2012) not reporting quality of life outcomes for participants younger than 17 years old. Hence, these studies do not sufficiently identify outcomes specifically for paediatric populations. Whilst a trial of surgery versus drug treatment would not be ethically justified, cohort studies (i.e. where some patients have undergone surgery and others have not) may be able to provide important insights within this research area when clinical trials are not appropriate. Evidence suggests early surgical intervention for an epilepsy in infancy can reduce seizure frequency (Duchowny et al., 1998), resulting in improved neurocognitive outcomes within this sensitive developmental period (Loddenkemper et al., 2007; Cross, 2011). Thus, studying neurosurgical outcome is vital to furthering knowledge regarding the prevention, limitation, and management of cognitive impairment in paediatric patients following treatment for temporal lobe epilepsy.

When evaluating intervention outcomes in a field where studies are overwhelmingly case series design, existing guidance states non-RCT studies can be considered providing that the reviewer acknowledges the greater likelihood of bias and the need for more attention to be paid when assessing selection and reporting bias within these sources (Reeves et al., 2008). While most systematic reviews use quality assessment tools designed for evaluating controlled trials, many assessment tools to evaluate risk of bias in non-RCT studies exist (Zeng et al., 2015). However, these are often only appropriate for assessing a specific type of study design, e.g. examining case-controlled or cohort study designs using the Newcastle-Ottawa Scale (NOS; Wells et al., 2016). There is a lack of specific guidance regarding the quality assessment of uncontrolled case series, with no widely validated tool existing to appraise studies of this kind (Sanderson et al., 2007). Extant quality assessment tools which can be used with mixed study designs, such as Mixed Methods Appraisal Tool (MMAT; Pluye et al., 2011), can only appraise methodological quality and not reporting quality.

A standard which is frequently assumed to denote research quality is the impact factor of the journal in which a given study is published. Previous studies have suggested that methodological quality of clinical research appears to increase as a function of journal impact status (Saha et al., 2003; Ali et al., 2017). However, the measurements of quality used were not always comprehensive in these studies e.g. Saha et al. (2003) asked participants to rate methodological quality of studies on a scale of 1–10 with no distinction between different types of potential bias. Equally, contradictory findings exist which highlight limitations of using journal impact factor as a gauge of methodological quality (Kurmis, 2003). Thus, evidence is lacking to support this assumption.

Undertaking quality assessment judgments based upon the Cochrane (2013) categories of bias (Cochrane Consumers and Communication Review Group, 2013), this paper aimed to appraise evidence for the cognitive outcomes following paediatric temporal lobe surgery for epilepsies, allowing greater insight into the potential biases of research areas dominated by non-RCT studies. This study also examined whether a meaningful correlation existed between common biases within publications and journal impact factor. Based on previous literature (e.g. Saha et al., 2003), it was hypothesised that there would be less bias observed within studies published in higher impact journals. To the authors' knowledge, this relationship has not been investigated within this selected research field. The current study explored the quality of research within this field and did not investigate or comment on surgery efficacy.

2. Methods

2.1. Data source

Seventy-three articles were sourced from a published systematic review (Flint et al., 2017), which examined evidence reporting neurocognitive outcomes following temporal lobe surgery for an epilepsy in childhood. Of these, 45 (62%) were uncontrolled retrospective case series, 20 (27%) were case reports, three (4%) were mixed longitudinal case series with cross-sectional data from comparison with a chronic epilepsy control group, two (3%) were longitudinal case series data with cross-sectional data from a comparison group of healthy young people, one (1%) was a single case study with a healthy control group, one (1%) was a single case study with the child's twin as a control participant and one (1%) study was a prospective cohort study with a chronic epilepsy control group.

2.2. Quality assessment

A number of quality assessment measures for observational studies exist (Downs and Black, 1998; Wells et al., 2000; Slim et al., 2003; Sterne et al., 2014). However, these generally contain items that were considered to be inapplicable for uncontrolled studies, such as allocation concealment. Therefore most sources would have received low scores, with difficulty in distinguishing quality between sources.

Due to a lack of specific guidance regarding the quality assessment of uncontrolled case series, risk of methodological and reporting bias were assessed based upon the main categories of bias recommended by Cochrane (Cochrane Consumers and Communication Review Group, 2013), which are suitable for all research designs. Each study's risk of bias was rated in a consistent and structured manner using the following Cochrane categories: sample bias (representativeness of cohort, selection bias); attrition bias (loss to follow-up); confounding bias (or performance bias e.g. comorbidities, concurrent treatments, poorly defined predictive factors); measurement bias (detection bias, validity of outcome measurement); validity of reporting/claims made and notable biases or threats to validity. Some methodological quality items commonly rated within reviews, like allocation concealment or masking of outcomes, were logically omitted due to these items being typically reported in RCTs and not present in any of the selected studies.

