The postural effects of load carriage on young people – a systematic review

Emily Steele*1, Andrea Bialocerkowski2 and Karen Grimmer1

Address: 1The Centre for Allied Health Research, University of South Australia, Adelaide, (5000), Australia and 2School of Physiotherapy, The University of Melbourne, Melbourne, (3010), Australia

Email: Emily Steele* - emily.steele@unisa.edu.au; Andrea Bialocerkowski - aebial@unimelb.edu.au;
Karen Grimmer - Karen.grimmer@unisa.edu.au
* Corresponding author

Abstract

Background: Spinal pain in young people is a significant source of morbidity in industrialised countries. The carriage of posterior loads by young people has been linked with spinal pain, and the amount of postural change produced by load carriage has been used as a measure of the potential to cause tissue damage. The purpose of this review was to identify, appraise and collate the research evidence regarding load-carryage related postural changes in young people.

Methods: A systematic literature review sought published literature on the postural effects of load carriage in young people. Sixteen databases were searched, which covered the domains of allied health, childcare, engineering, health, health-research, health-science, medicine and medical sciences. Two independent reviewers graded the papers according to Lloyd-Smith's hierarchy of evidence scale. Papers graded between 1a (meta-analysis of randomised controlled trials) and 2b (well-designed quasi-experimental study) were eligible for inclusion in this review. These papers were quality appraised using a modified Crombie tool. The results informed the collation of research evidence from the papers sourced.

Results: Seven papers were identified for inclusion in this review. Methodological differences limited our ability to collate evidence.

Conclusions: Evidence based recommendations for load carriage in young people could not be made based on the results of this systematic review, therefore constraining the use of published literature to inform good load carriage practice for young people.

Background

Spinal pain affecting the cervical, thoracic or lumbar regions, is one of the most costly and disabling problems affecting individuals in industrialized countries [1–3]. Population-based surveys of spinal pain variably report a point prevalence of 15%-30%, a one-year prevalence of 50%, and a lifetime prevalence of 60%-80% [4–6]. Furthermore, this type of pain places a significant economic burden on the individual and the community. In Australia, spinal pain is the leading musculoskeletal cause of health system expenditure, with an estimated total cost of $700 million in 1993–1994 [7].

The spine has been identified as a common site of pain in young people as well as adults [8–14]. Estimates of the point prevalence of spinal pain in 12–18 year-old school
students varies between 15 and 44% (age and gender
dependent) [8–14]. It has been hypothesised that spinal
pain experienced in childhood and adolescence is a signif-
icant risk for spinal pain experienced in later life [15,16],
although there are few longitudinal studies to verify this.
Nevertheless, the investigation and reduction of contrib-
uting factors to spinal pain in young people may be an
appropriate step towards reducing the burden of spinal
pain on our society.

Load carriage has been associated with spinal pain in both
adolescents and adults [17,18], although it is not ethically
possible to experimentally investigate the causal nature of
this relationship. A change in spinal posture has been
accepted by experts as a plausible intermediary measure of
the potential for spinal pain due to load carriage [19–21].
To our knowledge, measurement of load carriage induced
postural change is the only method of experimentally
approximating the potential for load carriage to induce
spinal pain that is currently used by researchers.

The use of postural change as a proximate measure
involves the hypothesis that larger postural displacement
from the unloaded position increase the likelihood of
developing spinal pain. By experimentally manipulating
loads and how they are carried, and measuring the degree
of induced postural change, researchers’ estimate the
potential of posterior load carriage to induce spinal pain.
The effect of load weight [19,20],[22–25], method of load
carriage (2 strapped backpack, 1 strapped backpack)
[21,22], position of the load on the spine [23], time of
load carriage [20,22], and distance of load carriage [19]
on young people have been investigated in this manner.

An understanding of load-carriage-induced postural dis-
placements, and their potential to produce spinal pain, is
needed to direct recommendations for load carriage by
young people with the aim of minimising their spinal
pain.

This article reports on a systematic review undertaken to
identify, appraise and collate the research evidence regard-
ing load-carriage related postural changes in young
people.

Methods

Literature search
We employed a comprehensive search strategy to source
papers that described the postural effects of load carriage
on the spinal health of young people under the age of 18
years. Allied health, child care, health, health-research,
health-science, medicine and medical sciences databases
were accessed, including Academic Search Elite, AEI,
AMED, AMI, APAIS, Ausport Med, AUSThealth, Australian
Public Affairs, Blackwell Science and Munksgaard Online
Journals, CINAHL, Cochrane Library, ERIC, FAMILY,
MEDLINE, Science Direct and Wiley InterScience. Our
search strategy consisted of three stages:

1. Combinations of specific keyword searches were used
   in each database; ('backpack' OR 'bag' OR 'load' OR 'ruck-
sack') AND ('youth' OR 'child/ren' OR 'adolescen/t/nce')
   AND ('back' OR 'posture' OR 'pain'). The appropriate
   truncation symbol for each database was used with each
   key word. The internet and library databases were also
   searched using these terms.

2. Time and language restraints were set;
   > Years of publication from January 1985 to November
     2002.
   > English language only.

3. Papers were excluded if the main outcome measured
   was not postural change.

Hierarchy of evidence
Two experienced research physiotherapists independently
assessed all papers sourced. First, the level of evidence of
each paper was determined according to the hierarchical
system of Lloyd-Smith [26] (Table 1). The level reflects the
degree to which bias has been considered within study
design, with a lower rating on the hierarchy indicating less
bias. Only papers that scored between 1a and 2b on
Lloyd-Smith’s scale [26] were included in this review. In
this way we could ensure that recommendations for load
 carriage advocated by this review were based on findings
of high-level evidence.
Quality appraisal
Second, we assessed the quality of these papers based on a modified version of a well-established quality appraisal tool recommended by Crombie [27]. An extra appraisal item, ‘Sensitivity of outcome tool’ was added to the published tool since the use of an insensitive outcome tool may have meant that differences in posture between conditions were not measured, significantly impacting on study outcomes [28]. The quality of each paper was scored according to factors shown in Figure 1. Prior to scoring, it was necessary to clarify one of the appraisal items to ensure that reviewers were consistent in their approach. Reviewers recognized that study design is unlikely to account for all potential biases, therefore appraisal item number 11 (Figure 1) ‘Attention to potential biases’ was scored positively if the paper acknowledged the potential impact of all likely biases. One point was allocated for fulfilment of each quality appraisal item. The maximum score, (indicating high quality), was 16, with the lowest possible score being zero. The methodological quality of each study was subsequently rated as low (0–5 points), moderate (6–11 points), or high (12–16 points), similar to the procedure outlined by Geytenbeek [29]. Any disagreements between the reviewers were resolved by consensus building. We reported on critical appraisal items which were poorly addressed in the papers.

Conditions (static/dynamic)
Third, we recorded the condition of bag carriage that was being assessed (static or dynamic). We hypothesized that the testing condition would impact on the study outcomes since walking requires different muscular activity compared to static standing [30].

Measurement methods
Fourth, we recorded the method of measurement of posture described in each of the papers sourced. This information helped identify whether the results of the studies could be synthesised.

Study outcomes
Information from our four stages of assessment was utilised to guide the collation of research evidence regarding load-carriage related postural changes in young people, for variables such as load weight, method of load carriage (2 strapped backpack, 1 strapped backpack), position of load on the spine, time of load carriage, and distance of load carriage.

Results

Literature search
Four hundred and eighty eight papers were identified from our initial search of the databases. Three hundred and twelve of these papers were excluded from our review, as they did not specifically measure postural effects of load carriage in young people. One appropriate paper was found through the internet and library database searches.

Hierarchy of evidence and quality appraisal
The remaining 177 papers were assessed for level of evidence. Only seven of these papers scored 1a to 2b [19–25] in the hierarchy of evidence [26]. None of the papers were meta-analyses of randomised controlled trials (level 1a), three of the papers were randomised controlled studies (level 1b) [19,23,24], and four were well-designed, non-randomised studies (level 2a) [20–22,25].

Table 2 provides the total number of papers (maximum of seven) that fulfilled the criteria for each appraisal item. Table 3 provides the following information relating to the publications included in this systematic review; hierarchy level, score achieved for the quality appraisal items most poorly addressed by the studies (items for which four or less studies scored a point), appraisal score and quality category. Based on the results of the quality appraisal process, one of the seven papers was ranked as high quality [23], with the remainder being of moderate quality [19–22,24,25].

Conditions (static/dynamic), measurement methods, study outcomes
Table 4 (additional file 1) summarises the conditions (static/dynamic), measurement methodology and outcomes of the seven papers. One paper measured the effects of static load carriage [23], four measured the effects of dynamic load carriage [19,20],[24,25], and two papers studied effects of both conditions [21,22]. There were similarities in methods of measurement of posture across the seven papers, although no two papers used identical approaches. Inconsistent results were found across the seven studies of the effects of load related variables that were investigated. Differences in study methodology and quality, condition under which posture was assessed (static/dynamic), and postural measurement should be considered as reasons for lack of consistency of outcomes (see Table 4). Table 4 highlights the differences in methodology and outcomes of the eligible studies, and underlies the difficulties of undertaking any comparisons or syntheses of results.