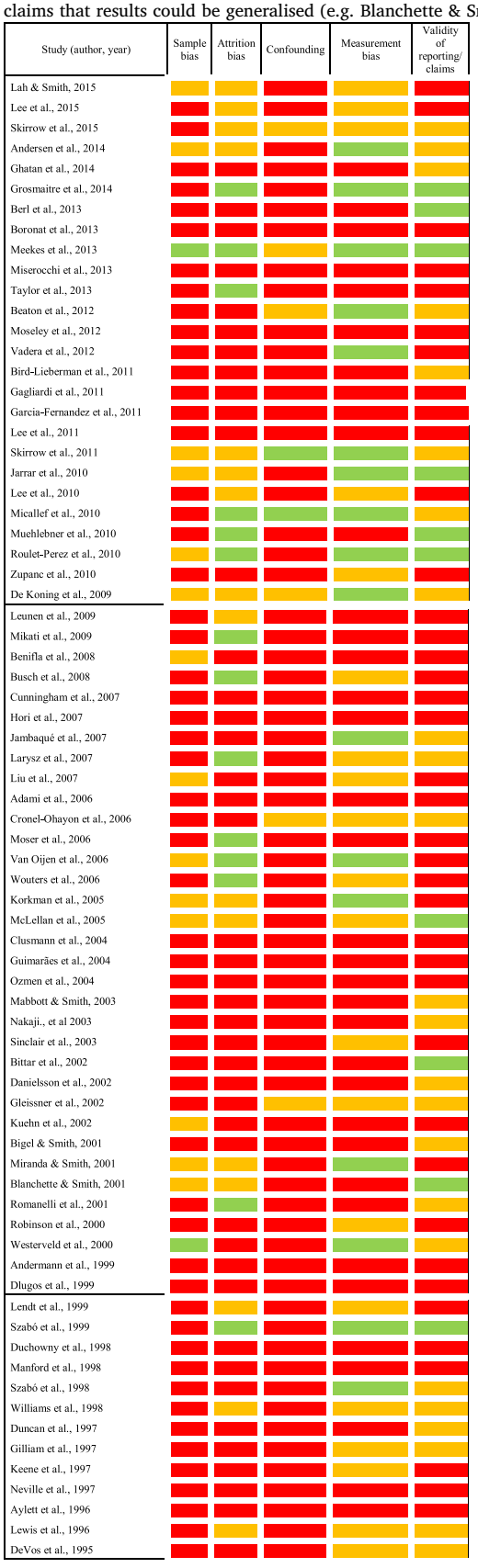
Each paper was scored as high, partial/unclear or low risk of bias, for each of the categories of bias. The frequency and severity of incidences of bias were used as a way of gauging the degree of bias present. This scoring method is consistent with that adopted in a previous published systematic review (Smithson et al., 2013) and follows the initial scoring structure present in existing quality assessment scales (Pluye et al., 2011; Guo et al., 2016). These ratings of bias were presented visually in a traffic light scheme (Table 1), which provided a visual overview of whether each study scored high (red), partial/unclear (amber) or low (green) for each of the categories of bias.

A number of specific potential sources of bias were observed repeatedly in studies, e.g. an absence of pre- and post-intervention assessment. The total number of these common biases present was calculated for each source (Table 2). Three non-blinded researchers undertook quality appraisal (AEF, MGW, MCHJM). Researchers performed their ratings independently prior to meeting and then discussed the ratings for each individual paper. Consensus was reached for all ratings.

2.3. Statistical analysis

A Spearman Rank-order correlation coefficient was used to

Table 1
Risk of bias ratings for all studies, presented using a traffic light scheme. Note red = high risk of bias; amber = unclear/partial risk of bias; green = low risk of bias. Those at low risk were only rated as such because they made no clear claims that results could be generalised (e.g. Blanchette & Smith, 2001).



investigate the association between the average number of common bias sources per study and the impact factor of the journal in which they were published. Impact factors were obtained from the 2014 Journal Citation Reports (Thomson Reuters, 2015). 1/73 studies was obtained from a book chapter (Berl et al., 2013) and thus was not included in the correlation analysis. The average number for common biases across papers published in the same journal was computed. Therefore, each journal had an average number of common biases. These values were then correlated with impact factor of the included journals.

3. Results

3.1. Assessment of main categories of bias

The ratings for risk of bias for each of the included studies are provided in Table 1.

3.1.1. Sample bias

The majority of studies (65/73; 89.04%) were retrospective uncontrolled observational case series or single case reports, with selection method not always indicated. 56/73 (76.71%) had small sample sizes of children having temporal lobe surgery (n < 30).

3.1.2. Attrition bias

As most studies were retrospective, the majority (63/73; 86.30%) did not report participants lost to follow-up. Nearly all studies (68/73; 93.15%) outlined the follow-up duration post-surgery. Follow-up durations differed greatly both between and within studies, ranging from two months to 27 years.

3.1.3. Confounding bias

Of the 73 studies, 16 (21.92%) reported post-surgical outcomes in the absence of any pre-surgical assessment. In contrast, 15/57 (26.32%) of the studies reported both pre- and post-surgical outcomes but did not report any statistical analysis of the differences, if any, between the two assessment points.