Discussion
Evidence-based practice focuses on finding consistencies across studies that have investigated the same interventions on the same study populations [31]. The collation of information from similar studies should lead to evidence-based recommendations. Load carriage by young people is a contentious issue [32,33], because of concerns for morbidity related to spinal pain. Although much has been written on the topic of load induced postural change and young people (as evidenced by the number of papers...
sourced for this review), very little of the available literature was of high quality. Only seven papers sourced met our requirements for level of evidence, and all contained methodological limitations, despite being graded as having moderate and high methodological quality. These limitations, such as a lack of randomisation of order of testing conditions, and insensitive, unreliable and invalidated outcome tools, limited our ability to draw definitive

## Table 1: Hierarchy of evidence (Lloyd-Smith 1997)

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>Study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Meta-analysis of randomised controlled trials</td>
</tr>
<tr>
<td>1b</td>
<td>One individual randomised controlled study</td>
</tr>
<tr>
<td>2a</td>
<td>One well-designed, non randomised controlled study</td>
</tr>
<tr>
<td>2b</td>
<td>Well – designed quasi-experimental study</td>
</tr>
<tr>
<td>3</td>
<td>Non-experimental descriptive studies-comparative/ case studies</td>
</tr>
<tr>
<td>4</td>
<td>Respectable opinion</td>
</tr>
</tbody>
</table>

## Table 2: Scores achieved for quality appraisal items

<table>
<thead>
<tr>
<th>Quality appraisal item</th>
<th>Number of papers which fulfilled the criteria (maximum = 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Clearly stated aims</td>
<td>a–g</td>
</tr>
<tr>
<td>2. Appropriateness of design to meet the aims</td>
<td>b–e</td>
</tr>
<tr>
<td>3. Adequate specifications of subject group given</td>
<td>b</td>
</tr>
<tr>
<td>4. Justification of sample size</td>
<td>b</td>
</tr>
<tr>
<td>5. Likelihood of reliable and valid measurements</td>
<td>b</td>
</tr>
<tr>
<td>6. Sensitivity of outcome tool</td>
<td>a,b</td>
</tr>
<tr>
<td>7. Adequate description of statistical methods</td>
<td>a,b,d,f,g</td>
</tr>
<tr>
<td>8. Adequate description of the data</td>
<td>a–g</td>
</tr>
<tr>
<td>9. Consistency in the number of subjects reported throughout the paper</td>
<td>a–e,g</td>
</tr>
<tr>
<td>10. Assessment of statistical significance</td>
<td>a–g</td>
</tr>
<tr>
<td>11. Attention to potential biases</td>
<td></td>
</tr>
<tr>
<td>12. Meaningful main findings</td>
<td>a–g</td>
</tr>
<tr>
<td>13. Interpretation of null findings</td>
<td>a,b,d,g</td>
</tr>
<tr>
<td>14. Interpretation of important effects</td>
<td>a–d,g</td>
</tr>
<tr>
<td>15. Comparison of results with previous reports</td>
<td>a–g</td>
</tr>
<tr>
<td>16. Implications in real life</td>
<td>b</td>
</tr>
</tbody>
</table>

* a refers to Chansirinukor et al (2001); b refers to Grimmer et al (2002); c refers to Hong and Brueggemann (2000); d refers to Hong and Cheung (2003); e refers to Kennedy et al (1999); f refers to Pascoe et al (1997); g refers to Wong and Hong (1997).

## Table 3: Results of the hierarchy of evidence and quality appraisal stages

<table>
<thead>
<tr>
<th>Paper</th>
<th>Hierarchy level</th>
<th>Appropriate design</th>
<th>Specification of subject group</th>
<th>Justification of sample size</th>
<th>Reliable and valid measurements</th>
<th>Sensitive outcome tool</th>
<th>Attention to potential biases</th>
<th>Real life implications</th>
<th>Appraisal score</th>
<th>Quality category</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>2a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10/16</td>
<td>Moderate</td>
</tr>
<tr>
<td>b</td>
<td>1b</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>15/16</td>
<td>High</td>
</tr>
<tr>
<td>c</td>
<td>1b</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8/16</td>
<td>Moderate</td>
</tr>
<tr>
<td>d</td>
<td>1b</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9/16</td>
<td>Moderate</td>
</tr>
<tr>
<td>e</td>
<td>1b</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8/16</td>
<td>Moderate</td>
</tr>
<tr>
<td>f</td>
<td>2a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9/16</td>
<td>Moderate</td>
</tr>
<tr>
<td>g</td>
<td>2a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9/16</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

* a refers to Chansirinukor et al (2001); b refers to Grimmer et al (2002); c refers to Hong and Brueggemann (2000); d refers to Hong and Cheung (2003); e refers to Kennedy et al (1999); f refers to Pascoe et al (1997); g refers to Wong and Hong (1997).
conclusions from their outcomes. In addition, differences in study design and inconsistencies in study outcomes constrained the production of evidence-based load carriage recommendations for young people.

Assessment of the quality of studies is a vital part of the systematic review process as it guides the interpretation of the results or outcome of each paper [34]. Many quality scales and checklists have been published [35]. However, these tools should be used with caution, as they are generally based on 'accepted' criteria, and often have not undergone validation, nor reflect key issues pertinent to the area under review. Therefore, it is possible for a paper to be scored moderately by a quality tool, yet still contain significant methodological limitations, as occurred in this systematic review. The quality appraisal items which received a score in four or less of the seven studies were (as seen in Table 3):

- **Appropriateness of design to meet the aims** The lack of randomization of the order of testing conditions in three papers reviewed is likely to have produced a systematic postural change, affecting the outcome of these studies [28,36].

- **Adequate specifications of the subject group** Musculoskeletal injury, or disease processes which affect musculoskeletal integrity, could affect postural response to load carriage [36]. Therefore, the lack of specific exclusion criteria in six reviewed papers may have decreased the validity of findings.

- **Justification of sample size** Unjustified sample sizes suggest that six of the seven studies may lack sufficient power to detect significant results [36]. Therefore, it is unlikely that results of these papers can be generalized to the wider population, even within the gender/age group of study subjects.

- **Likelihood of reliable and valid measures/sensitivity of outcome tool** It was not known, in the majority of the papers reviewed, whether the outcome measure assessed the construct it was supposed to measure (validity), assessed the construct in a consistent manner on repeated occasions of assessment (reliability) or could detect meaningful changes in the construct over time (sensitivity). These three factors are critical in determining whether study outcomes are meaningful [28].

- **Attention to potential biases** None of the seven papers acknowledged all potential biases to study outcomes. This information is critical in determining the relevance of study findings [36].

- **Implications in real life** As assessed throughout the quality appraisal stage, most papers (6/7) did not produce results which could be readily generalized to 'real life' due to small and/or unjustified sample size, non-randomised subject selection process, and subject group limited to one gender or to a specific age [36].

The appropriateness of the use of load-induced postural change as a proximate measure of spinal pain has not been widely discussed in published literature and needs to be given further attention by researchers in the future.

**Conclusion**

From a public health perspective, concerted efforts should be directed to decreasing spinal pain experienced by young people to lessen the future burden of adult spinal pain on modern societies. The first step in this process is to conduct rigorous research to determine the postural effects of load carriage in young people. This review outlines the areas which require more attention, these being: the inclusion of randomisation of order of conditions tested; adequate specifications of subject group (including appropriate exclusion criteria); justified sample sizes; reliable, valid and sensitive outcome tools; attention to potential biases, and the ability of study outcomes to be generalised to 'real life'.

This review has highlighted that there is currently no standardised approach to the study of load-induced postural change for young people. For example, static and dynamic postural conditions are being assessed, and within these postural domains the methodology used to measure changes in posture varies significantly. As a consequence, difficulty exists in comparing results, precluding meta-analysis [30]. This lack of consensus regarding standardised data collection may be the result of research being undertaken by different groups, over many continents. Therefore, it is important that researchers reach international consensus regarding the most appropriate method(s) used to assess postural change in order to hasten the development of evidence-based recommendations for the carriage of loads by young people.

**Competing interests**

Once the seven papers to be included in this review had been identified, we acknowledged that there was a potential for bias in our reviewing process. This potential arose from the fact that our third author, Dr Grimmer, is also an author of two of the seven reviewed papers (Chansirinukor et al 2001, Grimmer et al 2002). Therefore, she was not involved in the hierarchy of evidence or quality appraisal stages of this systematic review. It is important to note that the two reviewers involved in these stages (the first and second authors of this paper) were not involved.
in Dr Grimmer's two papers, or the research described within, which were included in the systematic review.

**Author's contributions**

ES undertook the literature search, and reviewed the level of evidence, methodological quality, conditions tested, and measurement methods of each of the sourced papers. ES coordinated the analysis of results of this review. AB played a significant role in the analysis of results of the review also. AB was the second reviewer for the level of evidence, methodological quality, conditions tested, and measurement methods of each of the sourced papers. ES coordinated the analysis of results of this review. ES undertook the literature search, and reviewed the level of evidence, methodological quality, conditions tested, and measurement methods of each of the sourced papers. ES coordinated the analysis of results of this review.

All authors participated in the writing of this review, and read and approved of the final manuscript.

**Additional material**

**Additional File 1**

Descriptive data. Results of conditions tested, measurement methods and study outcomes stages of the systematic review

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2474-4-12-S1.doc](http://www.biomedcentral.com/content/supplementary/1471-2474-4-12-S1.doc)

**References**

36. Elwood M: Critical Appraisal of Epidemiological Studies and Clinical Trials
37. Lucas G: General Orthopaedics In Essentials of Musculoskeletal Care
   Surgeons; 2001.

Pre-publication history
The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2474/4/12/prepub