3.1.4. Measurement bias

Psychosocial outcomes were reported in 13/73 (17.80%) studies without quantifying or operationalizing the outcome, based on clinical observation or parent report, for example stating that speech 'improved' (Romanelli et al., 2001) or 'behaviour improved dramatically' (Nakaji et al., 2003) without mention of where this information originated or providing examples of the changes observed in behaviour. Differing types of assessments were used for each outcome domain, each with their own validity, reliability and risks of bias. 32/73 studies (43.84%) explicitly stated that the interval between neurocognitive assessments were less than one year. 29 of these studies (90.63%) did not account for/acknowledge potential practice effects which may have occurred when neuropsychological assessments were repeated in a short period of time, which may also have introduced measurement bias (Chelune et al., 1993).

3.1.5. Reporting bias

Discounting studies that were single-case design (n = 23), 33/50 of the remaining studies (66%) did not report individual participant characteristics alongside individual outcome data (e.g. Jarrar et al., 2002). Two papers did not report the age range or mean age at surgery for children undergoing temporal resection (Gagliardi et al., 2011; Robinson et al., 2000). The nature of the temporal resection was not reported in 11 (15.07%) studies. Furthermore, despite commenting on cognitive skills and psychosocial outcomes, only four (5.48%) studies reported factors of key importance to these domains, such as family functioning and socioeconomic status of the children. Equally, only 39 (53.42%) studies reported whether or not children were prescribed antiepileptic drugs (AEDs) at follow-up, and in those studies reporting that children were prescribed AEDs at follow up, none of these reported any risk of confounding

Table 2
Common sources of bias identified within each study and journal of publication. *Note. X indicates presence of bias.*

Author, Year	Retrospective design	Unvalidated measures	No pre- and post-assessment	Variable follow-up	No control group	Pooling across surgery types	Patient details omitted	Sample size of N < 3	Journal
Lah & Smith 2015	X			X	X		X		Epilepsy & Behavior
Lee et al 2015	X			X	X				Pediatric Neurology
Skirrow et al 2015	X			X					Brain
Andreson et al 2014	X			X	X		X		Frontiers in Neurology
Ghatan et al 2014	X	X			X	X			Journal of Neurosurgery: Pediatrics
Grosmaitre et al 2014	X							X	Neurocase
Berl et al., 2013	X				X		X	X	Book chapter
Boronat et al 2013	X	X	X		X			X	Childs Nervous System
Meekes et al., 2013									Epilepsy Research
Miserocchi et al 2013	X			X	X				Journal of Neurosurgery: Pediatrics
Taylor et al 2013	X	X	X		X			X	Journal of Neurosurgery: Pediatrics
Beaton et al 2012	X			X	X				Seizure
Vadera et al 2012	X			X	X				Journal of Neurosurgery: Pediatrics
Moseley et al 2012	X	X	X		X			X	Journal of Child Neurology
Bird-Lieberman et al 2011	X	X	X	X	X			X	Journal of Neurosurgery: Pediatrics
Gagliardi et al., 2011	X			X	X		X		Arquiva Neuropsiquiatra
Garcia-Fernandez et al 2011	X	X		X	X	X	X		Seizure
Lee et al 2011	X	X	X	X	X		X		Childs Nervous System
Skirrow et al 2011	X			X					Neurology
Lee et al 2010	X			X	X				Childs Nervous System
Micallef et al 2010			X	X					Epilepsia
Muehlebner et al 2010	X	X			X			X	Epilepsy Research
Roulet-Perez et al 2010	X			X	X				Epilepsia
Zupanc et al., 2010	X			X	X	X	X		Pediatric Neurology
de Koning et al 2009	X				X				Epilepsia
Leunen et al 2009	X		X	X			X		Epilepsy & Behavior
Mikati et al 2009	X	X			X			X	Epilepsy & Behavior
Benifla et al 2008	X		X	X	X				Epilepsy Research
Busch et al 2008	X			X	X			X	Epileptic Disorders
Cunningham et al 2007	X	X			X		X	X	Journal of Developmental & Behavioral Pediatrics
Hori et al 2007	X			X	X			X	Journal of Neurosurgery
Jambaqué et al 2007	X			X	X		X		Neuropsychologia
Jarrar et al., 2007	X		X	X	X		X		Neurology
Larysz et al 2007	X	X			X	X		X	Childs Nervous System
Liu et al 2007	X			X	X		X		Brain & Development
Adami et al 2006	X	X	X		X			X	Acta Psychiatrica Scandinavica
Cronel-Ohayon et al 2006	X							X	Neuropediatrics
Moser et al 2006	X				X			X	Acta Paediatrica
Van Oijen et al 2006	X			X	X		X		European Journal of Pediatric Neurology
Wouters et al 2006	X				X			X	Developmental Medicine & Child Neurology
Korkman et al 2005	X				X		X		Pediatric Neurology
McLellan et al 2005	X	X		X	X		X		Developmental Medicine & Child Neurology
Clussman et al 2004	X				X				Neurosurgery
Guimarães et al 2004	X	X			X		X	X	Epilepsy & Behavior
Ozmen et al 2004	X	X	X		X			X	Epilepsy & Behavior

(*continued on next page*)

Table 2 (*continued*)

Author, Year	Retrospective design	Unvalidated measures	No pre- and post-assessment	Variable follow-up	No control group	Pooling across surgery types	Patient details omitted	Sample size of N < 3	Journal
Mabbott et al 2003	X			X	X		X		Neuropsychologia
Nakaji et al 2003	X	X	X		X			X	Pediatrics
Sinclair et al 2003	X				X				Pediatric Neurosurgery
Bittar et al 2002	X	X		X	X				Journal of Clinical Neuroscience
Danielsson et al 2002	X				X				Epilepsy & Behavior
Gleissner et al 2002	X			X	X		X		Epilepsy Research
Kuehn et al 2002	X			X	X		X		Child's Nervous System
Bigel & Smith 2001	X			X	X		X	X	Brain and Cognition
Blanchette & Smith 2001	X			X	X		X		Brain and Cognition
Miranda & Smith 2001	X			X	X				Epilepsy & Behavior
Romanelli et al 2001	X	X	X		X			X	Neurosurgery
Robinson et al., 2000	X				X		X		Journal of Neurosurgery
Westerveld et al 2000	X			X	X		X		Journal of Neuropsychology
Andermann et al 1999	X	X			X			X	Epilepsia
Dlugos et al 1999	X			X	X				Pediatric Neurology
Lendt et al., 1999	X								Epilepsia
Szabó et al 1999	X			X	X				Pediatric Neurology
Duchowny et al., 1998	X	X		X	X		X		Epilepsia
Manford et al 1998	X	X			X			X	J. Am. Acad. Child Adolesc. Psychiatry
Szabó et al 1998	X			X	X				Epilepsia
Williams et al 1998	X			X	X				Pediatric Neurology
Duncan et al 1997	X	X	X	X	X		X		Pediatric Neurosurgery
Gilliam et al 1997	X			X	X	X			Neurology
Keene et al 1997	X		X	X	X	X			Child's Nervous System
Neville et al 1997	X	X			X			X	Pediatric Neurology
Aylett et al 1996	X	X	X		X			X	European Child & Adolescent Psychiatry
Lewis et al 1996	X			X	X				Journal of Epilepsy
DeVos et al, 1995	X			X	X	X			Neurology

cognitive outcomes of participants.

3.1.6. Validity of claims made

39 (53.42%) studies were at high risk of overstating claims based on the results, 26 (35.62%) at medium risk and eight (10.96%) at low risk. Those at low risk were only rated as such because they made no clear claims that results could be generalised (e.g. Blanchette & Smith, 2001). Most papers overstated the extent to which the outcomes could be attributed to the surgery, without controlling adequately for confounding factors. For example, Zupanc et al. (2010) claimed that Quality of Life (QoL) improved after surgery, when there was no reported pre-surgical assessment of QoL.

3.2. Common sources of bias in included studies

A number of more specific potential sources of bias were observed repeatedly in studies. These are displayed in Table 2, presenting which studies had each common source of bias and the journal in which they were published.

Only one study (Meekes et al., 2013) had none of the common sources of bias displayed in Table 2 and one study (Lendt et al., 1999) had one. Nine studies had two of the common bias sources, 16 had three, 25 had four, 17 had five and four had six bias sources.

3.3. Common sources of bias and publication impact factor

No significant association [r_s (32) = -0.028, p = 0.878] was found between journal impact factor and the average number of common bias sources in their studies (Fig. 1).

4. Discussion

Utilising a quality assessment approach which could suitably evaluate risk of bias in non-RCTs, all selected studies reporting neurocognitive outcomes of paediatric temporal lobe epilepsy surgery were found to convey substantial risks of bias. Some bias may have been caused by features of the study design, some by the quality of reporting. The conclusions of uncontrolled case studies are not necessarily biased, providing they do not make claims of generalizability or causation, have a well-defined cohort and assessment procedures used are adequately reported, as defined by STROBE (Strengthening the reporting of observational studies in epidemiology (STROBE) checklist v.4, 2007). Equally, studies can have low risk of bias but still be low quality evidence.

4.1. Impact of bias in the selected field

Generally, study methods did not control adequately for other variables that may confound intervention outcomes. As Hermann et al. (2002) note, studies measuring cognitive outcome of paediatric epilepsy surgery rarely include a control group of age-matched children with comparable epilepsies, who undergo neuropsychological assessment at the same time-points as the surgical group. Therefore, it is not possible to disentangle the effects of surgery fully from expected neurocognitive development during that time without surgery. For some studies, there was a noticeable lack of pre-surgical assessment data, so the reported outcome data may have been due to pre-existing differences between young people. Additionally, 90.63% of the studies which measured neurocognitive abilities repeatedly within a period of less than one year did not account for potential practice effects, which could confound the results of intervention. Studies generally lacked enough participants for the sample to be considered representative of the wider cohort of paediatric patients with temporal lobe epilepsy. Follow-up durations differed greatly between studies; with period of follow-up often being inappropriately short for the variables being measured, for example, intellectual trajectory which is likely to change over the long term as the young person develops.

As individual participant data were absent frequently, it was not possible to disaggregate outcome information for participants with different resections, ages or other characteristics. If individual outcomes are not reported, there is the possibility of bias as it is not possible to know if other features of the patients, beyond those reported of the group as a whole, might have influenced the outcomes. Conceivably, in not reporting individual outcomes, there is a possibility of a general conclusion being derived which may not represent variability in outcomes between individuals. Key demographic variables, such as age and side of surgery, were occasionally not reported meaning that the results for these participants cannot be generalised. In a large proportion of studies, the nature of the surgical intervention or assessments administered was not reported in specific detail, limiting study replicability. Moreover, a thorough history of other interventions, including AEDs, was not presented. By not reporting whether or not children are prescribed AEDs at follow-up, this poses the risk of confounding the neurocognitive outcomes for these participants. For instance, cognitive outcomes could be influenced by unwanted side effects of AEDs in decreasing neurocognitive function (Loring et al., 2007; Ijff and Aldenkamp, 2013). Despite this, a high proportion of studies overstated claims regarding post-surgical outcomes based on their results.

The biases observed have specific implications for the treatment of paediatric epilepsies. The quality of the existing evidence base indicates

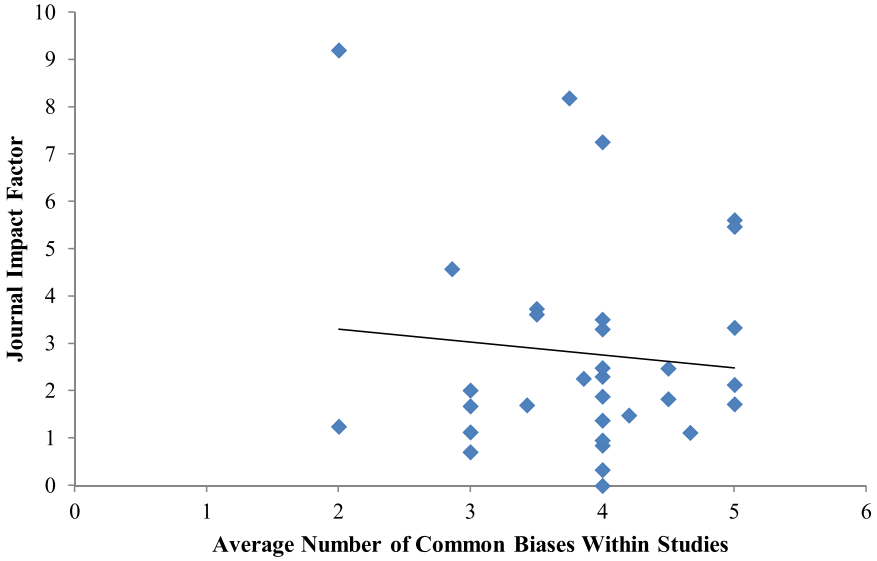


Fig. 1. Scatter graph with trend line to show the relationship between journal impact factor and the average number of common bias sources in their studies.

that caution should be applied when communicating previous findings to families, particularly when making inferences regarding whether an intervention is effective or the risks it poses. As 27 in one million children with a form of epilepsy might benefit from resective surgery (Berg et al., 2009), it is important that accurate evidence is used when clinicians and parents are considering whether or not to proceed with this treatment. Identifying sources of bias will help determine methodological solutions to improve the quality of future research, even if these cannot be implemented practically within non-RCT studies (but in these cases they should still be reported). Improved reporting will mean that replication is easier and transparent, and comparable evidence is used in clinical decision-making.

The average number of identified biases per paper from each journal did not correlate significantly with journal impact factor. This finding casts doubt upon the assumption that may be made by some that impact factor represents evidential quality (Saha et al., 2003), at least for literature concerning neurocognitive outcomes following paediatric epilepsy surgery. Publication in itself (as a study has successfully passed peer review) and the impact factor of the journal in which a study is published may lead readers to expect lower levels of bias within these sources. Future inquiries should examine the relationship between journal impact factor and risk of bias across diverse research domains. This would avoid impact factor being used habitually as a proxy for research quality. Quality assessments of individual studies, regardless of the journal in which they are published, should be undertaken when determining intervention outcomes.

4.2. Study limitations and wider implications

A potential limitation of the current study may be that risk of bias was identified as a function of quality judgment, but not in the context of a standardised 'tool'. However, this appraisal approach was performed systematically in accordance with the Cochrane (2013) categories of bias, due to extant tools being found to be unsuitable when sources retrieved were predominantly uncontrolled case studies. This raises the question as to whether or not current assessment tools are sufficient in the information that they can convey, when reviewing non-RCTs of varying designs. Although efforts have been made in recent years to validate a quality appraisal tool that can be used with diverse study designs (Pluye et al., 2011) and with exclusively case series studies (Guo et al., 2016), reviewers may still experience difficulties in selecting the most appropriate tool in research fields where a combination of non-RCT observational study designs must be evaluated. Therefore, a limitation in the literature is that there is no widely accepted and validated quality assessment tool for assessing both methodological and reporting quality of uncontrolled non-RCT studies.

This study also raises epistemological questions about what is considered as evidence and how this is disseminated. The nature of publication in clinical research may contribute to evidence being inappropriately reported or cited. Restrictions during the publication process (e.g. word limits) may result in acknowledgement of methodological flaws or study limitations being removed from final published works. Thus, reporting bias may occur in sound empirical research. Although most papers in the current study documented their limitations and used tentative language when drawing their conclusions, their results may then be cited as supporting evidence by further studies and clinicians communicating with families without reference to their methodological shortfalls (Spencer and Huh, 2008) and used to support stronger conclusions. This may contribute to the emergence of a consensus that has a spurious certainty which is not based on evidence at the ground level of research.

When publishing in scientific journals there is often a requirement for authors to provide novel answers to research questions, in order for such research to appear meaningful. This could lead to overstating claims or only presenting some outcomes with significant results (Ioannidis et al., 2014). As highlighted in the current study, conclusions of published case series may overstate their findings and imply causal links which cannot be validly inferred from their results due to the methods used, or to inadequate reporting. Shared outcome measures and open reviews, in which multiple trials make their datasets available for individual participant analysis (IPA) (Stewart and Tierney, 2002; Vale et al., 2015), would provide a way for individual centres to contribute to knowledge and report their outcomes without the need to overstate claims or present data selectively. This requires transparency and cooperation, which may require a culture shift within scientific publishing, as this field sometimes rewards competition and prestige (Ioannidis et al., 2014).

4.3. Recommendations

In view of the issues outlined above, clinical research in its entirety needs to improve the reporting of outcome data. A key mechanism by which this can be achieved is not to avoid the publication of uncontrolled case series data, but to ensure that vital pieces of information are included routinely in reporting. Specifically, the inclusion of supplementary materials containing anonymised individual participant data, to ensure claims are sufficiently supported. Reporting of outcomes from individual centres will enable the development of a larger dataset of surgical outcomes and should permit greater transparency for quality assurance and service improvement. Indeed there has been recent movement towards combining results from multiple research groups and clinical centres, e.g. the Human Connectome project (Van Essen et al., 2013). Numerous surgical centres were not represented in the included studies, suggesting that many centres do not publish their outcomes routinely, as noted by Shastin et al. (2015). The reporting of routine outcome data represents a valuable opportunity for practice-based evidence to complement the results of trials, which have strong internal validity but may lack external validity (Barkham et al., 2010). Thus, despite methodological and reporting limitations of the studies included in this review, centres should be commended for publishing their data. The following recommendations are suggested:

- Publications of case studies should aim to be concordant with STROBE (2007) so that study aims and methodology are transparent, with this consideration being undertaken by authors and the peer review process.
- Studies should provide individual participant data, ensuring claims are adequately supported and to enable findings to be related to homogenous samples.
- Abstracts should provide relevant participant information (e.g. age range, intervention).
- Studies should report their drop-out rate with reasons, if known, and characteristics of those who dropped-out where possible.
- Studies should report the method of outcome assessment and use agreed definitions when referring to concepts such as Quality of Life, rather than using terms informally.
- Authors should state if other outcomes of the reported cohort of patients have been reported elsewhere, so that this is readily accessible.
- Efforts should be made to conduct pre- and post-assessments where suitable, in order to accurately capture change in patient outcomes and control for pre-surgical ability.
- Authors should also consider the impact of practice effects where assessments are conducted on several occasions, particularly when repeated within a period of less than one year.
- Given restrictive word limits that may be required by some journals, a solution could be the use of online stores in order to adequately disseminate more detailed information (e.g. individual participant data).

While several of these recommendations echo extant best practices which are already widely communicated, this evaluation indicates that these are not always applied.

5. Conclusions

This evaluation highlights the risk of bias when sources are predominantly non-RCTs and provides evidence that journal impact factor

is not a reliable indicator of quality within this specific research domain. As selected papers had all been published in peer-reviewed journals, and in some cases high impact factor journals, reviewers should avoid using impact factor as a proxy for research quality when examining outcomes of surgical interventions for paediatric epilepsies. Authors should make the utmost effort to limit bias in their methods and reporting of results, to ensure high quality evidence is used to inform clinical decision-making. Adherence to STROBE guidelines and the recommendations proposed above should improve future dissemination quality. This is recommended within the field of paediatric epilepsy intervention research and generally in the clinical sciences.

Declarations of interest

None.

Acknowledgements

None.

The research did not receive any special grant from funding agencies in the public, commercial or not-for-profit sectors.

References

Ali, U.A., Reiber, B.M., Joren, R., van der Sluis, P.C., Gooszen, H.G., Boermeester, M.A., Besselink, M.G., 2017. Journal impact factor and methodological quality of surgical randomized controlled trials: an empirical study. Langenbeck's Arch. Surg. 1–8.

Barkham, M., Hardy, G.E., Mellor-Clark, J. (Eds.), 2010. Developing and Delivering Practice-Based Evidence. Wiley, Chichester.

Berg, A.T., Mathern, G.W., Bronen, R.A., Fulbright, R.K., DiMario, F., Testa, F.M., Levy, S.R., 2009. Frequency, prognosis and surgical treatment of structural abnormalities seen with magnetic resonance imaging in childhood epilepsy. Brain 132 (10), 2785–2797.

Berl, M., Kenealy, L., Salpekar, J., 2013. A Case Study: Seizure Disorders Psychopathology of Childhood and Adolescence: A Neuropsychological Approach. Springer Publishing Co; US., New York, NY, pp. 723–727.

Centre for Reviews & Dissemination (CRD), 2009. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. University of York, Centre for Reviews and Dissemination.

Chelune, G.J., Naugle, R.I., Lüders, H., Sedlak, J., Awad, I.A., 1993. Individual change after epilepsy surgery: Practice effects and base-rate information. Neuropsychology 7 (1), 41.

Cochrane Consumers and Communication Review Group, 2013. Cochrane Consumers and Communication Review Group: Meta-Analysis. Retrieved from. http://cccrg.cochrane.org.

Cross, J.H., 2011. Epilepsy Surgery in Children: Are We Doing Enough? Epilepsy Professional. Retrieved from. http://www.epilepsy.org.uk/campaigns/commissioning-paediatric-epilepsy-surgery/are-we-doing-enough-professor-helen-cross.

Downs, S.H., Black, N., 1998. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J. Epidemiol. Commun. Health 52 (6), 377–384.

Duchowny, M., Jayakar, P., Resnick, T., Harvey, A.S., Alvarez, L., Dean, P., Gilman, J., Yaylali, I., Morrison, G., Prats, A., Altman, N., 1998. Epilepsy surgery in the first three years of life. Epilepsia 39 (7), 737–743.

Engel, J., McDermott, M.P., Wiebe, S., Langfitt, J.T., Stern, J.M., Dewar, S., Sperling, M.R., Gardiner, I., Erba, G., Fried, I., Jacobs, M., 2012. Early surgical therapy for drug-resistant temporal lobe epilepsy: a randomized trial. Jama 307 (9), 922–930.

Flint, A.E., Waterman, M., Bowmer, G., Vadlamani, G., Chumas, P., Morrall, M.C.H.J., 2017. Neuropsychological outcomes following paediatric temporal lobe surgery for epilepsies: evidence from a systematic review. Seizure 52, 89–116.

Gagliardi, I.C., Guimarães, C.A., Souza, E.A., Schmutzler, K.M., Guerreiro, M.M., 2011. Quality of life and epilepsy surgery in childhood and adolescence? Arq Neuro-psiquiatr 69 (1), 23–26.

Guo, B., Moga, C., Harstall, C., Schopflocher, D., 2016. A principal component analysis is conducted for a case series quality appraisal checklist. J. Clin. Epidemiol. 69, 199–207.

Hermann, B., Seidenberg, M., Bell, B., Rutecki, P., Sheth, R., Ruggles, K., Wendt, G., O'Leary, D., Magnotta, V., 2002. The neurodevelopmental impact of childhood-onset temporal lobe epilepsy on brain structure and function. Epilepsia 43 (9), 1062–1071.

Higgins, J.P.T., Green, S. (Eds.), 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011]. The Cochrane Collaboration, 2011 Retrieved from. www.cochrane-handbook.org.

Ijff, D.M., Aldenkamp, A.P., 2013. Cognitive side-effects of antiepileptic drugs in children. Handb. Clin. Neurol. 111, 707–718.

Ioannidis, J.P., 2005. Why most published research findings are false. PLoS Med. 2 (8), e124.

Ioannidis, J.P., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R., Moher, D., Schulz, K.F., Tibshirani, R., 2014. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 383 (9912), 166–175.

Jarrar, R.G., Buchhalter, J.R., Meyer, F.B., Sharbrough, F.W., Laws, E., 2002. Long-term follow-up of temporal lobectomy in children? Neurology 59 (10), 1635–1637.

Krucoff, M.O., Chan, A.Y., Harward, S.C., Rahimpour, S., Rolston, J.D., Muh, C., Englot, D.J., 2017. Rates and predictors of success and failure in repeat epilepsy surgery: a meta-analysis and systematic review. Epilepsia 58 (12), 2133–2142.

Kurmis, A.P., 2003. Understanding the limitations of the journal impact factor. JBJS 85 (12), 2449–2454.

Lendt, M., Helmstaedter, C., Elger, C.E., 1999. Pre-and postoperative neuropsychological profiles in children and adolescents with temporal lobe epilepsy. Epilepsia 40 (11), 1543–1550.

Loddenkemper, T., Holland, K.D., Stanford, L.D., Kotagal, P., Bingaman, W., Wyllie, E., 2007. Developmental outcome after epilepsy surgery in infancy. Pediatrics 119 (5), 930–935.

Loring, D.W., Marino, S., Meador, K.J., 2007. Neuropsychological and behavioral effects of antiepilepsy drugs. Neuropsychol. Rev. 17 (4), 413–425.

Meekes, J., Braams, O., Braun, K.P., Jennekens-Schinkel, A., van Nieuwenhuizen, O., 2013. Dutch Collaborative Epilepsy Surgery Programme (DuCESP). Verbal memory after epilepsy surgery in childhood? Epilepsy Res. 107 (1), 146–155.

NHS Commissioning Board, 2013. 2013/14 NHS Standard Contract for Children'S Epilepsy Surgery Service. Retrieved from. http://www.england.nhs.uk/wp-content/uploads/2013/06/e09-child-epilepsy-surg.pdf.

OCEBM Levels of Evidence Working Group, 2011. The Oxford 2011 Levels of Evidence.

Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'cathain, A., Griffiths, F., Boardman, F., Gagnon, M.P., Rousseau, M.C., 2011. Proposal: A Mixed Methods Appraisal Tool for Systematic Mixed Studies Reviews Vol. 2. McGill University, Montréal, pp. 1–8.

Reeves, B.C., Deeks, J.J., Higgins, J.P., Wells, G.A., 2008. Including non-randomized studies. Cochrane handbook for systematic reviews of interventions. Version, 5(0).

Robinson, S., Park, T.S., Blackburn, L.B., Bourgeois, B.F., Arnold, S.T., Dodson, W.E., 2000. Transparahippocampal selective amygdalohippocampectomy in children and adolescents: efficacy of the procedure and cognitive morbidity in patients. J. Neurosurg. 93 (3), 402–409.

Saha, S., Saint, S., Christakis, D.A., 2003. Impact factor: a valid measure of journal quality? J. Med. Libr. Assoc. 91 (1), 42.

Sanderson, S., Tatt, I.D., Higgins, J., 2007. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int. J. Epidemiol. 36 (3), 666–676.

Shastin, D., Chandrasekaran, S., Ferrie, C., Vadlamani, G., Morrall, M., Warren, D., Macmullen-Price, J., Ray, M., Long, V., Goodden, J., Crimmins, D., Sivajumar, G., Chumas, P., 2015. Monitoring the changing pattern of delivery of paediatric epilepsy surgery in England—an audit of a regional service and examination of national trends. Child's Nerv. Syst. 31 (6), 931–939.

Slim, K., Nini, E., Forestier, D., Kwiatkowski, F., Panis, Y., Chipponi, J., 2003. Methodological index for non-randomized studies (MINORS): development and validation of a new instrument. ANZ J. Surg. 73 (9), 712–716.

Smithson, E.F., Phillips, R., Harvey, D.W., Morrall, M.C., 2013. The use of stimulant medication to improve neurocognitive and learning outcomes in children diagnosed with brain tumours: a systematic review. Eur. J. Cancer 49 (14), 3029–3040.

Spencer, S., Huh, L., 2008. Outcomes of epilepsy surgery in adults and children. Lancet Neurol. 7 (6), 525–537.

Sterne, J.A.C., Higgins, J.P.T., Reeves, B.C., on behalf of the development group for ACROBATNRSI, 2014. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBATNRSI). Version 1.0.0, 24 September 2014. Retrieved from. http://www.riskofbias.info.

Stewart, L.A., Tierney, J.F., 2002. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval. Health Prof. 25 (1), 76–97.

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Checklist Version 4. Retrieved from https://www.strobe-statement.org/index.php?id=available-checklists.

Thomson Reuters, 2015. 2014 Journal Citation Reports® Science Edition.

Vale, C.L., Rydzewska, L.H., Rovers, M.M., Emberson, J.R., Gueyffier, F., Stewart, L.A., 2015. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. BMJ 350, h1088.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, Wu-Minn H.C.P., 2013. The WU-minn human connectome project: an overview. Neuroimage 80, 62–79.

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., Tugwell, P., 2000. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses. Retrieved from. http://www.medicine.mcgill.ca/rtamblyn/Readings/The%20Newcastle%20%20Scale%20for%20assessing%20the%20quality%20of%20nonrandomised%20studies%20in%20meta-analyses.pdf.

Wells, G.A., Shea, B., Oconnel, D., Peterson, J., Welch, V., Losos, M., 2016. The Newcastle–Ottawa Scale (NOS) for Assessing the Quality of non-Randomized Studies in Meta-Analyses. 2009 Ottawa. Canada Ottawa Health Research Institute, Ontario.

Wiebe, S., Blume, W.T., Girvin, J.P., Eliasziw, M., 2001. A randomized, controlled trial of surgery for temporal-lobe epilepsy. N. Engl. J. Med. 345 (5), 311–318.

Zeng, X., Zhang, Y., Kwong, J.S.W., Zhang, C., Li, S., Sun, F., Niu, Y., Du, L., 2015. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J. Evid. Based Med. 8, 2–10.

Zupanc, M.L., dos Santos Rubio, E.J., Werner, R.R., Schwabe, M.J., Mueller, W.M., Lew, S.M., et al., 2010. Epilepsy surgery outcomes: quality of life and seizure control. Pediatr Neurol 42 (1), 12–20.